

Comparative Analysis of the 5S rRNA and Its Associated Proteins Reveals Unique Primitive Rather Than Parasitic Features in *Giardia lamblia*

Jin-Mei Feng^{1,2,3}, Jun Sun^{1,2,3}, De-Dong Xin³, Jian-Fan Wen^{1*}

1 State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, People's Republic of China, **2** Graduate School of Chinese Academy of Sciences, Beijing, People's Republic of China, **3** College of Chemistry and Life Sciences, Zhejiang Normal University, Jinhua, People's Republic of China

Abstract

Background: 5S rRNA is a highly conserved ribosomal component. Eukaryotic 5S rRNA and its associated proteins (5S rRNA system) have become very well understood. *Giardia lamblia* was thought by some researchers to be the most primitive extant eukaryote while others considered it a highly evolved parasite. Previous reports have indicated that some aspects of its 5S rRNA system are simpler than that of common eukaryotes. We here explore whether this is true to its entire system, and whether this simplicity is a primitive or parasitic feature.

Methodology/Principal Findings: By collecting and confirming pre-existing data and identifying new data, we obtained almost complete datasets of the system of three isolates of *G. lamblia*, two other parasitic excavates (*Trichomonas vaginalis*, *Trypanosoma cruzi*), and one free-living one (*Naegleria gruberi*). After comprehensively comparing each aspect of the system among these excavates and also with those of archaea and common eukaryotes, we found all the three *Giardia* isolates to harbor a same simplified 5S rRNA system, which is not only much simpler than that of common eukaryotes but also the simplest one among those of these excavates, and is surprisingly very similar to that of archaea; we also found among these excavates the system in parasitic species is not necessarily simpler than that in free-living species, conversely, the system of free-living species is even simpler in some respects than those of parasitic ones.

Conclusion/Significance: The simplicity of *Giardia* 5S rRNA system should be considered a primitive rather than parasitically-degenerated feature. Therefore, *Giardia* 5S rRNA system might be a primitive system that is intermediate between that of archaea and the common eukaryotic model system, and it may reflect the evolutionary history of the eukaryotic 5S rRNA system from the archaeal form. Our results also imply *G. lamblia* might be a primitive eukaryote with secondary parasitically-degenerated features.

Citation: Feng J-M, Sun J, Xin D-D, Wen J-F (2012) Comparative Analysis of the 5S rRNA and Its Associated Proteins Reveals Unique Primitive Rather Than Parasitic Features in *Giardia lamblia*. PLoS ONE 7(6): e36878. doi:10.1371/journal.pone.0036878

Editor: Paul Wrede, Charité-Universitätsmedizin Berlin, Germany

Received: February 1, 2012; **Accepted:** April 13, 2012; **Published:** June 7, 2012

Copyright: © 2012 Feng et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by grants from the National Natural Science Foundation of China (30830018; 31172081) and the Knowledge Innovation Program (KSCX2-EW-J-23). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: wenjf@mail.kiz.ac.cn

These authors contributed equally to this work.

Introduction

5S rRNA is a highly conserved component of large ribosomal subunit in all organisms, excepting only certain mitochondrial systems [1]. In prokaryotes, 5S rRNA is generally transcribed in a common transcription unit with 16S and 23S rRNAs, while in most eukaryotes, 5S rRNA gene is located separately from 18S, 5.8S and 28S rRNA genes and is transcribed alone by RNA polymerase III (RNA pol III) with the participation of a specific transcription factor IIIA (TFIIIA). Here TFIIIA not only acts as an essential transcription factor by binding to the 5S rRNA gene but also combines with 5S rRNA itself to form a 7S ribonucleoprotein (RNP) particle to stabilize the rRNA and facilitate its nuclear export [2,3]. Ribosomal protein L5 is also known to bind specifically to 5S rRNA and to be involved in its cytoplasmic-nucleolar transport. In eukaryotes, prior to ribosome assembly, the

5S rRNA transcript changes binding partners, trading TFIIIA for L5, and is then escorted to the nucleolus by the latter [4]. In the present work, we refer to 5S rRNA and these associated proteins as the 5S rRNA system. From the above, we find that the components and function of this system are very clear, making it a suitable candidate for investigation across diverse organisms for the purposes of a comparative evolutionary study.

Giardia lamblia is one of the most widespread intestinal protozoan parasites in the world. It was once thought to be the most primitive extant eukaryote because it has many so-called primitive molecular and cellular traits [5] and branches out very early from the eukaryote trunk on many phylogenetic trees [6–10]. But, actually *Giardia*'s evolutionary position has been under controversy since the proposition of its being extreme primitive eukaryote [11], especially after the discovery of its mitochondrial remnant

organelle mitosome [12]. Recently, phylogenomic analyses based on large phylogenomic datasets have placed *G. lamblia* within excavate, one of the three primary divisions of eukaryotes [13]. However, no matter it is primitive or not, it is amazing that it was reported that in this organism no 5S rRNA was detected in either experiments or BLAST searches against its genome database [14,15], since 5S rRNA is so ubiquitous and highly conserved in both prokaryotic and eukaryotic cells. For a long time, there were no 5S rRNA annotations in the *G. lamblia* genome database before we began the present study. Later, 5S rRNA genes, though there were only eight copies, far fewer than the number seen in most common eukaryotes, were annotated in *G. lamblia* genome database [16]. Studies on some *G. lamblia* 5S rRNA-associated proteins revealed that they are much simpler than those of typical eukaryotes. For example, only one subunit specific to RNA pol III, called C34, was found, and none of the transcriptional factors of RNA polymerase III (TFIIIA, TFIIIB, or TFIIIC) were found either. There were only two components of TFIIIB (TBP and BRF) identified in the *G. lamblia* genome database [17]. But it is unknown whether the entire *G. lamblia* 5S rRNA system is simpler than that of common eukaryotes yet.

Some authors consider the simplicity of many aspects of *G. lamblia*, including genomic reduction, to be the results of parasitic degeneration rather than primitive traits [16,18–20]. We doubt that this is the case for certain aspects of this organism, such as 5S rRNA system, which have no direct relationship with parasitic lifestyle but are rather involved in the fundamental functions of a eukaryotic cell. To confirm this, by collecting published and identifying new data of 5S rRNA system, we compared the *G. lamblia* 5S rRNA system to those of two parasitic species (*Trichomonas vaginalis* and *Trypanosoma cruzi*) and one free-living species (*Naegleria gruberi*), which here served as the representatives with genome database in other three subgroups of excavate, respectively, and also to those of archaea and common eukaryotes. In addition, the genome databases of two other *G. lamblia* isolates, assemblage B isolate GS and assemblage E isolate P15, are now available [21,22]. In this way, we identified their 5S rRNA system to confirm whether the simplicity of the 5S rRNA system is a common trait among different *G. lamblia* isolates.

Materials and Methods

Database

Data regarding genomes and the annotated transcripts and proteins of *G. lamblia* assemblage A isolate WB, B isolate GS, and E isolate P15 were downloaded from the GiardiaDB (<http://giardiadb.org/giardiadb/>). Genomic data regarding *T. vaginalis* (<http://trichdb.org/trichdb/>), *T. cruzi* (<http://tritrypdb.org/tritrypdb/>), and *N. gruberi* (<http://genome.jgi-psf.org/Naegr1/Naegr1.home.html>) were downloaded from their respective databases.

Data Collection and Identification

We collected previously reported data on the 5S rRNA system, including the 5S rRNA gene, RNA pol III, associated transcriptional factors, and L5, in *G. lamblia* (including isolates WB, GS, and P15), *T. vaginalis*, *T. cruzi*, and *N. gruberi*. They are summarized as follows: 1) The 5S rRNA gene and its promoters: Eight 5S rRNA genes were reported in the genome database of *G. lamblia* WB and six in the P15 database [16,22]. The 5S rRNA gene and its TATA box and internal control region (ICR) were reported in *T. vaginalis* [23]. 5S rRNA genes and their ICRs were identified in *T. cruzi* [24]; 2) transcriptional factors and RNA pol III-specific subunits: a TFIIIB with only two components and a RNA pol III with only

the specific subunit, C34, were previously reported in *G. lamblia* WB [17]; 3) L5 sequences of *G. lamblia* (including WB, GS, and P15), *T. vaginalis*, *T. cruzi*, and *N. gruberi* were all obtained through the Entrez system at NCBI (<http://www.ncbi.nlm.nih.gov/sites/gquery>).

The data not already published were identified as follows:

5S rRNA Gene

An *in silico* search for the 5S rRNA genes of *G. lamblia* GS and *N. gruberi* was performed using the BLASTn program against the *G. lamblia* GS genome database and *N. gruberi* genome database [21,25]. The 5S rRNA gene sequences from prokaryotes and eukaryotes that were used as query sequences were retrieved from the 5S rRNA database (<http://www.man.poznan.pl/5SData>) [26]. BLAST search default parameters were modified to reduce stringency: the penalty for a gap or a mismatch was changed from 3 to 1 to enable us to identify the 5S rRNA gene. The 5S rRNA genes of *G. lamblia* WB and P15 were used as queries to search for 5S rRNA gene against the *G. lamblia* GS genome database by using the BLAST program (<http://giardiadb.org/giardiadb/>).

Data regarding the internal control region (ICR) and TATA box were found in the *G. lamblia* 5S rRNA genes in the following manner: For ICR, the consensus sequences of its three elements: Box A [5'-(N(G/C)(C/T)(C/T)AANCNNNNNN-3')], intermediate element (IE) [5'-(C/G)NN(G/A)(G/A)N-3'], and Box C [5'-NNG(G/A)TGGGNG(T/A)CCN(C/T)NNG-3'] were searched manually within the *G. lamblia* 5S rRNA genes. For the TATA box, its highly conserved core sequence (5'-TATAAA-3') was used as a consensus sequences to analyze the upstream sequences of *G. lamblia* 5S rRNA genes. Because of the flexible position of the TATA box, an AT-rich region, being relative to the transcriptional start site in *G. lamblia*, the entire 60 bp upstream region of the transcriptional start site of *G. lamblia* 5S rRNA genes were analyzed. Using these same methods, the ICR of 5S rRNA gene in *N. gruberi* and TATA boxes in the 5S rRNA genes of *T. cruzi* and *N. gruberi* were identified.

TFIIIA

No TFIIIA had previously been identified in *G. lamblia* WB by homolog search against its partial genome assemblies [17]. Once the fully sequenced *G. lamblia* WB genome database was available, we tried to identify TFIIIA via BLASTp and PSI-BLAST searches against the database again. Considering the high primary sequence divergences of TFIIIA among different species, we used the C2H2 motif of TFIIIA as query. The C2H2 motif-containing sequences were obtained from the PROSITE database (release 32 of the PROSITE pattern ZINC-FINGER_C2H2, Accession No. PS00028). Then the best hits were filtered by the presence of C2H2 zinc finger domains using the Interproscan program [27]. TFIIIA was identified in *G. lamblia* GS and P15, *T. vaginalis*, *T. cruzi*, and *N. gruberi* by BLASTp and PSI-BLAST searching against their corresponding genome databases.

TFIIIB, TFIIIC, and Specific Subunits of RNA Pol III

The data regarding TFIIIB, TFIIIC and specific subunits of RNA pol III in *G. lamblia* were reported previously by BLAST searching against the *G. lamblia* WB partial genome assembly [17]. In this work, we again found them in the fully sequenced genome database. *Saccharomyces cerevisiae* sequences of subunits of TFIIIB and TFIIIC and of the specific subunits of RNA pol III were used as queries to BLASTp search against genome databases of all the three *G. lamblia* isolates, *T. vaginalis*, *T. cruzi*, and *N. gruberi*, to get putative homologs. Then the putative homologs were assessed using domain information in the Pfam database (<http://pfam>).

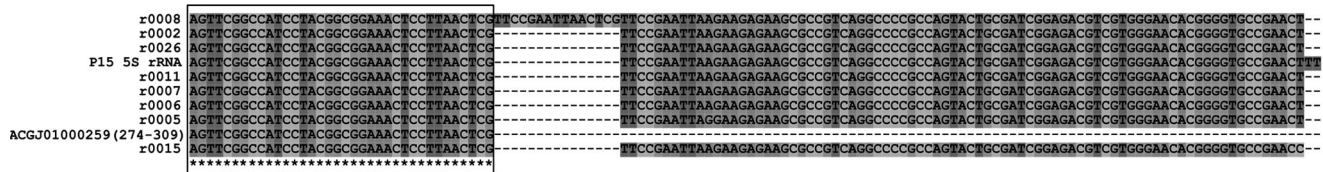


Figure 1. Sequence alignments of 5S rRNA genes of the three *G. lamblia* isolates. r0002, r0005, r0006, r0007, r0011, r0015, r0026, and r0008 represent the eight *G. lamblia* WB 5S rRNA genes. ACGJ01000259 (274–309) represents sequences from residue 274 to 309 of contig ACGJ01000259 in the *G. lamblia* GS genome. Box denotes the putative partial 5S rRNA gene identified in *G. lamblia* GS sharing 100% identity with 5' fragment of 5S rRNA gene in *G. lamblia* WB and P15.

doi:10.1371/journal.pone.0036878.g001

sanger.ac.uk) [28]. Another BLASTp search was performed, this time against the non-redundant (nr) NCBI protein database. Sequences lacking characteristic domains of the given proteins or indicating other proteins were discarded. For proteins for which no homologs could be found by BLASTp search, at least five iterations of PSI-BLAST search was carried out online to identify putative homologs. Then these putative homologs were confirmed by Pfam domain analysis and BLASTp against the NCBI nr database, as described above.

RT-PCR and Sequencing

To examine the transcription activities of the *G. lamblia* 5S rRNA genes, reverse transcription polymerase chain reaction (RT-PCR) experiments were carried out as follows: Small RNAs (≤ 200 nucleotides) were extracted from *G. lamblia* WB cells using a mirVanaTM miRNA Isolation Kit (Ambion, Austin, TX, U.S.) according to the manufacturer's instructions. Then the small RNAs were polyadenylated at 37°C for 30 min in 60 μ l of reaction mixture with 500 ng small RNAs, 3 mM rATP, 6 μ l 10 \times Buffer and 15 U *E. coli* poly(A) polymerase (New England BioLab), and the poly(A)-tailed RNAs were recovered by phenol/chloroform extraction and ethanol precipitation. Finally the purified RNAs were used as templates for RT-PCR, using a RNA PCR Kit (TaKaRa Biotechnology Co., Ltd., Dalian, China.) according to the manufacturer's instructions. The gene-specific primers were designed according to the sequence of our identified *G. lamblia* 5S rRNA genes. Primers (Pfr 1:5'-TCggCCATCCTACggCggAAACT-3'; Pfr 2:5'-CCCACgACgTCTCCgATCgCagT-3') were synthesized by Sangon (Sangon Ltd., Shanghai, China).

The purified products of these RT-PCR experiments were cloned into PMD 19-T vectors (TaKaRa). Twenty clones were selected at random and sequenced at Sangon Ltd.

Results

Identification of 5S rRNA Genes in *G. lamblia* (WB, GS, and P15) and *N. gruberi*

Our BLASTn search included modified parameters to reduce stringency. It confirmed that there are only eight 5S rRNA genes located separately in the *G. lamblia* WB database and six in the P15 database. This is consistent with the annotations in these two databases [16,22]. Based on the genome sequence, seven of the eight *G. lamblia* WB 5S rRNA genes each were found to have a putative 5S rRNA coding region 117 bp in length while the other had an extra 15 bp insert within the coding region. However, in our RT-PCR and sequencing experiments, the products detected were only from the 117 bp genes. Thus, the gene with the extra insert might not be transcribed.

No putative 5S rRNA gene was identified in the *G. lamblia* GS genome database upon the retrieval of various 5S rRNA sequences

from the 5S rRNA database as queries to BLASTn search against the genome database. However, one hit (Contig ID: ACGJ01000259), which was found to possess a 36 bp sequence sharing 100% identity with the 5' terminal region (from +1 bp to 36 bp) of 5S rRNA genes of *G. lamblia* WB and P15 was found when the 5S rRNA genes of the two isolates were used as queries (Figure 1). The region in contig ACGJ01000259 (309 bp long) sharing 100% identity with 5S rRNA gene (from +1 bp to 36 bp) of *G. lamblia* WB and P15 was found at the end of this contig (from 274 bp to 309 bp). Moreover, a 79 bp RNA fragment sharing 99% identity with the 3' terminal region (from +39 bp to +117 bp) of *G. lamblia* WB 5S rRNA was once reported in *G. lamblia* GS [21]. In this way, the lack of a *G. lamblia* 5S rRNA gene in the *G. lamblia* GS genome database is most probably due to the numerous gaps in the database and the small number genes in this organism.

Using the BLASTn program to search against the *N. gruberi* genome database, we found 279 potential 5S rRNA genes (Table S1).

Comparisons of 5S rRNA Genes among *G. lamblia*, and Other Three Excavates, Common Eukaryotes, and Archaea

According to the above identified 5S rRNA genes in *G. lamblia* and *N. gruberi*, and the reported many copies of 5S rRNA gene in *T. vaginalis* and *T. cruzi* [23,24], it can be said that in contrast to common eukaryotes and the other three investigated excavates, *G. lamblia* had a very small copy number of 5S rRNA genes and no tandem 5S rRNA gene regions, suggesting that it is very similar to archaea (Table 1).

Analysis on putative promoter elements revealed that there are no special relevance to the intermediate element [IE: 5'-(C/G)NN(G/A)(G/A)N-3'], the Box A [5'-N(G/C)(C/T)(C/T)AANCNNNNNN-3'], or Box C [5'-NNG(G/A)TGGGNG(T/A)CCN(C/T)NNG-3'] of the ICR in any part of the *G. lamblia* WB or P15 5S rRNA genes. This indicates that these genes do not possess ICR and would not bind TFIIIA specifically during transcription, as found in common eukaryotes. Our analysis showed that presence of ICR in the 5S rRNA genes of *N. gruberi* (Figure 2). The presences of ICR have been reported previously in the 5S rRNA genes of *T. cruzi* and *T. vaginalis* [23,24]. This suggests that the three other excavates, both parasitic and free-living, are similar to common eukaryotes in this respect.

However, although no canonical TATA box was found in *G. lamblia* 5S rRNA genes, we have identified a 16 bp AT-rich region between -58 and -43 bp upstream of the transcriptional initiation site of these genes, whose consensus sequence is 5'-AA(T/C)A(C/T/A)(A/G) C(A/T)(G/A)(T/A)AT(G/T)(T/A)A(T/C)-3' (Figure 3). Previous studies have indicated that the TATA boxes of *G. lamblia* genes are not as canonical as those of

r0007	-60	ACAATACACA ATATGA ATCT	CTAACAGTCT	CTGACGCATC	GCACAGCTCC	CTCAGCCGCC	-1
r0006	-60	ACAATACACA ATATGA ATCT	CTAACAGTCT	CTGACGCATC	GCACAGCTCC	CTCAGCCGCC	-1
r0011	-60	ACAATATACA GTATGA ATCT	CTAACAACTCT	CTGACACATC	GCACAGCTCC	CTCAGCCGCC	-1
r0026	-60	ACAATATACA GTATGA ATCT	CTAACAACTCT	CTGACACATC	GCACAGCTCC	CTCAGCCGCC	-1
r0002	-60	ACAATATACA GTATGA ATCT	CTAACAACTCT	CTAACACATC	GCACAGCCCT	GCCGGCCGCC	-1
r0015	-60	ACAATATACA GTATGA ATCT	CTAACAGTCT	CTGACACATC	GCACAGCTCC	CTCAGCCGCC	-1
r0005	-60	GTAACAAGCT AAATTT ACAA	GCTCAAGTCT	TTTGTGCATT	GCACAGCCCC	ACCGACCGCC	-1

Figure 3. TATA-like box identified in *G. lamblia* WB 5S rRNA gene. The regions between -58 and -43 bp upstream the transcriptional initiation site of *G. lamblia* WB 5S rRNA genes all contain an AT-rich region (denoted by box), which might be the TATA-like box of *G. lamblia* WB 5S rRNA gene.

doi:10.1371/journal.pone.0036878.g003

has been reported to have TATA box in its 5S rRNA genes [23] (Table 1).

Identification and Comparison of Transcriptional Factors and Specific Subunits of RNA Pol III

TFIIIA was found in *T. vaginalis* but not in *T. cruzi*, *N. gruberi*, or the three *G. lamblia* isolates upon BLASTp or PSI-BLAST search against their corresponding genome databases. We then used a C2H2 motif as a query for another BLASTp search against the *G. lamblia* WB genome database to confirm whether TFIIIA is really absent from it. After the BLASTp search, we found 1,959 best hits. Of these, 102 were predicted to be putative zinc finger proteins, and 18 out of these 102 proteins were further predicted to be C2H2-type zinc finger proteins. But 17 out of the 18 proteins each contained only 1–2 zinc finger repeats, and the other one (ORF: 17003) had a total of four repeats. In common eukaryotes, the number of TFIIIA finger repeats is usually nine. None of the 18 C2H2-type zinc finger proteins were annotated as TFIIIA in the NCBI database (Table S2).

Our investigation further confirmed that this organism lacks TFIIIB and TFIIIC, but two components of TFIIIB were present in *G. lamblia* WB as previously reported. In *G. lamblia* GS and P15, similarly, no TFIIIB or TFIIIC was found, but two components of TFIIIB were identified. However, both TFIIIB and TFIIIC were found in all the other three investigated excavates, *T. vaginalis*, *T. cruzi*, and *N. gruberi* (Table S3).

Of the five specific subunits of RNA pol III (C17, C31, C34, C53, and C82), two (C34 and C82) were identified in *G. lamblia* WB, GS, and P15 by our PSI-BLAST. Of these, only one, C34, had been reported in *G. lamblia* WB before this work [17]. Two, C34 and C53, were found in *T. cruzi*; three, C34, C53, and C82,

were identified in *T. vaginalis*; and four, C34, C53, C82, and C17, were found in the free-living species *N. gruberi* (Table S3).

Examination of Transcription of *G. lamblia* 5S rRNA Gene

Our RT-PCR experiment on the *G. lamblia* WB 5S rRNA genes showed positive results (Figure 4). The sequencing results indicated the RT-PCR products were the exact transcripts of the seven *G. lamblia* WB 5S rRNA genes with the normal length (117 bp), while no transcript corresponds to the other one which has a 15 bp insertion. Therefore, the seven normal-length *G. lamblia* WB 5S rRNA genes were found to be actively transcribed while the other was not.

Discussion

By collecting and confirming the existing data and identifying new data, we were able to construct almost complete datasets of the 5S rRNA systems of the three sequenced isolates of *G. lamblia* and three other excavates. Therefore, in this study, we were able to comprehensively compare each aspect of the 5S rRNA system among them and with those of archaea and common eukaryotes.

In the genomes of common eukaryotes, 5S rRNA genes are usually arranged in tandem arrays of several hundreds and thousands of copies in several different rDNA regions. The total gene copy number varies widely by species. Humans have 1700–2000 copies and *Xenopus* oocytes have \approx 20,000 [34,35]. Excavates, *Crithidia fasciculata* have \approx 250, *T. vaginalis* and *T. cruzi* each have hundreds and thousands of copies [23,24,36]. There are 279 putative 5S rRNA genes in *N. gruberi*. However, our study has indicated that in all the three sequenced *G. lamblia* isolates, only a few copies of 5S rRNA genes exist. The total gene copy number in *G. lamblia* is unexpectedly small, and the genes are not arranged in tandem. This is not consistent with the pattern observed in common eukaryotes and in other excavates (both free-living and parasitic) but rather with those in archaea. On the other hand, the *G. lamblia* 5S rRNA genes are located far from the 18S-5.8S-28S rRNA gene unit, as found in common eukaryotes, suggesting they are also transcribed independently by RNA pol III.

In this study we found that there are no ICR but rather a gene-external TATA-like box in each *G. lamblia* 5S rRNA gene. Most eukaryotes have both [37]. In the other three excavates studied, both well defined gene-external TATA box and gene-internal ICR appear in *T. vaginalis* 5S rRNA genes [23], and both were also found in 5S rRNA genes of *N. gruberi* in the present study, however, there is only gene-internal ICR in *T. cruzi* 5S rRNA genes [24]. The characteristics of 5S rRNA gene promoters in these excavates

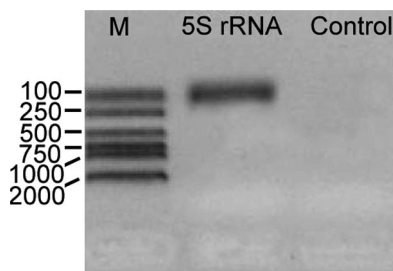


Figure 4. Results of the RT-PCR experiment of the identified *G. lamblia* WB 5S rRNA gene. M, marker.

doi:10.1371/journal.pone.0036878.g004

suggests which kind of promoter appears in a species has nothing to do with its lifestyle. However, the results of our study indicate that the *G. lamblia* 5S rRNA gene promoter might be a very primitive, and similar to that of archaea. Moreover, this might further supports the idea that during evolution of eukaryotes, the gene-external TATA box promoter of the 5S rRNA genes might have been inherited from archaea and emerge very early in the evolution of eukaryotes, while the gene-internal ICR arose at later point after the divergence of *G. lamblia* from the eukaryotic trunk.

Usually, three transcription factors are involved in the transcription of 5S rRNA by RNA pol III in common eukaryotes, and TFIIIA is the only one specific to this process. The binding of TFIIIA to the ICR of 5S rRNA gene is the first step of the transcription of this gene. This allows the recruitment of TFIIIC, TFIIIB, and RNA pol III [38]. TFIIIA also combines with 5S rRNA to export it from the nucleus to the cytoplasm. Up to now, knowledge about TFIIIA has mainly come from studies of higher eukaryotes, and no TFIIIA has ever been reported in protists. One previous study of *G. lamblia* reported finding no TFIIIA [17]. Considering that this previous study was based on partial *G. lamblia* genome assemblies, the present study used the complete *G. lamblia* genome databases and relatively conserved C2H2 domain of TFIIIA as query to search TFIIIA, but still identify no homolog to the typical eukaryotic TFIIIA in the three isolates of *G. lamblia*. Our sequence analysis also revealed that the absence of ICR in the *G. lamblia* 5S rRNA gene, which probably also implies the absence of TFIIIA. These findings further confirm the absence of TFIIIA in *Giardia*, and suggest that *G. lamblia* 5S rRNA transcription is TFIIIA-independent. Previously, TFIIIA-independent 5S rRNA transcription in eukaryotes was only reported in *Yarrowia lipolytica*, where 5S rRNA gene is co-transcribed with tRNA [39]. But no *G. lamblia* 5S rRNA genes have a neighboring location with tRNA gene as in *Y. lipolytica* according to our investigation. Furthermore, there is no homologs of TFIIIB and TFIIIC in *G. lamblia* [17], which was also further confirmed in the present study. These findings may indicate that the initiation mechanism of *G. lamblia* 5S rRNA transcription is different from those of all eukaryotes studied. Although it does not have TFIIIB, *G. lamblia* possesses two important subunits of TFIIIB: TATA box binding protein (TBP) and TFIIIB-related factor (BRF) [17]. In archaea, TBP and TFB can mediate the initiation of basal transcription [40]. Archaeal TFB and eukaryotic BRF are evolutionarily related by sharing a common ancestry [17]. For this reason, the initiation mechanism of *G. lamblia* 5S rRNA transcription should be similar to that of archaea, specifically, with respect to without the participation of TFIIIA, TFIIIB, and TFIIIC but including TBP and TFB/BRF. As for the other three investigated excavates, both TFIIIB and TFIIIC exist in all of them, while TFIIIA only exists in the parasitic excavate *T. vaginalis* and not in the free-living *N. gruberi*. This means that parasitism does not necessarily lead to the reduction of the transcription factors of RNA pol III. Therefore, the simplicity of the *G. lamblia* transcription factor system should not be interpreted as necessarily the results of parasitic degeneration but most probably as a primitive feature of this organism.

The three eukaryotic RNA polymerases (pol I, II, III) all have both common subunits and subunits specific to each of themselves, whereas *G. lamblia* RNA pol III has been reported to possess all the common subunits but lack all specific subunits except for C34 [17]. However, besides C34, we found the fully sequenced *G. lamblia* genome to include another specific subunit, C82. In yeast, subunit C34 was shown to interact with subunits C82 and C31 to form a subcomplex functioning in the initiation of RNA pol III transcription [41]. Subunit C34 and C31 were shown to play a key

role in determining the initiation specificity of RNA pol III transcription [42,43]. Archaeal RNA polymerase has recently been reported to have one ortholog to C34, which might be involved in the transcription of non-coding RNAs [44]. Our investigation revealed the presence of C34 and C82 in all the investigated excavates (except *T. cruzi*, which lacked C82, probably due to secondary loss), while C31 was absent from all of them. These mean that C34 might have been inherited from archaea and emerged very early in the evolution of eukaryotes, while C31 arose at least after the divergence of excavate from the eukaryotic trunk. Lacking C31, the RNA pol III of excavate might carry out the specific initiation of RNA pol III transcription using only C34 and C82. This is similar to the initiation of non-coding RNA transcription only with C34 in archaea. However, the parasitic excavates *T. vaginalis* and *T. cruzi* also have another more specific subunit that is absent from *G. lamblia*, C53. Therefore, the similarity of *G. lamblia* RNA pol III to those of archaea also seem not to be due to parasitic degeneration but to be the very primitive traits of this organism.

Since TFIIIA also functions as a 5S rRNA transcript chaperone required for the nuclear-cytoplasmic transport in common eukaryotes [3,45], without TFIIIA, the 5S rRNA nuclear-cytoplasmic transport in *G. lamblia* must be unique. Whereas, the presence of 5S rRNA-associated protein L5 in *G. lamblia*, however, suggests that the mechanisms of *G. lamblia* 5S rRNA cytoplasmic-nucleolar transport and incorporation into large ribosomal subunit are similar to that of common eukaryotes. No TFIIIA has yet been reported in archaea, but L5 protein has been reported [46,47]. The *G. lamblia* 5S rRNA transport system can then be said to be similar to that in archaea. The fact that TFIIIA is present in the parasitic excavate *T. vaginalis* but absent from the free-living *N. gruberi* further suggests that the absence of TFIIIA in *G. lamblia* is not necessarily due to parasitic degeneration but most probably to the primitive nature of this organism.

In summary, our results reveal a unique 5S rRNA system in *G. lamblia*: On one hand, some of its features, such as the separated location of the 5S rRNA genes and the transcription of 5S rRNA genes by RNA pol III, are consistent with that seen in common eukaryotes. On the other hand, many other traits, such as the lack of a tandem 5S rRNA gene region, the low gene copy number, the relatively simple transcription system (lacking the participation of TFIIIA, TFIIIB, and TFIIIC but inclusion of TBP, BRF, the RNA pol III with only two specific subunits, C34 and C82), and the TFIIIA-independent but L5-dependent 5S rRNA transport and incorporation system, are much simpler than those of most eukaryotes and very similar to those of archaea. Comparisons of the entire 5S rRNA system among the three sequenced *G. lamblia* isolates and with the other three investigated excavates showed 1) All the three *G. lamblia* isolates harbor the same simplified 5S rRNA system, which is the simplest one among all the investigated excavates in almost every aspect of the system; 2) parasitic species are not definitely simpler than that of free-living ones, conversely, the free-living species can be simpler in some respects of the system, such as transcription factors (Table 1). In fact, as an obligate parasite, though not an excavate, *Encephalitozoon cuniculi* was also compared to *G. lamblia* by us. We found the 5S rRNA system of *G. lamblia* is to be simpler than that of *E. cuniculi*, especially in the aspect of transcription factors (Table S4). In this way, the simplicity of *G. lamblia* 5S rRNA system should not be considered due to parasitic degeneration. If this is not true, then it would become difficult to explain why the system was definitely degenerated to be an archaeal form.

In conclusion, *G. lamblia* represents an ancient primitive eukaryotic 5S rRNA system intermediate to archaea and the

common eukaryotic model system. It may reflect the history of many aspects of the typical eukaryotic 5S rRNA system as it evolved from its archaeal form. Simultaneously, our work implies although the adaptation to a parasitic lifestyle has caused a number of profound changes to *G. lamblia*, such as a compact genome, mitochondrial relics, and simplified machinery for some metabolic pathways, some very primitive eukaryotic features have been retained by this organism. These have no direct relationship with the parasitic lifestyle but are rather involved in the fundamental functions of a eukaryotic cell. Therefore, we propose that *G. lamblia* is most likely a combination of eukaryotic primitiveness and secondary parasitic degeneration.

Supporting Information

Table S1 Putative 5S rRNA genes identified in *N. gruberi* genome database.
(DOC)

Table S2 The identified C2H2-type zinc finger proteins in *G. lamblia* WB genome.

References

- Gray MW, Burger G, Lang BF (1999) Mitochondrial evolution. *Science* 283: 1476–1481.
- Pelham HR, Brown DD (1980) A specific transcription factor that can bind either the 5S RNA gene or 5S RNA. *Proc Natl Acad Sci U S A* 77: 4170–4174.
- Guddat U, Bakken AH, Pieler T (1990) Protein-mediated nuclear export of RNA: 5S rRNA containing small RNPs in xenopus oocytes. *Cell* 60: 619–628.
- Szymanski M, Barciszewska MZ, Erdmann VA, Barciszewski J (2003) 5 S rRNA: structure and interactions. *Biochem J* 371: 641–651.
- Gillin FD, Reiner DS, McCaffery JM (1996) Cell biology of the primitive eukaryote *Giardia lamblia*. *Annu Rev Microbiol* 50: 679–705.
- Sogin ML, Gunderson JH, Elwood HJ, Alonso RA, Peattie DA (1989) Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. *Science* 243: 75–77.
- Cavalier-Smith T, Chao EE (1996) Molecular phylogeny of the free-living archaeozoan *Trepomonas agilis* and the nature of the first eukaryote. *J Mol Evol* 43: 551–562.
- Hashimoto T, Nakamura Y, Nakamura F, Shirakura T, Adachi J, et al. (1994) Protein phylogeny gives a robust estimation for early divergences of eukaryotes: phylogenetic place of a mitochondria-lacking protozoan, *Giardia lamblia*. *Mol Biol Evol* 11: 65–71.
- Hashimoto T, Nakamura Y, Kamaishi T, Nakamura F, Adachi J, et al. (1995) Phylogenetic place of mitochondrion-lacking protozoan, *Giardia lamblia*, inferred from amino acid sequences of elongation factor 2. *Mol Biol Evol* 12: 782–793.
- Inagaki Y, Ford Doolittle W (2000) Evolution of the eukaryotic translation termination system: origins of release factors. *Mol Biol Evol* 17: 882–889.
- Adam RD (2001) Biology of *Giardia lamblia*. *Clin Microbiol Rev* 14: 447–475.
- Tovar J, Leon-Avila G, Sanchez LB, Satak R, Tachezy J, et al. (2003) Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation. *Nature* 426: 172–176.
- Hampel V, Hug L, Leigh JW, Dacks JB, Lang BF, et al. (2009) Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups." *Proc Natl Acad Sci U.S.A.* 106: 3859–3864.
- Edlind TD, Chakraborty PR (1987) Unusual ribosomal RNA of the intestinal parasite *Giardia lamblia*. *Nucleic Acids Res* 15: 7889–7901.
- Chen XS, Rozhddestvensky TS, Collins LJ, Schmitz J, Penny D (2007) Combined experimental and computational approach to identify non-protein-coding RNAs in the deep-branching eukaryote *Giardia intestinalis*. *Nucleic Acids Res* 35: 4619–4628.
- Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, et al. (2007) Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science* 317: 1921–1926.
- Best AA, Morrison HG, McArthur AG, Sogin ML, Olsen GJ (2004) Evolution of eukaryotic transcription: insights from the genome of *Giardia lamblia*. *Genome Res* 14: 1537–1547.
- Lloyd D, Harris JC (2002) *Giardia*: highly evolved parasite or early branching eukaryote? *Trends Microbiol* 10: 122–127.
- Embley TM, Hirt RP (1998) Early branching eukaryotes? *Curr Opin Genet Dev* 8: 624–629.
- Manning G, Reiner DS, Lauwaet T, Dacre M, Smith A, et al. (2011) The minimal kinome of *Giardia lamblia* illuminates early kinase evolution and unique parasite biology. *Genome Biol* 12: R66.
- Franzen O, Jerlstrom-Hultqvist J, Castro E, Sherwood E, Ankarklev J, et al. (2009) Draft genome sequencing of *giardia intestinalis* assemblage B isolate GS: is human giardiasis caused by two different species? *PLoS Pathog* 5: e1000560.
- Jerlstrom-Hultqvist J, Franzen O, Ankarklev J, Xu F, Nohynkova E, et al. (2010) Genome analysis and comparative genomics of a *Giardia intestinalis* assemblage E isolate. *BMC Genomics* 11: 543.
- Torres-Machorro AL, Hernandez R, Sanchez J, Lopez-Villasenor I (2006) The 5S ribosomal RNA gene from the early diverging protozoa *Trichomonas vaginalis*. *Mol Biochem Parasitol* 145: 269–273.
- Hernandez-Rivas R, Martinez-Calvillo S, Romero M, Hernandez R (1992) *Trypanosoma cruzi* 5S rRNA genes: molecular cloning, structure and chromosomal organization. *FEMS Microbiol Lett* 71: 63–67.
- Fritz-Laylin LK, Prochnik SE, Ginger ML, Dacks JB, Carpenter ML, et al. (2010) The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. *Cell* 140: 631–642.
- Szymanski M, Barciszewska MZ, Erdmann VA, Barciszewski J (2002) 5S ribosomal RNA database. *Nucleic Acids Res*. pp 176–178.
- Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, et al. (2005) InterProScan: protein domains identifier. *Nucleic Acids Res* 33: W116–120.
- Finn RD, Mistry J, Tate J, Coggill P, Heger A, et al. (2010) The Pfam protein families database. *Nucleic Acids Res* 38: D211–222.
- Elmendorf HG, Singer SM, Pierce J, Cowan J, Nash TE (2001) Initiator and upstream elements in the alpha2-tubulin promoter of *Giardia lamblia*. *Mol Biochem Parasitol* 113: 157–169.
- Seshadri V, McArthur AG, Sogin ML, Adam RD (2003) *Giardia lamblia* RNA polymerase II: amanitin-resistant transcription. *J Biol Chem* 278: 27804–27810.
- Yee J, Mowatt MR, Dennis PP, Nash TE (2000) Transcriptional analysis of the glutamate dehydrogenase gene in the primitive eukaryote, *Giardia lamblia*. Identification of a primordial gene promoter. *J Biol Chem* 275: 11432–11439.
- Sun CH, Tai JH (1999) Identification and characterization of a ran gene promoter in the protozoan pathogen *Giardia lamblia*. *J Biol Chem* 274: 19699–19706.
- Holberton DV, Marshall J (1995) Analysis of consensus sequence patterns in *Giardia* cytoskeleton gene promoters. *Nucleic Acids Res* 23: 2945–2953.
- Sorensen PD, Frederiksen S (1991) Characterization of human 5S rRNA genes. *Nucleic Acids Res* 19: 4147–4151.
- Korn IJ (1982) Transcription of *Xenopus* 5S ribosomal RNA genes. *Nature* 295: 101–105.
- Kock J, Cornelissen AW (1990) The 5S ribosomal RNA genes of *Crithidia fasciculata*. *Mol Biochem Parasitol* 38: 295–298.
- Peng Z, Bateman E (2004) Analysis of the 5S rRNA gene promoter from *Acanthamoeba castellanii*. *Mol Microbiol* 52: 1123–1132.
- Bieker JJ, Martin PL, Roeder RG (1985) Formation of a rate-limiting intermediate in 5S RNA gene transcription. *Cell* 40: 119–127.
- Acker J, Ozanne C, Kachouri-Lafond R, Gaillardin C, Neugeglise C, et al. (2008) Dicotronic tRNA-5S rRNA genes in *Yarrowia lipolytica*: an alternative TFIIIA-independent way for expression of 5S rRNA genes. *Nucleic Acids Res* 36: 5832–5844.
- Bartlett MS (2005) Determinants of transcription initiation by archaeal RNA polymerase. *Curr Opin Microbiol* 8: 677–684.
- Werner M, Chaussivert N, Willis IM, Sentenac A (1993) Interaction between a complex of RNA polymerase III subunits and the 70-kDa component of transcription factor IIIB. *J Biol Chem* 268: 20,721–20,724.

42. Brun I, Sentenac A, Werner M (1997) Dual role of the C34 subunit of RNA polymerase III in transcription initiation. *EMBO J* 16: 5730–5741.
43. Thuillier V, Stettler S, Sentenac A, Thuriaux P, Werner M (1995) A mutation in the C31 subunit of *Saccharomyces cerevisiae* RNA polymerase III affects transcription initiation. *EMBO J* 14: 351–359.
44. Blombach F, Makarova KS, Marrero J, Siebers B, Koonin EV, et al. (2009) Identification of an ortholog of the eukaryotic RNA polymerase III subunit RPC34 in Crenarchaeota and Thaumarchaeota suggests specialization of RNA polymerases for coding and non-coding RNAs in Archaea. *Biol Direct* 4: 39.
45. Honda BM, Roeder RG (1980) Association of a 5S gene transcription factor with 5S RNA and altered levels of the factor during cell differentiation. *Cell* 22: 119–126.
46. Hatakeyama T (1990) Amino acid sequences of the ribosomal proteins HL30 and HmaL5 from the archaeobacterium *Halobacterium marismortui*. *Biochim Biophys Acta* 1039: 343–347.
47. Yang D, Kusser I, Kopke AK, Koop BF, Matheson AT (1999) The structure and evolution of the ribosomal proteins encoded in the *spc* operon of the archaeon (Crenarchaeota) *Sulfolobus acidocaldarius*. *Mol Phylogenet Evol* 12: 177–185.
48. Tourova TP (2003) Copy number of ribosomal operons in prokaryotes and its effect on phylogenetic analyses. *Microbiology* 72: 389–402.