

RESEARCH

Open Access

SINE Retrotransposon variation drives Ecotypic disparity in natural populations of *Coilia nasus*



Dong Liu^{1,3,4}, Jinqian Yang¹, Wenqiao Tang^{1,3,4*}, Xing Zhang², Clay Matthew Royster² and Ming Zhang^{2*}

Abstract

Background: SINEs are a type of nonautonomous retrotransposon that can transpose from one site to be integrated elsewhere in an organism genome. SINE insertion can give rise to genetic variants and regulate gene expression, allowing organisms to acquire new adaptive capacity. Studies on this subject have focused on the impacts of SINEs on genes. However, ecological disparities in fish have not yet been explained by SINEs.

Results: New SINEs were isolated from *Coilia nasus*, which has two ecotypes—migratory and resident—that differ in their spawning and migration behaviors. The SINEs possess two structures that resemble a tRNA gene and a LINE retrotransposon tail. Comparison of olfactory tissue transcriptomes, intact SINE transcript copies were detected in only the migratory fish at the initial retrotransposition stage. The SINE DNA copy numbers were higher in the resident type than in the migratory type, while the frequency of SINE insertion was higher in the migratory type than in the resident type. Furthermore, SINE insertions can lead to new repeats of short DNA fragments in the genome, along with target site duplications. SINEs in the resident type have undergone excision via a mechanism in which predicted cleavage sites are formed by mutations, resulting in gaps that are then filled by microsatellites via microhomology-induced replication.

Conclusions: Notably, SINEs in the resident type have undergone strong natural selection, causing genomic heteroplasmy and driving ecological diversity of *C. nasus*. Our results reveal possible evolutionary mechanisms underlying the ecological diversity at the interface between SINE mobilization and organism defense.

Keywords: Ecotypic diversity, Retrotransposable element, Short interspersed nuclear elements, Long interspersed nuclear elements, Natural selection

Introduction

Short interspersed elements (SINEs) are a type of retrotransposon frequently found in eukaryotic genomes; these elements can expand in the genome and generate multiple copies [1, 2]. Some inserted copies may directly affect the functions of individual genes via regulation of expression or creation of novel genes in response to environmental challenges [3]. SINE insertions have been shown to benefit *Drosophila melanogaster* during the spread of this organism out of Africa [4]. SINE insertions upstream of genes

can enhance gene expression and expand gene function [5]. Two functional enhancers in the *POMC* gene originated from ancient insertions in *D. melanogaster* [6]. In addition, a transposon-originated gene associated with high-latitude adaptation was identified in soybean plants [7]. Moreover, the industrial melanism of the peppered moth in Britain resulted from transposon insertion within a gene intron and provides a visible demonstration of an evolutionary response to environmental change [8].

SINEs have contributed to species evolution. Mobilization and nonhomologous recombination of SINEs have generated intraspecific polymorphisms and led to interspecific diversity [9]. The well-studied formation events of salmonid species were found to be correlated with a burst in the dispersion of retrotransposons [10]. In salmon, the mobilization of these SINEs remains ongoing and continues to drive the genomic

* Correspondence: wqtang@shou.edu.cn; mzhang01@uga.edu

¹Key Laboratory of Marine Animal Taxonomy and Evolution, Shanghai Universities, Shanghai 201306, China

²Department of Epidemiology and Biostatistics, University of Georgia, Athens, GA 30602, USA

Full list of author information is available at the end of the article



diversity of the species [11]. In two closely related puffer fish species, transposable elements are responsible for genome size variation, with 2% SINE content in one species and 0.2% in the other [12]. The abundance and diversification of transposable elements are among the major mechanisms driving genomic variation in teleosts [13].

SINEs originated from tRNAs, 7SL RNA or 5S rRNA [2]. A majority of reported SINEs were derived from tRNAs and consist of three regions: a tRNA-related region, a body and a tail region [2]. SINEs are a type of nonautonomous retrotransposon, and in RNA form, SINEs acquire their mobility and the ability to undergo bursts of retrotransposition from long interspersed element (LINE)-encoded proteins [14]. This protein recognizes the 3' tail sequence of the SINE, which is similar to that of a LINE [15], in a procedure that is required for both initiation of reverse transcription of the SINE RNA and integration of the SINE into a new genomic location. SINEs can integrate into gene-dense regions, into specific regions or throughout the genome [1].

SINE insertion may be deleterious to the organism genome [9]. However, these harmful insertions can be eliminated by multiple mechanisms in organism, such as breeding systems [16], purifying selection [17], ectopic recombination [18], and genetic drift [19]. The retrotransposons themselves, via mutation, can evolve to form a "fossil" state that is no longer mobile. In particular, the repeats within the 3' tail regions of SINEs shrink rapidly [20]. Furthermore, the promoter in the tRNA-related region can accumulate mutations to disrupt the initiation of SINE transcription. The changes in SINE length, caused by random deletions can alter SINE RNA folding, leading to loss of mobility [2].

Considering the evolutionary importance of SINEs in the generation of genomic diversity, it is reasonable to consider SINEs as suitable genetic markers in systematic biology and as a tool to track common ancestry among specific taxa [21, 22]. It is widely accepted that the evolution of SINEs is unidirectional and irreversible [23], and SINE markers are homoplasy free [21]. Therefore, information regarding the ancestral states (present or absent) of the SINE-inserted loci can be very useful for phylogenetics at both the species and population levels. The application of characterization of SINE insertions as markers has been largely limited to humans, primates, whales, and a few fish species [24–26]. The major obstacles impeding broader application of SINE insertions include both isolation of SINEs from organisms for which genomic sequences are unknown and identification of a large number of polymorphic loci in genomes [21].

Currently, the retrotransposons associated with life history disparity in natural animal populations remain poorly studied. Our study subject, *Coilia nasus*, is an anadromous fish that has undergone rapid ecotype

divergence following population expansion from the ocean to fresh water [27]. Two ecotypes of *C. nasus*—the resident type and the migratory type—are found in the Yangtze River in China, and these ecotypes differ in their spawning/migration behavior and exhibit distinct morphological divergence. During the spawning season from March to August, *C. nasus* migrate from coastal water up to the Yangtze River, even penetrate 1400 km upstream for breeding. The gonads of the fish are developing via migration, and matured fish spawn in the reaches of these rivers and adjacent lakes. After reproduction, these fish and their progeny migrate back to the ocean from September to November of the same year. Their phenotypic differences include body shape, vertebral counts, anal fin counts, eye size and gill raker counts. In particular, the maxilla of resident type is short and does not reach the edge of the gill cover, while the migratory type has a long maxilla that extends well beyond the gill cover [28]. A high level of genetic diversity has also been observed between the two types via mitochondrial DNA marker analyses [29, 30].

C. nasus is an economically important fish: catches of the migratory type were greater than 3000 tons in the lower reaches of the Yangtze River and accounted for 35–50% of the total fish catches in this area in the 1970s. However, the catches sharply decreased to 50 tons in 2005 [31]. The migratory fishes have been protected and the Chinese government stopped issuing special fishing licenses for this ecotype in 2019. In contrast, catches of the resident type from lakes increased from 640 tons in 1952 to more than 20,000 tons in 2004, making *C. nasus* the dominant species (~64% of the total fish catches) [32].

The Yangtze River is ~6000 km long, with some reaches and lakes interwoven. The overlapping habitats between the migratory and resident types of *C. nasus* in the Yangtze River have created challenges for the determination of gene flow between these two types. Molecular techniques have offered the opportunity to identify and delineate fish population structures that may not be apparent based on phenotypic or behavioral characteristics alone. The two ecotypes of *C. nasus* in the Yangtze River provide a good system for the study of SINE biology and determination of whether and how mobile elements have influenced the population ecology.

In this study, we selected representative samples from the collected 1200 samples, including both *C. nasus* ecotypes, which exhibit different reproductive behaviors. We examined the insertion patterns of 71 SINE loci to better understand the life history diversity of *C. nasus*. Our results provide strong evidence that SINE motility has driven the genomic heterogeneity of *C. nasus* and is associated with ecological diversity.

Materials and methods

Sample collection

During 2009–2013, we collected approximately 1200 individual samples of *C. nasus* from six locations in China. Of these locations, five were in the middle and lower reaches of the Yangtze River: Chongming (CM), Jingjiang (JJ), Taihu Lake (TH), Poyang Lake (PY), and Dongting Lake (DT). The 6th location was in Xiangshan (XS), which is in the coastal region of the East Sea in China (Fig. 1). Sampling from these six locations ensured the inclusion of all ecotypes of *C. nasus*.

The CM and JJ samples were captured by net when the fish returned home from the sea to freshwater habitats. After the anadromous individuals had left Poyang Lake and Dongting Lake and returned to the sea, resident individuals were still present in the lakes, and samples were caught by net. In order to avoid of using admixture/hybridization individuals of the two ecotypes as reported in our previous study [33], the representative samples of the anadromous ecotype from CM, JJ and XS populations were chose from the 1200 samples based on the maxilla extending well beyond the gill cover, and the captured time between March to April (Fishes began to migrate from sea back to river). The representative samples of the resident ecotype from PY and DT were selected from the 1200 samples based on the maxilla significantly shorter than the length of the gill cover, and the capture times within January (Fishes are resident in lakes). Ten

individual genomes from each population were examined for SINE insertion analysis. The genome of one sample from CM was used to isolate SINE insertion sites. Additionally, three samples of *Coilia mystus* were collected from the Minjiang River in Fujian Province and were used as an outgroup related to *C. nasus*. Muscle tissue for downstream analyses was preserved in 95% ethanol.

Isolation and identification of SINES

Genomic DNA enrichment was performed according to the amplified fragment length polymorphism (AFLP) technique as described previously [34]. This step was carried out to isolate and characterize the SINES in *C. nasus*. The genomic DNA was digested with EcoRI and BclI (Takara, China), purified, and ligated to either the EcoRI adaptor [34] or the modified MseI adaptor (Additional file 1: Table S1) using T4 DNA ligase (Takara, China). Genomic DNA was amplified by PCR with EcoRI and MseI primers (Additional file 1: Table S1). A total of 5 PCRs for genomic DNA of each individual were performed with 14.5 μ L of ddH₂O, 1.0 μ L of 10 mM dNTPs, 2.5 μ L of 10 \times PCR buffer, 0.5 μ L of each 10 μ M primer, 0.5 μ L of Taq, and 0.5 μ L of the adaptor-ligated DNA product. The PCR program was as follows: 94 $^{\circ}$ C for 3 min, followed by 17 cycles of 94 $^{\circ}$ C for 30 s, 55 $^{\circ}$ C for 30 s, and 72 $^{\circ}$ C for 1 min 30 s; and finally, a 10-min extension step at 72 $^{\circ}$ C. The 5 PCR products were electrophoresed on a 2% agarose gel. Fragments



Fig. 1 Sampling locations used for *C. nasus* collection. The numbered dots correspond to the following locations along the Yangtze River: 1, Xiangshan; 2, Chongming; 3, Jingjiang; 4, Taihu Lake; 5, Poyang Lake; and 6, Dongting Lake.

in the 500- to 1000-bp range were eluted using a DNA gel purification kit (Sangon, China).

A magnetic particle system was established using MagneSphere magnetic separation products (Promega, Germany) following the manufacturer's instructions. The procedure for SINE isolation from the genome of *C. nasus* was described in our previous study [35]. Briefly, the AFLP-enriched DNA fragments were denatured and then hybridized with the 5'-biotin-labeled probe sequence specific to a 120-bp internal region of the SINEs. The biotin present at the 5' end of the probe specifically attached to the magnetic particles. The probe-target DNA complexes were eluted from the magnetic particles.

The eluted DNA was used as a template for PCR with EcoRI and MseI primers (Additional file 1: Table S1). The PCR products were electrophoresed on a 2.0% agarose gel. Fragments in the 500–1000-bp range were eluted and ligated into the pGMD19-T vector (Takara, China), and the constructs were transformed into *E. coli* DH5 α competent cells. Recombinant colonies were screened by PCR with the Sc-F and Sc-R primers (Additional file 1: Table S1), which correspond to the sequence of the conserved region of SINEs. The positive colonies were sequenced using an automated DNA sequencer (ABI PRISM 3730).

Sequences of positive clones were aligned with the primary SINE consensus sequence (Additional file 2: Table S2). Then, the sequences with high scores were searched against zebrafish DNA sources using the RepeatMasker web server (www.repeatmasker.org) to classify the repetitive elements. Finally, sequences that were similar to known repeat elements in Repbase were retrieved via the CENSOR algorithm [36].

Tails of retrotransposon LINEs and secondary structures

To test whether the new SINEs transpose via a tail region similar to that of retrotransposon LINEs, the genome walking method was used to identify the tail sequence of the LINE using one arbitrary degenerate primer obtained from a kit (Takara, China) and a special primer designed specifically for the LINEs of *C. nasus* [35]. The entire PCR process was conducted according to the manufacturer's instructions for the kit. The secondary structures of the tail portions of the SINEs and LINEs were predicted using the Mfold tool as previously described [37].

The isolated SINE elements were used to determine the consensus sequence for the primary sequence of the SINE family. To determine the possible tRNA ancestry of SINEs, the tRNA-related regions of the SINEs were used to verify the tRNA-like secondary structure via tRNAscan-SE [38].

Transcriptomic analysis of SINEs

Transcriptomic analysis was used to determine whether the expression of the new SINEs differed in both ecotypes of *C. nasus*. Total RNA was extracted from

olfactory tissues of *C. nasus* using the TRIzol Kit (Invitrogen, USA). cDNA library construction and sequencing, sequence data processing, and de novo assembly for RNA-seq assay were performed as previously described [39]. In brief, individuals with phase III of gonad, similar ages and size were used for transcriptome analysis of ecotypes. One individual of JJ and a mixture with 3 individual of CM were used as the migratory transcriptome samples. A mixture of 3 individuals of PY and 3 individuals of DT was used as the resident transcriptome samples. The contig data were used to construct a local BLAST library. Blastn was used to search for significant hits in the library using the consensus SINE sequence as a query. When a contig matched the query with a high score (> 80), the function of a gene within the matched contig was annotated via Blastx against the GenBank database.

DNA copy number for SINE analysis

To determine whether the SINEs have undergone natural selection, the genomic copy numbers of the SINEs were determined by real-time PCR. The plasmids with SINE insertions and the genomic DNA of *C. nasus* samples collected from the six locations were prepared as standards and samples for real-time PCR. DNA concentrations were measured with a spectrophotometer, and ten-fold serial dilutions were prepared as templates for real-time PCR in an ABI 7500 instrument (ABI). Ten samples were used, and three replicates were included for each. The real-time PCR program was as follows: 95 °C for 5 min, followed by 40 cycles of 95 °C for 10 s, 55 °C for 20 s and 72 °C for 30 s. The 20 μ L PCR mixture included 0.5 μ L of each primer (Sc-F and Sc-R) and 10 μ L of HRM Master Mix 2 \times (Qiagen, China). Finally, a melting curve analysis was performed after amplification. Standard curve preparation and data analysis were performed with MJ Opticon Monitor (MJ Research, Waltham, MA). The average genome size of *C. nasus* considered to be 3.534 pg based on a previous report [40] is required to normalize the average size of population genomes. Significance in copy number difference was calculated by ANOVA in SPSS 16.0.

SINE insertion polymorphisms

During pre-examination, samples collected from the six sampling sites (10 individuals per site) (Fig. 1) were screened to analyze SINE insertion polymorphisms. The primers were designed with Primer Premier 6.0 [41] and were specific for the flanking sequences of each insertion. Finally, five of the 71 total insertion loci (obtained via SINE DNA capture and cloning, as described above) showed polymorphic bands (the PCR primers are listed in Additional file 3: Table S3). PCR was performed for 35 cycles in a 2:1 reaction volume using Taq Master Mix 2 \times (TianGen, China). The allele bands of the five insertion loci were separated, cloned, and sequenced. To

confirm the presence/absence of SINE insertions, three to five clones were selected for sequencing per allele at a single locus.

Allele frequencies, gene diversity and exact p -value tests for Hardy-Weinberg equilibrium departure were estimated using PowerMarker v.3.25 [42]. Statistical significance for the Hardy-Weinberg equilibrium departure test was considered to be reached at $p = 0.001$. The index for each locus showed the intensity and direction of deviation from the overall value. An AMOVA test was performed to clarify the hierarchical apportioning of SINE frequency variance using Arlequin ver. 3.5 [43]. To analysis genetic relationships among populations, a neighbor-joining (NJ) tree was constructed based on allele frequencies using PowerMarker v.3.25 [42].

The evolutionary history of SINEs in *C. nasus* was evaluated by utilizing *C. mystus* as a comparison outgroup. Three individuals per species were evaluated by PCR with these locus insertion primers. The presence of an expected insertion in *C. mystus* suggests a relatively old insertion in *C. nasus*, and the absence of an expected insertion in *C. mystus* suggests a relatively recent insertion in *C. nasus*.

Results

Molecular characterization of SINEs

A total of 259 SINE clones were identified in the *C. nasus* genome. After sequencing, redundant sequences were deleted. The resulting sequences were aligned with

the SINE consensus sequence. We obtained 71 validated positive clones with SINEs. These SINEs vary between 178 and 285 bp in size and contain the flanking sequences of the insertion locus, which have target site duplications (TSDs) ranging from 2 to 31 nucleotides, although a majority of the TSDs range in size between 3 and 8 nucleotides (Fig. 2). These SINE clones are 77–94% similar to the SINE consensus sequence, confirming that these clones belong to the SINE family. The sequences of the SINEs in the *C. nasus* genome showed that the tRNA-related regions of the SINEs, which are approximately 75 bp in length, originated from six ancient tRNA genes (Additional file 4: Table S4) and can be folded into perfect cloverleaf tRNA structures, despite low similarity (49%) between their primary sequences (Fig. 3). Overall, our results indicate that these SINEs originated from ancient tRNA genes.

The body regions of SINE elements vary in length from 60 to 170 bp because of either a 41–91-bp insertion or 3–31 bp indels (insertions and deletions). We found that the deletions occurred randomly, but the insertions occurred largely in the nuclear repeat regions. The 3' tail regions of the SINE elements exhibited 76–99% conservation and contained a 56-bp tail block similar to that of L2–2-DRE, a LINE retrotransposon from the zebrafish genome [44]. Considering the importance of the conserved poly(A) or TGTAA motif in the 3' tail region of SINE transposons [14], we searched for these motifs in SINEs with 3' tails similar to those of LINES. Three

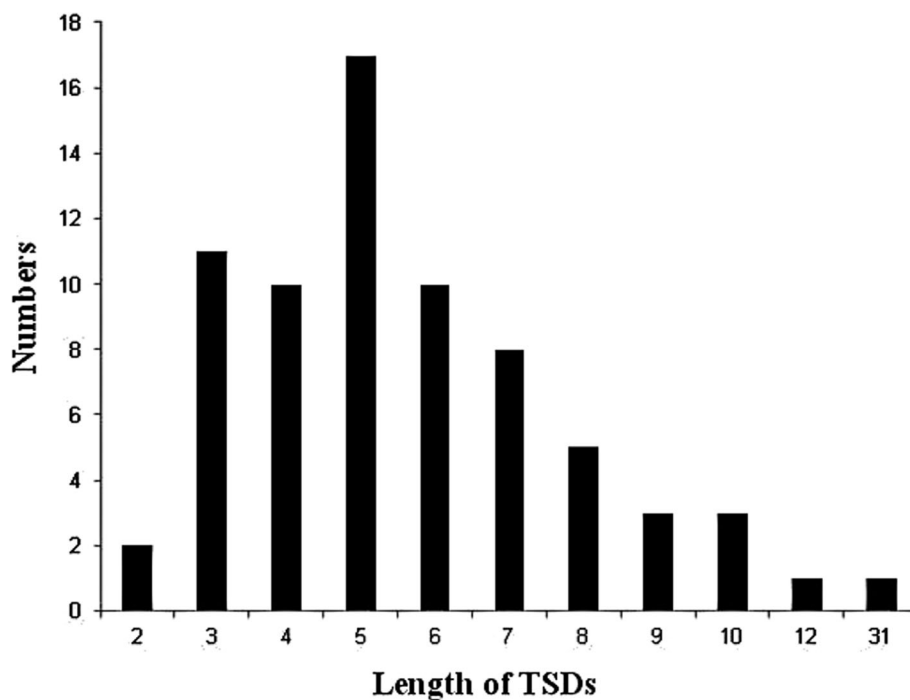
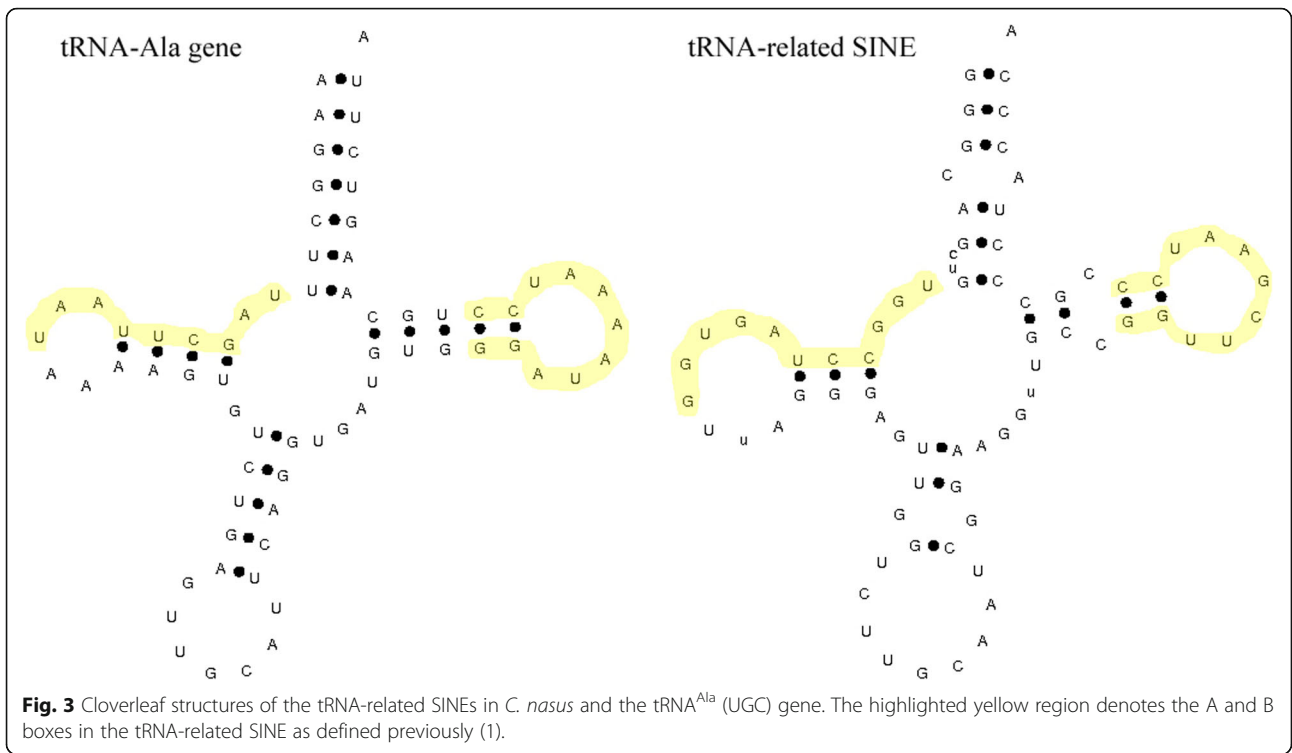


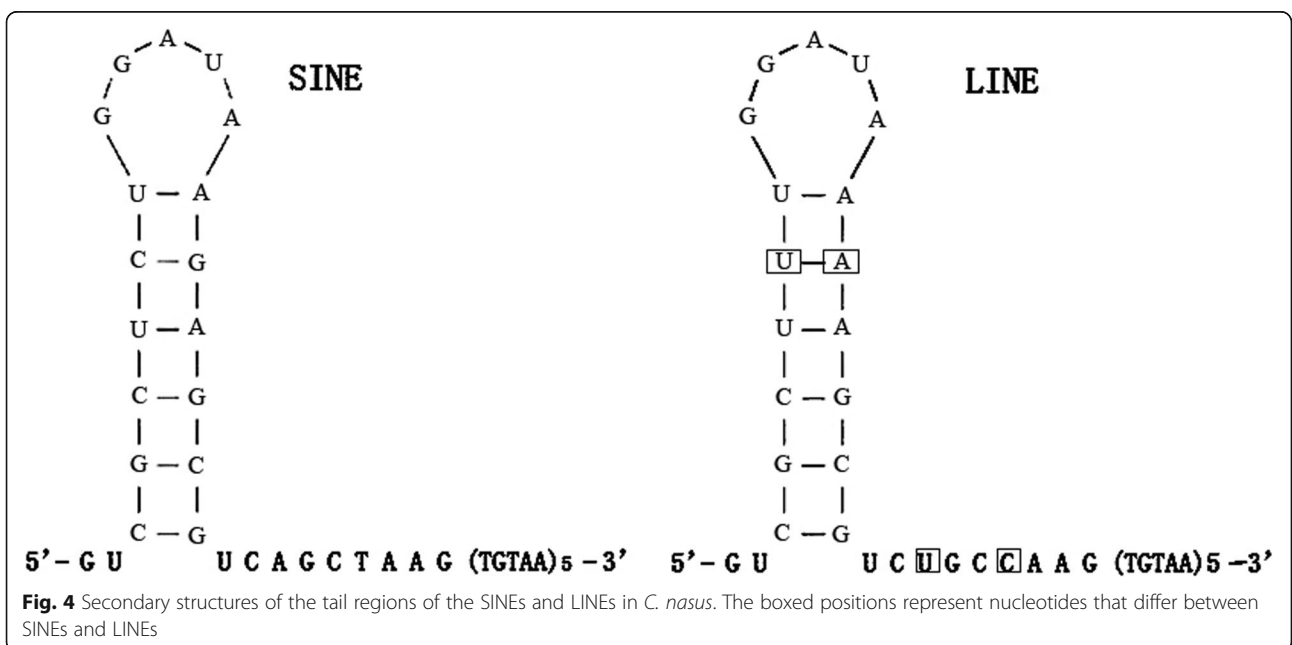
Fig. 2 Length distribution of target site duplications (TSDs) in the SINE insertion flanking regions of *C. nasus*



motif patterns, poly(A), TGTA A, and TGTA A-poly(A), were identified.

The transposition ability of these SINES in the *C. nasus* genome depends on whether the organism shares these motif patterns in the tails of its LINES. Interestingly, the 3' tails of LINES from *C. nasus* range from 337 to 402 bp and exhibit 58–64% similarity. This low

similarity is due to the variation in the 3' region of the LINES; however, this region harbors two types of motifs, namely, poly(A) and TGTA A motifs. Both motifs exhibited high similarity with the 53-bp tail of the SINES and shared a stem-loop secondary structure and five TGTA A repeats (Fig. 4). The results supported that the nonautonomous SINES can mobilize via both the slippage



reaction and recognition by the LINE reverse transposase [15].

SINE transcription differs between ecotypic populations

To investigate whether the RNA intermediates of SINEs that involve retrotransposition are transcribed at the replicative stage of mobilization, two transcriptomes of olfactory tissues from the migratory and resident types were constructed by de novo assembly. A data set of migratory transcriptome comprised of one individual of JJ (NCBI SRA: SRP035517) and a mixture with 3 individual of CM (NCBI SRA: SRP100816). The other data set of resident transcriptome consisted of a mixture with 3 individual of PY (SRP035517) and a mixture with 3 individual of DT (SRP100816). With the two data sets, 343,265 and 491,297 contigs were obtained from the combined reads in the transcriptomes of two ecotypes, migratory type and resident type, respectively. Using the SINE consensus sequence as a query, we obtained significant hits as SINE copy numbers from the two sets of contig data. Three complete transcript copies of the SINEs were found in the migratory type but not in the resident type (Fig. 5). A complete SINE element has three regions: a tRNA-related region (75 bp), a body

region (75 bp) and a 3' tail region (56 bp) (Fig. 5). The copy number of the tRNA-related region in the migratory type was slightly lower than that in the resident type. The copy number of the 3' tail region in the migratory type was significantly higher than that in the resident type (Fig. 5). The observed difference in SINE expression in these two *C. nasus* ecotypes may explain their genetic variation and species differentiation.

Out of the total number of contig hits with SINEs, 42 of the resident type and 34 of the migratory type, contain genes with annotated functions related to signal transduction, cell structure, and transporter activity (Additional file 5 and Additional file 6: Tables S5 and S6). Importantly, two genes within the SINE-hit contigs were found to be known genes that are important for migrating *C. nasus* fish, namely, the genes encoding the S100 calcium-binding protein and the interferon regulatory factor. These genes that exhibited differences between the two transcriptomes may explain the different life histories of *C. nasus*.

SINEs underwent strong natural selection that resulted in genomic disparity between populations

Given our finding that the complete RNA transcripts of SINEs were present in the migratory type but not in the

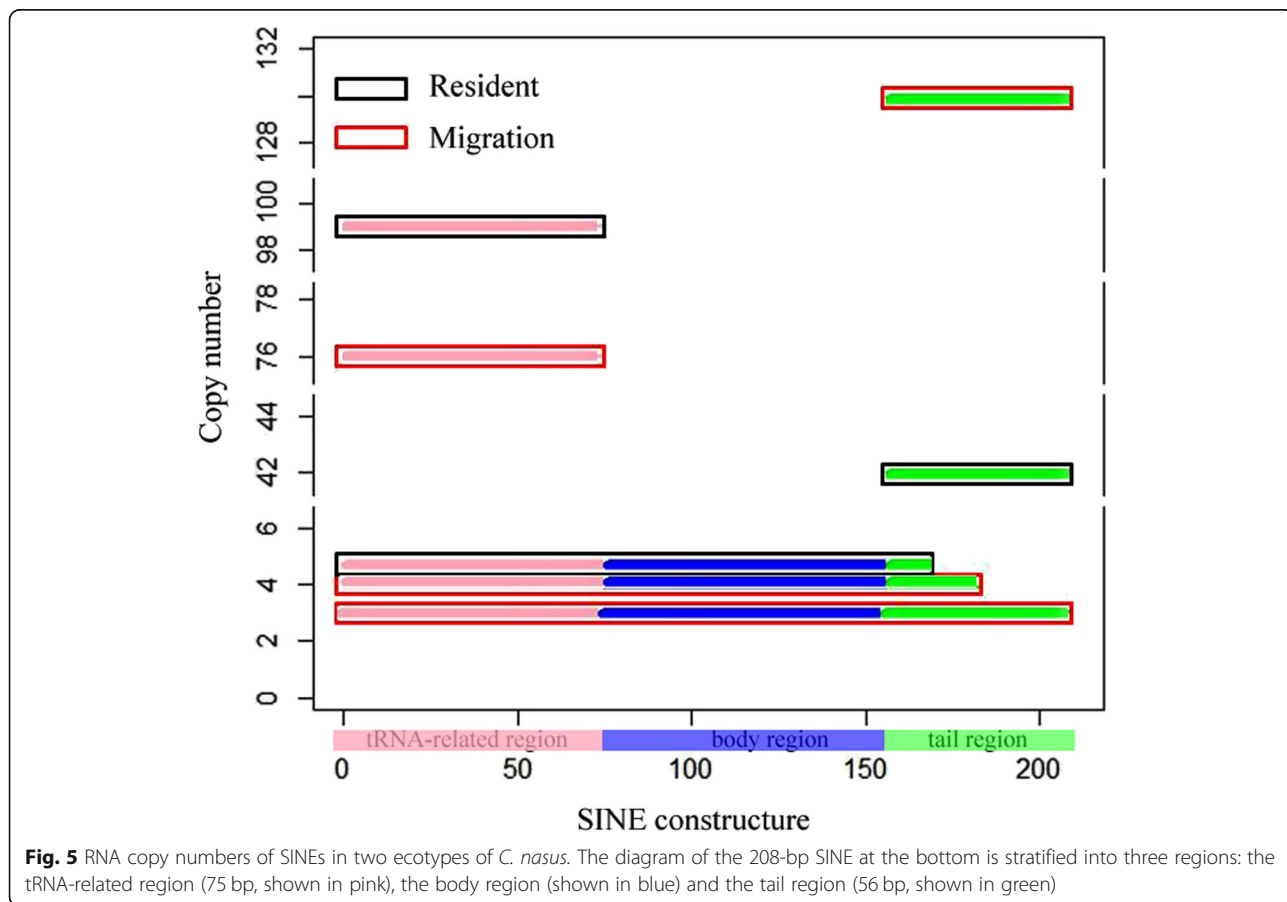


Fig. 5 RNA copy numbers of SINEs in two ecotypes of *C. nasus*. The diagram of the 208-bp SINE at the bottom is stratified into three regions: the tRNA-related region (75 bp, shown in pink), the body region (shown in blue) and the tail region (56 bp, shown in green)

resident type, we next focused on examining whether SINEs had undergone natural selection in the resident type via DNA copy number analysis of SINEs. The DNA copy numbers of the SINEs were quantified by real-time PCR. The DNA copy number in each sample was calculated by generating a standard curve between the cycle threshold (Ct) and the log of the initial template copy number using the regression equation $y = -3.11x + 40.838$ ($R^2 = 0.999$).

The average DNA copy numbers from the six sampling sites varied between 1.8×10^5 in XS to 3.8×10^5 in TH (Fig. 6). The average DNA copy number was 3.5×10^5 in the resident type and 2.3×10^5 in the migratory type. The DNA copy numbers of the SINEs showed no significant difference within the ecotypes ($p > 0.05$) but a significant difference between the ecotypes ($p < 0.05$).

We next examined the SINE insertion polymorphisms among the ecological populations. Five loci, namely, the Ls5, Ls29, Ls40, Ls58 and Ls60 loci, from 71 validated positive clones, were found to have insertion polymorphisms in the two *C. nasus* ecotypes via PCR. For Ls5, three bands were observed in all the samples from the six populations (Fig. 7a). By sequencing verification, we

found that the largest band (644 bp) contained an insertion, the second band (358 bp) was an insertion-free fragment, and the smallest band (282 bp) represented a nonspecific PCR amplification product. The Ls40 locus was polymorphic, showing the presence or absence of a SINE insertion in all samples (Fig. 7b). The polymorphisms at loci Ls29, Ls58 and Ls60, which were caused by the presence/absence of insertions, were also confirmed via a similar cloning and sequencing procedure.

We also found that the heterozygosity of the SINE insertions—a value that indicates the genetic diversity—was low in all the samples. The heterozygosity value was zero for the Ls5, Ls29 and Ls58 loci and slightly higher than 0 for the Ls40 and Ls60 loci in all six populations (Table 1). This result demonstrates a low heterozygosity level in the *C. nasus* genome. Next, we compared the SINE heterozygosity between the two *C. nasus* ecotypes. Although the heterozygosity value was not significantly different within ecotypes, the value for the migratory type was ~1.5-fold higher than that of the resident type across all five loci, and three loci (Ls5, Ls29 and Ls58) deviated significantly from Hardy-Weinberg equilibrium (Table 2). These results suggest the existence of a

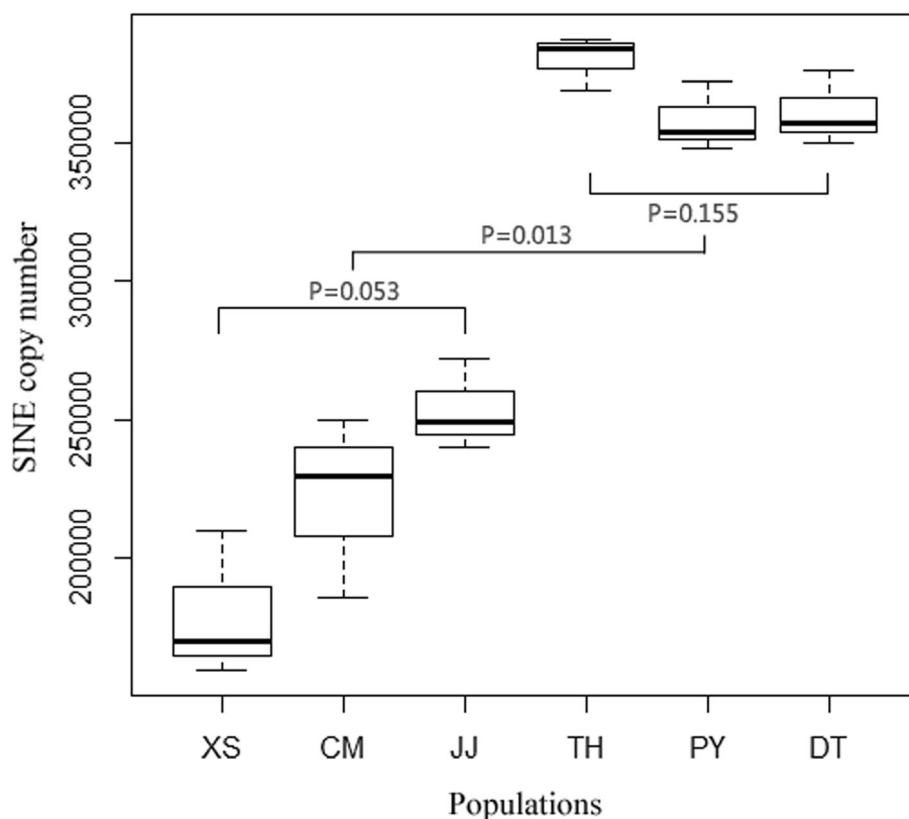
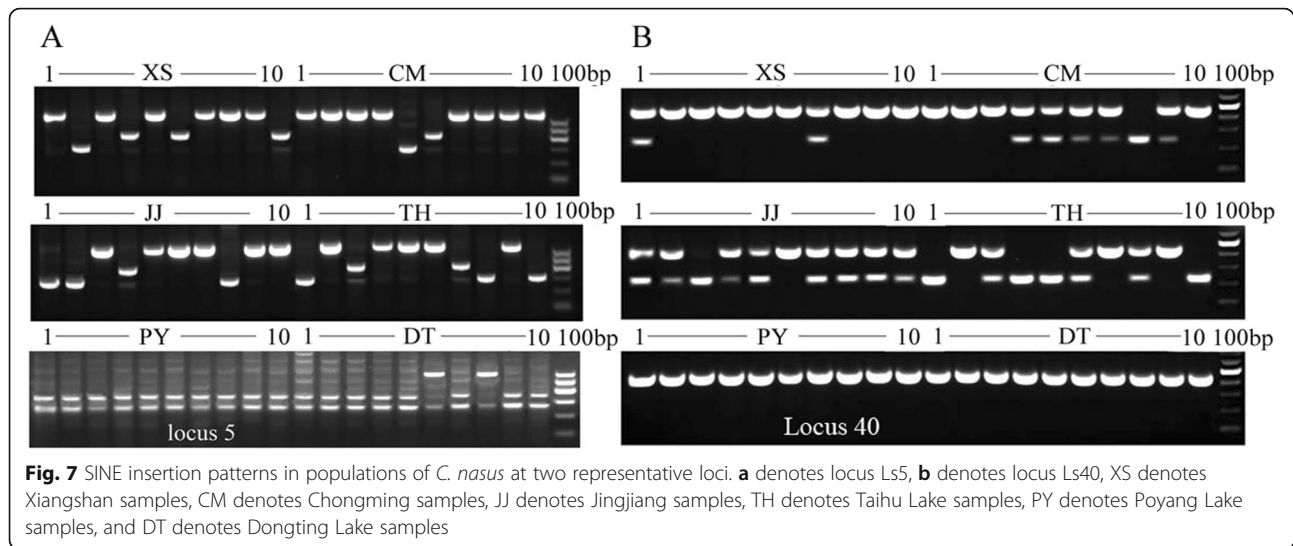


Fig. 6 SINE DNA copy number estimates in the genomes of the *C. nasus* populations. Abbreviation: XS, Xiangshan; CM, Chongming; JJ, Jingjiang; TH, Taihu Lake; PY, Poyang Lake; DT, Dongting Lake. P denotes a significant difference. The migratory ecotype includes XS, CM and JJ, while the resident ecotype includes TH, PY and DT



disparity distribution of SINE insertions in the *C. nasus* genome due to natural selection.

To examine the differentiation between these populations, a neighbor-joining (NJ) tree was constructed based on the allele frequencies of SINE insertion (Table 1). The NJ tree clearly illustrated a lineage of two ecotypes. XS, JJ, CM and TH grouped as the migratory type, while DT and PY grouped as the resident type. Of the migratory populations, TH was most closely related to the resident ecotype (Fig. 8). Although the migratory type displayed genetic variation, populations of the migratory type (TH excluded) were to compare the genetic variance of TH population, and no difference was observed between populations ($P > 0.05$), suggesting no significantly genetic differentiation between the migratory type (TH excluded) and TH population. In contrast, the genetic variance in populations of the resident type compared with TH population was 75–25% ($P < 0.5$). Therefore, this implies that there is genetic diversification between migratory and resident *C. nasus* in the Yangtze River. The lack of SINEs in the resident type (Ls29 and Ls58) and migratory type (Ls60), together with the SINE insertions in the migratory type (Fig. 8),

may have contributed to the genetic diversity of *C. nasus*. Thus our findings may assist in the examination of the life history diversity of *C. nasus*.

SINE excision revealed microhomology-induced replication

SINE insertions that underwent natural selection led to a disparity distribution in these populations. To determine the evolutionary scale of SINE insertions, including those at the Ls5, Ls29, Ls40, Ls58 and Ls60 loci, we utilized *C. mystus*, a species that is closely related to *C. nasus*, as an outgroup for comparison. We found insertions at the Ls29, Ls58 and Ls60 loci in both species. Insertions at the Ls5 and Ls40 loci were present in *C. nasus* but absent in *C. mystus*, suggesting that these insertions originated from a recent transposition activity in *C. nasus*, and the Ls29 and Ls58 insertions, which were absent in the resident type of *C. nasus*, had undergone excision.

To test whether the SINE-specific burst in *C. nasus* had occurred recently, we sequenced the flanking region of the Ls5 locus; a “TGT” TSD was observed at the terminus of the insertion, and a 17-bp duplication was observed at the preinsertion sites at the 5′ terminal flanking region of the

Table 1 Allele frequencies and heterozygosities of SINE insertions in *C. nasus* populations

Locus	XS		CM		JJ		TH		PY		DT		Total Pop	
	Het	f_{SINE}	Het	f_{SINE}	Het	f_{SINE}	Het	f_{SINE}	Het	f_{SINE}	Het	f_{SINE}	f_{SINE}	Het
Ls40	0.90	0.20	0.65	0.50	0.50	0.80	0.45	0.30	1.00	0.00	1.00	0.00	0.75	0.30
Ls5	0.60	0.00	0.80	0.00	0.60	0.00	0.50	0.00	0.00	0.00	0.20	0.00	0.45	0.00
Ls58	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.10	0.00	0.00	0.00	0.68	0.00
Ls60	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50	1.00	0.50	1.00	0.17	0.33
Ls29	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.67	0.00
Mean	0.70	0.04	0.69	0.01	0.62	0.16	0.59	0.06	0.32	0.20	0.34	0.20	0.54	0.13

Abbreviations: XS Xiangshan; CM Chongming; JJ Jingjiang; TH Taihu Lake; PY Poyang Lake; DT Dongting Lake. The f_{SINE} values donate SINE frequency, and Het donates genomic heterozygosity

Table 2 Allele frequencies and heterozygosity values of five SINE insertion loci in *C. nasus* ecotypes. The f_{SINE} values denote SINE frequency, Het denotes genomic heterozygosity, and P -value denotes the statistical significance of the Hardy-Weinberg equilibrium departure test

Locus	Migration		Residence		P -value
	f_{SINE}	Het	f_{SINE}	Het	
Ls40	0.57	0.65	1.00	0.00	0.855
Ls5	0.70	0.00	0.10	0.00	< 0.001
Ls58	1.00	0.00	0.05	0.00	< 0.001
Ls60	0.00	0.00	0.50	1.00	0.035
Ls29	1.00	0.00	0.00	0.00	< 0.001
Mean	0.65	0.13	0.33	0.20	

insertion site (Fig. 9). Our results provide evidence of SINE insertion in *C. nasus* because TSDs are present at these loci. For SINE insertion-excision, we expected footprints that would be present in the corresponding insertion-lacking locus. We observed incomplete SINE excision at the Ls29, Ls58 and Ls60 loci. For Ls29, the SINE 5'-flanking portion was excised. The gap was filled with a fragment (up to 115 bp) consisting of a microsatellite of 45 (TG) repeats and an 8-bp microhomology domain with the SINE 3' terminal flanking sequences (Fig. 10). For Ls58 and Ls60, various

deletions in the flanking sequences and short microhomologies in the repaired DNA were observed (Fig. 10). Our findings support the idea that chromosomal breakpoints can be joined via microhomology-induced replication and that double-stranded break repair followed by template switching between microhomologous sequences can lead to the generation of a new sequence to refill the excision site [45].

This observation suggested that if mutations occurred in the flanking sequences of a SINE insertion locus, a new cleavage site could be formed, and the insertion could be incompletely excised, and followed by sequence changes from DNA repair. A region with a repeat motif should be prone to mutations, as observed in the 3' tails of the SINEs via transcriptomic analysis. Upon examining whether the 5' flanking regions of the SINEs have repeat motifs at nearby insertion loci, we observed that the 50-bp sequences immediately upstream of the SINE 5' flanking regions were enriched in A and T residues (mean = 56.6%). The AT content was significantly higher than the GC content ($P < 0.05$). The 50-bp sequences immediately downstream of the SINE 3' flanking regions were similarly AT rich (mean = 57%). However, tandem repeats were predominantly found in the upstream region. These tandem repeats, which were up to 100 bp in length, were found at 13 loci. In contrast, only one locus was found to have a tandem repeat located in the downstream region (Table 3). The properties of the repeats in the 5' flanking and 3' tail sequences of SINEs provide crucial clues regarding the mechanism of SINE insertion-excision and replication by the organism.

Discussion

A novel SINE family isolated from *C. nasus*

In this study, we isolated new SINE family members from *C. nasus*. Based on GenBank or Rebase database searches, the 208-bp consensus sequence of the new SINE members showed no hits from other organisms [36]. However, in another search, these SINEs were identified in 19 species of the family Engraulidae (data not shown). Thus, this type of retrotransposon was first identified in this family.

These SINEs are conserved in the tRNA-related region and can fold to form a perfect cloverleaf tRNA structure, which is important for the initiation of SINE RNA transcription (Fig. 3) [15]. Generally, the SINE tRNA-related region is followed by a strongly conserved region called the body region, which was previously thought to represent a common origin in the same species [2]. However, our results showed that the body regions of the *C. nasus* SINEs share < 50% sequence identity, a value that is less than the defined threshold (76%) of the conserved regions in the same SINE family [46–49]. Therefore, the body regions of the SINEs identified in this study are not homologous to those of the existing SINE families, suggesting a different origin from those of the known SINEs.

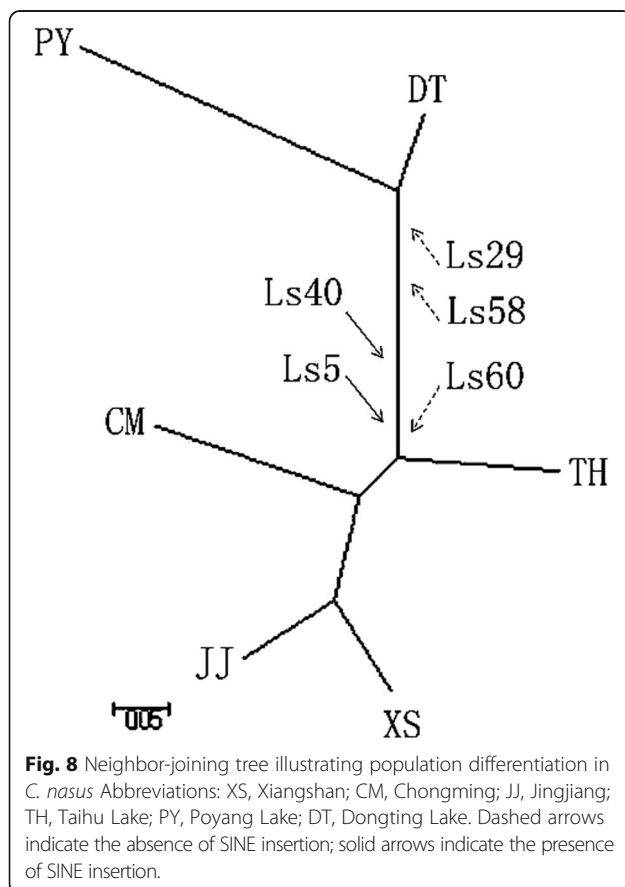
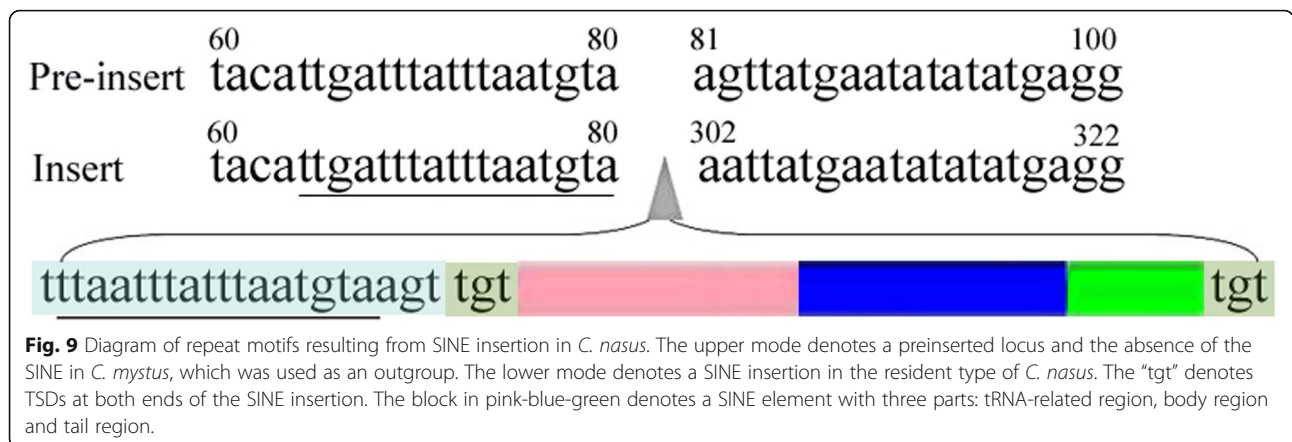


Fig. 8 Neighbor-joining tree illustrating population differentiation in *C. nasus*. Abbreviations: XS, Xiangshan; CM, Chongming; JJ, Jingjiang; TH, Taihu Lake; PY, Poyang Lake; DT, Dongting Lake. Dashed arrows indicate the absence of SINE insertion; solid arrows indicate the presence of SINE insertion.



Expression divergence of SINEs in *C. nasus* populations

To determine whether the transcript copy numbers of SINEs differ between the two ecotypes of *C. nasus* at the RNA level, we compared the transcriptomic profiles of the two ecotypes. The migratory type contains more SINE contig hits than the resident type (Fig. 5). Most of the hit contigs showed variations due to nucleotide substitution, insertions and deletions. The sequence divergence of SINEs as mobile elements, as previously reported, can be targeted by uniquely mapping short interfering RNAs (siRNAs), which are involved in efficient siRNA-mediated methylation at the mobile elements [50]. Subsequently, the methylated mobile element can strongly regulate the expression of adjacent genes [50–52], leading to further species divergence [53, 54]. Therefore, it is reasonable to propose a significant role for SINEs in the ecotypic divergence of *C. nasus*.

In this study, the effects of SINEs on gene expression were also observed in two transcripts associated with SINE insertions. The first transcript was the interferon regulatory factor gene in the resident type, which contains a SINE insertion in its 3' downstream region; this gene is

known to respond to viral infection [55]. The second transcript was the S100 gene in the migratory type, which is involved in the regulation of cytosolic calcium concentration during migration from the ocean to fresh water [56]. Our results suggest that SINEs make important contribution to gene expression variation in *C. nasus* and likely influence the differences in life history of this species.

Implications of the skewed frequency and copy number for purifying selection

Purifying selection in evolution leads to a transposition selection equilibrium that occurs via elimination of deleterious insertions to limit population frequencies [57]. A disparity insertion indicates an adaptive event [18]. In this study, we found disparity SINE insertions among the *C. nasus* populations sampled from six sites, and some insertions were associated with life history factors in *C. nasus*. At the population level, the mean insertion frequency of SINEs in three populations of the migratory type (0.62–0.7) was higher than that in the two populations with resident behavior (0.32–0.34, see Table 1). Similarly, the mean insertion frequency of the migratory

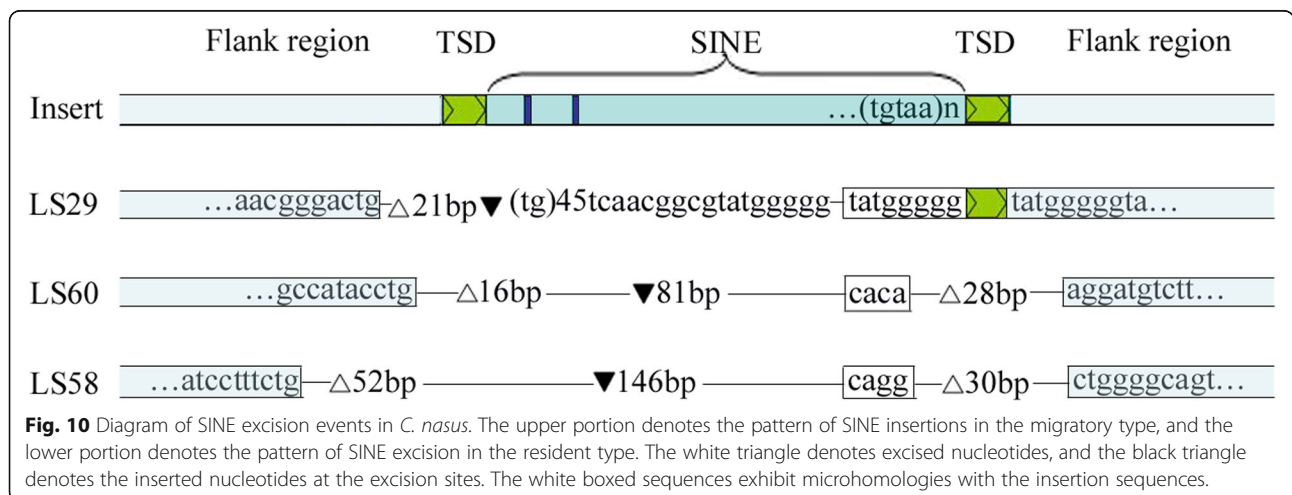


Table 3 Characteristics of the simple repeats in the SINE flanking regions

Loci	Flanks	Repeat unit	Sites
Ls2	5'	(G)38	-76
Ls61	5'	(A)17	-1
Ls11	5'	(TG)26	-55
Ls26	5'	(CA)16	-56
Ls69	5'	(CA)14	-48
Ls29	5'	(CG)18	-16
Ls62	5'	(CT)7	-85
Ls52	5'	(AAC)17(AAT)8	-3
Ls32	5'	(ATTAC)7	-2
Ls18	5'	(TTATTTAA)3	-10
Ls12	5'	(TTGACCAGC)6	-3
Ls43	5'	(ACTAGGGAACACCAGGGG)14	-8
Ls44	3'	(GT)10	+ 115

type (0.65) was higher than that of the resident type (0.33, see Table 2). This disparity distribution of insertion frequency suggested that the SINE sites have undergone purifying selection to further polarize the populations [58].

It is plausible that these SINE insertions purified via natural selection have impacted on the differences in life history of the *C. nasus* ecotypes. *C. nasus* initially originated in the ocean and migrated into rivers and lakes for reproduction [59]. Therefore, this species faced environmental challenges during the reproductive process. However, SINEs can undergo transposition bursts in the genome, allowing *C. nasus* to adapt to the environment via SINE insertions that disrupt gene expression [9]. Even insertions that are separated by large distances in the genome (> 2000 bp) can lead to disruption of promoter-enhancer interactions [50]. These retrotranspositions can be instrumental in increasing the frequency of beneficial insertions into the genome and decreasing deleterious insertions to avoid insertion-related damage [5]. As a result, populations that have achieved beneficial insertions exhibit high SINE copy numbers, especially small populations [19]. In fact, the resident type was a small population relative to the migratory type in terms of fishery catches in the 1950s. The SINE copy number in the *C. nasus* resident type was higher than that in the migratory type (3.5×10^5 vs 2.3×10^5 , see also Fig. 6). This result is consistent with results from other studies that have shown how copies of mobile elements have contributed to genomic divergence during population expansions [29, 30, 60].

Phylogenetic analyses of *C. nasus* populations based on SINE insertion frequency readily distinguished two ecotypes of *C. nasus* (Fig. 8). The DT and PY populations were defined as the resident types and were sorted into one group. The XS, JJ and CM populations were sorted into the other

group, which was regarded as the migratory type in previous studies [30]. The TH population, which was identified as a subspecies of *C. nasus* in previous studies [59], displayed diversification from both the migratory and resident ecotypes. This result is consistent with our prior study, which showed that the genomic disparity of *C. nasus* was based on polymorphisms at three insertion sites [61]. Taken together, these results show that *C. nasus* can be distinguished into two ecotypes based on the insertion frequency and copy number of SINEs. Our results, together with mitochondrial DNA [29], nuclear DNA and AFLP [62], could help us to clarify the evolutionary history of *C. nasus*.

Genetic polymorphism driven by SINEs

SINEs can retrotranspose in mammalian genomes because these elements have a 3' tail structure similar to that of LINE sequences [14, 63]. We observed that the *C. nasus* SINEs and LINES shared the stem-loop structure of the 3' tail region (Fig. 4). As in other organisms [15], it is plausible that this stem-loop structure functions as a recognition site for the retrotransposase proteins encoded by LINES. Retrotransposition of SINEs would have resulted in TSDs at the insertion site via the copy-and-paste mechanism [2]. The SINEs in *C. nasus* showed perfect TSDs at some insertion sites, and a few TSDs were composed of simple repeats in the insertion flanking regions of the SINEs (Table 3). Such repeats are regarded as generators of microsatellites [64]. About 23% of total minisatellites/satellites are derived from transposons in human genome [65].

Although SINE insertion can increase SINE copy numbers in genomes, this process is restricted by various mechanisms in the genome, such as insertion/deletion, genetic drift, and ectopic recombination [17–19, 58]. In this study, we observed incomplete excision of SINEs in *C. nasus* (Fig. 10). Deletion of these SINEs may have given rise to the genetic diversity of this species and led to the further development of a different population via an insertion-selection process, as suggested by others [8, 66].

However, the deletion mechanism has not been explained to date, with the exception of element mutations. In this regard, the SINE evolutionary events that occurred in *C. nasus* may help elucidate the mechanism underlying SINE insertion/deletion. First, the TGTA short repeat in the SINE 3' tail could permit template slippage during initiation of DNA replication [14] and generate various numbers of the short repeat in DNA fragments containing SINEs. Second, replication slippage mispairing can increase the rate of mutation [67], which could be used to generate possible cleavage sites for SINEs. Therefore, the SINE element could be incompletely excised, and the gap at the cleavage site could be replicated by the organism. This SINE-based mechanism can be used to explain genetic polymorphism in *C. nasus*.

Conclusions

In conclusion, novel SINEs were isolated from an anadromous fish, *C. nasus*, which includes a freshwater-resident ecotype. The two ecotypes of *C. nasus* differ in their reproductive behavior. These SINEs are active and underwent a transposition burst in the genome of the anadromous ecotype, producing polymorphic insertions and further influencing gene expression and function. As part of life history adaptation to the freshwater environment, incomplete excisions occurred at a small spatial scale in individuals with habitat specialization, and the benefit conferred by SINE insertions resulted in greater SINE copy numbers in the resident ecotype, eventually facilitating population divergence and speciation. Therefore, SINE activity and incompletely excision led to the ecotypic diversity of the *C. nasus* populations. Our results provide a valuable clue to understanding speciation and population structure within commercially important species.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13100-019-0198-8>.

Additional file 1: Table S1. Adaptors used to structure the AFLP library and primer sequences used to scan SINE insertions in the library.

Additional file 2: Table S2. The consensus sequence of SINE family from genome of *C. nasus*.

Additional file 3: Table S3. Primers specific to the flanking sequences of the SINE insertion loci used to detect insertion polymorphisms in *C. nasus* populations.

Additional file 4: Table S4. Types, origins and percent similarities of the tRNA-related portions of SINEs in *C. nasus*.

Additional file 5 Table S5. Genes annotated in SINE-hit contigs from the migratory type.

Additional file 6 Table S6. Genes annotated in SINE-hit contigs from the resident type.

Acknowledgments

We thank all members of Dr. Tang's group for helping with sample collection and for constructive discussion during the development of this project.

Authors' contributions

DL performed all experiments and wrote the paper; WQ and JQ designed the experiments; XZ and CMR helped improve the manuscript; MZ performed analysis of the gene annotations and revised the manuscript. All authors read and approved the final manuscript.

Funding

This research was funded by the National Key R&D Program of China (No. 2018YFD0900802), the Central Public-interest Scientific Institution Basal Research Fund, CAFS (No. 2018HY-XKQ02-07), Science and Technology Commission of Shanghai Municipality (No. 19050501900) and the Shanghai Universities First-class Disciplines Project of Fisheries.

Availability of data and materials

Sequencing data have been deposited in the NCBI database under the following accession numbers: MK621683–753, SINE elements at the inserted loci; MK621754–757, tail portions of LINE elements; and MK621758–769, sequences of the different bands at each insertion locus.

Ethics approval and consent to participate

All experiments were approved by the Animal Ethics Committee of Shanghai Ocean University. The experimental protocol for fish was performed in strict accordance with the requirements of Animal Ethics Procedures and Guidelines of the People's Republic of China.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Key Laboratory of Marine Animal Taxonomy and Evolution, Shanghai Universities, Shanghai 201306, China. ²Department of Epidemiology and Biostatistics, University of Georgia, Athens, GA 30602, USA. ³Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources, Ministry of Education, Shanghai 201306, China. ⁴National Demonstration Center for Experimental Fisheries Science Education, Shanghai Ocean University, Shanghai 201306, China.

Received: 18 May 2019 Accepted: 27 December 2019

Published online: 08 January 2020

References

- Elbarbary RA, Lucas BA, Maquat LE. Retrotransposons as regulators of gene expression. *Science*. 2016;351:7247002E.
- Kramerov DA, Vassetzky NS. Origin and evolution of SINEs in eukaryotic genomes. *Heredity*. 2011;107:487.
- Elena C, Josefa G. The impact of transposable elements in environmental adaptation. *Mol Ecol*. 2013;22:1503–17.
- González J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA. High rate of recent transposable element-induced adaptation in *Drosophila melanogaster*. *PLoS Biol*. 2008;6:e251.
- Goodier JL Jr. HHK: Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell*. 2008;135:23–35.
- Nicolás F, Davis GK, Diego V, Shu W, Fran?Ois P, stern DL: phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature*. 2010;466:490–3.
- Liu B, Akira K, Hisakazu M, Ryoji T, Kyuya H, Jun A. Genetic redundancy in soybean photoresponses associated with duplication of the phytochrome a gene. *Genetics*. 2008;180:995–1007.
- van't Hof AE, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, Hall N, Darby AC, Saccheri IJ. The industrial melanism mutation in British peppered moths is a transposable element. *Nature*. 2016;534:102.
- Kazazian HH. Mobile elements: drivers of genome evolution. *Science*. 2004;303:1626–32.
- Kido Y, Aono M, Yamaki T, K-i M, Murata S, Saneyoshi M, Okada N. Shaping and Reshaping of Salmonid Genomes by Amplification of tRNA-Derived Retroposons During Evolution Proceedings of the National Academy of Sciences of the United States of America, vol. 88; 1991. p. 2326–30.
- Matveev V, Okada N. Retroposons of salmonoid fishes (Actinopterygii: Salmonoidei) and their evolution. *Gene*. 2009;434:16–28.
- Guo B, Zou M, Gan X, He S. Genome size evolution in pufferfish: an insight from BAC clone-based *Diodon holocanthus* genome sequencing. *BMC Genomics*. 2010;11:396.
- Gao B, Shen D, Xue S, Chen C, Cui H, Song C. The contribution of transposable elements to size variations between four teleost genomes. *Mob DNA*. 2016;7:4.
- Kajikawa M, Okada N. LINEs mobilize SINEs in the eel through a shared 3' sequence. *Cell*. 2002;111:433–44.
- Kajikawa M, Ichianagi K, Tanaka N, Okada N. Isolation and characterization of active LINE and SINEs from the eel. *Mol Biol Evol*. 2005;22:673–82.
- Valizadeh P, Crease TJ. The association between breeding system and transposable element dynamics in *daphnia pulex*. *J Mol Evol*. 2008;66:643–54.
- Wright SI, Le QH, Schoen DJ, Bureau TE. Population dynamics of an ac-line transposable element in self- and cross-pollinating arabidopsis. *Genetics*. 2001;158:1279–88.

18. Petrov DA, Aminetzach YT, Davis JC, Douda B, Hirsh AE. Size matters: non-LTR retrotransposable elements and ectopic recombination in *Drosophila*. *Mol Biol Evol*. 2003;20:880–92.
19. Rouzic A, Deceliere G. Models of the population genetics of transposable elements. *Genet Res*. 2005;85:171–81.
20. Prescott D. Alu elements: know the SINEs. *Genome Biol*. 2011;12:236.
21. Ray DA. SINEs of progress: Mobile element applications to molecular ecology. *Mol Ecol*. 2010;16:19–33.
22. Shedlock AM, Okada N. SINE insertions: powerful tools for molecular systematics. *BioEssays*. 2000;22:148–60.
23. Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ. African origin of human-specific polymorphic Alu insertions. *Proc Natl Acad Sci U S A*. 1994;91:12288–92.
24. Nikaido M, Piskurek O, Okada N. Toothed whale monophyly reassessed by SINE insertion analysis: the absence of lineage sorting effects suggests a small population of a common ancestral species. *Mol Phylogenet Evol*. 2007;43:216–24.
25. Ray DA, Walker JA, Hall A, Llewellyn B, Ballantyne J, Christian AT, Turteltaub K, Batzer MA. Inference of human geographic origins using Alu insertion polymorphisms. *Forensic Sci Int*. 2005;153:117–24.
26. Tong C, Gan X, He S. Multiple source genes of HAmo SINE actively expanded and ongoing retroposition in cyprinid genomes relying on its partner LINE. *BMC Evol Biol*. 2010;10:115.
27. Liu W. Biochemical and morphological comparison and interspecific relationships of four species of the genus *Coilia* in China. *Oceanologia et Limnologia Sinica*. 1995;26:558–64.
28. Yang J, Arai T, Liu H, Miyazaki N, Tsukamoto K. Reconstructing habitat use of *Coilia mystus* and *Coilia ectenes* of the Yangtze River estuary, and of *Coilia ectenes* of Taihu Lake, based on otolith strontium and calcium. *J Fish Biol*. 2006;69:1120–35.
29. Ma C, Cheng Q, Zhang Q, Ping Z, Zhao Y. Genetic variation of *Coilia ectenes* (Clupeiformes: Engraulidae) revealed by the complete cytochrome b sequences of mitochondrial DNA. *J Exp Marine Biol Ecol*. 2010;385:14–9.
30. Yang J, Xuelian HU, Tang W. Genetic structure and population history of *Coilia nasus* in Yangtze River and its south adjacent waters. *J Shanghai Fish Univ*. 2008;17:513–9.
31. Zhang M, Xu D, Liu K, Shi W. Studies on biological characteristics and change of resource of *Coilia nasus schlegel* in the lower reaches of the Yangtze River. *Resour Environ Yangtze Basin*. 2005;14:694–8.
32. Liu ES, Bao CH, Wan Q. The changes of the catch and age structure of Lake anchovy and the relationship with environmental variation in Lake Taihu. *J Fish China*. 2008;32:229–35.
33. Liu D, Guo HY, Tang WQ, Yang JQ. Comparative evolution of S7 intron 1 and ribosomal internal transcribed spacer in *Coilia nasus* (Clupeiformes: Engraulidae). *Int J Mol Sci*. 2012;13:3085–100.
34. Vos P, Hogers R, Bleeker M, Reijns M, Lee Tvd, Hornes M, Friters a, pot J, Paleman J, Kuiper M, Zabeau M: AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res*. 1995;23:4407–14.
35. Liu D, Zhu G, Tang W, Yang J, Guo H. PCR and magnetic bead-mediated target capture for the isolation of short interspersed nucleotide elements in fishes. *Int J Mol Sci*. 2012;13:2048–62.
36. Bao W, Kojima KK, Kohany O. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob DNA*. 2015;6:11.
37. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*. 2003;31:3406–15.
38. Lowe TM, Chan PP. tRNAscan-SE on-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res*. 2016;44:54–7.
39. Zhu G, Wang L, Tang W, Liu D, Yang J. De novo transcriptomes of olfactory epithelium reveal the genes and pathways for spawning migration in Japanese grenadier anchovy (*Coilia nasus*). *PLoS One*. 2014;9:e103832.
40. Xu G, Dong J, Nie Z, Xu P, Gu R. Studies on lactate dehydrogenase isozymes and DNA content in different tissues of *Coilia nasus*. *J Shanghai Ocean Univ*. 2012;21:481–8.
41. Lalitha S. Primer premier 6 Biotech Software & Internet Report. *Comp Software J Scient*. 2000;1:270–2.
42. Liu K, Muse S. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*. 2005;21:2128–9.
43. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and windows. *Mol Ecol Resour*. 2010;10:564–7.
44. Sugano T, Kajikawa M, Okada N. Isolation and characterization of retrotransposition-competent LINEs from zebrafish. *Gene*. 2006;365:74–82.
45. Celia P, Romain K, Bernard D, Gilles F. Segmental duplications arise from Pol32-dependent repair of broken forks through two alternative replication-based mechanisms. *PLoS Genet*. 2008;4:e1000175.
46. Akasaki T, Nikaido M, Nishihara H, Tsuchiya K, Segawa S, Okada N. Characterization of a novel SINE superfamily from invertebrates: “Ceph-SINEs” from the genomes of squids and cuttlefish. *Gene*. 2010;454:8–19.
47. Gilbert N, Labuda D. CORE-SINEs: eukaryotic short interspersed retroposing elements with common sequence motifs. *Proc Natl Acad Sci*. 1999;96:2869–74.
48. Nishihara H, Smit A, Okada N. Functional noncoding sequences derived from SINEs in the mammalian genome. *Genome Res*. 2006;16:864–74.
49. 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.
50. Wang X, Detlef W, Smith LM. Transposon variants and their effects on gene expression in *Arabidopsis*. *PLoS Genet*. 2013;9:e1003255.
51. El Baidouri M, Kim KD, Abernathy B, Li YH, Qiu LJ, Jackson SA. Genic C-methylation in soybean is associated with gene paralogs relocated to transposable element-rich pericentromeres. *Mol Plant*. 2018;11:485–95.
52. Hirsch CD, Springer NM. Transposable element influences on gene expression in plants. *BBA - Gene Regulatory Mechanisms*. 2016;1860:157.
53. Hollister JD, Smith LM, Ya-Long G, Felix O, Detlef W, Gaut BS. Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proc Natl Acad Sci U S A*. 2011;108:2322–7.
54. Lavagnino N, Arya J, Korovaichuk A, Fanara J. Genetic architecture of olfactory behavior in *Drosophila melanogaster*: differences and similarities across development. *Behav Genet*. 2013;43:348–59.
55. Huang B, Qi ZT, Xu Z, Nie P. Global characterization of interferon regulatory factor (IRF) genes in vertebrates: glimpse of the diversification in evolution. *BMC Immunol*. 2010;11:22.
56. Heizmann CW. The multifunctional S100 protein family. *Methods Mol Biol*. 2002;172:69–80.
57. Bustamante CD, Wakeley J, Sawyer S, Hartl DL. Directional selection and the site-frequency spectrum. *Genetics*. 2001;159:1779–88.
58. Lockton S, Ross-Ibarra J, Gaut B. Demography and weak selection drive patterns of transposable element diversity in natural populations of *Arabidopsis lyrata*. *Proc Natl Acad Sci*. 2008;105:13965–70.
59. Yuan C, Qing A, Liu R, Lin J. Discussion on subspecific taxonomy of the genus *Coilia* in middle and lower reaches of Changjiang and southeast coastal China Sea. *J Nanjing Univ*. 1980;3:67–82.
60. Christian B, Christiane N, Grégory D, David L, Catherine L, Cristina V. Worldwide distribution of transposable element copy number in natural populations of in natural populations of *Drosophila simulans*. *Evolution*. 2003;57:159–67.
61. Liu D, Li Y, Tang W, Yang J, Guo H, Zhu G, Li H. Population structure of *Coilia nasus* in the Yangtze River revealed by insertion of short interspersed elements. *Biochem Systematics Ecol*. 2014;54:103–12.
62. Han Z, Han G, Wang Z, Gao T. The possible physical barrier and coastal dispersal strategy for Japanese grenadier anchovy, *Coilia nasus* in the East China Sea and Yellow Sea: evidence from AFLP markers. *Int J Mol Sci*. 2015;16:3283–97.
63. Okada N, Hamada M, Ogiwara I, Ohshima K. SINEs and LINEs share common 3' sequences: a review. *Gene*. 1997;205:229–43.
64. Luchetti A, Mantovani B. Talua SINE biology in the genome of the Reticulitermes subterranean termites (Isoptera, Rhinotermitidae). *J Mol Evol*. 2009;69:589.
65. Ahmed M, Liang P. Transposable elements are a significant contributor to tandem repeats in the human genome. *Comp Func Genomics*. 2012;2012:1–7.
66. Charlesworth B, Langley CH. The population genetics of *Drosophila* transposable elements. *Annu Rev Genet*. 1989;23:251.
67. Levinson G, Gutman GA. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol*. 1987;4:203–21.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.