Pseudogene-derived small interference RNAs regulate gene expression in African *Trypanosoma brucei*

Yan-Zi Wen[a,1], Ling-Ling Zheng[b,1], Jian-You Liao[b], Ming-Hui Wang[a], Ying Wei[a], Xue-Min Guo[a], Liang-Hu Qu[b,2], Francisco J. Ayala[c,2], and Zhao-Rong Lun[a,2]

[a]Center for Parasitic Organisms, State Key Laboratory of Biocontrol, School of Life Sciences and Key Laboratory of Tropical Diseases and Control of the Ministry of Education, Zhongshan Medical School, Sun Yat-Sen University, Guangzhou 510275, People's Republic of China; [b]State Key Laboratory of Biocontrol and Key Laboratory of Gene Engineering of the Ministry of Education, School of Life Sciences, Sun Yat-Sen University, Guangzhou 510275, People's Republic of China; and [c]Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697

Contributed by Francisco J. Ayala, March 16, 2011 (sent for review February 1, 2011)

Pseudogenes have been shown to acquire unique regulatory roles from more and more organisms. We report the observation of a cluster of siRNAs derived from pseudogenes of African *Trypanosoma brucei* using high through-put analysis. We show that these pseudogene-derived siRNAs suppress gene expression through RNA interference. The discovery that siRNAs may originate from pseudogenes and regulate gene expression in a unicellular eukaryote provides insights into the functional roles of pseudogenes and into the origin of noncoding small RNAs.

sleeping sickness | Nagana | high through-put sequencing | noncoding RNAs | gene regulation

Pseudogenes have long been considered dysfunctional genes with gene-like features because most of them have lost the ability of protein-coding or are otherwise no longer expressed because of frame-shift or stop-codon mutations (1, 2). Therefore, pseudogenes were conventionally thought to be accidental by-products of genome evolution that did not call for special attention by researchers. Over the past years, however, it has become evident that pseudogenes may have diverse functions, mainly in regulating gene expression (3–7). Since 1999, several studies have demonstrated that a nitric oxide synthase pseudogene is functional in snail (3, 5, 7). The retrotransposed pseudogene *Makorin1-p1* in the mouse ostensibly destabilizes the expression of its homologous gene *Makorin1* (MKRN1) (6). This result showed that pseudogenes are able to play a transregulatory role in gene expression, which motivated other scientists to investigate the functions of pseudogenes in other organisms. Indeed, a contemporaneous review by Balakirev and Ayala (4) summarized a number of examples in *Drosophila* and other organisms where pseudogenes are involved in gene modulation, indicating that pseudogenes could play important biological roles in eukaryotic organisms. Using deep sequencing techniques and suitable experimental evidence, two recent publications have documented that a subset of siRNAs derived from pseudogenes could regulate gene expression in mouse oocytes (8, 9). Subsequently, in rice, a computational analysis found small RNAs, originated from pseudogenes, that exhibited regulatory roles (10). These observations suggested that functional pseudogenes, through the synthesis of small RNAs, might be ubiquitous in eukaryotic organisms.

The protozoan parasite *Trypanosoma brucei* in Africa causes sleeping sickness in humans and Nagana disease in cattle, camels, and other animals (11). There are two main developmental stages in the life cycle of *T. brucei*: the bloodstream forms in the mammalian host, including humans, and the procyclic forms in the midgut of the tsetse fly. To adapt to the disparate environments in mammalian and insect hosts, *T. brucei* must undergo complex differentiation during the life cycle. Various researchers have revealed that a diversity of transcripts is involved in the pathogen's differentiation across its life stages (12–14). However, the mechanisms that control expression shifts remain undetermined. Owing to polycistronic transcription and the absence of RNA polymerase II promoters for protein-coding genes, the genome of *T. brucei* is transcribed constitutively (15). Consequently, the modulations of gene expression in *T. brucei* should take place at the post-transcriptional level, rather than at the transcriptional level.

RNA interference (RNAi) is an important component of the RNA modulation pathway in eukaryotic organisms—a process by which specific double-stranded RNA (dsRNA) is cleaved into siRNAs or microRNAs guiding the RNase-mediated cleavage of the homologous mRNA. In protozoan parasites, the RNAi pathway has been thoroughly demonstrated and applied as a mature tool for functional analysis (16). Mallick et al. (17), using computational analysis, have suggested that microRNAs are excited in *T. brucei*, but they did not provide any experimental evidence to support it. As far as it is known, siRNAs are the key regulators in *T. brucei*.

Pseudogenes are pervasive in the genome of *T. brucei*: ~10% of predicted genes are pseudogenes (900 pseudogenes of 9,068 genes) (18). As early as 1989, research on pseudogenes in *T. equiperdum*, a genetically close species to *T. brucei*, revealed that recombination processes occur between variable surface glycoprotein (*VSG*) genes and their pseudogenes by reassorting sequence. Since the publication of this research, the pseudogenes in trypanosomes have been considered as chimeric genes and have been associated with antigenic diversity through recombination (19). However, additional questions have arisen due to the discovery of new pseudogene functions in other organisms. Are the pseudogenes the potential source of small RNAs in this unicellular eukaryotic organism as well? Do the small RNAs act as siRNAs and modulate gene expression by an RNAi pathway? Seeking to answer these questions, we have used high through-put analysis of small noncoding RNAs in *T. brucei*.

Results

Pseudogene-Derived Small RNAs in *T. brucei* Match the Features of siRNA. Sequences and amounts of small noncoding RNAs were obtained from the bloodstream forms of *T. brucei* by deep sequencing, and 16.17% small RNAs were categorized as originating from pseudogenes (Fig. 1*A*). This suggests that in *T. brucei* the sources of small RNAs would be pseudogenes, rather than just transcriptions from distinctive sequences in the genome. The lengths of most pseudogene-derived small RNAs cluster about 23–26 nt (87.49%) (Fig. 1*B*), consistent with a notable feature of known siRNAs in *T. brucei* (20). Additionally, the 5′ terminal nucleotide of most of these small RNAs preferred uridine (75.89%) to the other nucleotides (Fig. 1*C*), as did the siRNAs in other organisms, such as *Arabidopsis* (21, 22).

Author contributions: Y.-Z.W., L.-L.Z., L.-H.Q., F.J.A., and Z.-R.L. designed research; Y.-Z.W., L.-L.Z., M.-H.W., and Y.W. performed research; X.-M.G. and L.-H.Q. contributed new reagents/analytic tools; Y.-Z.W., L.-L.Z., J.-Y.L., L.-H.Q., and Z.-R.L. analyzed data; and Y.-Z.W., L.-L.Z., F.J.A., and Z.-R.L. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]Y.-Z.W. and L.-L.Z. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: fjayala@uci.edu, lssqlh@mail.sysu.edu.cn, or lsslzr@mail.sysu.edu.cn.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1103894108/-/DCSupplemental.

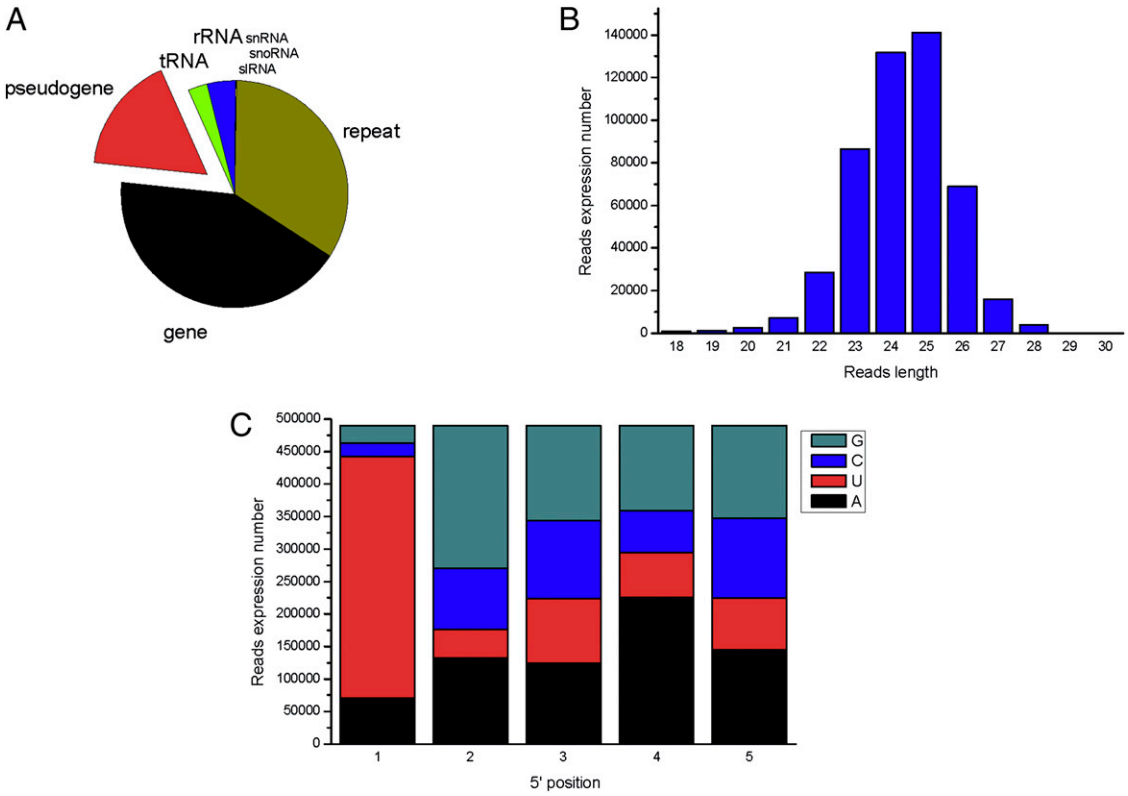

Fig. 1. Characteristics of pseudogene-derived siRNAs in *T. brucei* bloodstream form. (*A*) A small RNA library (≤30 nt) of the STIB 920 bloodstream form was deeply sequenced, and the reads were mapped to various categories. (*B*) Length distribution of pseudogene-derived small RNAs. (*C*) 5′ terminal nucleotide bias of pseudogene-derived small RNAs (A, adenine; G, guanine; C, cytosine; U, uridine).

Expression of Pseudogene-Derived Small RNAs Is Dependent on TbDCL1. siRNAs are processed from long dsRNAs under the action of dicer, which is a way to distinguish them from random degradation products. The activities of two dicer-like proteins (TbDCL1 and TbDCL2) are involved in the *T. brucei* RNAi pathway (23). TbDCL2 is responsible for the RNAi pathway in the nucleus, whereas TbDCL1 is in charge in the cytoplasm (24, 25). Because antisense transcript-mediated RNA destabilization occurs mostly in the cytoplasm (26), we considered that TbDCL1 should be the direct enzyme catalyzing pseudogene-derived siRNA synthesis. Seeking to ascertain whether pseudogene-derived small RNAs result from the production of dicer, we knocked down *TbDCL1* in *T. brucei* Lister 427 (see *Materials and Methods* for details). Although TbDCL1 fuels the RNAi pathway, leading to only partial suppression (about 50%) in the tetracycline-induced cell lines (Fig. 2*A*), the pseudogene-derived small RNAs show almost a half decrease (Fig. 2*B*). This result shows that pseudogene-derived small RNAs in *T. brucei* originate from the long dsRNAs in response to the activity of TbDCL1 and that they are indeed siRNAs rather than degraded fragments.

Pseudogene-Derived siRNAs from the Natural RNA Duplex Formed by Pseudogenes and Protein-Coding Genes. There is good evidence from mouse oocytes showing the production and regulatory potential of pseudogene-derived siRNAs, based on the simultaneous accumulation of sense-strand small RNAs from the protein-coding genes and antisense-strand small RNAs from the pseudogenes in the complementary region (8). Following this idea, we designed a strategy to analyze the small RNA dataset of *T. brucei* bloodstream forms (Fig. 3). Nine gene–pseudogene pair candidates were obtained with the capability of producing siRNAs (Table 1). In one case, the gene Tb927.10.16200 on chromosome 10 paired with pseudogene Tb09.v4.0143 on tryp_IXb-218d07.p1c. Blasting showed that nearly 94% of the gene and pseudogene nucleotides are identical in the 5′ region (209 of 222 bp). At the Tb927.10.16200 locus, 22 unique small RNAs with expression number at 30.93 were mapped to the 5′ protein-coding region; all these small RNAs were oriented in the same direction as the gene. In comparison, 20 unique small RNAs with expression number at 18.75 were mapped in the pseudogene Tb09.v4.0143, which complements Tb927.10.16200; their directions were antisense to the coding gene (Fig. 4). This is consistent with the idea that dsRNAs are the only source of siRNAs (29). Further results from quantitative real time PCR (qRT-PCR) detection demonstrated that, with the associated pseudogene-derived siRNAs decreasing (p-siRNA-3 from Tb927.5.4710 and p-siRNA-4 from Tb09.v4.0143) (Fig. 2*B*), the Tb927.10.16200 transcripts were up-regulated to 1.8-fold in conditional *TbDCL1* knockdown cell lines, and another correspondingly regulated protein-coding gene, Tb09.211.4940, increased 1.2-fold as well (Fig. 2*A*). These results show that some siRNAs in *T. brucei* are produced from the regions where dsRNAs were formed between antisense pseudogenes and their cognate coding genes. Simultaneously, the coding genes are suppressed through an RNAi pathway.

Pseudogene-Derived siRNAs from the Natural RNA Duplex Formed by Pseudogene–Pseudogene Pairs. Natural RNA duplexes can form between pseudogenes and their cognate expressed genes. Similarly, we presumed that pairs could also be formed from two transcribed pseudogenes with complementary regions. Using the method shown in Fig. 3, three groups of pseudogene–pseudogene pairs were generated (Table 2); no small RNAs from their parental genes were detected. Interestingly, most of these pseudogenes have been annotated as retrotransposon hot-spot (*RHS*) pseudogenes. siRNAs will block protein synthesis by partial complementarity or lead to mRNA cleavage by full complementarity (30). To investigate further whether all siRNAs derived from pseudogenes in *T. brucei* function in gene expression regu-

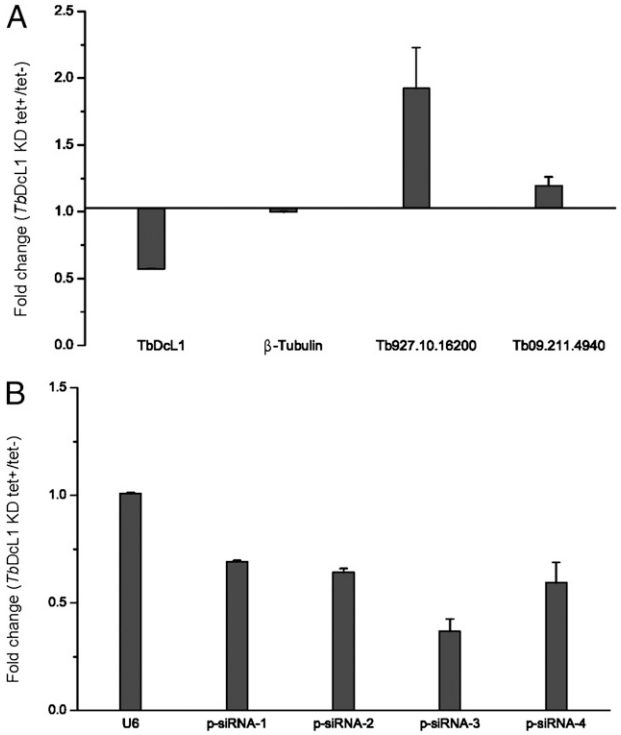

Fig. 2. Quantitative RT-PCR analysis of RNA levels in *TbDCL1* knockdown cell lines. (*A*) The mRNA level changes in conditional *TbDCL1* knockdown cell lines (tet+) compared with noninduced cells (tet–). Housekeeping β-tubulin genes were added as an external control. Error bars represent SD (n = 3). (*B*) Fold change of selective pseudogene-derived siRNAs in conditional *TbDCL1* knockdown cell lines (tet+) compared with noninduced cells (tet–). U6 small RNAs were added as an external control to both induced and noninduced cell lines. Error bars represent SD (n = 3). The sequences of siRNAs are listed in Table S1.

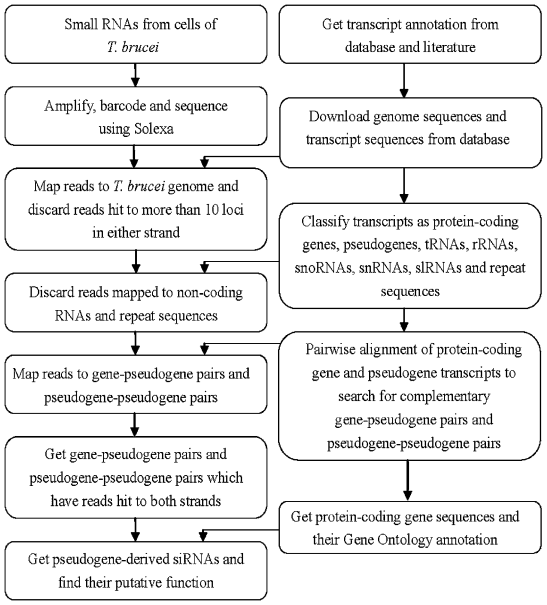

Fig. 3. Overall workflow for identifying siRNAs derived from gene–pseudogene pairs or pseudogene–pseudogene pairs.

lation, including those from gene–pseudogene pairs and pseudogene–pseudogene pairs, we mapped the siRNAs to potentially relevant regulatory targets. Many of these siRNAs were found to align to their targets located in regions essential for slicer cleavage, either with no mismatches or with few mismatches (Fig. 5*A*). Gene Ontology (GO) analysis on these candidate genes revealed that pseudogene-derived siRNA may regulate the expression of genes with various functions (Fig. 5*B*).

Discussion

A function of pseudogenes in mouse oocytes has been recently shown to be modulating gene expression (8, 9). In *T. brucei*, pseudogenes have long been considered to be involved only in ectopic recombination because no other functions have been ascertained for these genes (19).

With the proof of sense-strand and antisense-strand small RNAs coexisting in the complementary regions, we have found that some pseudogenes in *T. brucei* are able to base-pair with the corresponding protein-coding genes. When *TbDCL1* was knocked down, the expression of these protein-coding genes was up-regulated. In contrast, the amount of siRNAs originated from the pairs declined. These results provide strong evidence indicating that the unique role of siRNAs generated from pseudogenes could be to modulate gene expression. Additionally, the production of dsRNAs in *T. brucei* by interaction between sense and antisense transcripts from different genomic loci parallels the similar finding in mouse oocytes, the only tissue where this phenomenon has been observed in this animal (8). The direct binding between the sense and antisense transcripts in eukaryotes causes various effects, including the inhibition of translation and the formation of RNA duplexes that are precursors of siRNAs (31, 32). We have demonstrated that a set of siRNAs derives from gene–pseudogene pairs (Table 1). This result suggests that antisense pseudogene transcripts in *T. brucei* prominently function by inducing the degradation of sense protein-coding genes, similarly

Table 1. Gene–pseudogene pairs with significant numbers of siRNAs

Gene				Pseudogene			
Gene ID	Paired region	Sense RNA expression no.	Annotation	Gene ID	Paired region	Antisense RNA expression no.	Annotation
Tb927.10.16200	1…333	30.93333	Hypothetical protein	Tb927.5.4710	421…751	14.08333	*VSG* pseudogene
				Tb09.v4.0143	1…222	18.75	*VSG* pseudogene
Tb09.211.4940	149.0.395	5.0	Hypothetical protein	Tb927.5.4710	515…762	14.08333	*VSG* pseudogene
				Tb09.v4.0143	1…227	18.75	*VSG* pseudogene
Tb927.10.16210	756…888	124.0595	Hypothetical protein	Tb09.v4.0009	737…858	35.93452	*VSG* pseudogene
Tb11.01.0560	119…357	2.295238	Hypothetical protein	Tb927.5.4710	519…751	14.08333	*VSG* pseudogene
				Tb09.v4.0143	1…225	18.75	*VSG* pseudogene
Tb927.10.90	1…145	48.34444	Hypothetical protein	Tb09.142.0300	1…145	125.481	Hypothetical protein pseudogene
				Tb927.3.3420	1…145	83.15238	Hypothetical protein pseudogene

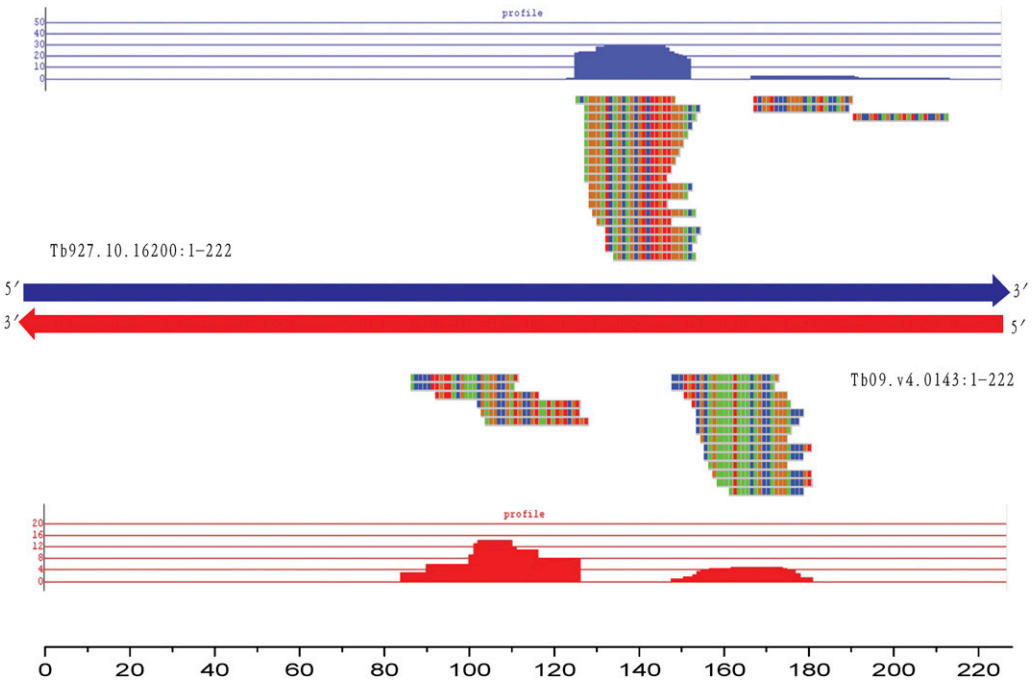

Fig. 4. Small RNAs originated from sense and antisense of gene–pseudogene interaction. Arrowed lines indicate the gene (blue) and its paired pseudogene (red); arrows represent the transcription direction (5′ to 3′). Locations of small RNAs mapped to the overlapped region of the two transcripts are drawn by two software programs, IGV (27) and DeepView (28). Short line segments above and below the arrowed lines indicate the reads originating from genes and pseudogenes, respectively. The colors indicate the types of nucleotide in the locus of the transcripts (green, A; blue, C; yellow, G; red, U). The profiles at the top and the bottom, respectively, indicate the expression number of small RNAs generated from genes and pseudogenes. The ruler at the bottom marks the paired region of the protein-coding gene.

as in a conventional RNAi pathway. So far, we do not have any idea whether other mechanisms are also involved in this processing; and therefore additional investigations are needed.

Earlier results have indicated that the ablation of argonaute (AGO) protein could cause an increase of transcripts of the *RHS* family (33), including protein-coding genes and their pseudogenes with integration of retro-elements (34). However, the cause of this change is unknown. AGO protein is one of the key components in the RNAi pathway. It binds unwound siRNAs and carries out nuclease activity directly against mRNAs that are complementary to their bound siRNAs (35). This indicates that the *RHS* family may be under the control of RNAi and that the siRNAs are essentially involved. Intriguingly, we found that an abundance of pseudogene-derived siRNAs originates from pseudogene–pseudogene pairs whose major components are *RHS* pseudogenes (Table 2). Therefore, we hypothesize that *RHS* pseudogenes in *T. brucei* are exploited to produce antisense siRNAs, which in return regulate the expression of the whole *RHS* family. Following the partial deletion of the *TbDCL1*, the expression of siRNAs derived from *RHS* pseudogenes (p-siRNA-1 and p-siRNA-2) is correspondingly down-regulated. However, we were surprised that similar change could not be detected in several other selected transcripts of *RHS* in our *TbDCL1* RNAi cell lines. It seems likely that the siRNAs dispersedly target other genes in addition to those in the *RHS* family, resulting in low regulatory efficiency that is beyond the detectable range in our study.

T. brucei is one of the most ancient eukaryotes in which RNAi is operational and is the only unicellular pathogen where RNAi has been extensively studied and used as a model for genetic analysis (16). So far as we know, in *T. brucei* most of the small nucleolar RNAs (snoRNAs) and siRNAs derived from retrotransposons or repeats have been considered to be endogenous, functional small noncoding RNAs (20, 25). Because of their specific origination, the function of the snoRNAs is thought to be focused on repeat silencing, whereas the siRNAs are considered to be limited to impact genome stability. No endogenous small RNAs in *T. brucei* had been demonstrated to control gene expression on the level of posttranscription until the discovery of the pseudogene-derived siRNAs in our work. Unlike the retrotransposons or repeats, the high degree of homology to the protein-coding genes determines the expression modulation potential of pseudogene-derived siRNAs in *T. brucei*. The present study has strongly verified this characteristic function of pseudogene-derived siRNAs. The function of pseudogene-derived siRNAs in gene expression reg-

Table 2. Pseudogene–pseudogene pairs with significant numbers of siRNAs

Pseudogene				Pseudogene			
Gene id	Paired region	Sense RNA expression no.	Annotation	Gene id	Paired region	Antisense RNA expression no.	Annotation
Tb927.1.450	5627…5828	656.4583	*RHS* pseudogene	Tb11.v4.0009	657…858	40.96786	*VSG* pseudogene
Tb927.1.430	5630…5831	721.9583	*RHS* pseudogene				
Tb927.1.300	5022–5223	484.119	*RHS* pseudogene				
Tb927.1.500	5019…5220	810.7917	*RHS* pseudogene				
Tb927.1.240	5631…58321	810.7917	*RHS* pseudogene				
Tb927.1.500	466…808	114.0	*RHS* pseudogene	Tb927.5.4710	421…762	14.08333	*VSG* pseudogene
Tb927.1.240	1031.0.1371	5.672619	*RHS* pseudogene				
Tb927.1.430	1091…1413	96.0	*RHS* pseudogene				
Tb927.1.350	468…715	4.6667	*RHS* pseudogene				
Tb927.1.350	471…697	4.66667	*RHS* pseudogene	Tb09.v4.0143	1…227	18.75	*VSG* pseudogene
Tb927.1.500	469…697	9.0	*RHS* pseudogene				
Tb927.1.240	1034…1260	5.672619	*RHS* pseudogene				

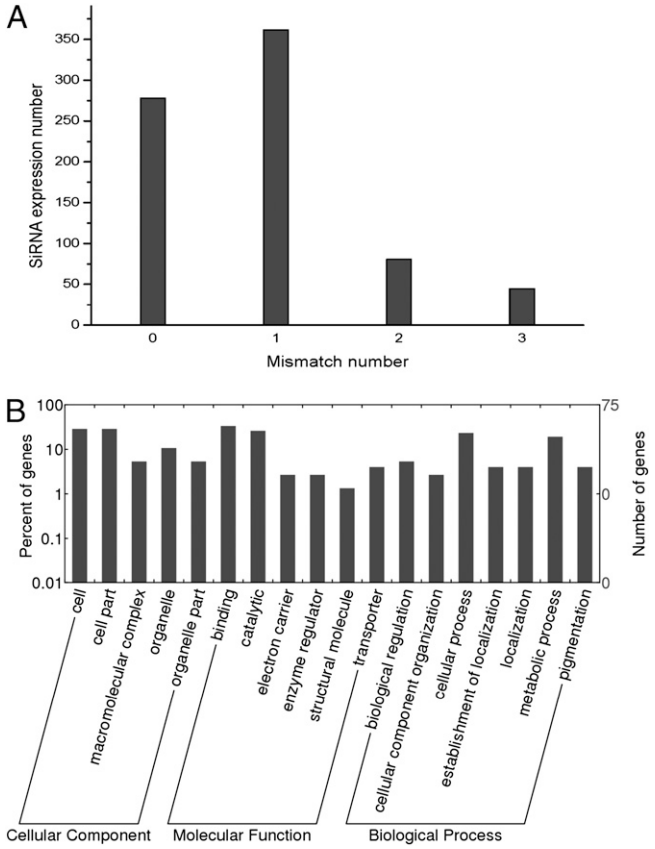

Fig. 5. The status of siRNAs originated from pseudogene pairs targeted to protein-coding genes. (*A*) siRNAs counted by the number of mismatches to protein-coding genes. (*B*) GO (gene ontology) annotations on protein-coding genes targeted by siRNAs.

ulation has been confirmed by experimental data in mouse oocytes (8, 9), and a similar function has been reported in rice on the grounds of computational analysis (10). The discovery of pseudogene-derived siRNAs in *T. brucei* suggests that siRNAs that have originated from pseudogenes should exist in all eukaryotic organisms, from single cell to mammals, and that this derivation might date back to the early unicellular eukaryotes. This consideration should prompt investigations into the origin of small noncoding RNAs in eukaryotic cells.

The analysis of our results shows that nucleotide bias exists in the 5′ terminal of pseudogene-derived siRNAs. This phenomenon has not been reported in other siRNAs in *T. brucei*, although it has been reported as an important feature of siRNAs in plants (21, 22). siRNAs are usually loaded to different AGO complexes on the basis of the specific 5′ terminal nucleotides to carry out distinct interference functions. For example, in *Arabidopsis*, 86% of small RNAs associated with AGO2 use uridine as the specific 5′ terminal nucleotide (21). Interestingly, our high through-put sequencing results indicate that not only pseudogene-derived small RNAs, but also 75% of total small RNAs found in *T. brucei* display the characteristics of 5′ terminal uridine. Because only one AGO complex (TbAGO1) has been found in *T. brucei* (35–37), the bias on the 5′ terminal nucleotide might not be useful for recognition of different AGO complexes. Nevertheless, the consistency in the great majority of siRNAs in *T. brucei* clearly indicates that the 5′ terminal uridine preference in this parasite is a signal for TbAGO1 to distinguish functional from dysfunctional siRNAs. Indeed, this trait of siRNAs likely has an ancient root in the eukaryotic lineage.

RNA-dependent RNA polymerase (RDRP) catalyzes the replication of RNA from an RNA template. In *Caenorhabditis elegans* and plants such as tomatoes, accumulation of siRNAs requires RDRP activity (38–41). One hypothesis is that RDRP uses the siRNAs as primers to amplify the target mRNAs to build up dsRNAs, thus generating more substrates for dicer cleavage and triggering secondary and long-lasting RNAi. However, RDRP does not exist in all organisms but was found in some species with tissue specificity. In mouse oocytes, for example, RDRP is absent, but pseudogenes and their homologous genes are able to form natural duplex RNAs and generate siRNAs (8, 9). In *T. brucei*, Patrick and colleagues (24) have reported that they could not detect RDRP, although they carefully examined the database. Our experimental results demonstrate that pseudogenes generate siRNAs in *T. brucei*. The similar origins and patterns of pseudogene-derived siRNAs in diverse organisms suggest the hypothesis that pseudogenes producing siRNAs would be an alternative way to compensate for the absence of RDRP. In both mouse oocytes and *T. brucei*, siRNAs are the key regulators for gene expression, and they are abundantly needed. Without a secondary wave of siRNAs made by RDPR, the pseudogenes are stimulated to generate enough siRNAs to fulfill the demands of development and differentiation. To further demonstrate this, additional experiments should be carried out, and more pseudogenes from various organisms need to be studied.

In conclusion, we have now shown that pseudogene-derived siRNAs in the parasitic protozoan *T. brucei* regulate protein-coding genes by the RNAi pathway. This not only represents a unique function of pseudogenes, but also enriches the source of siRNAs in this organism. However, the mechanism for the modulation of differentiation in *T. brucei* is still not completely clear. Here, GO analysis has shown that pseudogene-derived siRNAs target several kinds of functional genes, including genes involved in metabolic processes, indicating the regulatory potential of these siRNAs in differentiation. The discovery of pseudogene-derived siRNAs regulating gene expression at the posttranscriptional level highlights a mechanism for understanding the development and differentiation of this African trypanosome.

Materials and Methods

Cultivation of Trypanosome Cell Lines and Vector Construction. *T. brucei brucei* STIB920 (Swiss Tropical Institute, Basel) bloodstream forms were isolated from infected mice at a peak parasitemia by DEAE cellulose (DE-52; Whatman) purification (42). The cells were abundantly collected and suspended in TRIZOL (Invitrogen). Bloodstream forms of *T. b. brucei* Lister 427, which carried integrated genes for T7 polymerase and the tetracycline repressor, were grown in HMI-11 medium in the presence of G418. RNA interference of *TbDCL1* has been reported in a former study (23). About 500-bp-long nucleotides of TbDCL1-coding regions were PCR-amplified using the primers TbDCL1-Fw (5′-CTCGAGGATGAAGTTGGTGC-3′) and TbDCL1-Rv (5′-CTGCAG-TTTAATACAGCAGG-3′) (added XhoI and PstI restriction sites are underlined) from the genome and inserted between the two opposing tetracycline-inducible T7 RNA polymerase promoters of plasmid vector p2T7-177. The construct was linearized with NotI and transfected into the bloodstream form of strain 427, as previously described (43). Cells were cloned and selected against phleomycin.

Deep Sequencing. Total RNAs were extracted from cells by TRIZOL reagent following the manufacturer's protocol and run on agarose gels to check quality. Small RNA library preparation and Solexa sequencing were performed by the Beijing Genomics Institute in ShenZhen, China.

Data Source. Genomic sequences and scaffold sequences of *T. brucei* were downloaded from the TriTrypDB (release-2.1, update time: 04-Mar-2010 15:23), one component of EuPathDB (http://tritrypdb.org/tritrypdb/) (44). Kinetoplast maxicircle sequences with GenBank no. M94286.1 were obtained from the National Center for Biotechnology Information (NCBI). Transcript sequences and annotation information were also from TriTrypDB and rearranged by the in-house program Perl script. Sequences of SLAC and CIP147 repeat elements were retrieved from the genome according to their genome loci annotated in the literature, and sequences of ingi element were from NCBI (GI 162134).

Sequence Profiling and Analysis of Small RNAs. The 3′ adaptors were removed from the high through-put reads using a dynamic programming algorithm that required at least 5 nt overlapping between 35-nt reads and the 3′ adaptor sequence (45). After rejecting the adaptor sequences, poly(A) reads,

and 5′ adaptor contaminants, we mapped the remaining 18- to 30-nt reads to the *T. brucei* genome without mismatch, using the bowtie program (46). For the reads mapped to multiple loci, their expressions were divided on the basis of the hit times on the genome. Only the reads whose hit numbers were less than 10 were left. The reads that mapped to tRNA, rRNA, snoRNA, snRNA, slRNA, and repeat sequences were also eliminated.

Potential Source of Pseudogene-Derived siRNAs. Pairwise alignments were done between gene and pseudogene transcript sequences or between pseudogenes and pseudogenes to search out complementary pairs for siRNAs source. Pairs of transcript sequences were identified as sources of siRNAs if they satisfied the following criteria: (*i*) continuous pairing regions were longer than 30 nt, (*ii*) the complementation between the two transcripts was greater than 90%, and (*iii*) there were sequencing reads hit to both strands of the putative RNA duplex. We then counted the numbers of small RNAs mapped to the alignment region(s) as the expression number. One counterpart of the pair was defined as sense strand and the other one as antisense strand.

Target Prediction and GO Analysis. To predict the putative targets, we mapped these gene–pseudogene pair-derived or pseudogene–pseudogene pair-derived siRNAs to the protein-coding gene's transcript sequences, allowing for three mismatches. We then analyzed the GO function of these target genes. The program WEGO was used to facilitate histogram creation of GO annotation against cell component, biology process, and molecular function. A"GO level" of 3 was selected (47).

Reverse Transcription and Quantitative Real-Time PCR. Aliquots of 20 μg of total RNA separately extracted from noninduced *TbDCL1* RNAi cells, as well as from cells following 5 days of RNAi induction, were treated with DNase I (Takara). To prepare the templates for quantitative real time PCR, cDNAs were synthesized using the PrimeScript RT reagent kits (Takara), separately following the standard protocol for genes and the stem-loop reverse transcription protocol for siRNAs (48). A parallel reverse transcription reaction was also performed with the same amount of RNA without reverse transcriptase as a control for effective DNA contamination. All cDNAs were diluted properly and ready for quantitative PCR (qPCR). Primers used in qPCR are listed in Table S1 (Sangon). The qPCR reactions were performed with SYBR Premex Ex Taq kits (Takara) on the Bio-Rad iQ5 Gradient Real Time PCR Amplification System. The data were analyzed by iQ5 Optical System software.

ACKNOWLEDGMENTS. This work was supported by Grant 31071995 (to Z.-R.L.) from the National Science Foundation of China and by Grant 2011CB811300 (to L.-H.Q.) from the National Basic Research program ("973" program).

1. Vanin EF (1985) Processed pseudogenes: Characteristics and evolution. *Annu Rev Genet* 19:253–272.
2. Wilde CD (1986) Pseudogenes. *CRC Crit Rev Biochem* 19:323–352.
3. Korneev SA, Park JH, O'Shea M (1999) Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene. *J Neurosci* 19:7711–7720.
4. Balakirev ES, Ayala FJ (2003) Pseudogenes: Are they "junk" or functional DNA? *Annu Rev Genet* 37:123–151.
5. Korneev SA, O'Shea M (2002) Evolution of nitric oxide synthase regulatory genes by DNA inversion. *Mol Biol Evol* 19:1228–1233.
6. Hirotsune S, et al. (2003) An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature* 423:91–96.
7. Korneev SA, et al. (2005) Timed and targeted differential regulation of nitric oxide synthase (NOS) and anti-NOS genes by reward conditioning leading to long-term memory formation. *J Neurosci* 25:1188–1192.
8. Tam OH, et al. (2008) Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* 453:534–538.
9. Watanabe T, et al. (2008) Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* 453:539–543.
10. Guo X, Zhang Z, Gerstein MB, Zheng D (2009) Small RNAs originated from pseudogenes: *cis*- or *trans*-acting? *PLOS Comput Biol* 5:e1000449.
11. Brun R, Blum J, Chappuis F, Burri C (2010) Human African trypanosomiasis. *Lancet* 375: 148–159.
12. Koumandou VL, Natesan SK, Sergeenko T, Field MC (2008) The trypanosome transcriptome is remodelled during differentiation but displays limited responsiveness within life stages. *BMC Genomics* 9:298.
13. Jensen BC, Sivam D, Kifer CT, Myler PJ, Parsons M (2009) Widespread variation in transcript abundance within and across developmental stages of *Trypanosoma brucei*. *BMC Genomics* 10:482.
14. Queiroz R, Benz C, Fellenberg K, Hoheisel JD, Clayton C (2009) Transcriptome analysis of differentiating trypanosomes reveals the existence of multiple post-transcriptional regulons. *BMC Genomics* 10:495.
15. Ouellette M, Papadopoulou B (2009) Coordinated gene expression by post-transcriptional regulons in African trypanosomes. *J Biol* 8:100.
16. Ullu E, Tschudi C, Chakraborty T (2004) RNA interference in protozoan parasites. *Cell Microbiol* 6:509–519.
17. Mallick B, Ghosh Z, Chakrabarti J (2008) MicroRNA switches in *Trypanosoma brucei*. *Biochem Biophys Res Commun* 372:459–463.
18. Berriman M, et al. (2005) The genome of the African trypanosome Trypanosoma brucei. *Science* 309:416–422.
19. Thon G, Baltz T, Eisen H (1989) Antigenic diversity by the recombination of pseudogenes. *Genes Dev* 3:1247–1254.
20. Djikeng A, Shi H, Tschudi C, Ullu E (2001) RNA interference in *Trypanosoma brucei*: Cloning of small interfering RNAs provides evidence for retroposon-derived 24-26-nucleotide RNAs. *RNA* 7:1522–1530.
21. Mi S, et al. (2008) Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5′ terminal nucleotide. *Cell* 133:25–26.
22. Montgomery TA, et al. (2008) Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. *Cell* 133:128–141.
23. Shi H, Tschudi C, Ullu E (2006) An unusual Dicer-like1 protein fuels the RNA interference pathway in *Trypanosoma brucei*. *RNA* 12:2063–2072.
24. Patrick KL, et al. (2009) Distinct and overlapping roles for two Dicer-like proteins in the RNA interference pathways of the ancient eukaryote *Trypanosoma brucei*. *Proc Natl Acad Sci USA* 106:17933–17938.
25. Gupta SK, et al. (2010) Small nucleolar RNA interference in *Trypanosoma brucei*: Mechanism and utilization for elucidating the function of snoRNAs. *Nucleic Acids Res* 38:7236–7247.
26. Hildebrandt M, Nellen W (1992) Differential antisense transcription from the Dictyostelium EB4 gene locus: Implications on antisense-mediated regulation of mRNA stability. *Cell* 69:197–204.
27. Robinson JT, et al. (2011) Integrative genomics viewer. *Nat Biotechnol* 29:24–26.
28. Yang JH, Shao P, Zhou H, Chen YQ, Qu LH (2010) deepBase: A database for deeply annotating and mining deep sequencing data. *Nucleic Acids Res* 38(Database issue): D123–D130.
29. Hannon GJ (2002) RNA interference. *Nature* 418:244–251.
30. Doench JG, Petersen CP, Sharp PA (2003) siRNAs can function as miRNAs. *Genes Dev* 17:438–442.
31. Vanhée-Brossollet C, Vaquero C (1998) Do natural antisense transcripts make sense in eukaryotes? *Gene* 211:1–9.
32. Militello KT, Refour P, Comeaux CA, Duraisingh MT (2008) Antisense RNA and RNAi in protozoan parasites: Working hard or hardly working? *Mol Biochem Parasitol* 157: 117–126.
33. Durand-Dubief M, et al. (2007) The Argonaute protein TbAGO1 contributes to large and mini-chromosome segregation and is required for control of RIME retroposons and RHS pseudogene-associated transcripts. *Mol Biochem Parasitol* 156:144–153.
34. Bringaud F, et al. (2002) A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*. *Eukaryot Cell* 1:137–151.
35. Durand-Dubief M, Bastin P (2003) TbAGO1, an argonaute protein required for RNA interference, is involved in mitosis and chromosome segregation in *Trypanosoma brucei*. *BMC Biol* 1:2.
36. Shi H, Djikeng A, Tschudi C, Ullu E (2004) Argonaute protein in the early divergent eukaryote *Trypanosoma brucei*: Control of small interfering RNA accumulation and retroposon transcript abundance. *Mol Cell Biol* 24:420–427.
37. Shi H, Tschudi C, Ullu E (2007) Depletion of newly synthesized Argonaute1 impairs the RNAi response in Trypanosoma brucei. *RNA* 13:1132–1139.
38. Vaucheret H (2006) Post-transcriptional small RNA pathways in plants: Mechanisms and regulations. *Genes Dev* 20:759–771.
39. Pak J, Fire A (2007) Distinct populations of primary and secondary effectors during RNAi in C. elegans. *Science* 315:241–244.
40. Sijen T, Steiner FA, Thijssen KL, Plasterk RH (2007) Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. *Science* 315:244–247.
41. Kasschau KD, et al. (2007) Genome-wide profiling and analysis of Arabidopsis siRNAs. *PLoS Biol* 5:e57.
42. Lanham SM, Godfrey DG (1970) Isolation of salivarian trypanosomes from man and other mammals using DEAE-cellulose. *Exp Parasitol* 28:521–534.
43. Hashimi H, Cicová Z, Novotná L, Wen YZ, Lukes J (2009) Kinetoplastid guide RNA biogenesis is dependent on subunits of the mitochondrial RNA binding complex 1 and mitochondrial RNA polymerase. *RNA* 15:588–599.
44. Aslett M, et al. (2010) TriTrypDB: A functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res* 38(Database issue):D457–D462.
45. Liao JY, et al. (2010) Deep sequencing of human nuclear and cytoplasmic small RNAs reveals an unexpectedly complex subcellular distribution of miRNAs and tRNA 3′ trailers. *PLoS ONE* 5:e10563.
46. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25.
47. Ye J, et al. (2006) WEGO: A web tool for plotting GO annotations. *Nucleic Acids Res* 34(Web Server issue):W293–W297.
48. Chen C, et al. (2005) Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res* 33:e179.