Russian Journal of Genetics, Vol. 39, No. 11, 2003, pp. 1292–1305. Translated from Genetika, Vol. 39, No. 11, 2003, pp. 1527–1541.

ANIMAL GENETICS

Molecular Evolution of Satellite DNA CLsat in Lizards from the Genus *Darevskia* (Sauria: Lacertidae): Correlation with Species Diversity

D. G. Ciobanu[1], V. V. Grechko[1], and I. S. Darevsky[2]

[1] *Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 119991 Russia; fax: (095) 135-14-05; e-mail: dia@id.ru; grechko@genome.eimb.relarn.ru;*
[2] *Zoological Institute, Russian Academy of Sciences, St. Petersburg, 199034 Russia*

Received April 8, 2003

Abstract—The structure and evolution of a satellite DNA family was examined in lizards from the genus *Darevskia* (family Lacertidae). Comparison of tandem units of repeated DNA (satDNA), CLsat, in all species from the genus *Darevskia* has shown that their variability is largely explained by single-nucleotide substitutions, which form about 50 diagnostic positions underlying classification of the family into three subfamilies. Maximum differences between the subfamilies reached 25%. At this level of tandem unit divergence in the subfamilies, no cross-hybridization between them was observed (at 65°C). The individual variability within one subfamily within the species was on average 5% while the variability between species consensuses within a subfamily was 10%. The presence of highly conserved regions in all monomers and some features of their organization show that satellites of all *Darevskia* species belong to one satDNA family. The organization of unit sequences of satellites CLsat and Agi160 also detected by us in another lizard genus, *Lacerta* s. str. was compared. Similarity that was found between these satellites suggests their relatedness and common origin. A possible pathway of evolution of these two satDNA families is proposed. The distribution and content of CLsat repeat subfamilies in all species of the genus was examined by Southern hybridization. Seven species had mainly CLsatI (83 to 96%); three species, approximately equal amounts of CLsatI and CLsatIII (the admixture of CLsatII was 2–5%); and five species, a combination of all three subfamilies in highly varying proportions. Based on these results as well as on zoogeographic views on the taxonomy and phylogeny of the *Darevskia* species, hypotheses on the evolution of molecular-genetic relationships within this genus are advanced.

INTRODUCTION

SatDNA (noncoding tandem repeats) has been found virtually in all eukaryotic taxa examined. This DNA is represented by sequences differing in organization, size, length of repeated units, chromosome localization, and taxon specificity [1, 2]. For a long time their function was unknown and no apparent general properties were found, these repeats (together with another class, dispersed repeats) have long been thought to be "junk" DNA, which is not involved in evolution and speciation propagating "selfishly" irrespective of the general metabolism and requirements of the organism [3]. Today this simplistic view seems obsolete. It may have stemmed from insufficient understanding of the role of repeats at the time when it was first advanced. Apparently, the presence or absence of repeats in a DNA region, their mutation and movement, interaction with proteins can affect the regulation of metabolic processes, the formation of high-order structures and DNA–protein complexes in the cell, which has been conclusively shown in many cases (see Trifonov [4] for review).

SatDNA (tandem repeats) has been examined in a plethora of studies containing both factual material (examination of numerous satellite forms and families) and theoretical analysis of the mechanisms underlying their formation and evolution (see [1, 4] for review). The evidence on repeated sequences (in particular, satellite repeats) accumulated over the last 10–15 years has revealed their marked taxon specificity not only at the high-rank taxon level but also at the level of genera and lower [2, 5]. These data suggest a causative relationship between the process of morphological speciation and molecular evolution of repeats of various types, including satellite ones. This relationship cannot be studied experimentally for obvious reasons, but some parallels between morphological hierarchy of taxa and properties of DNA repeats confirm this view [2].

There is virtually no data on reptiles among the tandem repeats studied in various taxa (mammals, birds, fish, amphibians, insects, crustaceans, etc.). This gap hinders data generalization and formation of the concept on the relationship of satDNA molecular evolution with evolution of biological taxa. We had an opportunity to study this issue on a quite large set (about 30) of species and subspecies of lizards from the genus *Darevskia* (order Sauria, family Lacertidae) dwelling in the Caucasus and until recently included in a huge and poorly structured genus *Lacerta* [6]. At present morphological taxonomy of lacertids has been extensively

1022-7954/03/3911-1292$25.00

studied and improved, but many issues remain unresolved because of the limitations of this approach. This group of reptiles is a good model for studying problems of speciation. Taxonomists distinguished among the lizards from the genus *Lacerta* s. lato (in the broad sense) several groups referred to as "complexes," which consisted of populations with different degrees of relatedness, partly unclear status, and internal subdivision. One of these groups, the *Lacerta saxicola* complex containing about 20 species, has been recently recognized as a separate genus *Darvevskia* [7].

Another group of lacertids, the so-called *Lacerta agilis* complex, contains five major species under a generic name of *Lacerta* s. str. (sensu stricto: in the narrow sense) [12]. According to views of zoologists, the divergence of these species inferred from paleontological and immunological data, occurred over 25 Myr ago [8–10]. The degree of their relatedness is still unclear due to the aforementioned vague understanding of their phylogeny and taxonomy. The development of these issues in the context of molecular genetics seems of interest. Until recently, we have focused on the genus *Darevskia*, studying the relationships of its species and populations with the use of the DNA taxonoprint [11] and RAPD [12, 13] techniques. In the present work, we used satDNA of this genus as molecular markers.

In the past decade, studies on isolation and characterization of tandem repeats in lizards from the family Lacertidae were undertaken by us and a group of authors from Italy. The Italian authors have discovered tandem repeated DNA of lizards from the Mediterranean genus *Podarcis* and several other species from the same family [14, 15]. We have reported two novel families of tandem repeats. One of them, CLsat, is specific for the genus *Darevskia* [11, 12, 16, 17]; the other, Agi160, for members of the genus *Lacerta* s. str. [Ciobanu, unpublished data]. The tandem repeats described by the Italian authors are completely different from the repeats that we found in sequence and size of the unit. However, we have shown that CLsat and Agi160 display a degree of similarity, which testifies to their relatedness [Ciobanu, unpublished data]. Studying DNA repeats in reptiles in general and lizard taxa in particular is crucial for understanding lacertid phylogeny and correlating it with the molecular evolution of satDNA.

In the present work, we present the data on the structure and organization of CLsat tandem repeats and their distribution in virtually all of the known species of the genus *Darevskia*. In addition, we present comparative analysis of the structure of CLsat and Agi160, from which their divergence and evolution pathways are inferred.

MATERIALS AND METHODS

Biological material. The sampling localities of the biological species examined are given in Table 1.

DNA isolation. DNA was isolated from blood of the animals after they were anesthesized by means of nembutal. The samples were fixed in an EDTA solution (final concentration 0.05 M, pH 7.4) and stored at –55°C. The erythrocyte lysate was centrifuged, and DNA isolated from the nucleus precipitate by the standard deproteinization program using proteinase K and phenol–chloroform mixture [18]. The DNA concentration was determined by comparing intensity of DNA bands at different dilutions (electrophoresis in 1% agarose gel) with the intensities of marker bands of genomic murine DNA or human DNA with known concentration.

Cloning and sequencing. The genomic DNA was digested with restriction endonucleases *Hind*III, *Cla*I, *Sau*IIIA, and *Taq*I (Fermentas, Lithuania). After the fractionation in 2% agarose gel, the DNA fragment about 150 bp in size were transferred to a nitrocellulose DEAE membrane (NA-45, Schleicher and Schuel). The fragments were ligated into plasmid vectors pGEM (Promega) and cloned using *Escherichia coli* strain XL-1 Blue.

Positive clones were selected by the white-blue (β-galactosidase) test. The plasmids were isolated using a standard program [19]. The clones were sequences by Sanger's method, using a cyclic sequencing kit (Promega) and [γ-^{32}P]ATP labeled universal primer pUC/M13 or direct labeling using [α-^{32}P]dATP, following the instructions of the manufacturer.

Hybridization was conducted using Southern and dot hybridization. Intensity of hybridization signals was estimated using a Phosphoimager radioscanner supplemented with the OptiQuant program. The satDNA amounts was transformed to percentages from hybridization intensity signal values expressed in light units using the OptiQuant program. Hybridization satDNA probes were prepared by PCR using specific primers and earlier obtained plasmid constructions containing a unit of a specific satDNA type. The template was purified from possible admixture of other PCR products by electrophoresis in agarose gel. The labeling was also conducted by PCR using specific primers and [α-^{32}P]dATP. First we used for CLsatI and CLsatII primers and template designed by Rudykh *et al.* [16, 17]; then primers 5'-AAGCTTCATTTTAGC-3', 5'-GAAACACAACTACAT-3', and template obtained by us (*L. caucasica Hind*III-1pl) were used for CLsatII and primers 5'-AACCTTCATTTTAGCTGATT-3', 5'-TCAAAACACAAAGACATCCG-3', and template *L. dryadaTaq*I-5pl for CLsatIII template.

RESULTS

The CLsat Tandem Repeat Family: Variability and Conservatism of the Structure

To examine rates and direction of satDNA evolution and to determine possibilities of using tandem repeats in phylogenetic studies, we have analyzed the structure and organization of CLsat unit sequences and measured

Table 1. Species examined and sampling localities of lizards from the family Lacertidae

Species	Abbreviation	Sampling locality
Genus *Darevskia*		
D. armeniaca (P)	arm	Armenia, Tezh
D. armeniaca (P)		Armenia, Sevan Lake
D. alpina	alp	Russia, Kabardino-Balkaria, Baksanskoe Canyon, settlement of Azau
D. bendimachiensis (P)	ben	Turkey, Muradije
D. caucasica	cau	Russia, Kabardino-Balkaria, Golubye Ozera, canyon of Chelek River
D. chlorogaster	chl	Azerbaijhan, Dashdadyuk, woods
D. clarkorum	cla	Turkey, Magden
D. daghestanica	dag	Russia, Dagestan, village of Kvarchi
D. dahli (P)	dah	Armenia, Dilizhan, Shagali
D. derjugini derjugini	der	Georgia, Akhaldaba
D. derjugini abhazica		Georgia, Adzharia, Batumi
D. dryada	dry	Georgia, Adzharia, Girgami
D. lindholmi	lin	Ukraine, Crimea, Yalta
D. mixta	mix	Georgia, Akhaldaba
D. nairensis	nai	Armenia, Lchap
D. parvula	par	Georgia, Adzharia, Batumi
D. parvula		Georgia, Akhaldaba
D. portschinskii	por	Armenia, Gosh
D. praticola pontica	pra	Russia, Sochi
D. praticola praticola		Russia, Kabardino-Balkaria, Golubye Ozera
D. raddei	rad	Armenia, Gosh
D. raddei		Azerbaijan, Southeast, Dob-Achach
D. rostombekovi (P)	ros	Armenia, Gosh
D. rudis obscura	rud	Georgia, Akhaldaba
D. sapphirina (P)	sap	Turkey, Patnos
D. saxicola darevskia	dar	Russia, Sochi
D. saxicola szczerbaki	szc	Ukraine, Crimea, Anapa
D. unisexualis (P)	uni	Armenia, Lchap
D. uzzelli (P)	uzz	Turkey, Horason
D. valentini	val	Armenia, Lchashen
Genus *Lacerta* s. str.		
L. agilis boemica	a. boe	Russia, Kabardino-Balkaria, Nal'chik

Note: P, parthenogenetic unisexual species.

their content in members of approximately 25 species of the genus *Darevskia*. At least five, and in many cases over ten, sequences (in all, 270) were analyzed to estimate individual, interpopulation, and interspecific variability. This data on two CLsat subfamilies was partly published with Rudykh as coauthor [11, 12, 16, 17]. These results are reproduced in the present study with the use of a larger number of species and sequenced clones containing repeats. In the present work, the third subfamily has been found and a complex analysis of the previous and new data is presented.

Analysis of nucleotide sequences of all isolated repeated units in 20 species of the genus *Darevskia* showed that the repeats varied in a wide range. Individual intraspecific variation, as a rule, was 0–10% (data not presented). Some monomer units from one animal contained rare variants that exhibited higher variation of 10–15%. Although their presence did not affect the species consensus sequence, detailed analysis of their structure and significance will be presented in a separate paper. One "species" consensus was constructed on the basis of groups of monomers from animals of the

same species with 90–100% similarity. Groups whose differences attained 25% were found in the genome of one species. A detailed analysis of all of the 270 sequences revealed about 50 diagnostic nucleotide positions, based on which all monomers could be grouped at least in three subfamilies. A comparison of species consensuses of the three CLsat subfamilies is presented in Fig. 1.

Interspecific variability of consensus monomers within subfamilies CLsatI and CLsatIII reached 15–20%; in CLsatII it ranged from 0 to 12% (Fig. 2). At the same time, variability of species consensuses among different subfamilies was on average 25%. According to some literature data, this level of divergence in other taxa is characteristic of different but related families of satDNA [14, 15, 20]. However, we believe that in this case, the term *subfamily* better reflects the common origin and relatedness of CLsat repeats. Since cross-hybridization between sequences of different CLsat subfamilies does not occur at this variability level, we could carry out screening of the content of each of the subfamilies in the *Darevskia* species genome using specific labeled probes (see below).

Figure 1 shows that all three subfamilies have repeated units of similar size, frequent (12 repeats, which is approximately one-third of the monomer) homopolymer tracks, and a number of conserved regions. These regions are represented by the following similarity sequences: two, between monomers CLsatI and CLsatII (48 bp: 85% and 20 bp: 95%); one, between CLsatI and CLsatIII (40 bp: 82–87%); one, between CLsatI and CLsatIV (29 bp: 93%); and other two, between CLsatII and CLsatIII (25 bp: 95% and 30 bp: 80–87%).

Thus, in addition to the high (up to 25%) nucleotide sequence variability of the family monomers, CLsat repeats have similar organization and internal structure of monomers, which suggests their relatedness and common origin. The common origin of CLsat repeats is illustrated by the fact that the similarity between the consensuses of each of subfamilies CLsatI, CLsatII, and CLsatIII and the total consensus of the CLsat family is 5–12% higher than the similarity between them (Fig. 3, Table 2). The similarity between the total consensuses of the three subfamilies is also higher (by about 5–10%) than the similarity between species consensuses of different subfamilies.

Hybridization analysis has shown that all species of the genus have the CLsatI subfamily although in four of them (*D. parvula*, *D. praticola*, *D. daghestanica*, and *D. caucasica*), the amount of this monomer is low. In addition to CLsatI, several species have monomers CLsatII and CLsatIII in different proportions. This issue is considered below, with the results of Southern hybridization.

Table 2. Similarity (%) between the CLsat consensus, consensuses of three CLsat families, and the Agi160M2 monomer

	CLsatI	CLsatII	CLsatIII	Agi160M2
CLsat	85	85	89	71
Deletions	0	1	1	2
CLsatI	–	80	77	62
Deletions		0	2	3
CLsatII		–	78	65
Deletions			2	4
CLsatIII			–	63
Deletions				1
Agi160M2				–
Deletions				

Note: Figures in the Deletions row show percentage of single deletions resulting from alignment of monomer sequences.

Comparison of Mutation Accumulation Rates and Homogenization of Three CLsat Subfamilies

The species consensus sequences within each CLsat subfamily are presented in Fig. 2. The examination of variability of individual monomers has shown that the interspecific variation of consensus monomers within a subfamily is proportional to the intraspecific divergence of individual monomers (data not presented). The average values of the intraspecific to interspecific CLsat variation ratio is as follows: CLsatI ~ 5/10%; CLsatII ~ 2/3%; CLsatIII ~ 7/15%.

Individual and interspecific monomer variability within each CLsatI subfamily is characterized by a high accumulation rate of mutations, among which transitions C $\rightleftharpoons$ T and G $\rightleftharpoons$ A prevail (Table 3). The GC $\rightleftharpoons$ AT transitions are known to be among the most common spontaneous nucleotide substitutions caused by deamination of the C base. However, the divergence of some CLsatIII monomers results from insertions/deletions that may be caused by replication slippage. This is indicated by the presence of similar DNA regions upstream of the amplified sequence as well as short repeats flanking this zone, which may serve as mispairing sites.

Analysis of direction of nucleotide substitutions showed that the CLsat sequences are generally characterized by prevalence of transversions over transitions, but T $\rightleftharpoons$ C transitions are the most common substitutions (Table 3). This transition/transversion ratio suggests that the nucleotide substitution direction in CLsat is not subjected to selection and that CLsat monomer sequences are under random mutation. Nevertheless, this does not exclude the possibility of natural selection at the other organization of the CLsat tandem repeat family.

Comparison of the nucleotide substitution numbers in various CLsat subfamilies shows that CLsatII monomers have the lowest levels of interspecific and individ-

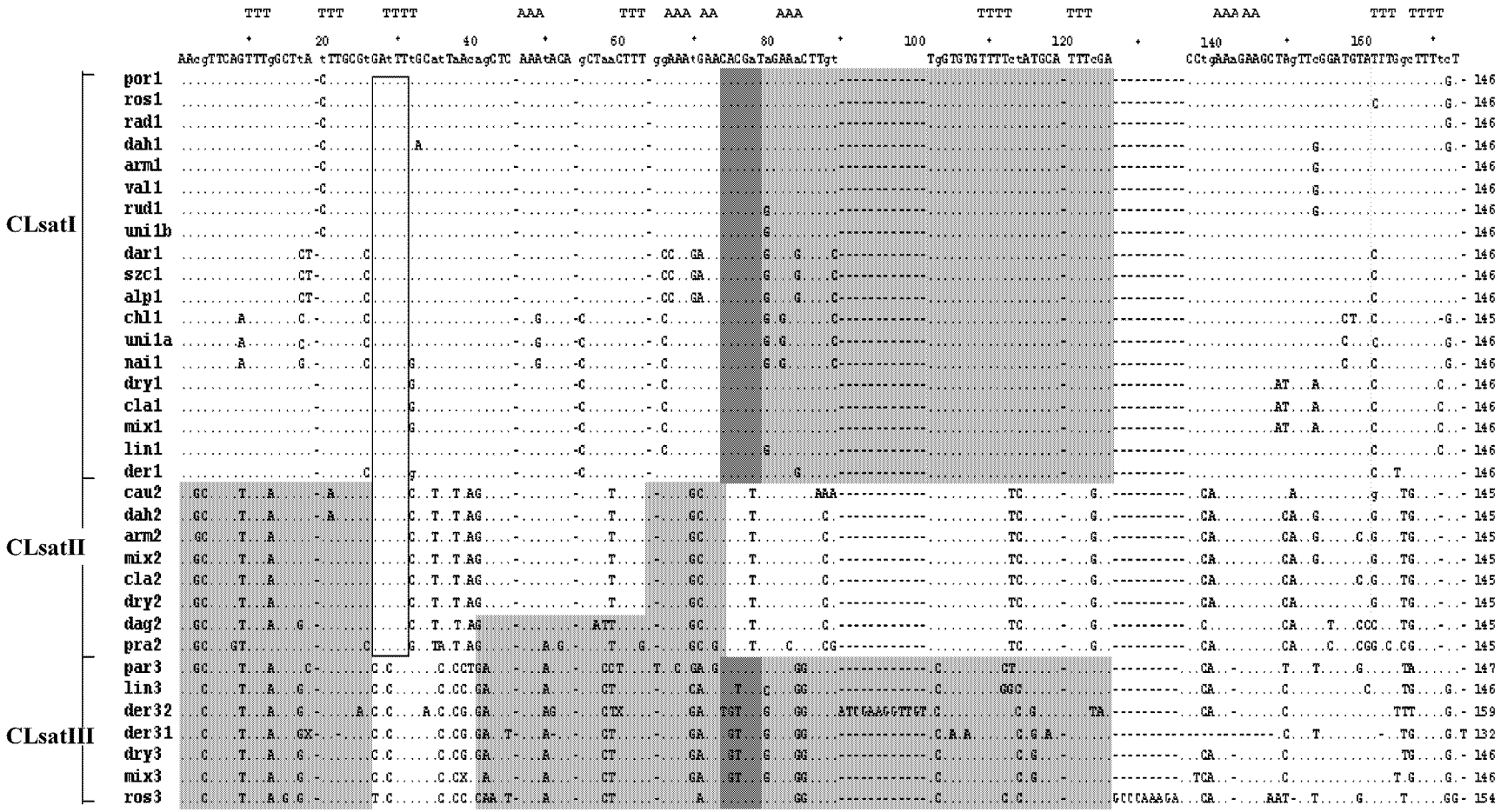

Fig. 1. Comparison of species consensus sequences of monomers of the CLsat satDNA family of lizards from the genus *Darevskia*. The general consensus is above the aligned monomers; low-case letters indicate varying positions. Homopolymer tracks are above the general consensus. The species affiliation of the sequencies is given in the left column (see Abbreviations in Table 1); monomers of subfamilies CLsatI, CLsatII, and CLsatIII are designated by corresponding numerals. Nucleotides of the compared monomers identical to the general consensus are shown by dots; deletions are shown by dashes. Light, moderate, and dark hatching indicates conserved zones for respectively CLsatI and CLsatII; CLsatII and CLsatIII, and CLsatI and CLsatIII.

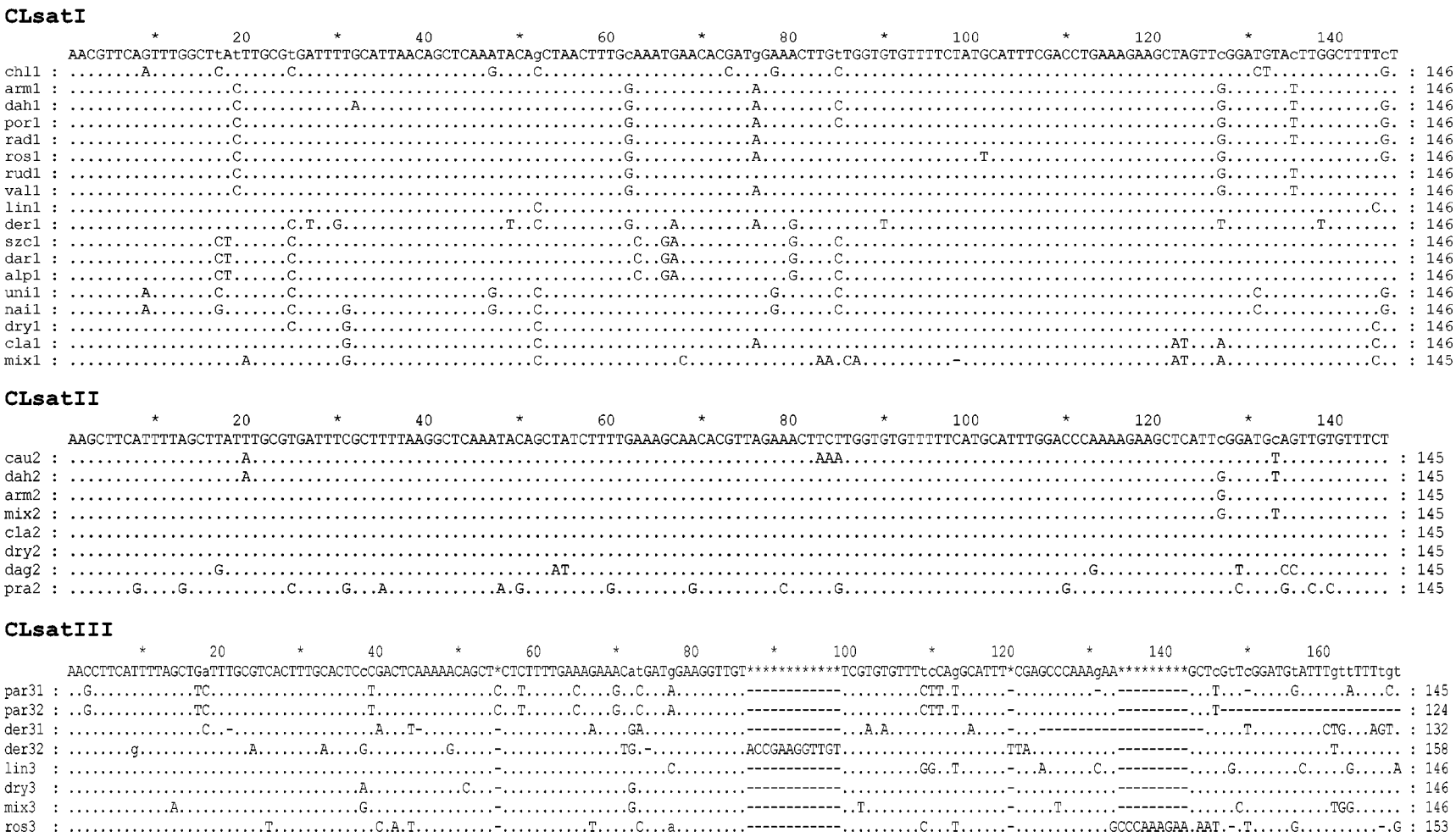

Fig. 2. Comparison of species consensus sequences within each of the CLsat subfamily. The consensus of each of the subfamilies is given above. See Fig. 1 for designations.

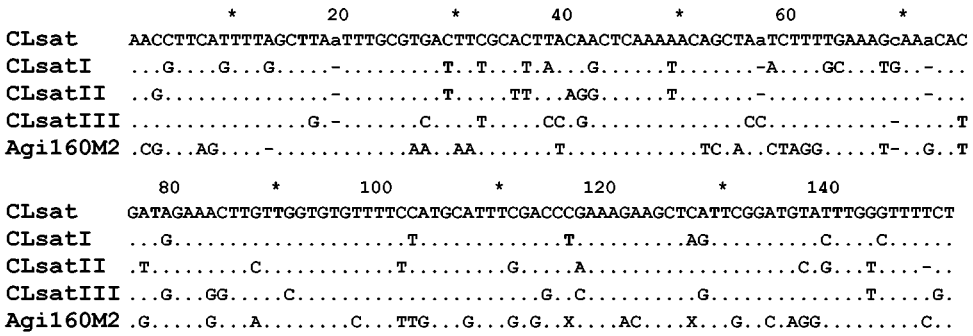

Fig. 3. Comparison of sequences of general consensus of the CLsat family (above), consensuses of three subfamilies (CLsatI, CLsatII, and CLsatIII), and single monomer Agi160M2(AY184825). See Fig. 1 for designations. Similarity value in percent is given in Table 2.

ual variability. The transition/transversion ratio in this subfamily (0.13) is much higher than in CLsatI (0.78) and CLsatIII (0.78). However, the most common substitutions in the CLsatII subfamily are T ⇄ A transversions. They may be caused by pyrimidine dimers (e.g., T–T) caused by UV radiation of sunlight or reactive metabolites (e.g., active oxygen forms) followed by T to A substitution. CLsatI monomers are characterized by a combination of high mutation accumulation rate with such level of homogenization that can lead to the formation subgroups both within the specimen of species (not presented) and at the species consensus level (Fig. 2).

Based on the concept of concerted evolution [5], that assumes dynamic interaction of several molecular mechanisms as a driving force changing tandem DNA repeats, we can suggest that the ratio between the mutation accumulation rate and homogenization indicate concerted evolution of this satDNA family. The direction of the evolution vector depends on the rate of "switching" of a particular molecular mechanism and on the ratio of these rates. A high rate of mutation accumulation and low homogenization enhance intraspecific variability and decrease species-specificity of repeats, whereas high rate of mutation accumulation together with high frequency of gene conversion and amplification, may lead to a "substitution" of a repeat variant by another one. CLsat repeats are characterized by a high mutation accumulation rate and a high homogenization level, which is indirectly indicated by the formation of three CLsat subfamilies and species-specificity of these repeats within each subfamily.

Comparison of Subfamilies of CLsat and Agi160 Repeats and a probably Origin of satDNA in Lizards from the Genus Darevskia

Agi160 SatDNA, which we have found in the genome of species from the genus *Lacerta* s. str., was not revealed in other species of the family Lacertidae by means of hybridization technique [Ciobanu, unpublished data]. In the present work, we have shown that monomers Agi160, in contrast to CLsat, exhibit higher variability of the repeated unit due to different number (two to four) of internal 10-nucleotide repeats. The sequences of different CLsat and Agi160 monomers have 74% similarity over a long region constituting a little less that half of the monomer; similarity of complete sequences varies from 50 to 69% depending on the monomer.

In the present study, we compared consensus sequences of each of the CLsat subfamilies with one of the most similar to CLsat monomers of the Agi160 family, Agi160M2(AY184825) (Fig. 3). This monomer has highest similarity with species consensus monomers of CLsat, as compared to nine other Agi160 monomers examined [Ciobanu, unpublished data]. The similarity between the general consensuses of CLsatI, CLsatII, and CLsatIII is about 15% higher than their similarity to Agi160M2. At the same time, the similarity between

Table 3. Numbers of nucleotide substitutions in monomers of species consensuses within CLsat subfamilies

Subfamily	Transitions			Transversions				
	T ⇄ C	C ⇄ A	Total	C ⇄ G	A ⇄ C	A ⇄ T	G ⇄ T	total
CLsat I	10	7	17	4	3	7	8	22
CLsat II	1 + 3	1 + 4	2 + 7	2 + 4	2 + 1	5 + 2	2 + 2	11 + 9
CLsat III	12	9	21	8	8	2	9	27

Note: Figures after the plus sign indicate the number of nucleotide substitutions in the strongly diverged monomer pra2 (CLsatII).

the Agi160M2 monomer and the general consensus of the CLsat family is lower than the similarity between the consensuses of the CLsat subfamilies and the general family consensus by about the same value (Fig. 3). Thus, the similarity between the Agi160M2 monomer and the general CLsat consensus is close to the similarity between the CLsatI, CLsatII, and CLsatIII consensuses, which is an additional evidence for the common origin of Agi160 and CLsat DNA. Nevertheless, the remaining nine Agi160 monomers more strongly diverged from the common Agi160/CLsat ancestor, which is explained mainly by relatively recent amplification of the 10-bp repeats and stronger divergence of the A-rich sequence located downstream of these repeats [Ciobanu, unpublished data].

Comparison of the distribution of maximum Agi160–CLsat similarity regions with the internal structure of the Agi160 monomer [Ciobanu, unpublished data] and CLsat (Fig. 4a) revealed a correlation between the distribution of the maximum similarity regions and "subdivision" of monomers into two similar halves (Fig. 4b). This figure shows an Agi160(AY184833) monomer containing four short (10-bp) internal repeats (underlined). This region coincides with the beginning of the A' region as well as with the beginning and the most similar part of the internal repeat. As in other Agi160 monomers [Ciobanu, unpublished data] showing variation in the 10-bp repeat number, no such coincidence was observed, monomers with two and three 10-bp repeats have even lower similarity between A and A'. This indicates stronger and, probably, secondary divergence of the A' region in the Agi160 monomers. Although this "hidden repeat" is highly degenerate, no other possible repeats of the same size with the same or higher similarity have been found.

The correspondence of position of the most conserved region for Agi160 and CLsat (74% for about 70 bp) with one part of the "repeat" suggests that the divergence between Agi160 and CLsat has occurred after the amplification of the precursor of the "74% similarity zone" about 70 bp in size (Fig. 4c). This level of divergence of these two satDNA families compared with the divergence of *Darevskia* and *Lacerta* s. str. species suggests that the extant Agi160 and CLsat forms did not appear de novo after the divergence of the genera but have evolved in concert with the speciation.

Hybridization Analysis of satDNA of the CLsat Family

In the present study, we have examined by means of DNA hybridization the content and distribution of the three variants of CLsat repeats in all species of the genus *Darevskia* and in members of other genera of the family Lacertidae. First, hybridization in mild conditions (58°C) of three CLsat subfamilies with the genomic DNA of the closest genera *Lacerta* s. str., *Podarcis*, and *Zootoca*, as well as more distant genera *Eremias*, *Ophisops*, and *Gallotia* was shown to be lacking (data not presented).

Earlier, Rudykh *et al.* [16] reported the results of hybridization of genomic DNA from a fewer species with the CLsatI monomer isolated from the *D. saxicola darevskii* genome. In the present work we repeated, verified and expanded these data by including in the study probes of all the three CLsat subfamilies and the species that had not been examined from this viewpoint (*D. chlorogaster, D. dryada, D. clarkorum, D. parvula, D. alpina, D. caucasica, D. saxicola szczerbaki*) and employing a more precise estimation of hybridization signal (see Materials and Methods). The former two species of this list are regarded as crucial in speciation of Caucasian lacertids from the genus *Darevskia* [8, 21, 22]. The examination of the third CLsat subfamily, which had been discovered by us and is described here for the first time, yielded new significant data on genetic relatedness of the species as well as on correlation of satDNA evolution with speciation.

Figure 5 presents histograms of the content of each of the three CLsat subfamilies in the genomes of bisexual *Darevskia* species, obtained by Southern hybridization. This largest group contains five species of rock lizards, in whose genomes the CLsatI repeat prevails while CLsatII is virtually absent (Fig. 5a). The second group join species whose genomes mainly have two types of repeats, CLsatI and CLsatIII, at the background of a very low CLsatII content (Fig. 5b). The third group comprises lizards having all three subfamilies in different proportions. This group (Fig. 5c) includes *D. dryada* preserved in a tertiary refugium, which is supposedly among the most ancient *Darevskia* species [21]. Finally, the fourth group consists of three and presumably, evolutionarily older species *D. chlorogaster, D. derjugini*, and *D. particola* [8, 22], in the genome of each of which one of the three subfamilies (CLsatI or CLsatII) prevails as well as *D. parvula*, which is similar to the former species in this respect (CLsatIII) (Fig. 5d).

DISCUSSION

We used the obtained data to find an association between phylogeny of the *Darevskia* species and molecular evolution of their satDNA by examining specificity of different CLsat variants for particular clusters of the species. Based on this, we propose a possible scenario of speciation (phylogeny) in the genus *Darevskia.*

Taxonomists believe that rock lizard species are very young, their gradual formation lasting no more than 10 000–20 000 years (within precision limits of estimation of the time of the last Pleistocene (quaternary) glaciation in the Caucasus) [23]. From this viewpoint, the terrestrial species are older than the rock ones. The former species survived the glaciation of the Caucasus mountainous regions and cooling of the climate in warm and humid refugia, which preserved the tertiary flora until present. These refugia were situated in valleys of the rivers that descended to the Black and Caspian seas: in the Kolkhida lowland and in the Kurinskaya hollow (Lenkoran') [24], divided by the water-

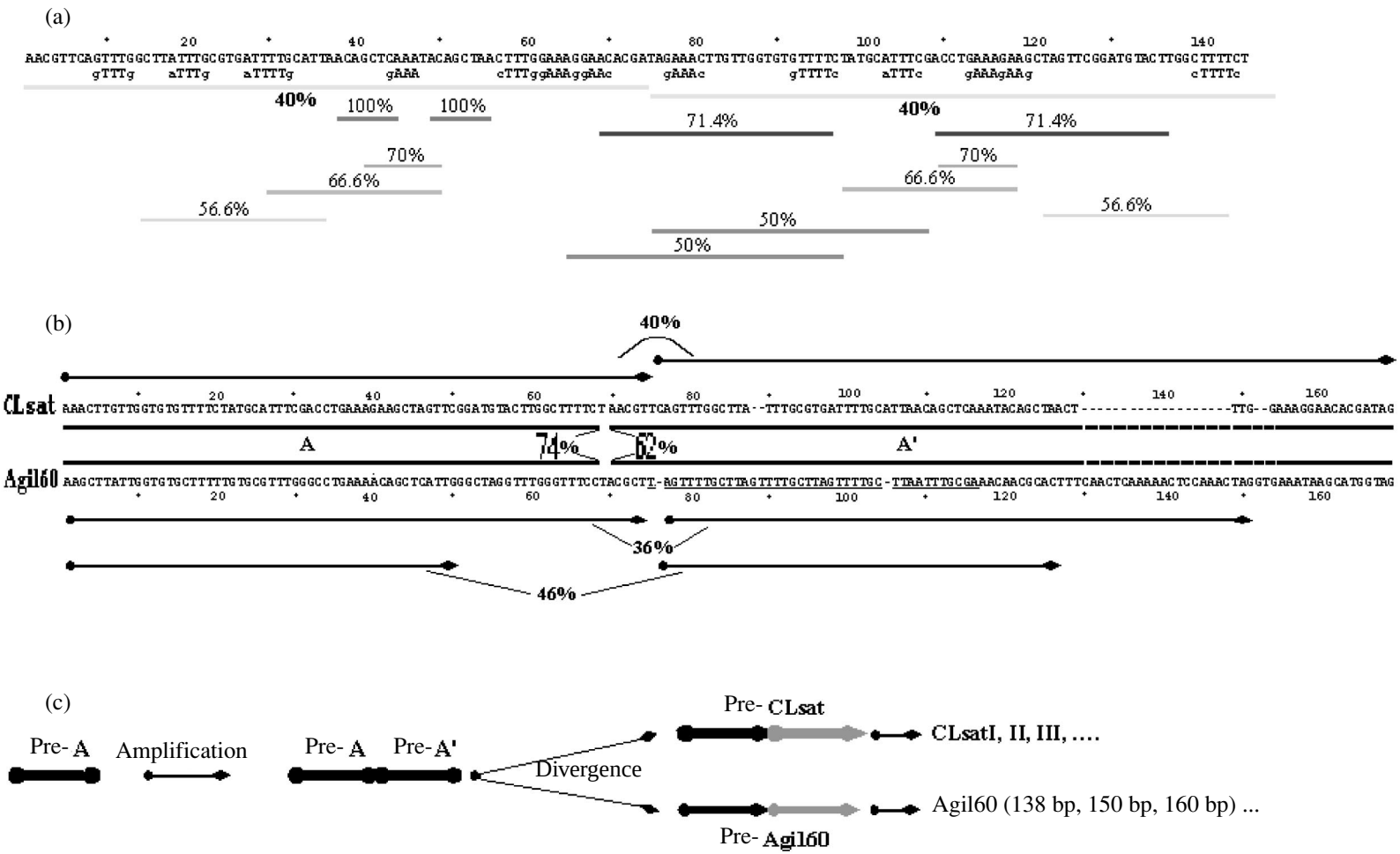

Fig. 4. Comparison of monomers CLsat of *Darevskia* and Agi160 of *Lacerta* s. str. A probable evolution pathway of satDNA CLsat and Agi160. (a) Internal structure of CLsat monomers. Tracks with flanking nucleotides are shown below the monomer sequence. Repeats are shown by lines of the same length and color. Similarity values are given in percent. (b) Correspondence between the similarity of CLsat consensus monomers and individual monomer Agi160 (AY184833) with their internal structure. The monomer sequences are presented in low-case letters. The longest "internal repeats" of the monomer are indicated by arrows. Bold lines between monomers indicate two regions of different similarity ("conserved," A; "variable," A') of CLsat and Agi160 monomers. Similarity values are given in percent. (c) Total scheme of molecular evolution of CLsat and Agi160 satDNA. Bold black segments indicate precursor regions of the "conserved" (A) and "variable" (A') zones. The "variable" zone in the precursor monomer of CLsat and Agi160 (Pre-Slsat and PreAgi160) is hatched (as the most diverged zone). This figure presents monomer Agi160 containing four short (10-bp) internal repeats, which are underlined in the sequence. This region coincides with the beginning of the A' sequence. In other Agi160 monomers (Ciobanu, unpublished data) this zone contains two or three 10-bp repeats. In these monomers, similarity between A and A' is even higher, indicating stronger (probably, secondary) divergence of A' in monomers Agi160.

shed Suramskii ridge. Kolkhida valley is still inhabited by *D. derjugini, D. parvula, D.dryada*, and *D. clarkorum*, and Lenkoran' hollow, by *D. chlorogaster* and several populations of *D. raddei*. Hence, we can assume that the species preserved in refugia could serve as speciation centers for younger modern species. Consider the eastern refugium, which harbors an ancient population of *D. chlorogaster* that contains virtually one variant of satellite CLsatI (Fig. 5e). This repeat variant may represent the oldest, "concertedly" preserved form. Exactly this form could be inherited by populations that dispersed from the eastern refugium. If this is so, the modern young rock species (*D. raddei, D. rudis, D. nairensis, D. alpina, D. saxicola*), which also have practically only CLsatI (Fig. 5a), could have a common predecessor with *D. chlorogaster.* According to the number of mutations in the CLsatI monomer sequences, all of the above species are approximately equidistant from *D. chlorogaster* (12–19 substitutions), except a closer species *D. nairensis* (four substitutions). The terrestrial species *D. derjugini* (15 substitutions) can be assigned to the same group. In this case, the CLsatI-carrying populations could spread northward, along the Caspian Sea coast, by southern (*D. derjugini*, subspecies *derjugini, orlowae, barani*) and northern (*D. derjugini*, subspecies *silvatica*: *D. saxicola* and its subspecies: *D. alpina*) slopes of the Caucasus ridge. The other CLsatI-carrying populations could disperse westward and southwestward forming species *D. valentini, D. rudis, D. raddei*, which are currently distributed over huge territories of Armenia around Sevan Lake, Van Lake in Turkey, and Urmia Lake in Iran.

The second group of species that we distinguished, which contained CLsatI and CLsatII in approximately equal proportions (*D. valentini, D. portschinskii, D. lindholmi*, Fig. 5b) could arise by acquiring or explosively amplifying the CLsatIII repeat. Species *D. valentini* and *D. portschinskii* practically do not differ in the CLsatI monomer sequence from *D. rudis*, which also carries only CLsatI. This means that these species have a common close precursor of CLsatI; they cluster in a closely related group (Fig. 6). These three species cluster together also on the basis of other molecular markers, which were examined previously: taxonorint [11], allozyme [25], and mtDNA [26]. Apparently, they stem from a common ancestor and have diverged only recently, after the CLsatIII amplification. Unfortunately, no data are available yet on the sequences of these monomers in the species considered. The third species, *D. lindholmi* (until recently regarded as a subspecies of *D. saxicola*), is similar to *D. valentinini* and *D. portschinskii* in the predominance of CLsatI and CLsatIII (Fig. 5b) but differs from these species in the sequence of the monomer of this subfamily far more (four to ten substitutions) than they differ from each other (one or two substitutions), which could be caused by an earlier origin of *D. lindholmi* from a common ancestor. The species *D. lindholmi* is a Crimean endemic, from the time of its description considered a subspecies of rock lizard *D. laxicola* [8]. Later, molecular data [11, 13] have shown that the relatedness of *D. lindholmi* with subspecies of *D. saxicola* is substantially more distant than the relatedness of the four Caucasian subspecies with one another. Recently, the Crimean subspecies was assigned the rank of a species [7], which was also based on our data [11, 13].

The species carrying all three monomers (Fig. 5c) tend to the western refugium (Kolkhida) and the oldest (and, in view of taxonomists, relict) species *D. dryada* that inhabits it. This species group (*dryada, clarkorum, mixta, caucasica, daghestanica*) may have arisen from an ancient form of the relict species of its ancestor. In fact, this species dwells in lowlands (up to 300 m above sea level), its population is small and depressed. Species *D. clarkorum*, which is closest to *D. dryada* morphologically and in CLsatI and CLsatII monomer sequences, inhabits mountainous regions up to 1700 m above sea level and is currently undergoing an explosive population growth; its number is about 50 times higher than that of *D. dryada* [21]. The genetic similarity of *D. dryada, D. clarkorum*, and *D. mixta* is truly striking: they differ only by zero to two nucleotides in the CLsatI and CLsatII monomer sequences. The differences of *D. caucasica* and *D. daghestanica* with these species is somewhat greater constituting four to five substitutions, but this also testifies to their close relatedness (if compared to the differences in monomer sequences between more distance species, which can reach 40–45 substitutions). In other words, the similarity between the species from the western refugium group is exceptionally high. This can mean that all of the three species examined share a common ancestor, but their separation into three geographically isolated populations have occurred so recently as it is required for maintaining absolute genetic similarity with regard to the examined satellites. If this is true, then the morphologically related species *caucasica* and *daghestanica* branched off earlier than the separation of the three above species. It would be reasonable to consider these three species (*dryada, clarkorum*, and *mixta*) as morphs of a single ancient species, since they in essence do not differ genetically. This close relatedness is formally reflected in the tree (Fig. 6). Note that in light of this hypothesis inferred from new data, the suggestion of Darevsky and Tuniev [21] on the origin of the *saxicola* species group directly from the *D. dryada* predecessor seems unlikely.

The appearance of additional satellite variants in speciation may be related to amplification (for some reasons) of rare monomer variants followed by their homogenization with preservation of the previous variant. However, there is another explanation, according to which some species of the genus could arise via interspecific hybridization (in other terms, reticulate speciation [12]) of species carrying a single variant of CLsat (Fig. 5d) producing species with a mixed satellite content. In that case, species of the *dryada* group (*clarkorum, mixta, caucasica, daghestanica*) and *valentini*

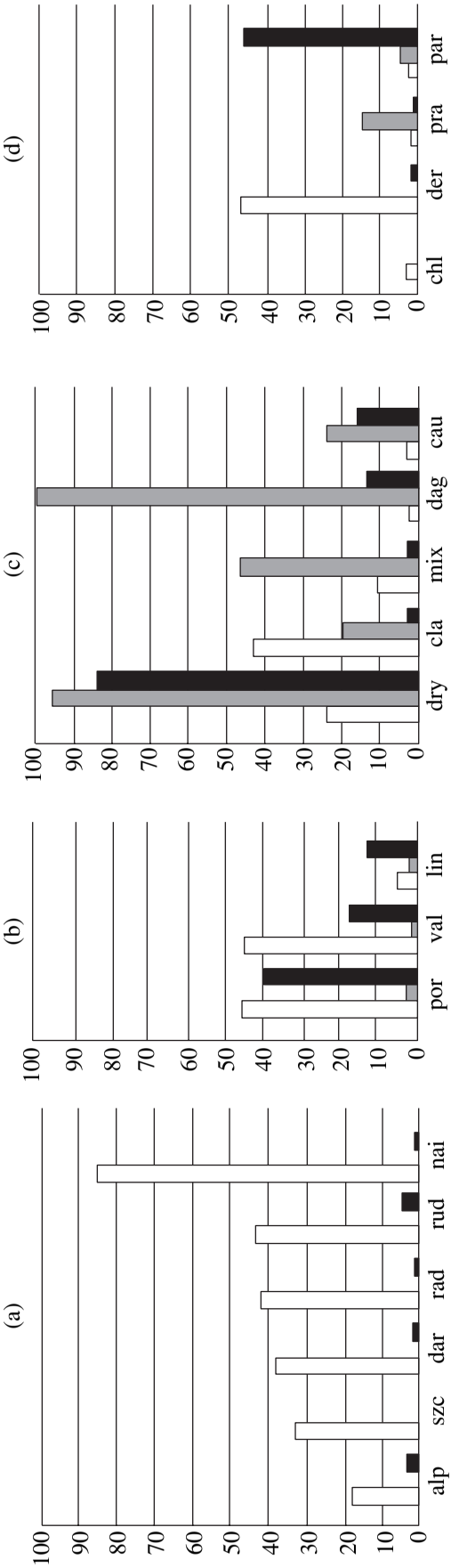

Fig. 5. Distribution of relative content of CLsat subfamilies in bisexual *Darevskia* species. Open, hatched, and solid histograms indicate CLsatI, CLsatII, and CLsatIII, respectively. Ordinate: relative content of CLsat, percent; the satDNA content is transformed in percent from intensities of hybridization signals (see Matetrials and Methods). Species names are abbreviated (see Table 1). (a) Group of rock lizards containing mainly CLsatI; (b) group of rock lizards containing mainly CLsatI and CLsatIII; (c) species of the *D. dryada* group containing mostly CLsatII along with CLsatI and CLsatIII; (d) separate species, each of which contains mainly one of the CLsat types.

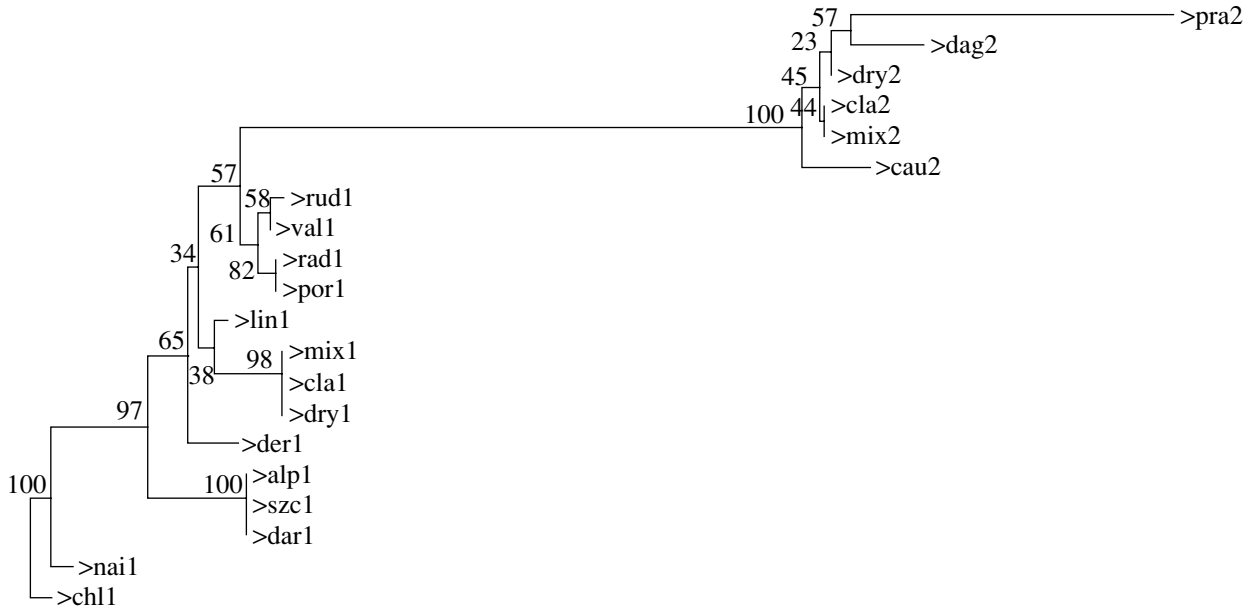

Fig. 6. Cladogram of the species consensus sequences of subfamilies CLsatI and CLsatII constructed by the neighbor-joining (NJ) method and algorithms of Gin and Nei [32] and Kimura [33]. Consensus species sequences of monomers of subfamilies CLsatI and CLsatII of the bisexual species presented in Figs. 1 and 2 are used. Numerals indicate bootstrap values. The given rooting is arbitrary and does not reflect the hierarchy of phylogenetic events. Bootstrap values exceeding 50% show clearly discernible clades; lower bootstrap values indicate sequences of closely related species.

(portschinskii, lindholmi) would be considered as descendants of a hybrid predecessor(s).

The idea of reticulate speciation in bisexual species in unpopular among zoologists, regardless of numerous (especially botanical) facts that this speciation mode could occur in evolution. These data are thoroughly discussed by Arnold [27] and Barton [28]; the latter author considers adaptive traits and evolutionary perspectives of model hybrid systems. The facts of hybrid origin of diploid and polyploid populations of agamous species are well known in fish, amphibians, and reptiles (including lizards examined in the present work) and conform to the views of Astaurov on the possibility of this type of speciation in nature [29]. In what follows, we present some data showing a potential possibility of this speciation way in Caucasian lacertids and favoring our assumption that we have observed exactly this mode of appearance of some hybrid diploid species.

It is known that the extant Caucasian lizard species in many cases easily intercross at the margins of their ranges forming hybrid zones. For instance, hybrids *D. clarkorum* × *D. rudis*, *D. dryada* × *D. rudis*, and *D. derjugini* with *D. parvula*, *D. mixta*, *D. saxicola*, and *D. caucasica* that possess the properties of both their parents are found in nature [8, 30] and sometimes produced in laboratory [8]. Animals from the hybrid populations exhibit very high variability. In hybrid zones of *D. derjugini* × *D. parvula*, one or both parental forms were "dissolved" in the abundance of hybrid animals [30]. This suggests that the hybrids are fertile, which was actually shown in a number of cases [8]. In this connection, note also that the authors who described *Lacerta* (now *Darevskia*) *mixta* suggested its hybrid origin, which is reflected in the name of this species [8]. Later this view was rejected [31] on the basis of the data that we regard as rather inconclusive. However, recently other authors have restored this hypothesis after detecting in the modern *D. mixta* similarity with *D. alpina* at mtDNA genes [26].

Thus, the hypothesis on a hybrid origin of the precursor species of the *dryada* group is in essence plausible and can be tested by examining the hybrids and the parental species using specific molecular markers.

Both hypotheses can be tested by various means. For instance, it is of interest to trace the dynamics of molecular DNA markers in populations and subspecies of *D. derjugini* along the putative pathway of its spreading: particularly, in populations of the eastern and western coasts and in marginal populations. It would also be of interest to check the CLsat inheritance by experimental hybridization of species that supposedly carry the original forms of CLsat subfamilies as well as using other molecular markers. Another promising line of research is tracing satellite similarity in major limited populations of such "strong" and widespread species as *D. praticola*, *D. derjugini*, and *D. raddei.*

The results presented in this study suggest that satDNA evolution is strongly associated with speciation. Studying satDNA can serve as a powerful tool of investigating species divergence irrespective of their causal relationships.

APPENDIX

The accession numbers of used sequences (GenBank database):

chl1 AY262941-9; ros1 AY262950-66; szc1 AY262967-71; dry1 AY262972-76; cla1 AY262977-81; mix1 AY262982-89; lin1 AY256930-43; dry2 AY262990-96; ros3 AY262997-99; der31 AY263000-1; der32 AY263002; lin3 AY256944-53; arm1 7687993; dah 1 gi7687991; val1 gi7688051; por1 gi7688047; rad1 gi7688046; lin1 gi7688015; rud1 gi7688049; rud1 gi7688048; alp1 gi7687990; dar1 gi3087812; por1 gi7688047; der1 gi7688014; uni1 gi7688050; nai1 gi7688045; cau2 gi7687994; dah2 gi7687992; cla2 gi18073593; mix2 gi7688052; arm2 gi5457399; dag2 gi5457400; pra2 gi5457401; dry3 gi18073592; par3 gi18073594; mix3 gi18073595; lin3 gi7688016.

ACKNOWLEDGMENTS

We thank F.D. Danielyan, B.S. Tuniev, N.G. Kan, M.A. Arakelyan, K.D. Mil'to, N.L. Ryabinina, I.A. Martirosyan, S.A. Kosushkin, D.M. Ryabinin, I.A. Rudykh for assistance in collecting the field material and N.S. Vasetskii. V.F. Orlova, A.A. Bannikova, N.L. Ryabinina, I.A. Rudykh, and S.A. Kosushkin for constant and fruitful discussion of our work.

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

REFERENCES

1. Charlesworth, B., Sniegowski, P., and Stephan, W., The Evolutionary Dynamics of Repetitive DNA in Eukaryotes, *Nature*, 1994, vol. 371, pp. 215–220.
2. Grechko, V.V., Molecular DNA Markers in Phylogenetic and Taxonomic Studies, *Genetika* (Moscow), 2002, vol. 38, no. 8, pp. 1013–1033.
3. Miklos, G.L.G., Localized Highly Repetitive DNA Sequences in Vertebrate and Invertebrate Genomes, *Molecular Evolutionary Genetics*, McIntyre, R.J., Ed., New York: Plenum, 1985, pp. 241–321.
4. Trifonov, E.N., Tuning Function of Tandemly Repeating Sequences: A Molecular Device for Fast Adaptation, *Proc. Int. Conf. Evol. Genomics*, Costa Rica, 1999, p. 74.
5. Dover, G.A., Brown, S., Coen, E., *et al.*, The Dynamics of Genome Evolution and Species Differentiation, *Genome Evolution*, Dover, G.A. and Flavel, R.B., Eds., London: Academic, 1982, pp. 351–367.
6. Arnold, E.N., 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.*, 1989, vol. 55, pp. 209–257.
7. Arribas, O.J., 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.*, 1999, vol. 6, pp. 1–22.
8. Darevsky, I.S., *Skal'nye yashcheritsy Kavkaza* (Caucasian Rock Lizards), Leningrad: Nauka, 1967.
9. Darevsky, I.S., Lizard of the Genus *Lacerta* from the Middle Sarmatian Lake Deposits of the Northern Caucasus, *Proc. Zool. Inst. Leningrad*, 1990, vol. 207, pp. 139–142.
10. Mayer, W. and Benyr, G., Albumin Evolution und Phylogenese in der Familie Lacertidae, *Ann. Naturhist. Mus. Wien*, 1994, vol. 96, pp. 621–648.
11. Grechko, V.V., Ryabinin, D.M., Fedorova, L.V., *et al.*, Molecular Genetic Classification and Phylogenetic Relationships of Several Lizard Species of the Family Lacertidae as Inferred from Specific Distribution Patterns of Restriction Sites in Repetitive DNA (Taxonoprinting), *Mol. Biol.* (Moscow), 1998, vol. 32, no. 1, pp. 172–183.
12. Ciobanu, D., Rudykh, I.A., Ryabinina, N.L., *et al.*, Reticulate Evolution of Parthenogenetic Rock Lizard Species of the Family Lacertidae: Inheritance of the CLsat Tandem Repeat and Anonymous RAPD Markers, *Mol. Biol.* (Moscow), 2002, vol. 36, no. 2, pp. 296–306.
13. Ryabinina, N.L., Grechko, V.V., and Darevsky, I.S., RAPD Polymorphism in Lizard Populations of the Family Lacertidae, *Genetika* (Moscow), 1998, vol. 34, no. 1, pp. 1661–1667.
14. Capriglione, T., Satellite DNA and Phylogeny of Lacertid Lizards, *Sci. Herpetol.*, 1995, pp. 68–70.
15. Capriglione, T., Repetitive DNA as a Tool to Study the Phylogeny of Cold-Blooded Vertebrates, *Chromos. Today*, 2000, vol. 13, pp. 183–194.
16. Rudykh, I.A., Grechko, V.V., Kramerov, D.A., and Darevsky, I.S., Genomic Distribution of the *Hin*dIII Repeat Reflects the Phylogenetic Relationships of Caucasian Lizard of the Genus *Lacerta*, *Dokl. Ros. Akad. Nauk*, 1999, vol. 367, pp. 563–566.
17. Rudykh, I.A., Grechko, V.V., Ciobanu, D.G., *et al.*, Variation of satDNA Restriction Sites as a Molecular Basis of Taxonoprinting (as Demonstrated with the Example of Caucasian Rock Lizard), *Genetika* (Moscow), 2002, vol. 38, no. 8, pp. 1110–1114.
18. Blin, N. and Stafford, D.W., A General Method for Isolation of High-Molecular-Weight DNA from Eukaryotes, *Nucleic Acids Res.*, 1976, vol. 3, pp. 2303–2308.
19. Sambrook, J., Fritsch, E.F., and Maniatis, T., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York: Cold Spring Harbor Lab., 1989.
20. Li, Y., Lee, C., Hsu, T.H., *et al.*, Direct Visualization of the Genomic Distribution and Organization of Two Cervid Centromeric Satellite DNA Families, *Cytogenet. Cell Genet.*, 2000, vol. 89, pp. 192–198.
21. Darevsky, I.S. and Tuniev, B.S., A New Lizard Species from the *Lacerta saxicola* Group, *Lacerta dryada* sp. nov. (Sauria, Lacertidae), and Some Comments Relative to *Lacerta clarkorum* Darevsky et Vedmederja, 1977, *Russ. J. Herpetol.*, 1997, vol. 4, pp. 1–7.
22. Tuniev, B.S., On the Mediterranean Influence on the Formation of Herpetofauna of the Caucasian Isthmus and Its Main Xerophylous Refugia, *Russ. J. Herpetol.*, 1995, vol. 2, pp. 95–119.
23. Dumitrashko, N.V., Milanovskii, E.E., Bal'yan, S.P., and Saadyan, Yu.V., Ancient Glaciation of the Caucasus, in *Geologiya chetvertichnogo perioda (pleistotsen)* (Geol-

ogy of the Quaternary Period (The Pleistocene)), Yerevan, 1977, pp. 26–33.

24. Gvozdetskii, N.A., *Kavkaz. Ocherk prirody* (The Caucasus: An Essay on the Nature), Moscow: GIGL, 1963.
25. Murphy, R.W., Darevsky, I.S., MacCulloch, R.D., Fu, J., and Kupriyanova, L., Evolution of the Bisexual Species of Caucasian Rock Lizards: A Phylogenetic Evaluation of Allozyme Data, *Russ. J. Herpetol.*, 1996, vol. 3, pp. 18–31.
26. Fu, J. and Murphy, R.W., Toward the Phylogeny of Caucasian Rock Lizards: Implications from Mitochondrial DNA Gene Sequences (Reptilia: Lacertidae), *Zool. J. Linnaean Soc.*, 1997, vol. 121, pp. 463–477.
27. Arnold, M.L., *Natural Hybridization and Evolution*, New York: Oxford Univ. Press, 1997.
28. Barton, N.H., The Role of Hybridization in Evolution, *Mol. Ecol.*, 2001, vol. 10, pp. 551–568.
29. Astaurov, B.L., Heredity and Development, *Aktual'nye voprosy sovremennoi genetiki* (Topical Problems of Modern Genetics), Alikhanyan, S.I., Ed., Moscow: Mosk. Gos. Univ., 1966, pp. 65–114.
30. Orlova, V.F., Geographic Variation and Description of the Lectotype of Artvinian Lizard *Lacerta derjugini* Nik., *Sb. Tr. Zool. Muzeya Mosk. Gos. Univ.*, 1979, vol. 17, pp. 188–203.
31. Uzzell, I. and Darevsky, I.S., Electrophoretic Examination of *Lacerta mixta*, a Possible Hybrid Species, *Russ. J. Herpetol.*, 1973, vol. 7, pp. 11–15.
32. Jin, L. and Nei, M., Limitations of the Evolutionary Parsimony Method of Phylogenetic Analysis, *Mol. Biol. Evol.*, 1990, vol. 7, pp. 82–102.
33. Kimura, M., A Simple Method for Estimating Evolutionary Rates of Base Substitutions through Comparative Studies of Nucleotide Sequences, *J. Mol. Evol.*, 1980, vol. 16, pp. 111–120.