Вавиловский журнал генетики и селекции. 2023;27(4):349-356
DOI 10.18699/VJGB-23-42

Comparison of the evolutionary patterns of DNA repeats in ancient and young invertebrate species flocks of Lake Baikal

Wang Yuxiang[1], T.E. Peretolchina[1], E.V. Romanova[1], D.Y. Sherbakov[1, 2, 3] ✉

[1] Limnological institute of the Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
[3] Irkutsk State University, Irkutsk, Russia
✉ sherb@lin.irk.ru

Abstract. DNA repeat composition of low coverage (0.1–0.5) genomic libraries of four amphipods species endemic to Lake Baikal (East Siberia) and four endemic gastropod species of the fam. Baicaliidae have been compared to each other. In order to do so, a neighbor joining tree was inferred for each quartet of species (amphipods and mollusks) based on the ratio of repeat classes shared in each pair of species. The topology of this tree was compared to the phylogenies inferred for the same species from the concatenated protein-coding mitochondrial nucleotide sequences. In all species analyzed, the fraction of DNA repeats involved circa half of the genome. In relatively more ancient amphipods (most recent common ancestor, MRCA, existed approximately sixty millions years ago), the most abundant were species-specific repeats, while in much younger Baicaliidae (MRCA equal to ca. three millions years) most of the DNA repeats were shared among all four species. If the presence/absence of a repeat is regarded as a separate independent trait, and the ratio of shared to total numbers of repeats in a species pair is used as the measure of distance, the topology of the NJ tree is the same as the quartet phylogeny inferred for the mitogenomes protein coding nucleotide sequences. Meanwhile, in each group of species, a substantial number of repeats were detected pointing to the possibility of non-neutral evolution or a horizontal transfer between species occupying the same biotope. These repeats were shared by non-sister groups while being absent in the sister genomes. On the other hand, in such cases some traits of ecological significance were also shared.

Key words: DNA repeats; Lake Baikal; phylogeny; Baicaliidae; amphipods; evolution of repeats; repeatome.

For citation: Wang Yuxiang, Peretolchina T.E., Romanova E.V., Sherbakov D.Y. Comparison of the evolutionary patterns of DNA repeats in ancient and young invertebrate species flocks of Lake Baikal. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(4):349-356. DOI 10.18699/VJGB-23-42

Сравнение эволюционных паттернов ДНК-повторов у представителей древних и молодых букетов видов из озера Байкал

Ван Юйсян[1], Т.Е. Перетолчина[1], Е.В. Романова[1], Д.Ю. Щербаков[1, 2, 3] ✉

[1] Лимнологический институт Сибирского отделения Российской академии наук, Иркутск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
[3] Иркутский государственный университет, Иркутск, Россия
✉ sherb@lin.irk.ru

Аннотация. Исследованы ДНК повторы, присутствующие в геномных библиотеках с низким покрытием (0.1–0.5) четырех видов амфипод, эндемичных для озера Байкал (Восточная Сибирь), и четырех эндемичных видов брюхоногих моллюсков семейства Baicaliidae. Для этого были построены деревья методом объединения ближайших соседей для каждого квартета видов (амфиподы и моллюски) на основе соотношения повторяющихся классов, общих для каждой пары видов. Топология этих деревьев была сопоставлена с филогениями, полученными для тех же видов на основе сцепленных белок-кодирующих митохондриальных нуклеотидных последовательностей. У всех проанализированных видов в долю повторов ДНК вовлечено около половины генома. У относительно более древних амфипод (самый последний общий предок, MRCA, существовал приблизительно шестьдесят миллионов лет назад) наиболее распространенными были видоспецифичные повторы, тогда как у гораздо более молодых байкалиид (MRCA приблизительно равен трем миллионам лет) большинство повторов ДНК были общими для всех четырех видов. Если наличие/отсутствие повтора рассматривать как отдельный независимый признак, а отношение общего числа повторов в паре видов использовать в качестве меры расстояния, топология дерева NJ такая же, как и филогения квартета, выведенная для белков митогеномов, кодирующих нуклеотидные последовательности. Между тем в каждой группе видов было обнаружено значительное количество повторов, указывающих на возможность ненейтральной эволюции или горизонтального переноса между видами, занимающими один и тот же биотоп. Эти повторы были общими для неродственных групп, но отсутствовали в сестринских геномах. С другой стороны, в таких случаях некоторые черты, имеющие экологическое значение, также были общими.

Ключевые слова: повторы ДНК; озеро Байкал; филогения; Baicaliidae; амфиподы; эволюция повторов; репитом.

Introduction

In Metazoa, approximately half of all genomic DNA is made up of repeated DNA sequences, which are otherwise called "non-genic DNA" (Cavalier-Smith, Beaton, 1999; Bird et al., 2006) or repeatome (Titievsky et al., 2021). The already known functions of this fraction of the genome are very diverse. Most of it is satellite DNA (Biscotti et al., 2015; Silva et al., 2019; Thakur et al., 2021). A significant proportion of DNA repeats account for mobile elements belonging to different classes. There is evidence that highly repeated mobile elements may play a certain role in the regulation of genetic activity (see for example (Rocha et al., 2022)), their distribution must also be taken into account in the epigenetic analysis (Lerat et al., 2019). It is important to note that evidence is accumulating about the important role that repeated mobile elements may play in horizontal gene transfer between phylogenetically distant species (Ahmad et al., 2021; Athanasouli, Rödelsperger, 2022; Kejnovsky, Jedlicka, 2022). Dodsworth et al. (2015) have shown that a set of repeated elements contains a significant phylogenetic signal. They also noted the presence of a repeat fraction, which was inconsistent with the phylogenies inferred from individual nucleotide sequences and from the repeats, but treated this fraction more like an obstacle rather than an interesting phenomenon. It has been shown that horizontal gene transfer is a common mechanism of transmission of traits involved in adaptation processes in bacteria (Lee et al., 2022) and fungi (Steensels et al., 2021). Recently, there has been more and more evidence that similar mechanisms are likely to be involved in the adaptive evolution of Metazoa (Boto, 2014; Chen et al., 2017; Ahmad et al., 2021; Li et al., 2022). Since then their work though the main focus of studies of the repeated DNA shifted mostly towards their potential structural role and was performed mostly of plant models (see for example (Titievsky et al., 2021)).

Here we apply the repeats analysis to the two species flocks of Baikalian invertebrates. Genetic studies of invertebrates from Lake Baikal allow to unravel many problems of their evolutionary history (Romanova et al., 2016; Peretolchina et al., 2020), mechanisms of speciation (Naumenko et al., 2017; Gurkov et al., 2019; Drozdova et al., 2022) and adaptation (Lipaeva et al., 2021), diversity and conservation (Butina et al., 2019; Yakhnenko, Itskovich, 2020). We use the features of the evolution of two species flocks of endemic Baikal invertebrates – amphipods (Bazikalova, 1945; Kamaltynov, 1999; Takhteev, 2019) and gastropods of the family Baicaliidae (Sitnikova et al., 2001; Hausdorf et al., 2003; Peretolchina et al., 2020) to study the evolution of the maximum diversity of repeats in their genomes. These two groups are attractive models for this kind of research for the following reasons:

1. Both groups of organisms have been well and comprehensively studied (see (Kozhov, 1963)).
2. Both groups evolved within Baikal, therefore all possible genome transformations were minimally, if at all, dependent on the introduction of genetic information from outside the ecosystem, and speciation processes occurred mainly by sympatric mechanisms.
3. The evolutionary histories of amphipods and Baicaliidae in Baikal are fundamentally different: if the former is represented by at least two branches that independently penetrated Baikal, the common ancestor of which existed at least 60 million years ago, then the maximum age of the common ancestor of Baikal species is at least 3 million years (Sherbakov, 1999; Mats et al., 2011).

Thus, the above-mentioned properties of evolutionary histories allow us to conduct a comparative analysis of sets of DNA repeats in two species-rich groups of invertebrates and assess the potential benefits of such a comparison for a deeper understanding of the evolutionary mechanisms that have shaped their modern diversity.

Materials and methods

In this work, the genomic libraries of gastropods were used: *Baicalia turriformis*, *Maackia herderiana*, *Korotnewia korotnewi* and *Godlewskia godlewskia*, the collection of samples and genome-wide sequencing of which is described in Peretolchina et al. (2020). Obtaining the genomic libraries of amphipods *Acanthogammarus victorii*, *Brachyuropus grewingkii*, *Garjajewia cabanisi* and *Macrohectopus branickii* is described in (Romanova et al., 2016).

Random sets of reads were prepared from the source libraries using Seqtk-1.3 (r106) (Shen et al., 2016) on the Galaxy (Jalili et al., 2020) platform. The size of a library was set to 5×10^5 reads. The search for repeating genetic elements was performed using the RepeatExplorer (Novák et al., 2013) pipeline implemented on the Galaxy platform.

Quality control and library filtering were performed using the standard Galaxy FastQC (de Sena Brandine, Smith, 2019) tool. Cluster analysis requires files containing sequences of reads in FASTA format as input data.

The search for repeated sequences was performed using RepeatExplorer2 clustering.

Launch Parameters:

Paired-end reads True, Read sampling false, Sample size 0,
Select taxon and protein domain database version (REXdb) Viridiplantae version 3.0,
Advanced options false,
Select queue basic_fast_queue,
Modify parameters (optional) –l
select=1:ncpus=10:mem=32gb:scratch_local=50gb –l walltime=48:00:00 –q elixirre@pbs.elixir-czech.cz –v TAREAN_MAX_MEM=4000000,TAREAN_CPU=4

The search for repeats was limited to those that occur more often than 0.01 % of the input reads. In addition to satellite repeats, the output data contains LTR-retrotrans-posons, 45S, 5S rDNA and all other repeats, the number of which exceeds the threshold value.

Comparative analysis of the composition of repeats was performed using a set of original scripts in Python 3.10 and Biopython ver. 1.79. nblast (Costa et al., 2016) was used to compare the nucleotide sequences of DNA repeat contigs from different species.

Results

Each of the groups of organisms is represented in this paper by four species selected in such a way that they cover the maximum range of evolutionary distances within their branch. The phylogenetic relationships of four gastropod species (*Baicalia turriformis*, *Maackia herderiana*, *Korotnewia korotnewi* and *Godlewskia godlewskia*) are shown in Fig. 1, *b*. The lifetime of their common ancestor does not exceed 3 million years

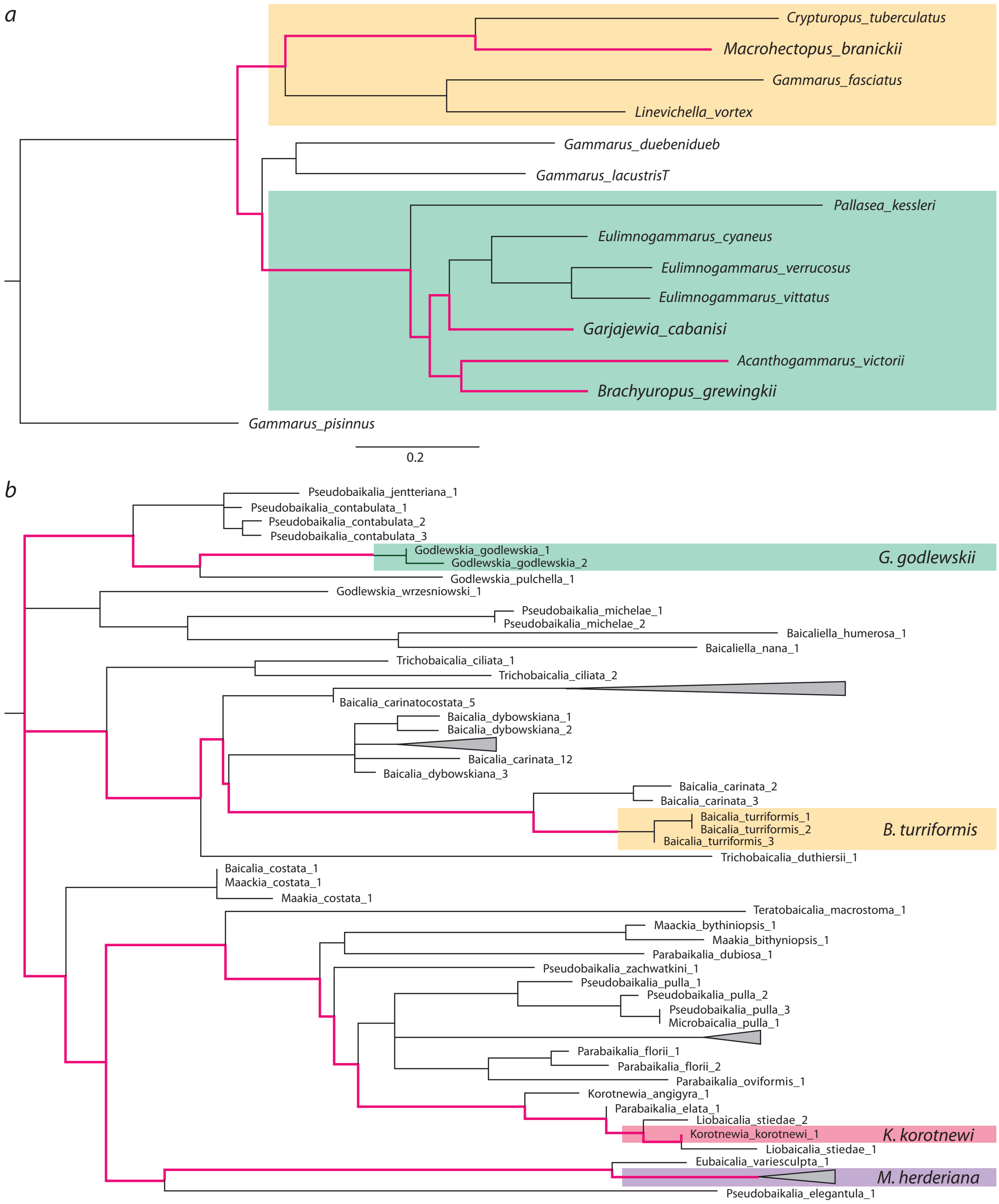

Fig. 1. *a*, A phylogenetic tree of Baikal amphipods and some representatives of the genus *Gammarus*, obtained on the basis of a comparison of concatenated nucleotide sequences on which the species studied in this work are isolated. Two branches represented in Baikal are highlighted in color, as well as phylogenetic relationships between *Acanthogammarus victorii*, *Brachyuropus grewingkii*, *Garjajewia cabanisi* and *Macrohectopus branickii*.
b, A phylogenetic tree of Baicaliidae, rooted at the midpoint, built on the basis of the analysis of the sequences of the Folmer fragment. The phylogenetic relationships between *Baicalia turriformis*, *Maackia herderiana*, *Korotnewia korotnewi* and *Godlewskia godlewskia* are highlighted in color.

(Zubakov et al., 1997; Sherbakov, 1999), and the connections between the selected four species pass through the root of the tree if the tree is rooted at the midpoint. It should be noted that the selected species differ dramatically in their important ecological characteristics and distribution.

Amphipods in Baikal belong to at least two large branches, both within the genus *Gammarus* (Sherbakov, 1999; Hou, Sket, 2016; Romanova et al., 2016; Naumenko et al., 2017). Of the species selected for the study, only *M. branickii* – a representative of the monotypic family – belongs to the branch 'Micruropus', the rest belong to the branch 'Acanthogammarus' – the most diverse in both species and ecology. *B. grewingkii* and *G. cabanisi* are abyssal species, *A. victorii* lives at shallow and medium depths, and *M. branickii* is a unique

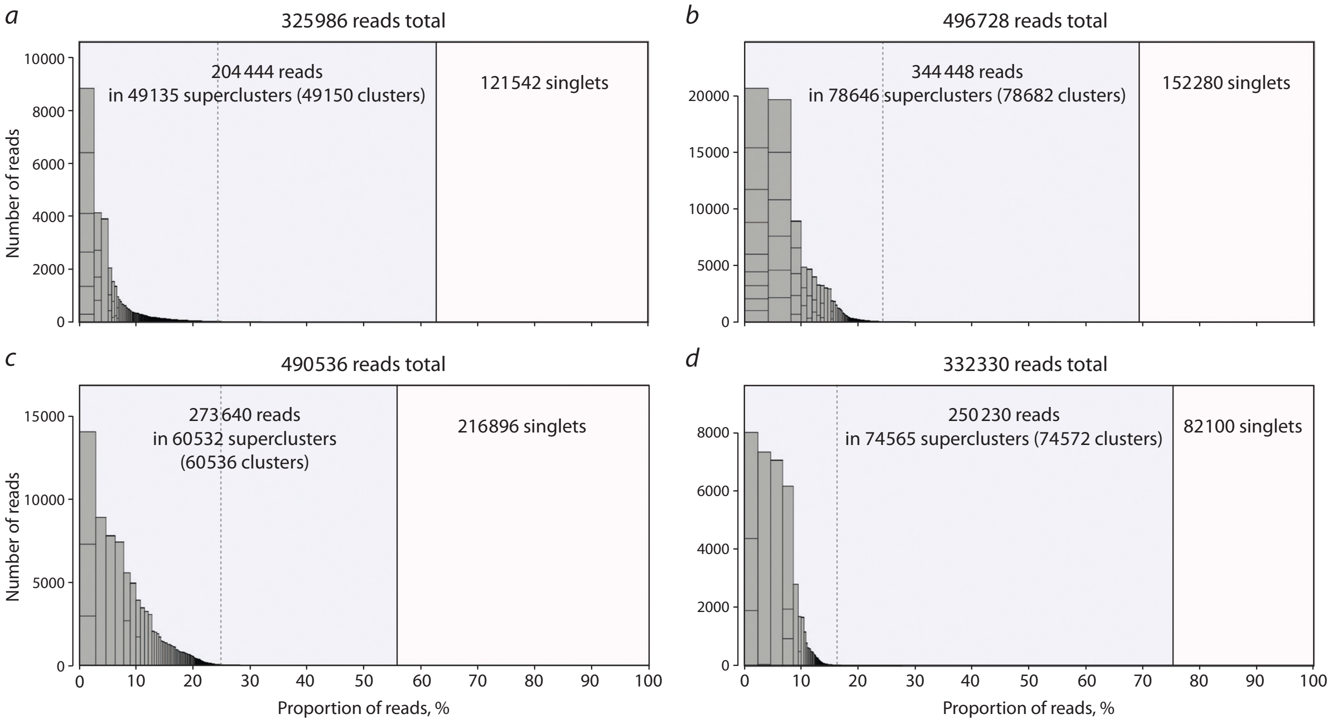

Fig. 2. Representation of the repeats among the species of amphipods sorted by their abundance: *a*, *Acanthogammarus victorii*; *b*, *Brachyuropus grewingkii*; *c*, *Garjajewia cabanisi*, and *d*, *Macrohectopus branickii*.

pelagic species distributed throughout the water column of the lake, including the maximum depths (Bazikalova, 1945) and as part of the quartet of species considered in this work is a distant outer group (see Fig. 1, *a*).

Libraries of repeated contigs were constructed from genome-wide libraries of four species of amphipods and four species of gastropods of the Baikal endemic family Baicaliidae, the production of which is described in (Romanova et al., 2016) and (Peretolchina et al., 2020), respectively. 0.5×10^6 reads were randomly selected from each library and without return, resulting in depleted libraries with a coverage degree of less than 0.5, as a result of which the representation of unique sequences in them turned out to be very low. These subsets of genomic libraries were used to search and annotate DNA repeats using repeatexplorer (Novák et al., 2013). In all cases, the repetitions included approximately 50 % of the reads, which accounted for from 5×10^3 to 10^4 of unique contigs (Fig. 2).

The distributions of contigs by representation in genomes were also approximately the same in all cases; however, if *A. victorii* and *G. cabanisi* had a single dominant repeat (refers to simple DNA repeats, SSR) (see Fig. 2, *a*, *c*), then *B. grewingkii* and *M. branickii* had several dominant repeats (see Fig. 2, *b*, *d*).

The comparison of the compositions of repeats in the species within each of the groups was carried out by concatenating the output files – lists of contigs resulting from a search through genomic libraries. To distinguish between contigs belonging to different species, prefix indexing was used, specific to each of the species. After converting a copy of the concatenated list into a library in blast format, we used nblast to search in the "all against all" mode. At the same time, deletions/insertions were allowed and the similarity threshold of sequences was set to 80 %.

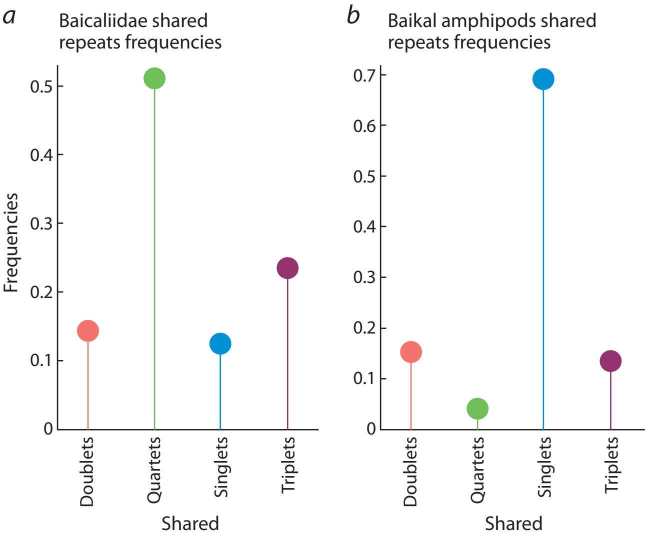

Fig. 3. The proportions of repeats common to several species and found as a result of the blast search for gastropods Baicaliidae and amphipods.

For each of the studied groups, libraries of repeat contigs were concatenated after adding species-specific tags to sequence names, then groups consisting of at least five sequences were selected and a nblast search was performed "all against all". The search conditions allowed 20 % differences and indels. The analysis of intragroup distributions of repeats revealed significant differences between amphipods and molluscs (Fig. 3, see the Table). In a much younger group of Bai-

Patterns of repeats shared in amphipods and gastropods (Baicaliidae)

Species	Total number of contigs	Unique contigs (ratio)	Fully shared contigs (ratio)
	Amphipods		
Brachyuropus grewingkii	21070.0	0.233	0.135
Acanthogammarus victorii	30695.0	0.315	0.093
Garjajewia cabanisi	58335.0	0.605	0.049
Macrohectopus branickii	10674.0	0.455	0.267
	Baicaliidae		
Baicalia turriformis	58491.0	0.036	0.626
Maackia herderiana	57525.0	0.044	0.637
Korotnewia korotnewi	51841.0	0.040	0.707
Godlewskia godlewskia	57872.0	0.037	0.633

kal, the repeats that occur in all four species turned out to be the most represented. Conversely, a relatively small proportion is accounted for by species-specific sequences (see Fig. 3, *a*). The representation of repetitions common to several species also turned out to be very similar in different Baikal species.

In amphipods, on the contrary, most of the repeats are unique (species-specific), and there are very few common ones for all four species (see Fig. 3, *a*). Repeats common to two, three and four species are also not equally represented in different genomes of amphipods (see the Table and Fig. 3, *b*). Interestingly, the largest proportion of common repeats (‘quartets’) was found in the genome of *M. branickii*, which is a very remote external group in relation to the other three species and, unlike the rest of the Baikal amphipods, lives in the pelagic zone of the lake.

In general, it should be noted that all possible patterns of repeat propagation are present in the genomes of both groups of species: there are both those present in only two species in all possible combinations, and all variants of absence in only one of the species (see Fig. 3).

A comparison of the distribution of repeats belonging to different classes according to their distribution in Figure 4 also does not reveal any interspecific variation in Baikal and rather significant differences between amphipod species. However, in both, all possible combinations of the two species are detected, which have repeated elements in common only for them (up to the sensitivity of detection and identification conditions).

Sets of common repeats were used to cluster species. To do this, as a measure of the distance between species, we used

$$d_{ij} = 1 - \frac{N_{\text{shared}}}{N_i + N_j},$$

where d_{ij} is the distance between species (genomes), i and j are species or genomes numbers, N_{shared}, N_i and N_j are the numbers of repeat types in the respective species. Note that the abundances of repeats of each type are not taken into account, but the denominator N_{common} involves all types of repeats found in the pulled repeats library of the species compared. These distances were used to construct the distance matrix, and it, in turn, was used to build a tree by combining the nearest neighbors (Saitou, Nei, 1987).

For the same species and both groups, maximum likelihood trees were inferred based on a comparison of concatenated protein-coding nucleotide sequences of mitochondrial genomes. The topologies of the trees coincided, but the ratio of nucleotide distances and distances calculated by common repetition is not linear (data not shown).

A more detailed examination of pairs of species with common repeats (Fig. 5) shows that a noticeable, albeit relatively small number of repeats is shared by species that are not sister species and thus are not consistent with the phylogeny. This property is present in both amphipods and gastropods.

Discussion

The libraries of the NGS reads of four species of Baikal endemic gastropods (*Baicalia turriformis*, *Maackia herderiana*, *Korotnewia korotnewi* and *Godlewskia godlewskia*) belonging to the Baikal endemic Baicaliidae, and four Baikal amphipods (*Acanthogammarus victorii*, *Brachyuropus grewingkii*, *Garjajewia cabanisi* and *Macrohectopus branickii*) were used to *de novo* search for repeated DNA elements using the repeatexplorer algorithm. All taxa whose genomic libraries are analyzed in this work, despite various evolutionary histories, evolved within the limits of the reservoir that continuously existed on the site of modern Baikal.

The gastropods of Baicaliidae are a relatively young group, the time of the most recent common ancestors (t_{MRCA}) of modern species is no more than 2.5 million years old. They are found at depths of no more than 100 m on a variety of soil types (Zubakov et al., 1997; Sitnikova, 2006). The amphipods in Baikal are represented by at least two large branches that diverged no earlier than about 60 million years ago (Sherbakov, 1999; Mats et al., 2011; Naumenko et al., 2017). The variety of ecological niches occupied by them is exceptionally large, they are found at all depths.

By their distribution between species, all theoretically possible combinations of repeat classes were found. They ranged from species-specific ones to those found in all the genomes studied. This circumstance made it possible to use the distribution of repeats between genomes as a tool for clustering the corresponding species and comparing the topology of the obtained quartets with the results of clustering of the same species based on a comparison of the nucleotide sequences of concatenated protein-coding fragments of mitochondrial genomes. The topologies coincided, but the ratio of the lengths

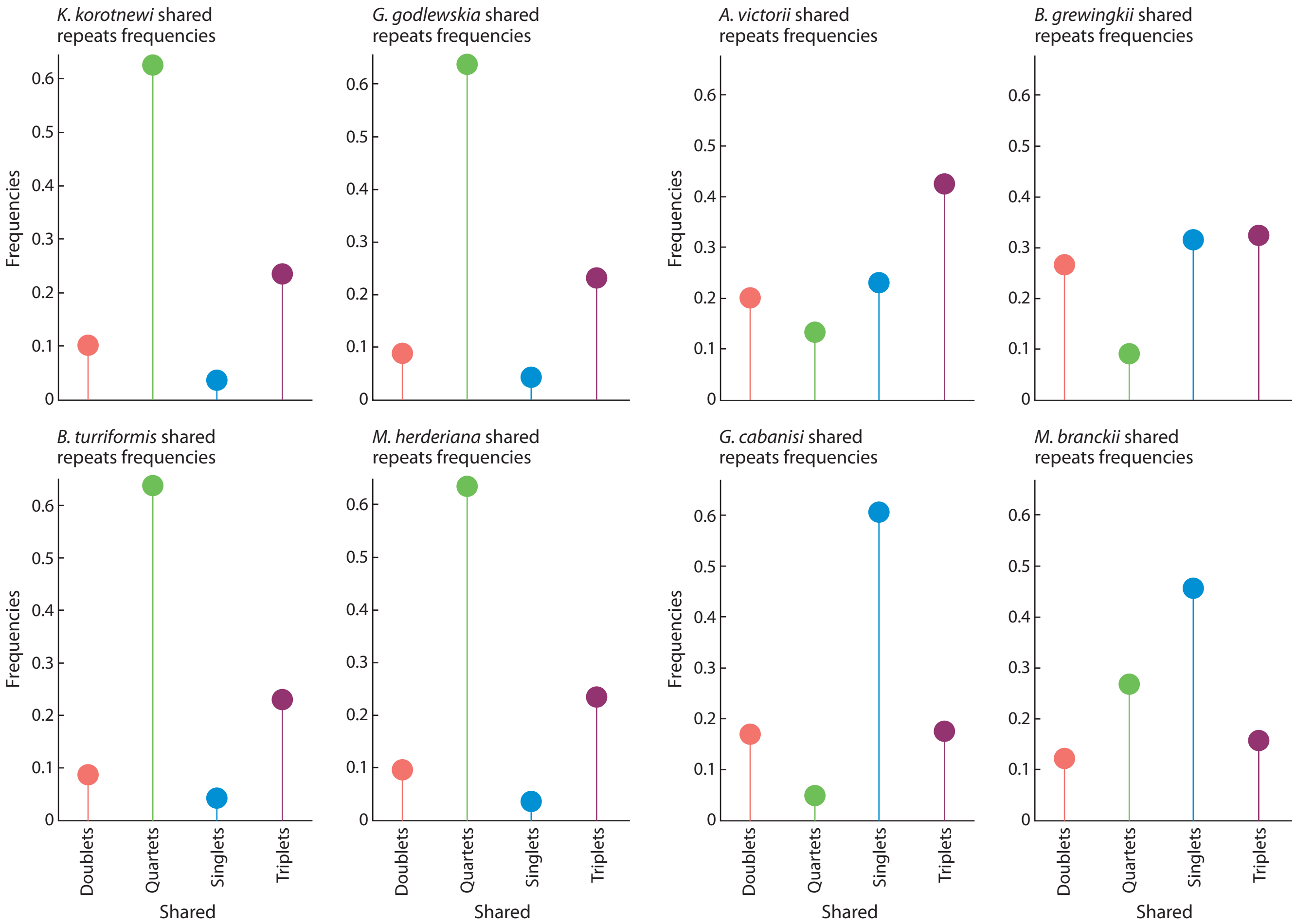

Fig. 4. Species-specific patterns of repeats shared in gastropods (the left four panels) and gastropods (the right four panels).

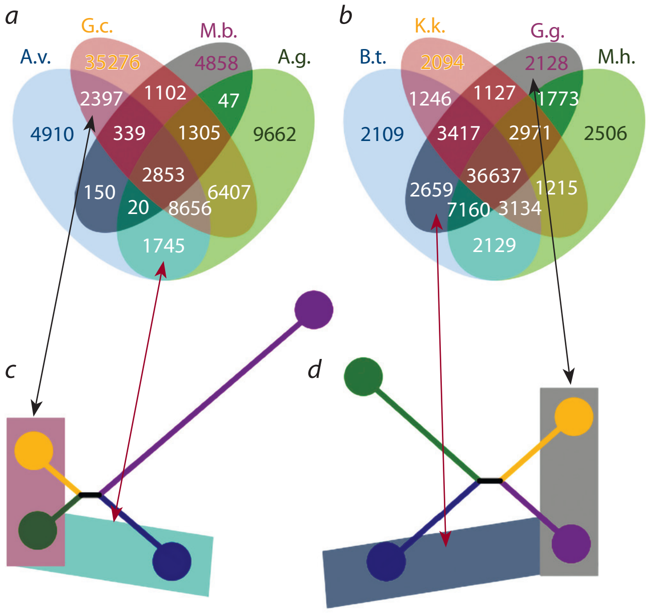

Fig. 5. Venn diagrams of the distribution of repeated DNA between amphipod (*a*) and Baicaliidae (*b*) species. *c*, *d* are unrooted trees inferred from concatenated protein-coding mitochondrial DNA of the corresponding species.

The colors of the branches correspond to the colors of the ovals on the Venn diagrams.

of the branches turned out to be different. In other words, the proportions of common repeats and the degree of differences in nucleotide sequences turned out to be independent, albeit partially correlated features.

The method we used to identify highly repeated sequences and a set of search parameters allow us to identify those that are repeated in the genome with at least 50–100 copies per haploid genome. The detected repeats make up approximately 50 % of the genome and are very diverse ($1 \times 10^4 \ldots 6 \times 10^4$ varieties per genome, see the Table). Therefore, at the present stage of the study, we focused on the integral characteristics of this repetition and the comparison of these characteristics in two flocks of invertebrate species.

The distance tree was inferred from the repeats data using the distance metric calculated from the presence/absence of a repeat class in a sample as justified by blast search under a mild set of parameters. This differs from the parsimony approach employed by (Dodsworth et al., 2015). Its advantage was in avoiding the assumption of strict homology. Nevertheless, like in their study, we obtained the same tree topology to the one inferred from mitogenome sequences for both animal groups studied. Although the topologies were the same, the ratios in branch lengths differed dramatically. We believe that these differences result from the peculiarities of the evolution of the presence/absence of repeats in genomes. The main feature is that in order to appear in the genome as a repeat, the nucleotide sequence starts as a single copy and must be amplified to such an extent that it can be detected by the repeatexplorer algorithm. The loss of repetition should also go through a gradual decrease in the number of copies of it.

Over time, differences in the compositions of repeated sequences accumulate. This confirms the spread of species-specific sequences in amphipods compared to Baicaliidae and vice versa, with a decrease in the proportion of repeats that occur in all four species. Therefore, the comparison of repeat spectra in a large number of species can be an interesting tool for phylogenetic analysis due to the high diversity of repeats and the fact that a large proportion of the genome is used in such an analysis, which gives hope for obtaining a more adequate and stable picture of evolution. A more detailed examination of pairs of species with common repeats (see Fig. 5) shows that a noticeable, albeit relatively small number of repeats, is common between species that are not sister species. This fraction, if "inconsistent", is present in both amphipods and gastropods.

Conclusion

Of particular interest are the repeats, the distribution of which between species contradicts the topology of phylogenetic trees, but corresponds to the ecological or geographical confinement of species. Such repeats are found in both groups (see the Table and Fig. 5), and in a significant (from hundreds to thousands) amount. From the point of view of phylogenetic analysis, they reduce its resolution but allow us to make an intriguing assumption that some part of them is involved in horizontal transfer between sympatrically inhabiting species. This requires an annotation of this part of the contigs, the results of which will be described elsewhere.

References

Ahmad S.F., Singchat W., Panthum T., Srikulnath K. Impact of repetitive DNA elements on snake genome biology and evolution. *Cells.* 2021;10(7):1707. DOI 10.3390/cells10071707.

Athanasouli M., Rödelsperger C. Analysis of repeat elements in the *Pristionchus pacificus* genome reveals an ancient invasion by horizontally transferred transposons. *BMC Genomics.* 2022;23(1):523. DOI 10.1186/s12864-022-08731-1.

Bazikalova A.Ya. Amphipoda of Lake Baikal. In: Proceedings of the Baikal Limnological Station. Vol. 11. Moscow, 1945. (in Russian)

Bird C.P., Stranger B.E., Dermitzakis E.T. Functional variation and evolution of non-coding DNA. *Curr. Opin. Genet. Dev.* 2006;16(6):559-564. DOI 10.1016/j.gde.2006.10.003.

Biscotti M.A., Olmo E., Heslop-Harrison J. Repetitive DNA in eukaryotic genomes. *Chromosome Res.* 2015;23(3):415-420. DOI 10.1007/s10577-015-9499-z.

Boto L. Horizontal gene transfer in the acquisition of novel traits by metazoans. *Proc. Biol. Sci.* 2014;281(1777):20132450. DOI 10.1098/rspb.2013.2450.

Butina T.V., Bukin Yu.S., Khanaev I.V., Kravtsova L.S., Maikova O.O., Tupikin A.E., Kabilov M.R., Belikov S.I. Metagenomic analysis of viral communities in diseased Baikal sponge *Lubomirskia baikalensis. Limnol. Freshw. Biol.* 2019;1:155-162. DOI 10.31951/2658-3518-2019-A-1-155.

Cavalier-Smith T., Beaton M. The skeletal function of non-genic nuclear DNA: new evidence from ancient cell chimaeras. *Genetica.* 1999;106(1-2):3-13. DOI 10.1023/a:1003701925110.

Chen S.C., Sun G.X., Rosen B.P., Zhang S.Y., Deng Y., Zhu B.K., Rensing C., Zhu Y.G. Recurrent horizontal transfer of arsenite methyltransferase genes facilitated adaptation of life to arsenic. *Sci. Rep.* 2017;7(1):7741. DOI 10.1038/s41598-017-08313-2.

Costa M., Manton J.D., Ostrovsky A.D., Prohaska S., Jefferis G.S. NBLAST: rapid, sensitive comparison of neuronal structure and construction of neuron family databases. *Neuron.* 2016;91(2):293-311. DOI 10.1016/j.neuron.2016.06.012.

de Sena Brandine G., Smith A.D. Falco: high-speed FastQC emulation for quality control of sequencing data. *F1000Res.* 2019;8:1874. DOI 10.12688/f1000research.21142.2.

Dodsworth S., Chase M.W., Kelly L.J., Leitch I.J., Macas J., Novák P., Piednoël M., Weiss-Schneeweiss H., Leitch A.R. Genomic repeat abundances contain phylogenetic signal. *Syst. Biol*. 2015;64(1):112-126. DOI 10.1093/sysbio/syu080.

Drozdova P., Saranchina A., Madyarova E., Gurkov A., Timofeyev M. Experimental crossing confirms reproductive isolation between cryptic species within *Eulimnogammarus verrucosus* (Crustacea: Amphipoda) from Lake Baikal. *Int. J. Mol. Sci*. 2022;23(18):10858. DOI 10.3390/ijms231810858.

Gurkov A., Rivarola-Duarte L., Bedulina D., Fernández Casas I., Michael H., Drozdova P., Nazarova A., Govorukhina E., Timofeyev M., Stadler P.F., Luckenbach T. Indication of ongoing amphipod speciation in Lake Baikal by genetic structures within endemic species. *BMC Evol. Biol*. 2019;19(1):138. DOI 10.1186/s12862-019-1470-8.

Hausdorf B., Röpstorf P., Riedel F. Relationships and origin of endemic Lake Baikal gastropods (Caenogastropoda: Rissooidea) based on mitochondrial DNA sequences. *Mol. Phylogenet. Evol*. 2003;26(3):435-443. DOI 10.1016/s1055-7903(02)00365-2.

Hou Z., Sket B. A review of gammaridae (crustacea: Amphipoda): the family extent, its evolutionary history, and taxonomic redefinition of genera. *Zool. J. Linn. Soc.* 2016;176(2):323-348. DOI 10.1111/zoj.12318.

Jalili V., Afgan E., Gu Q., Clements D., Blankenberg D., Goecks J., Taylor J., Nekrutenko A. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Res*. 2020;48(W1):W395-W402. DOI 10.1093/nar/gkaa434.

Kamaltynov R.M. On the higher classification of Lake Baikal amphipods. *Crustaceana.* 1999;72(8):933-944.

Kejnovsky E., Jedlicka P. Nucleic acids movement and its relation to genome dynamics of repetitive DNA: is cellular and intercellular movement of DNA and RNA molecules related to the evolutionary dynamic genome components? *BioEssays.* 2022;44(4):e2100242. DOI 10.1002/bies.202100242.

Kozhov M. Lake Baikal and Its Life. Monographiae Biologicae. Vol. 11. Dordrecht: Springer, 1963. DOI 10.1007/978-94-015-7388-7.

Lee I.P.A., Eldakar O.T., Gogarten J.P., Andam C.P. Bacterial cooperation through horizontal gene transfer. *Trends Ecol. Evol*. 2022;37(3):223-232. DOI 10.1016/j.tree.2021.11.006.

Lerat E., Casacuberta J., Chaparro C., Vieira C. On the importance to acknowledge transposable elements in epigenomic analyses. *Genes.* 2019;10(4):258. DOI 10.3390/genes10040258.

Li Y., Liu Z., Liu C., Shi Z., Pang L., Chen C., Chen Y., Pan R., Zhou W., Chen X.X., Rokas A., Huang J., Shen X.X. HGT is widespread in insects and contributes to male courtship in lepidopterans. *Cell.* 2022;185(16):2975-2987.e10. DOI 10.1016/j.cell.2022.06.014.

Lipaeva P., Vereshchagina K., Drozdova P., Jakob L., Kondrateva E., Lucassen M., Bedulina D., Timofeyev M., Stadler P., Luckenbach T. Different ways to play it cool: transcriptomic analysis sheds light on different activity patterns of three amphipod species under long-term cold exposure. *Mol. Ecol*. 2021;30(22):5735-5751. DOI 10.1111/mec.16164.

Mats V.D., Shcherbakov D.Y., Efimova I.M. Late Cretaceous–Cenozoic history of the Lake Baikal depression and formation of its unique biodiversity. *Stratigr. Geol. Correl.* 2011;19(4):404-423. DOI 10.1134/S0869593811040058.

Naumenko S.A., Logacheva M.D., Popova N.V., Klepikova A.V., Penin A.A., Bazykin G.A., Etingova A.E., Mugue N.S., Kondrashov A.S., Yampolsky L.Y. Transcriptome-based phylogeny of endemic Lake Baikal amphipod species flock: fast speciation accompanied by frequent episodes of positive selection. *Mol. Ecol.* 2017; 26(2):536-553. DOI 10.1111/mec.13927.

Novák P., Neumann P., Pech J., Steinhaisl J., Macas J. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics.* 2013;29(6):792-793. DOI 10.1093/bioinformatics/btt054.

Peretolchina T., Sitnikova T.Y., Sherbakov D.Y. The complete mitochondrial genomes of four Baikal molluscs from the endemic family Baicaliidae (Caenogastropoda: Truncatelloida). *J. Molluscan Stud.* 2020;86(3):201-209. DOI 10.1093/mollus/eyaa004.

Rocha A., Dalgarno A., Neretti N. The functional impact of nuclear reorganization in cellular senescence. *Brief. Funct. Genomics*. 2022; 21(1):24-34. DOI 10.1093/bfgp/elab012.

Romanova E.V., Aleoshin V.V., Kamaltynov R.M., Mikhailov K.V., Logacheva M.D., Sirotinina E.A., Gornov A.Y., Anikin A.S., Sherbakov D.Y. Evolution of mitochondrial genomes in Baikalian amphipods. *BMC Genomics*. 2016;17(Suppl.14):1016. DOI 10.1186/s12864-016-3357-z.

Saitou N., Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 1987;4(4):406-425. DOI 10.1093/oxfordjournals.molbev.a040454.

Shen W., Le S., Li Y., Hu F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PloS One.* 2016;11(10): e0163962. DOI 10.1371/journal.pone.0163962.

Sherbakov D.Y. Molecular phylogenetic studies on the origin of biodiversity in Lake Baikal. *Trends Ecol. Evol.* 1999;14(3):92-95. DOI 10.1016/s0169-5347(98)01543-2.

Silva B.S.M.L., Heringer P., Dias G.B., Svartman M., Kuhn G.C.S. *De novo* identification of satellite DNAs in the sequenced genomes of *Drosophila virilis* and *D. americana* using the RepeatExplorer and TAREAN pipelines. *PLoS One.* 2019;14(12):e0223466. DOI 10.1371/journal.pone.0223466.

Sitnikova T.Y. Endemic gastropod distribution in Baikal. *Hydrobiologia.* 2006;568(1):207-211. DOI 10.1007/s10750-006-0313-y.

Sitnikova T., Roepstorf P., Riedel F. Reproduction, duration of embryogenesis, egg capsules and protoconchs of gastropods of the family Baicaliidae (Caenogastropoda) endemic to Lake Baikal. *Malacologia.* 2001;43(1-2):59-85.

Steensels J., Gallone B., Verstrepen K.J. Interspecific hybridization as a driver of fungal evolution and adaptation. *Nat. Rev. Microbiol.* 2021; 19(8):485-500. DOI 10.1038/s41579-021-00537-4.

Takhteev V. On the current state of taxonomy of the Baikal Lake amphipods (Crustacea: Amphipoda) and the typological ways of constructing their system. *Arthropoda Selecta.* 2019;28(3):374-402. DOI 10.15298/arthsel.28.3.03.

Thakur J., Packiaraj J., Henikoff S. Sequence, chromatin and evolution of satellite DNA. *Int. J. Mol. Sci.* 2021;22(9):4309. DOI 10.3390/ijms22094309.

Titievsky A., Putintseva Y.A., Taranenko E.A., Baskin S., Oreshkova N.V., Brodsky E., Sharova A.V., Sharov V.V., Panov J., Kuzmin D.A., Brodsky L., Krutovsky K.V. Comparative genomics analysis of repetitive elements in ten gymnosperm species: "dark repeatome" and its abundance in conifer and gnetum species. *Life.* 2021;11(11):1234. DOI 10.3390/life11111234.

Yakhnenko A., Itskovich V. Analysis of mtDNA variability in closely related Baikal sponge species for new barcoding marker development. *Limnology.* 2020;21(1):49-57. DOI 10.1007/s10201-019-00599-7.

Zubakov D.Y., Shcherbakov D.Y., Sitnikova T.Y. Phylogeny of the endemial Baicaliidae molluscs inferred from partial nucleotide sequences of the CO1 mitochondrial gene. *Mol. Biol.* 1997;31(6): 935-939.

ORCID ID

T.E. Peretolchina orcid.org/0000-0002-2950-9762
D.Y. Sherbakov orcid.org/0000-0002-1410-392X

Acknowledgements. The research was supported by the State Project No. 0279-2021-0010.

Conflict of interest. The authors declare no conflict of interest.

Received December 23, 2022. Revised February 20, 2023. Accepted February 20, 2023.