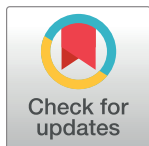RESEARCH ARTICLE

# Identification and functional characterization of the German cockroach, *Blattella germanica*, short interspersed nuclear elements

**Sergei Yu. Firsov, Karina A. Kosherova, Dmitry V. Mukha** [ID] *

Vavilov Institute of General Genetics Russian Academy of Sciences, Moscow, Russia

* dmitryVmukha@gmail.com

## Abstract

In recent decades, experimental data has accumulated indicating that short interspersed nuclear elements (SINEs) can play a significant functional role in the regulation of gene expression in the host genome. In addition, molecular markers based on SINE insertion polymorphisms have been developed and are widely used for genetic differentiation of populations of eukaryotic organisms. Using routine bioinformatics analysis and publicly available genomic DNA and small RNA-seq data, we first described nine SINEs in the genome of the German cockroach, *Blattella germanica*. All described SINEs have tRNA promoters, and the start of their transcription begins 11 bp upstream of an "A" box of these promoters. The number of copies of the described SINEs in the *B. germanica* genome ranges from several copies to more than a thousand copies in a SINE-specific manner. Some of the described SINEs and their degenerate copies can be localized both in the introns of genes and loci known as piRNA clusters. piRNAs originating from piRNA clusters are shown to be mapped to seven of the nine types of SINEs described, including copies of SINEs localized in gene introns. We speculate that SINEs, localized in the introns of certain genes, may regulate the level of expression of these genes by a PIWI-related molecular mechanism.

## Introduction

Short interspersed nuclear elements (SINEs) are nonautonomous retrotransposons transcribed by RNA polymerase III. The conversion of SINE RNA into DNA and the subsequent process of integration into different locations of the genome are controlled by the molecular machinery of autonomous retrotransposons [1–4]. Similar to other transposable elements (TEs), SINEs are ancient components of the genome [5], although young and highly active families of these TEs have also been described [6]. Interestingly, SINEs may account for a significant portion of the genome in some eukaryotic species but are absent in others [1, 4, 7].

Typical SINEs are 150–600 bp in length with a composite structure consisting of three parts: "head", "body" and "tail", sequentially localized starting from the 5'- end of the TE. "The head" is a fragment of one of the types of RNAs transcribed by RNA polymerase III: tRNA, 5S, or 7SL. Usually, SINEs contain tRNA fragments [8, 9]. Some SINEs may contain fragments of

several RNAs of one type [10, 11] and combinations of different RNAs, for example, tRNA + 7SL [12, 13] or tRNA + 5S [14, 15]. A fundamental aspect of the functional activity of a SINE is the presence of a promoter in this part of TE, which is recognized by RNA polymerase III, which, in turn, initiates the start of transcription upstream from the promoter location. To date, the features of the structural and functional organization of the promoters of the tRNA, 5S, and 7SL RNA genes have been described in detail [16, 17]. In particular, the transcription of eukaryotic tRNA genes has been shown to use a promoter system comprised of an "A" box and a "B" box (length of 11 bp each), variably spaced from one another by ~30–60 bp in a gene-specific manner [18–21]. The initiation of transcription typically begins ~7–20 bp upstream of the "A" box promoter element [16, 18, 20].

The main structural and functional component of "the body" of SINEs is the relatively extended region responsible for the binding of the protein product of the "partner" autonomous retrotransposons with the SINE RNA and subsequent reverse transcription of this RNA and SINE integration into the genome. Note that it is far from always possible to trace the similarity of the nucleotide sequence between SINE and its autonomous "partner" [2]. Moreover, relatively recently, so-called CORE sequences, which are evolutionarily conserved nucleotide sequences with a high degree of similarity between SINEs described in the genomes of organisms that are evolutionarily distant from each other, have been identified in the "body" of some SINEs. The evolutionary value of these conserved regions is still debated [22–27].

"The tail" of typical SINEs represents repetitive microsatellite motifs or poly (A) sequences. The molecular mechanism of the formation of these sequences remains largely unclear to date.

During the integration of retrotransposons, sequential cutting occurs first at the bottom and then at the upper strand of the target site. Depending on the particular retrotransposon, the second strand break can occur downstream, upstream, or in line with the bottom strand nick, resulting in target site duplications (TSDs), target site deletions or blunt insertions, respectively [28]. Since the integration of SINEs is conditioned by the peculiarities of the partner autonomous retrotransposon, during the integration of nonautonomous retrotransposons, any of the changes described above in the target sites are theoretically possible; however, to the best of our knowledge, only the formation of TSDs has been described thus far.

Whereas transposable elements (including SINEs) have been considered selfish or junk DNA [29–31], recent findings in genomic and epigenomic studies suggest that some of their copies have functional roles in gene regulation and/or chromatin organization. SINEs are assumed to insert in genes, providing new splicing sites resulting in the generation of a SINE-containing isoform; transcription factor binding sites in the SINE sequence may affect neighboring genes; SINEs can regulate gene expression at a distance as a tissue-specific enhancer; SINEs can stabilize nucleosome positioning in neighboring regions; and SINEs can mediate the methylation of the surrounding DNA and mediate histone modifications in the region of the integration site [for review, see 32, 33]. Thus, transposable elements play an important regulatory role in ensuring the functional activity of the genome; in addition, it has been shown that the structural and functional elements that determine the activity of transposable elements can be used to create vector constructs for biotechnological purposes [for review, see 34].

Since the target-site specificity of the integration of SINEs is dictated by the molecular machine of their autonomous partners, it is logical to assume that the distribution of autonomous and nonautonomous retrotransposons in the genome will be similar. However, the retrotransposons in the genome have been shown to occupy distinct parts of the genome, and their regional densities are negatively correlated with each other [32]. SINEs are clustered in gene-rich regions, while their autonomous partners are concentrated in gene-poor regions and are depleted from promoters. It was suggested that positive selection has been operating on SINEs inserted in or close to genes during evolution [35, 36]. In this regard, interesting

recent studies show that SINEs have undergone strong natural selection, causing genomic heteroplasmy and driving ecological diversity. Possible evolutionary mechanisms underlying ecological diversity at the interface between SINE mobilization and organism defense have been revealed [37, 38].

Of particular interest, from our point of view, are studies of the role of a recently discovered class of small RNAs in the regulation of the activity of TEs and the epigenetic modification of the integration sites of TEs mediated by these small RNAs. piRNAs (PIWI-interacting RNAs) are the largest class of small noncoding RNAs of 26 to 31 nucleotides in length expressed in germinal and somatic cells; they are found in complexes with proteins of the PIWI family, for which they were named [39]. piRNAs are involved in the control of TEs as part of an evolutionarily conserved mechanism [40–42]. Most of the piRNAs originate from loci known as piRNA clusters. These loci are enriched with inactive transposon sequences and are required to prevent the spread of active TEs throughout the genome. piRNAs from piRNA clusters direct PIWI proteins to cleave TE transcripts and induce their processing into new piRNAs. These new piRNAs themselves can act as PIWI guides to cleave complementary transcripts, inducing the production of piRNAs identical to initiator piRNAs derived from piRNA clusters, leading to the process of amplification of piRNA, which is called the ping-pong cycle [41, 43], reviewed by Ozata et al. [44].

piRNAs have been shown to regulate transpositional activity through RNA decay (posttranscriptional level) and/or through DNA methylation and histone modification (transcriptional level). DNA methylation and histone modification are known to be able to lead to repressive heterochromatin formation and change the dynamics of expression of a gene localized in this region [45–49]. It was shown that DNA methylation could be detected in all insect orders examined except Diptera (flies) [50]. We speculate that SINEs, localized in the introns of certain genes responsible for the performance of specific functions, by PIWI-related molecular mechanisms may change the local conformation of chromatin, which in turn will lead to a change (adjusting?) of the level of expression of these genes.

In recent decades, the German cockroach, *Blattella germanica*, has been increasingly considered a model object for studying the molecular genetic organization of eukaryotes and may serve as a suitable reference model for studying the molecular biology of insects [51–56].

In this study, we first described the structural and functional organization of SINEs of *B. germanica*. We examined a representative sample of approximately a thousand genes previously annotated in the genome of the German cockroach and identified the genes containing SINE copies in the introns. Degenerate copies of seven of the nine described SINEs were shown to be localized in piRNA clusters, and the corresponding piRNAs are mapped to copies of SINEs localized in gene introns. We consider the results obtained as the first step toward studying the possible regulatory function of SINEs of *B. germanica*.

Finally, since the number of SINEs in the genome, as a rule, is large enough and their integration occurs into random sites of the genome, the pattern of integrated copies of these TEs can be considered a polymorphic molecular genetic marker that allows solving the problems of population genetics, in particular, determining the genetic distances between populations of living organisms.

## Materials and methods

### SINEs identification

Recently, the 2-Gb genome of *B. germanica* was reported [54], which gives rise to the possibility of studying the features of the structural organization of SINEs of this insect. We used the

sequence data presented in a public database (https://www.ncbi.nlm.nih.gov/Traces/wgs/PYGN01).

To identify SINEs in the *B. germanica* genome, we used two methodological approaches: the first approach was based on an automatic search using the previously published program SINE_Scan 1.1.1 [57]; the second approach was based on the development of our own search algorithm.

The local SINE_Scan 1.1.1 package for Linux (https://github.com/maohlzj/SINE_Scan) was used with the default parameters (S1 File).

The essence of our new algorithm is as follows. In the first step, the prediction of tRNA sequences in the *B. germanica* genome was performed using the previously published program tRNAscan 2.0.3 [58]. The local tRNAscan 2.0.3 package for 64-bit Linux (http://lowelab.ucsc.edu/tRNAscan-SE/) was used with the default parameters (S1 File).

As a result, the positions of the first nucleotides for each predicted tRNA sequence were determined. The regular Python script (S1 Script), with the possibility of saving fragments with a length of 1000 nucleotides, was used for creation of a database containing the sequences corresponding to determined tRNAs at the 5'-end and an extended region corresponding to the sequence adjacent to tRNA in genomic DNA at the 3' end. The resulting pool of sequences was analyzed using CodonCode Aligner 8.0.2 software (https://www.codoncode.com/aligner). The "Assemble" command of this software was run with the following parameters of "Assembly" settings tab: Algorithm—Local alignments, Min. percent identity = 90, Min. overlap length = 150, with the following visual analysis of the sequences in "Contigs" folder. The candidates for SINE should have no less than three aligned sequences; the beginning of tRNA should have a shift no more than five nucleotides in length; overlapping sequences should have a length no less than 150 nucleotides (for an example see S1 Fig).

For each of the sequences considered a candidate for SINE, sequences were found in the genome of *B. germanica* with the maximum similarity to the analyzed sequence. The local tools from the BLAST 2.9.0+ package for 64-bit Linux (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/) [59] were used to perform searches with the default parameters, followed by the analysis and filtering of the output table. Only hits with E-values $< 1x10^{-25}$ were used for the following analysis.

At the final stage, based on the comparison of similar nucleotide sequences corresponding to each of the SINE types, consensus sequences were predicted using the Jalview V.2 program [60].

To identify evolutionarily conserved sequences within the SINEs we described, we used two approaches: 1) the CENSOR program [61, 62], available at https://www.girinst.org/censor/index.php, and 2) regular online Blastn search, which allowed to compare the sequences we analyzed with all the sequences presented in the GenBank database (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome)

The duplications of the integration sites were predicted online by TSD (target site duplication) Search program (https://sines.eimb.ru).

The number of copies of each of the SINE types and the coordinates of each specific copy in the genome of *B. germanica* were determined using the DotPlot ability in the UGENE 34.0 software package [63]. The following settings were used: identity = 90% or 80% and min. length = from the first to the last nucleotide.

## SINE piRNA analysis

A publicly available small RNA-seq database (https://www.ncbi.nlm.nih.gov/gds/?term=GEO%20GSE87031) was used for analysis. To detect the piRNAs localized in piRNA clusters, we

used previously described methodological approaches [52, 53, 55] with minor changes (S2 File).

Low-quality reads were filtered out, and adapters were removed from the 22 small RNA-Seq libraries of *B. germanica* using Trimmomatic 0.39 software [64]. Read pairs were then merged using the Pear 0.9.11 tool [65]. Clusters of piRNAs were identified using proTRAC 2.4.3 [66] with the default parameters (S1 File).

As a result, the sequences of piRNA clusters and unique piRNA sequences with a length of 26–31 bases located in piRNA clusters were obtained.

To create a database containing all piRNA repeating reads corresponding to the sequence of unique piRNA reads, the unique piRNA sequences were compared with the pool of reads presented in the original database, from which a new database was obtained containing all piRNA repeating sequences by definition presented in piRNA clusters.

Since the copies of each type of SINE localized in the genome may differ from each other and the corresponding consensus sequence due to the accumulation of random mutations, we created a database containing all copies of Sbg1-Sbg9 localized in the genome, and the degree of similarity between the sequences of each SINE type accounted for 80% or more. Moreover, tRNA sequences, which are part of both SINEs and TE-independent tRNA clusters, are known to be able to be a source of multiple short RNAs, which are products of the PIWI-independent pathway [67, 68]. For this reason, the sequences of SINEs included in the created database did not contain sequences complementary to tRNA. piRNAs localized in piRNA clusters were mapped to the described SINE sequences using Bowtie2 software [69], forcing zero mismatches on the read length; then, this piRNA fraction was retrieved, and the accumulated pool of piRNAs (9261 reads) was used for subsequent analysis (S2 Fig).

The mapping of the described pool of SINE-related piRNAs (9261 reads) to SINEs consensus sequences and genes of *B. germanica* was performed by Bowtie2 software, and the result was visualized by the UGENE 34.0 program [63].

## Results and discussion

### Structural and functional organization of SINEs of *Blattella germanica*

In our study, sequences referred to SINE had to meet the following criteria: 1) had a length of at least 150 bp; 2) contained a sequence corresponding to the RNA polymerase III promoter; 3) were represented in the genome by at least three copies; and 4) were flanked by target site duplications.

We identified nine SINEs of *B. germanica* (Sbg1 –Sbg9). The lengths of the identified SINEs were Sbg1–306 bp, Sbg2–313 bp, Sbg3–613 bp, Sbg4–456 bp, Sbg5–582 bp, Sbg6–359 bp, Sbg7–564 bp, Sbg8–355 bp, and Sbg9–658 bp.

To identify SINEs, we used two methodological approaches (see "Materials and Methods"). We first used the automatic search of SINEs by SINE_Scan 1.1.1 program which is considered to be fast and robust, and found six SINEs. Next, to double-check the results, we used our own search algorithm and found the same six and three additional SINEs, Sbg4, Sbg5 and Sbg9 that went unrecognized by the initial SINE_Scan 1.1.1 program.

In Fig 1, the schemes of the structural organization of the consensus sequences of the SINEs identified by us are shown; in S3 Fig—both the consensus sequences and the sequences of twelve similar to the consensus sequence SINE copies, presented in the genome, along with the sequences of the nearest environment are shown. Note that Sbg9 is a dimeric SINE composed of the Sbg1 and Sbg8 sequences (see S3 Fig).

The "heads" of all identified SINEs are represented as tRNA fragments. For six SINEs (Sbg1 –Sbg6), the tRNAscan 2.0.3 program [58] predicted tRNA sequences in the consensus
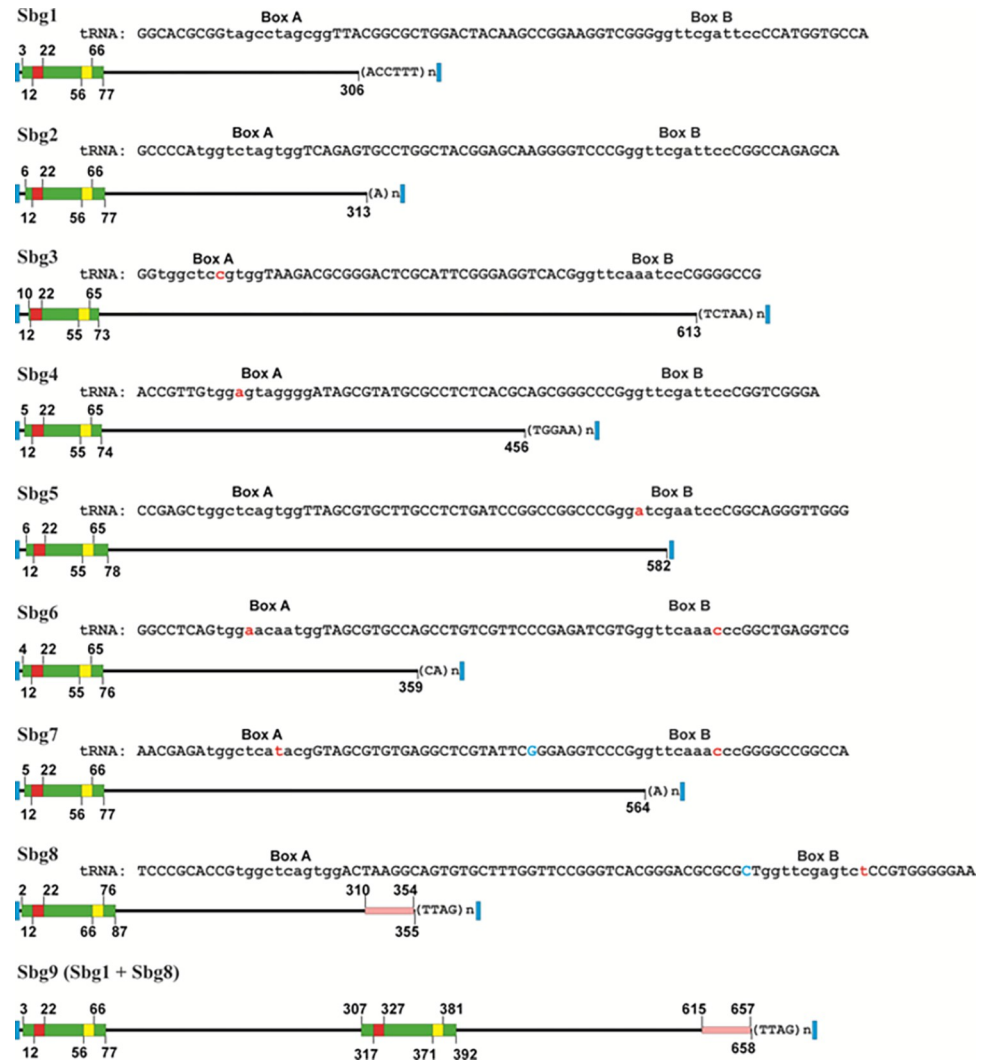
**Fig 1. Schemes of the structural organization of the consensus sequences of SINEs of *Blattella germanica* (Sbg1-Sbg9).** Areas corresponding to the following nucleotide sequences are highlighted with a colored background: tRNA-green; "A" box- red; "B" box- yellow; the region similar to the nonLTR retrotransposon-pink; target site duplication- blue. Repetitive microsatellite motifs are indicated in parentheses. "A" and "B" boxes are shown in lower case within the tRNA nucleotide sequences. Nucleotides other than canonical nucleotides are highlighted in red and blue (explanation in the text). The numbers indicate the positions of the nucleotides.

https://doi.org/10.1371/journal.pone.0266699.g001

sequences of SINEs; however, in the consensus sequences of Sbg7 and Sbg8, the tRNA sequences were not predicted. To determine the nature of Sbg7 and Sbg8 "heads", we analyzed by the tRNAscan 2.0.3 program all copies of SINEs identified in the genome of *B. germanica* that have at least 90% similarity both with each other and with the consensus sequences corresponding to Sbg7 or Sbg8. The consensus sequences of Sbg7 and Sbg8 were shown to have one nucleotide substitution each (in Figs 1 and S3 highlighted in blue and red fonts, respectively), distinguishing them from the canonical tRNA structures identified by the tRNAscan 2.0.3 program (S3 Fig). Note that some copies of both Sbg7 and Sbg8 contained tRNA sequences that were detectable by the tRNAscan 2.0.3 program.

The transcriptional mechanism of eukaryotic tRNA genes is known to use a promoter system comprised of an "A" box and a "B" box, which have the following canonical nucleotide

compositions: TRGYNNARNNG and RGTTCRANTCC, respectively [70]. An "A" box and a "B" box identified in the consensus sequences of Sbg1 and Sbg2 have canonical structures (in Figs 1 and S3 "A" and "B" boxes highlighted in lowercase letters). All other SINEs identified in the genome of *B. germanica* have one nucleotide substitution within the "A" or/and "B" boxes (in Figs 1 and S3, the substitutions highlighted in red fonts and yellow background, respectively).

The SINEs we have described have a different number of nucleotides from their 5'-end to the beginning of the tRNA sequence detected by the tRNAscan 2.0.3 program; at the same time, the distance from the beginning of SINEs to the start of the corresponding "A" box was always 11 bp (Figs 1 and S3). We call it the "eleven nucleotide rule", indicating that the start of transcription of all we have described SINEs begins 11 bp upstream of an "A" box of RNA polymerase III promoter.

As mentioned above, the initiation of transcription of eukaryotic tRNA genes typically begins ~7 to 20 bp upstream of the "A" box promoter element in a species-specific manner. Whether the "eleven nucleotide rule" applies to all tRNA genes of *B. germanica*, or is characteristic only for SINEs of this insect species remains unclear and will be the subject of a separate study.

For SINEs of other species presented in the RepBase (https://www.girinst.org) [8] and SINE BASE (https://sines.eimb.ru) [9] databases, the distances from the start of transcription sites to the corresponding "A" boxes range widely (data not shown). In addition, it was shown that in *Reticulitermes lucifugus*, a closely related species of *B. germanica*, these distances are different for SINEs Talua, Talub, Taluc and Talud (11 bp, 10 bp, 11 bp and 5 bp, respectively; S4 Fig).

The structure of "the body" of *B. germanica* SINEs was investigated by comparing their nucleotide sequences with previously described TEs of various species presented in the following databases: RepBase [8], SINE BASE [9], and NCBI (https://www.ncbi.nlm.nih.gov) using the program CENSOR [61, 62] and regular online Blastn search.

We were unable to identify any new or previously described CORE sequences. For most of the SINEs of *B. germanica*, sequences similar to the sequences of autonomous retrotransposons, potential partners of the studied SINEs, were also not identified. The exceptions are Sbg8 and Sbg9, which contain the same sequences as the sequences of the *Locusta migratoria* RTE retrotransposon (Figs 1 and S5). Note that Sbg9 is a dimeric SINE containing the Sbg8 nucleotide sequence.

"The tails" of Sbg2 and Sbg7 represent poly(A) sequences, whereas all other *B. germanica* SINEs contained short tandem repeats (STRs) in this structural part of retrotransposons. Interestingly, each of the Sbg1, Sbg4, Sbg6, Sbg8, and Sbg9 copies identified in the *B. germanica* genome contains perfect microsatellite motifs: "accttt", "tggaa", "ca", "ttag", and "ttag", respectively. However, only a few copies of Sbg5 contained the motif "tcaga". Sbg3 copies contain different repeat variants, apparently representing different variants of duplications of short nucleotide sequences of the parent SINE copy (Figs 1 and S3).

The molecular mechanisms responsible for changes in the lengths of STRs have been the subject of study over the past several decades since the discovery of these structural formations in eukaryotic genomes. The main mechanism leading to an increase/decrease in the number of repeats in the STR locus is now generally accepted to be slipped strand mispairing that occurs during genomic DNA replication [71]. The number of STRs in SINE copies integrated into the genome can change according to this mechanism. At the same time, in the process of transpositions of retrotransposons, in particular SINEs, there is an additional stage at which duplications can occur, namely, the synthesis of a cDNA copy of TE, due to the activity of the reverse transcriptase enzyme. At this stage, erroneous reinitiation of the start of cDNA synthesis may occur. The presence of 3'-end repeats of different lengths and nucleotide compositions

in different Sbg3 copies (S3 Fig) might be an indication that secondary initiation of synthesis of cDNA of a TE copy may be the primary cause of STR formation in SINEs.

All copies of Sbg1-Sbg9 retrotransposons identified in the genome of *B. germanica* had duplications of the integration site, represented by direct repeats 10–15 bp long, flanking the retrotransposons (Figs 1 and S3). Comparison of the nucleotide sequences of direct repeats flanking various copies of each of the described types of SINEs integrated into the genome did not reveal any peculiarities of their nucleotide composition, which indicates the sequence-independent (random) type of integration of these TEs.

It is known that in the process of transposition of TEs in the genome of eukaryotic organisms, the number of their copies increases; however, due to the accumulation of random mutations, a part of the TE copies integrated into the genome turns into degenerate copies that have lost the ability to transposition. Obviously, the greater the percentage of degenerate copies relative to the total number of copies of a given TE type, the greater the evolutionary age of a particular TE.

To determine the number of copies of described SINEs, the consensus sequences of each of the SINEs were compared with the nucleotide sequence of the *B. germanica* genome at the given parameters—90% and 80% similarity over the entire length of the SINE consensus sequence. This approach allows the determination of both the number of copies that are most similar to the consensus sequence and the percentage of relatively degenerate copies. The absolute values of the number of copies of the studied SINEs and graphical representations of their distribution in the genome are presented in S3 Fig; Fig 2A is a graph showing the copy number variability of each type of SINE depending on the degree of similarity with consensus sequences.

In general, the number of copies varies considerably depending on the SINE type: from 3 (Sbg9) to 1849 (Sbg4) copies with 90% or more similarity to consensus sequences. The number of "degenerate" copies, reflecting the evolutionary age of a particular SINE type, also varies widely depending on the SINE type (Fig 2). Fig 2B presents a diagram showing the proportion of copies with the highest (90% or more) similarity to consensus sequences among copies with 80% or more similarity. The copies of Sbg2 were shown to have the highest level of DNA sequence variability, demonstrating that they are among the oldest SINEs present in the *B. germanica* genome. Based on the described criterion for assessing the evolutionary age of SINEs, we should conclude that the most evolutionarily young variants of SINEs represented in the *B. germanica* genome are Sbg4, Sbg7, and Sbg9 (Fig 2B).

SINEs can be subdivided into distinct subfamilies by specific diagnostic nucleotide changes. Older subfamilies are generally very abundant, while younger subfamilies have fewer copies [1, 4]. Note that we were unable to identify any subfamilies within the SINEs described by us.

Sbg9 deserves special consideration. This TE is a dimeric SINE formed by the combination of two SINEs: Sbg1 and Sbg8. Apparently, among the SINEs described by us, Sbg9 is the most evolutionarily young, formed relatively recently, and, probably, for this reason it is represented in the genome by such a small number of copies. Note that dimeric SINEs are quite widespread in the described genomes of living organisms [for review, see 1, 4].

From our point of view, the SINEs described in the genome of the German cockroach, *B. germanica*, may have not only general scientific value but also applied value. *B. germanica* is a synanthropic species of organisms that can live both in human residential premises and in public institutions and on livestock farms. The spectrum of the negative influence of these insects on human life is unusually wide. A particular danger is the ability of these insects to be carriers of pathogenic microorganisms and cause severe allergic diseases in humans [72–75]. In this regard, understanding the structure of *B. germanica* populations and the possibility of determining the migration flows of this insect species are of particular importance. Currently,
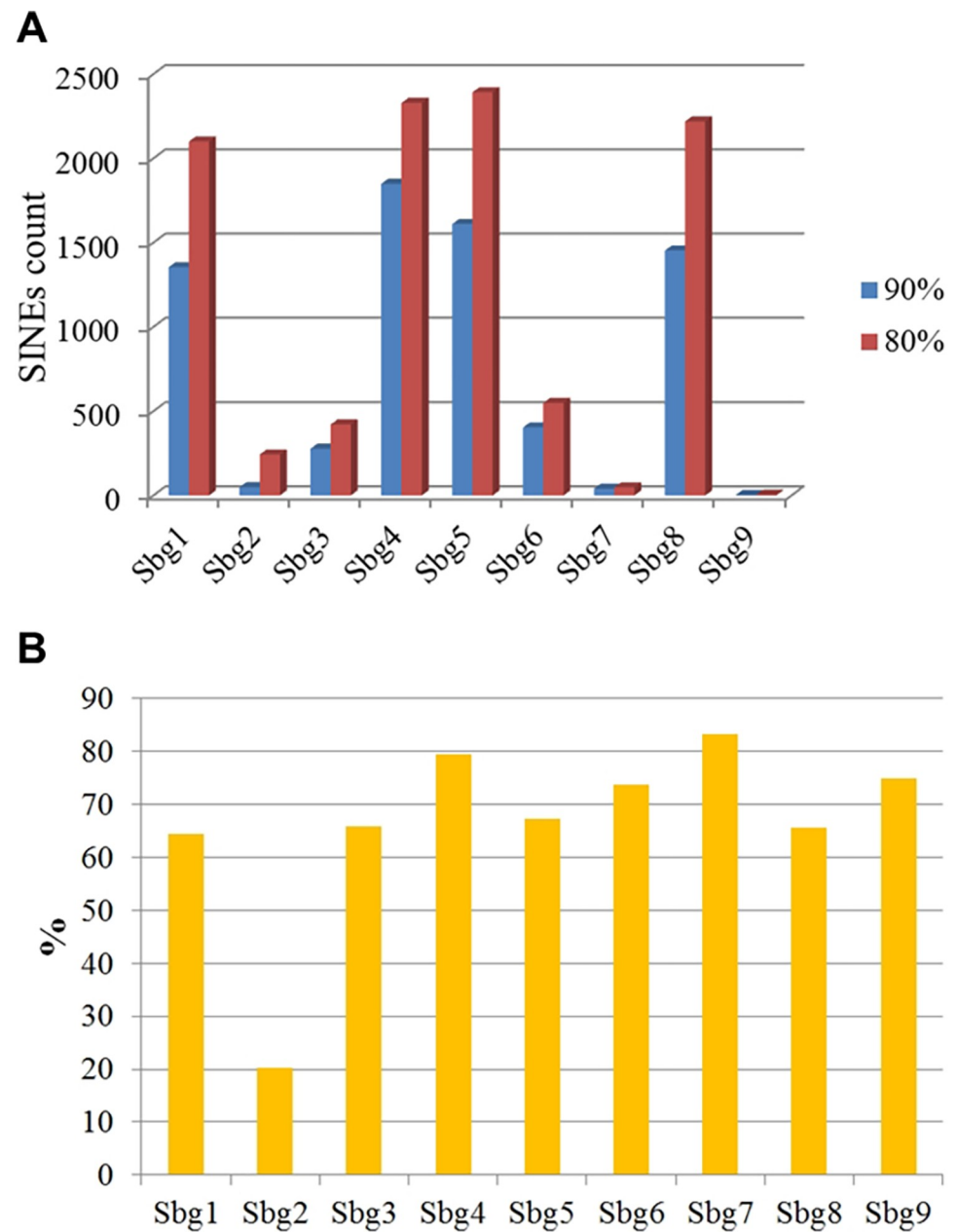
**Fig 2. Analysis of the copy number of the described SINEs.** (A) The ratio of the copy numbers of SINEs of each type (Sbg1-Sbg9) having different degrees of similarity in relation to their consensus sequences: blue and red bars-90% or more and 80% or more similarity, respectively. The abscissa is the SINE type; the ordinate is the number of SINE copies; (B) For each type of SINE, proportion of copies with the highest (90% or more) similarity to consensus sequences among copies with 80% or more similarity. The abscissa is the SINE type; the ordinate shows the percentage.

https://doi.org/10.1371/journal.pone.0266699.g002

several types of molecular genetic markers are used to solve the problems of population genetics of *B. germanica*: polymorphism of the ribosomal RNA gene cluster [76]; polymorphism of the length of microsatellite loci [77–79]; pattern of 5'-truncated copies of R2 retrotransposon [80]. Despite the rather high resolution of the methods used, the development of new molecular genetic markers remains relevant.

SINEs are represented in the genomes by many copies, and the transposition of young copies occurs constantly at a certain level. Based on these properties, we can assume that SINEs

can be considered unique informative molecular genetic markers that make it possible to differentiate populations of eukaryotic organisms. Indeed, at present, this type of marker is actively used for these purposes [for example, see 81–87]. In this study, we first described the structure of SINEs of the German cockroach, *B. germanica*. Over a thousand copies of some types of SINEs were shown to be present in the genome of *B. germanica*. We can assume that, as has been shown for other species of organisms studied in this regard, the SINEs described by us can be used to analyze the polymorphism of the integration sites of these TEs and, as a consequence, to develop a new type of molecular genetic marker that allows differentiation of populations of *B. germanica*. However, it is obvious that the resolution of the proposed markers must be verified experimentally.

## piRNA and SINEs of *B. germanica*

piRNAs play an important role in the control of the transpositional activity of mobile elements [39–49]. Do piRNAs take part in the regulation of the transposition activity of the SINEs of *B. germanica*?

To address this question, we checked if piRNAs sequences would map to the SINEs of *B. germanica*. Fig 3A shows the result of piRNAs mapping to the Sbg1 consensus sequence (205 reads); S6 Fig shows the result of the piRNA reads mapping to the Sbg1, Sbg3-Sbg6, and Sbg8 consensus sequences. In addition, S6 Fig shows the exact number of piRNA reads mapped and the relative value of the number of reads, reflecting the number of reads per 100 bases of SINE sequence per one thousand mapped reads. Overall, mapping of piRNA reads to the consensus sequences of SINEs of *B. germanica* showed that two of the nine SINEs described by us (Sbg3 and Sbg7) do not contain nucleotide sequences that are complementary to piRNAs selected in the above-described way; at the same time, from 46 (Sbg2) to 2290 (Sbg4) piRNA reads were mapped to the consensus sequences of the remaining SINEs. The results obtained indicate that the transposition activity of at least some of the SINEs described by us could be regulated by piRNAs.

As noted above, piRNAs can regulate the transpositional activity of TEs not only through RNA decay but also through DNA methylation and histone modification of TE sites. DNA methylation and histone modification lead to a change in chromatin conformation and cause a change in the dynamics of transcription of a specific TE [44–49]. Obviously, in the case of localization of TEs in gene introns, a change in the chromatin conformation in a given region of the genome can lead to a change in the level of transcription of the gene in the introns of which the TE is located.

Are SINEs (or their degenerate copies) localized in the introns of *B. germanica* genes, and, if so, are piRNAs, potentially capable of causing local changes in chromatin conformation, mapped to these SINE copies?

To search for answers to these questions, we identified all genes localized in the first hundred contigs of the *B. germanica* genome, for which the potential functional activity has been determined and for which the exon-intron structure has been described. A total of 801 genes were identified. Mapping of piRNAs (included in the above-described pool of 9261 piRNA reads–S2 Fig) to these genes showed that 100 or more piRNA reads were localized within each of 241 genes, and more than 1000 piRNA reads were localized within each of 100 genes. Note that all piRNAs were mapped only within the introns of the genes described. We did not find any piRNA cluster sequences within the introns of the genes described, only fragments of degenerate copies of SINEs were found in the introns described. Table 1 shows the results of the analysis of 20 genes to which the largest number of piRNA reads was mapped; S1 Table shows the result of the analysis of 100 genes, for each of which 1000 or more piRNA reads
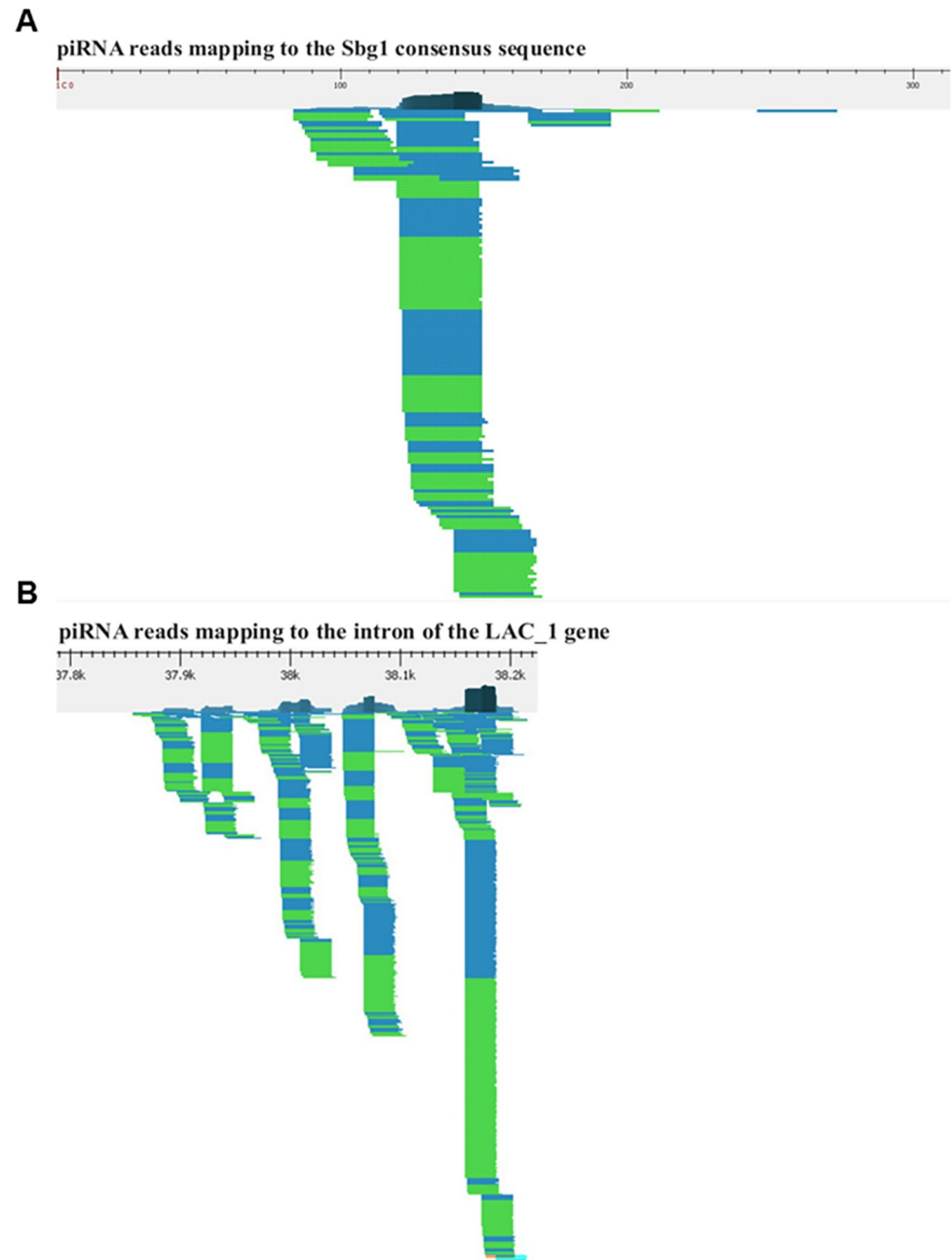
**A**

piRNA reads mapping to the Sbg1 consensus sequence

**B**

piRNA reads mapping to the intron of the LAC_1 gene

**Fig 3.** The result of piRNA reads mapping to the (A) Sbg1 consensus sequence; (B) intron of the LAC_1 gene. Reads mapped in direct orientation are highlighted in blue, and reads mapped in reverse complement orientation are green.

were mapped. The names of the genes, their coordinates in the genomic DNA sequence, the number of mapped piRNA reads are presented in Table 1 and S1 Table. For example, 2978 piRNA reads were mapped to the introns of the LAC_1 gene; Fig 3B shows the fragment of the distribution of piRNA reads within this gene.

Thus, the results obtained show that SINEs of varying degrees of degeneracy relative to their consensus sequences are localized within the introns of some of the genes of *B. germanica*. In addition, a significant number of SINE-related piRNAs derived from piRNA clusters

**Table 1. The result of the analysis of twenty genes to which the largest number of SINE-related piRNA reads was mapped.**

| Contig ID | Strand | Start | End | Gene ID | Gene full name | Reads count |
|---|---|---|---|---|---|---|
| PYGN01000005 | R | 832095 | 893324 | C1TC | C-1-tetrahydrofolate synthase | 3557 (F—1798; R—1759) |
| PYGN01000073 | F | 440373 | 516715 | LAC_1 | Lachesin | 2978 (F—1450; R—1528) |
| PYGN01000033 | F | 355625 | 439785 | Tdc-1_0 | Tyrosine decarboxylase | 2889 (F—1547; R—1342) |
| PYGN01000033 | F | 656364 | 694669 | K11H3-3 | Putative tricarboxylate transport protein | 2717 (F—1343; R—1374) |
| PYGN01000022 | F | 2406772 | 2456620 | lhx3 | LIM/homeobox protein Lhx3 | 2716 (F—1396; R—1320) |
| PYGN01000072 | R | 549086 | 581815 | Diap1 | Death-associated inhibitor of apoptosis 1 | 2655 (F—1375; R—1280) |
| PYGN01000005 | R | 2480629 | 2498464 | Or88 | Odorant receptor 88 | 2488 (F—1127; R—1361) |
| PYGN01000049 | R | 1701575 | 1765331 | Cher_2 | Filamin-A | 2478 (F—1267; R—1211) |
| PYGN01000083 | R | 266347 | 296944 | Pex1 | Peroxisome biogenesis factor 1 | 2469 (F—1167; R—1302) |
| PYGN01000050 | R | 461156 | 514648 | Ir76b | Ionotropic receptor 76b | 2433 (F—1150; R—1283) |
| PYGN01000086 | R | 1254016 | 1267262 | SCD | Acyl-CoA desaturase | 2416 (F—1247; R—1169) |
| PYGN01000005 | F | 3153139 | 3200419 | IFT140 | Intraflagellar transport protein 140 | 2414 (F—1183; R—1231) |
| PYGN01000060 | F | 519801 | 563025 | APN1_0 | Aminopeptidase N | 2400 (F—1127; R—1273) |
| PYGN01000016 | F | 1455716 | 1479561 | Ir41a12 | Ionotropic receptor 41a12 | 2290 (F—1109; R—1181) |
| PYGN01000056 | F | 343658 | 370413 | Rcbtb1 | RCC1 domain-containing protein 1 | 2290 (F—1181; R—1109) |
| PYGN01000029 | R | 1139008 | 1166057 | March6 | E3 ubiquitin-protein ligase MARCH6 | 2104 (F—1032; R—1072) |
| PYGN01000024 | R | 2118061 | 2140175 | CCND2 | G1/S-specific cyclin-D2 | 2081 (F—1089; R—992) |
| PYGN01000005 | F | 1745668 | 1812688 | Tle4_1 | Transducin-like enhancer protein 4 | 2072 (F—1020; R—1052) |
| PYGN01000010 | F | 1604309 | 1616070 | Smg9 | Protein SMG9 | 2055 (F—999; R—1056) |
| PYGN01000038 | F | 2126408 | 2158714 | HIRA | Protein HIRA | 2047 (F—1051; R—996) |

**F** and **R** denote forward and reverse complement orientations, respectively.

https://doi.org/10.1371/journal.pone.0266699.t001

are complementary to the preRNA sequences of certain genes and, therefore, are potentially capable of regulating the transcriptional activity of these genes.

In the course of further studies using RNA interference methods, we assume that we will block the synthesis of proteins that make up the PIWI complex and determine the level of the expression of targeted genes in the introns of which the SINEs are located. Note that a similar approach was recently implemented when the role of siRNA in the development of *B. germanica* was studied [56].

## Conclusions

In this study, we first described the structural and functional organization of nine SINE types in the German cockroach, *B. germanica*, genome. For each type of SINE, the number of copies of varying degrees of degeneracy in the genome of this insect species was determined.

It was shown that the transpositional activity of at least some of the SINEs we have described potentially can be regulated by piRNA. We suggest that the regulation of the transcriptional activity of genes with introns containing SINEs of varying degrees of degeneracy can be determined by DNA methylation and/or histone modifications in certain regions of these introns, followed by the formation of repressive heterochromatin using PIWI-related molecular mechanisms. A simple experimental model was proposed to test this hypothesis.

## Supporting information

**S1 File. The description of command lines used for bioinformatics analysis.**
(PDF)

**S2 File. Detection scheme for SINE-related piRNAs.**
(PDF)

**S1 Script. The regular Python script, with the possibility of saving fragments with a length of 1000 nucleotides, used for creation of a database containing the sequences corresponding to determined tRNAs at the 5'-end and an extended region corresponding to the sequence adjacent to tRNA in genomic DNA at the 3' end.**
(PY)

**S1 Fig. The result of alignment of the extended sequences containing one of the variants of the described SINEs.** Areas of DNA sequences corresponding to SINEs are highlighted in blue. Gray background–DNA sequences corresponding to the SINE environment and having a low level of similarity. Vertical lines in red, green, black, and bright blue indicate single nucleotide substitutions.
(PDF)

**S2 Fig. The reads of piRNA from the created database.** piRNAs localized in piRNA clusters were mapped to the described SINE sequences; then, this piRNA fraction was retrieved, and the accumulated pool of piRNAs (9261 reads) was used for subsequent analysis.
(PDF)

**S3 Fig. Nucleotide sequences and structural features of SINEs of *Blattella germanica* (Sbg1-Sbg9). (A)**–The consensus nucleotide sequence of the corresponding SINE with the designation of the tRNA structure. The tRNA nucleotide sequences are highlighted in gray; "A" and "B" boxes are highlighted in blue; nucleotides other than canonical nucleotides are highlighted in yellow background and red font (explanation in the text). **(B)**–Distribution of the corresponding SINE copies in the *B. germanica* genome. Blue vertical lines indicate the positions of SINE localized in direct orientation, red vertical lines—in inverted orientation. **(C)**–The result of the alignment of the consensus sequences and the sequences of twelve similar to the consensus sequence SINE copies presented in the genome, along with the nearest environment. Direct repeats flanking the retrotransposon are highlighted with a green background. Poly(A) sequences and short microsatellite repeats are highlighted in yellow, blue and pink backgrounds. Variable nucleotides are highlighted in red fonts. **(D)**–Alignment of Sbg1, Sbg8 and Sbg9, demonstrating that Sbg9 was formed by combining Sbg1 and Sbg8.
(PDF)

**S4 Fig. Nucleotide sequences and structural features of four SINEs (Talua. Talub, Taluc and Talud) of *Reticulitermes lucifugus*.** The consensus nucleotide sequences of the corresponding SINEs with the designation of the tRNA structure are shown. The tRNA nucleotide sequences are highlighted in gray; "A" and "B" boxes are highlighted in yellow; nucleotides other than canonical are highlighted in red font. The distances from the start of transcription of SINEs to the corresponding "A" boxes are highlighted in pink.
(PDF)

**S5 Fig. The result of the comparison of the consensus Sbg8 nucleotide sequence with the sequences presented in the RepBase database (https://www.girinst.org). (A)**–Masked Sbg8 sequence and the local sequence alignment of Sbg8 and RTE retrotransposons of *Locusta migratoria*. **(B)**–Sequence of RTE retrotransposon of *L. migratoria*. The sequence fragment similar to the 3'-end of Sbg8 is highlighted in blue.
(PDF)

**S6 Fig. The result of piRNA reads mapping to the Sbg1, Sbg3 –Sbg6, and Sbg8 consensus sequences.** Reads mapped in direct orientation are highlighted in blue, and reads mapped in reverse complement orientation are green.
(PDF)

**S1 Table. The analysis of 100 genes, for each of which 1000 or more piRNA reads were mapped. F** and **R** represent forward and reverse complement orientations, respectively.
(PDF)

## Author Contributions

**Conceptualization:** Dmitry V. Mukha.

**Data curation:** Sergei Yu. Firsov, Dmitry V. Mukha.

**Formal analysis:** Sergei Yu. Firsov.

**Investigation:** Karina A. Kosherova.

**Methodology:** Sergei Yu. Firsov, Dmitry V. Mukha.

**Project administration:** Dmitry V. Mukha.

**Resources:** Sergei Yu. Firsov.

**Supervision:** Dmitry V. Mukha.

**Writing – original draft:** Dmitry V. Mukha.

**Writing – review & editing:** Sergei Yu. Firsov, Karina A. Kosherova, Dmitry V. Mukha.

## References

1. Kramerov DA, Vassetzky NS. Short retroposons in eukaryotic genomes. Int Rev Cytol. 2005; 247: 165–221. https://doi.org/10.1016/S0074-7696(05)47004-7 PMID: 16344113

2. Ohshima K, Okada N. SINEs and LINEs: symbionts of eukaryotic genomes with a common tail. *Cytogenet*. Genome Res. 2005; 110: 475–490. https://doi.org/10.1159/000084981 PMID: 16093701

3. Deragon JM, Zhang X. Short interspersed elements (SINEs) in plants: origin, classification, and use as phylogenetic markers. Syst. Biol. 2006; 55: 949–956. https://doi.org/10.1080/10635150601047843 PMID: 17345676

4. Kramerov DA, Vassetzky NS. Origin and evolution of SINEs in eukaryotic genomes. Heredity. 2011; 107: 487–495. https://doi.org/10.1038/hdy.2011.43 PMID: 21673742

5. Okada N. SINEs: short interspersed repeated elements of the eukaryotic genome. Trends Ecol Evol. 1991; 6: 358–361. https://doi.org/10.1016/0169-5347(91)90226-N PMID: 21232509

6. Munemasa M, Nikaido M, Nishihara H, Donnellan S, Austin CC, Okanda N. Newly discovered young CORE-SINEs in marsupial genomes. Gene. 2008; 407: 176–185. https://doi.org/10.1016/j.gene.2007.10.008 PMID: 17988807

7. Kramerov DA, Vassetzky NS. SINEs. Wiley Interdiscip Rev RNA. 2011; 2: 772–786.

8. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA. 2015; 6: 11. https://doi.org/10.1186/s13100-015-0041-9 PMID: 26045719

9. Vassetzky NS, Kramerov DA. SINEBase: a database and tool for SINE analysis. Nucleic Acids Res. 2013; 41(D1): D83–D89. https://doi.org/10.1093/nar/gks1263 PMID: 23203982

10. Kosushkin SA, Vassetzky NS. Extreme diversity of SINE families in amphioxus. Branchiostoma belcheri. Biopolym. Cell Biopolymers and Cell. 2020; 36: 14–22.

11. Wenke T, Döbel T, Sörensen TR, Junghans H, Weisshaar B, Schmidt T. Targeted identification of short interspersed nuclear element families shows their widespread existence and extreme heterogeneity in plant genomes. Plant Cell. 2011; 23: 3117–3128. https://doi.org/10.1105/tpc.111.088682 PMID: 21908723

12. Roos C, Schmitz J, Zischler H. Primate jumping genes elucidate strepsirrhine phylogeny. Proc Natl Acad Sci U S A. 2004; 101: 10650–10654. https://doi.org/10.1073/pnas.0403852101 PMID: 15249661

13. Veniaminova NA, Vassetzky NS, Kramerov DA. B1 SINEs in different rodent families. Genomics. 2007; 89: 678–686. https://doi.org/10.1016/j.ygeno.2007.02.007 PMID: 17433864

14. Nishihara H, Smit AF, Okada N. Functional noncoding sequences derived from SINEs in the mammalian genome. Genome Res. 2006; 16: 864–874. https://doi.org/10.1101/gr.5255506 PMID: 16717141

15. Gogolevsky KP, Vassetzky NS, Kramerov DA. 5S rRNA-derived and tRNA-derived SINEs in fruit bats. Genomics. 2009; 93: 494–500. https://doi.org/10.1016/j.ygeno.2009.02.001 PMID: 19442632

16. Orioli A, Pascali C, Pagano A, Teichmann M, Dieci G. RNA polymerase III transcription control elements: themes and variations. Gene. 2012; 493: 185–194. https://doi.org/10.1016/j.gene.2011.06.015 PMID: 21712079

17. Tatosyan KA, Stasenko DV, Koval AP, Gogolevskaya IK, Kramerov DA. TATA-Like Boxes in RNA Polymerase III Promoters: Requirements for Nucleotide Sequences. Int J Mol Sci. 2020; 21: 3706. https://doi.org/10.3390/ijms21103706 PMID: 32466110

18. Paule MR, White RJ. Survey and Summary: Transcription by RNA polymerases I and III. Nucleic Acids Res. 2000; 28: 1283–1298. https://doi.org/10.1093/nar/28.6.1283 PMID: 10684922

19. Hamada M, Huang Y, Lowe TM, Maraia RJ. Widespread use of TATA elements in the core promoters for RNA polymerases III, II, and I in fission yeast. Mol Cell Biol. 2001; 21: 6870–6881. https://doi.org/10.1128/MCB.21.20.6870-6881.2001 PMID: 11564871

20. Schramm L, Hernandez N. Recruitment of RNA polymerase III to its target promoters. Genes Dev. 2002; 16: 2593–2620. https://doi.org/10.1101/gad.1018902 PMID: 12381659

21. Dieci G, Fiorino G, Castelnuovo M, Teichmann M, Pagano A. The expanding RNA polymerase III transcriptome. Trends Gene. 2007; 23: 614–622. https://doi.org/10.1016/j.tig.2007.09.001 PMID: 17977614

22. Gilbert N, Labuda D. CORE-SINEs: Eukaryotic short interspersed retroposing elements with common sequence motifs. Proc Natl Acad Sci U S A. 1999; 96: 2869–2874. https://doi.org/10.1073/pnas.96.6.2869 PMID: 10077603

23. Ogiwara I, Miya M, Ohshima K, Okada N. V-SINEs: a new superfamily of vertebrate SINEs that are widespread in vertebrate genomes and retain a strongly conserved segment within each repetitive unit. Genome Res. 2002; 12: 316–324. https://doi.org/10.1101/gr.212302 PMID: 11827951

24. Nilsson MA, Janke A, Murchison EP, Ning Z, Hallström BM. Expansion of CORE-SINEs in the genome of the *Tasmanian devil*. BMC Genomics. 2012; 13: 172. https://doi.org/10.1186/1471-2164-13-172 PMID: 22559330

25. Luchetti A, Mantovani B. Conserved domains and SINE diversity during animal evolution. Genomics. 2013; 102: 296–300. https://doi.org/10.1016/j.ygeno.2013.08.005 PMID: 23981965

26. Luchetti A, Šatović E, Mantovani B, Plohl M. RUDI, a short interspersed element of the V-SINE superfamily widespread in molluscan genomes. Molecular Genetics and Genomics. 2016; 291: 1419–1429. https://doi.org/10.1007/s00438-016-1194-z PMID: 26987730

27. Luchetti A, Plazzi F, Mantovani B. Evolution of Two Short Interspersed Elements in Callorhinchus milii (Chondrichthyes, Holocephali) and Related Elements in Sharks and the Coelacanth. Genome Biology and Evolution. 2017; 9: 1406–1417.

28. Han JS. Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions. Mobile DNA. 2010; 1: 15. https://doi.org/10.1186/1759-8753-1-15 PMID: 20462415

29. Ohno S. So much 'junk' DNA in the genome. In Evolution of genetic systems; Editor Smith H.H.; Brookhaven Symposia in Biology. NewYork: Gordon & Breach; 1972. Volume 23, pp. 366–370.

30. Doolittle WE, Sapienza C. Selfish genes, phenotype paradigm and genome evolution. Nature. 1980; 284: 617–618. https://doi.org/10.1038/284601a0 PMID: 6245369

31. Orgel LE, Crick FHC. Selfish DNA: the ultimate parasite. Nature. 1980; 284: 604–607. https://doi.org/10.1038/284604a0 PMID: 7366731

32. Ichiyanagi K. Epigenetic regulation of transcription and possible functions of mammalian short interspersed elements (SINEs). Genes Genet Syst. 2013; 88: 19–29. https://doi.org/10.1266/ggs.88.19 PMID: 23676707

33. Ichiyanagi T, Katoh H, Mori Y, Hirafuku K, Boyboy BA, Kawase M, et al. B2 SINE Copies Serve as a Transposable Boundary of DNA Methylation and Histone Modifications in the Mouse. *Molecular Biology and Evolution*. 2021; 38: 2380–2395. https://doi.org/10.1093/molbev/msab033 PMID: 33592095

34.   Palazzo A, Marsano RM. Transposable elements: a jump toward the future of expression vectors. Crit Rev Biotechnol. 2021; 41: 792–808. https://doi.org/10.1080/07388551.2021.1888067 PMID: 33622117

35.   Britten RJ. DNA sequence insertion and evolutionary variation in gene regulation. Proc Natl Acad Sci USA. 1996; 93: 9374–9377. https://doi.org/10.1073/pnas.93.18.9374 PMID: 8790336

36.   Tsirigos A, Rigoutsos I. Alu and B1 repeats have been selectively retained in the upstream and intronic regions of genes of specific functional classes. PLoS Comput Biol. 2009; 5: e1000610. https://doi.org/10.1371/journal.pcbi.1000610 PMID: 20019790

37.   Dong L, Jinquan Y, Wen-Qiao T, Xing Z, Clay R, Ming Z. SINE Retrotransposon variation drives Eco-typic disparity in natural populations of *Coilia nasus*. Mobile DNA. 2020; 11: 4. https://doi.org/10.1186/s13100-019-0198-8 PMID: 31921363

38.   Wang X, Liu S, Zuo H, Zheng W, Zhang S, Huang Y, et al. Genomic basis of high-altitude adaptation in Tibetan Prunus fruit trees. Current Biology. 2021; 31: 3848–3860.e8. https://doi.org/10.1016/j.cub.2021.06.062 PMID: 34314676

39.   Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, et al. A novel class of small RNAs bind to MILI protein in mouse testes. Nature. 2006; 442: 203–207. https://doi.org/10.1038/nature04916 PMID: 16751777

40.   Aravin A, Hannon GJ, Brennecke J. The Piwi-piRNA pathway provides an adaptive defense in the trans-poson arms race. Science. 2007; 318: 761–764. https://doi.org/10.1126/science.1146484 PMID: 17975059

41.   Brennecke J, Aravin A, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete small RNA-generat-ing loci as master regulators of transposon activity in *Drosophila*. Cell. 2007; 128: 1089–1103. https://doi.org/10.1016/j.cell.2007.01.043 PMID: 17346786

42.   Ernst C, Odom DT, Kutter C. The emergence of piRNAs against transposon invasion to preserve mam-malian genome integrity. Nature Communications. 2017; 8: 1411. https://doi.org/10.1038/s41467-017-01049-7 PMID: 29127279

43.   Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, et al. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. Science. 2007; 315: 1587–1590. https://doi.org/10.1126/science.1140494 PMID: 17322028

44.   Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD. PIWI-interacting RNAs: small RNAs with big functions. Nat Rev Genet. 2019; 20: 89–108. https://doi.org/10.1038/s41576-018-0073-3 PMID: 30446728

45.   Le TA, Rogers AK, Webster A, Marinov GK, Liao SE, Perkins EM, et al. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. Genes Dev. 2013; 27: 390–399. https://doi.org/10.1101/gad.209841.112 PMID: 23392610

46.   Li Z, Tang X, Shen EZ. How mammalian piRNAs instruct de novo DNA methylation of transposons. Sig-nal Transduct Target Ther. 2020; 5: 190. https://doi.org/10.1038/s41392-020-00294-5 PMID: 32895365

47.   Huang X, Wong G. An old weapon with a new function: PIWI-interacting RNAs in neurodegenerative diseases. Transl Neurodegener. 2021; 10: 9. https://doi.org/10.1186/s40035-021-00233-6 PMID: 33685517

48.   Wang C, Lin H. Roles of piRNAs in transposon and pseudogene regulation of germline mRNAs and lncRNAs. Genome Biol. 2021; 22: 27. https://doi.org/10.1186/s13059-020-02221-x PMID: 33419460

49.   Sienski G, Batki J, Senti KA, Dönertas D, Tirian L, Meixner K, et al. Silencio/CG9754 connects the Piwi-piRNA complex to the cellular heterochromatin machinery. Genes Dev. 2015; 29:2258–2271. https://doi.org/10.1101/gad.271908.115 PMID: 26494711

50.   Bewick AJ, Vogel KJ, Moore AJ, Schmitz RJ. Evolution of DNA Methylation across Insects. Mol Biol Evol. 2017; 34:654–665. https://doi.org/10.1093/molbev/msw264 PMID: 28025279

51.   Zhou X, Qian K, Tong Y, Zhu JJ, Qiu X, Zeng X. De Novo Transcriptome of the Hemimetabolous Ger-man Cockroach (Blattella germanica). PLoS ONE. 2014; 9: e106932. https://doi.org/10.1371/journal.pone.0106932 PMID: 25265537

52.   Ylla G, Fromm B, Piulachs MD, Belles X. The microRNA toolkit of insects. Scientific Reports. 2016; 6: 37736. https://doi.org/10.1038/srep37736 PMID: 27883064

53.   Ylla G, Piulachs MD, Belles X. Comparative analysis of miRNA expression during the development of insects of different metamorphosismodes and germ-band types. BMC Genomics. 2017; 18: 774. https://doi.org/10.1186/s12864-017-4177-5 PMID: 29020923

54.   Harrison MC, Jongepier E, Robertson HM, Arning N, Bitard-Feildel T, Chao H, et al. Hemimetabolous genomes reveal molecular basis of termite eusociality. Nat Ecol Evol. 2018; 2: 557–566. https://doi.org/10.1038/s41559-017-0459-1 PMID: 29403074

55. Llonga N, Ylla G, Bau J, Belles X, Piulachs MD. Diversity of piRNA expression patterns during the ontogeny of the German cockroach. J Exp Zool B Mol Dev Evol. 2018; 330: 288–295. https://doi.org/10.1002/jez.b.22815 PMID: 29975449

56. Montañés JC, Rojano C, Ylla G, Piulachs MD, Maestro JL. siRNA enrichment in Argonaute 2-depleted *Blattella germanica*. Biochim Biophys Acta Gene Regul Mech. 2021; 1864: 194704. https://doi.org/10.1016/j.bbagrm.2021.194704 PMID: 33895310

57. Mao H, Wang H. SINE_scan: an efficient tool to discover short interspersed nuclear elements (SINEs) in large-scale genomic datasets. Bioinformatics. 2017; 33: 743–745. https://doi.org/10.1093/bioinformatics/btw718 PMID: 28062442

58. Chan PP, Lowe TM. tRNAscan-SE: Searching for tRNA genes in genomic sequences. Methods Mol Biol. 2019; 1962: 1–14. https://doi.org/10.1007/978-1-4939-9173-0_1 PMID: 31020551

59. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: Architecture and applications. BMC Bioinform. 2009; 10: 421. https://doi.org/10.1186/1471-2105-10-421 PMID: 20003500

60. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009; 25: 1189–1191. https://doi.org/10.1093/bioinformatics/btp033 PMID: 19151095

61. Jurka J, Klonowski P, Dagman V, Pelton P. CENSOR—a program for identification and elimination of repetitive elements from DNA sequences. Computers and Chemistry. 1996; 20: 119–122. https://doi.org/10.1016/s0097-8485(96)80013-1 PMID: 8867843

62. Kohany O, Gentles AJ, Hankus L, Jurka J. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics. 2006; 7: 474. https://doi.org/10.1186/1471-2105-7-474 PMID: 17064419

63. Okonechnikov K, Golosova O, Fursov M. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics. 2012; 28: 1166–1167. https://doi.org/10.1093/bioinformatics/bts091 PMID: 22368248

64. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404

65. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina paired-end reAd mergeR. Bioinformatics. 2014; 30: 614–620. https://doi.org/10.1093/bioinformatics/btt593 PMID: 24142950

66. Rosenkranz D, Zischler H. proTRAC:A software for probabilistic piRNA cluster detection, visualization and analysis. BMC Bioinformatics. 2012; 13: 5. https://doi.org/10.1186/1471-2105-13-5 PMID: 22233380

67. Raina M, Ibba M. tRNAs as regulators of biological processes. Front Genet. 2014; 5: 171. https://doi.org/10.3389/fgene.2014.00171 PMID: 24966867

68. Su Z, Wilson B, Kumar P, Dutta A. Noncanonical Roles of tRNAs: tRNA Fragments and Beyond. Annu Rev Genet. 2020; 54: 47–69. https://doi.org/10.1146/annurev-genet-022620-101840 PMID: 32841070

69. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology. 2009; 10: R25. https://doi.org/10.1186/gb-2009-10-3-r25 PMID: 19261174

70. Berg MD, Brandl CJ. Transfer RNAs: diversity in form and function. RNA Biol. 2021; 18: 316–339. https://doi.org/10.1080/15476286.2020.1809197 PMID: 32900285

71. Levinson G, Gutman GA. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol. 1987; 4: 203–221. https://doi.org/10.1093/oxfordjournals.molbev.a040442 PMID: 3328815

72. Schal C, Hamilton RL. Integrated suppression of synanthropic cockroaches. Annu Rev Entomol. 1990; 35: 521–551. https://doi.org/10.1146/annurev.en.35.010190.002513 PMID: 2405773

73. Brenner RJ. Economics and medical importance of German cockroaches. In Understanding and Controlling the German Cockroach; Rust M.K., Owens J.M., Reierson D.A., Eds.; Oxford University Press: New York, NY, USA; 1995. pp. 72–92.

74. Rosenstreich DL, Eggleston P, Kattan M, Baker D, Slavin RG, Gergen P, et al. The role of cockroach allergy and exposure to cockroach allergen in causing morbidity among inner-city children with asthma. New Engl J Med. 1997; 336: 1356–1363. https://doi.org/10.1056/NEJM199705083361904 PMID: 9134876

75. Gore JC, Schal C. Cockroach allergen biology and mitigation in the indoor environment. Annu Rev Entonol. 2007; 52: 439–463. https://doi.org/10.1146/annurev.ento.52.110405.091313 PMID: 17163801

76. Mukha DV, Kagramanova AS, Lazebnaya IV, Lazebnyl OE, Vargo EL, Schal C. Intraspecific variation and population structure of the German cockroach, *Blattella germanica*, revealed with RFLP analysis of

the non-transcribed spacer region of ribosomal DNA. Med Vet Entomol. 2007; 21: 132–140. https://doi.org/10.1111/j.1365-2915.2007.00670.x PMID: 17550432

77. Crissman JR, Booth W, Santangelo RG, Mukha DV, Vargo EL, Schal C. Population genetic structure of the German cockroach (Blattodea: Blattellidae) in apartment buildings. J Med Entomol. 2010; 47: 553–564.

78. Booth W, Santangelo RG, Vargo EL, Mukha DV, Schal C. Population genetic structure in German cockroaches, *Blattella germanica*: Differentiated islands in an agricultural landscape. J Hered. 2011; 102: 175–183. https://doi.org/10.1093/jhered/esq108 PMID: 20980363

79. Vargo EL, Crissman JR, Booth W, Santangelo RG, Mukha DV, Schal C. Hierarchical genetic analysis of German cockroach (*Blattella germanica*) populations from within buildings to across continents. PLoS ONE. 2014; 9: e102321. https://doi.org/10.1371/journal.pone.0102321 PMID: 25020136

80. Zagoskina A, Firsov S, Lazebnaya I, Lazebny O, Mukha DV. R2 and Non-Site-Specific R2-Like Retrotransposons of the German Cockroach, *Blattella germanica*. Genes. 2020; 11: 1202. https://doi.org/10.3390/genes11101202 PMID: 33076367

81. Tatout C, Warwick S, Lenoir A, Deragon JM. SINE Insertions as Clade Markers for Wild Crucifer Species. Molecular Biology and Evolution. 1999; 16: 1614.

82. Prieto JL, Pouilly N, Jenczewski E, Deragon JM, Chèvre AM. Development of crop-specific transposable element (SINE) markers for studying gene flow from oilseed rape to wild radish. Theor Appl Genet. 2005; 111: 446–455. https://doi.org/10.1007/s00122-005-2017-5 PMID: 15942756

83. Ray DA. SINEs of progress: Mobile element applications to molecular ecology. Mol Ecol. 2007; 16: 19–33. https://doi.org/10.1111/j.1365-294X.2006.03104.x PMID: 17181718

84. Liu D, Li Y, Tang W, Yang J, Guo H, Zhu G, et al. Population structure of *Coilia nasus* in the Yangtze River revealed by insertion of short interspersed elements. Biochem Syst Ecol. 2014; 54: 103–112.

85. Brown H, Thompson R, Murphy G, Peters D, La Rue B, King J, et al. Development and validation of a novel multiplexed DNA analysis system, InnoTyper® 21. Forensic Sci Int Genet. 2017; 29: 80–99. https://doi.org/10.1016/j.fsigen.2017.03.017 PMID: 28391141

86. Chen C, Wang X, Zong W, D'Alessandro E, Giosa D, Guo Y, et al. Genetic Diversity and Population Structures in Chinese Miniature Pigs Revealed by SINE Retrotransposon Insertion Polymorphisms, a New Type of Genetic Markers. Animals (Basel). 2021; 11: 1136. https://doi.org/10.3390/ani11041136 PMID: 33921134

87. Chen C, D'Alessandro E, Murani E, Zheng Y, Giosa D, Yang N, et al. SINE jumping contributes to large-scale polymorphisms in the pig genomes. Mob. DNA. 2021; 12: 17. https://doi.org/10.1186/s13100-021-00246-y PMID: 34183049