FEBS Letters 588 (2014) 783–788

journal homepage: www.FEBSLetters.org

A small portion of plastid transcripts is polyadenylated in the flagellate *Euglena gracilis*

Kristína Záhonová [a,b], Lucia Hadariová [b], Rostislav Vacula [b], Vyacheslav Yurchenko [a], Marek Eliáš [a], Juraj Krajčovič [b], Matej Vesteg [a,b,*]

[a] Department of Biology and Ecology, Faculty of Science, University of Ostrava, Ostrava, Czech Republic
[b] Department of Genetics, Faculty of Natural Sciences, Comenius University, Bratislava, Slovakia

ARTICLE INFO

Article history:
Received 11 December 2013
Revised 8 January 2014
Accepted 9 January 2014
Available online 1 February 2014

Edited by Ulf-Ingo Flügge

Keywords:
Plastid
EST
Euglenozoa
Polyadenylation
Quantitative PCR
Trans-splicing

ABSTRACT

Euglena gracilis **possesses secondary plastids of green algal origin. In this study, *E. gracilis* expressed sequence tags (ESTs) derived from polyA-selected mRNA were searched and several ESTs corresponding to plastid genes were found. PCR experiments failed to detect SL sequence at the 5′-end of any of these transcripts, suggesting plastid origin of these polyadenylated molecules. Quantitative PCR experiments confirmed that polyadenylation of transcripts occurs in the *Euglena* plastids. Such transcripts have been previously observed in primary plastids of plants and algae as low-abundance intermediates of transcript degradation. Our results suggest that a similar mechanism exists in secondary plastids.**

1. Introduction

Euglena gracilis is a fresh-water photosynthetic flagellate belonging to the order Euglenida and to the protist phylum Euglenozoa [1,2]. This phylum includes also the orders Kinetoplastida, Diplonemida and Symbiontida [3,4], which comprise exclusively heterotrophic species. Most euglenid species are free-living heterotrophic flagellates, but some of them possess secondary plastids that arose via secondary endosymbiosis of a green alga [5]. Phylogenetic analysis of plastid genome sequences revealed that euglenid plastids are derived from a *Pyramimonas*-related prasinophyte alga [6].

A common euglenozoan feature is the processing of primary nuclear transcripts by spliced leader (SL) RNA-mediated *trans*-splicing [7]. This process includes replacement of the 5′-end of pre-mRNA by the 5′-end of SL-RNA, donating identical 5′-termini to the mRNA molecules. Similar to nuclear *cis*-splicing, *trans*-splicing process is also mediated by spliceosomes, but a Y-branch intron structure is formed instead of a lariat [8]. The only currently known nuclear mRNA lacking the SL sequence in *E. gracilis* is that of the fibrillarin gene [9]. Since SL-*trans*-splicing does not occur in organelles (mitochondria, plastids), the presence (or absence) of an SL sequence at the 5′-end of a euglenozoan mRNA is diagnostic for its synthesis in the nucleus (or organelles, respectively).

E. gracilis cell possesses approximately eight secondary plastids bounded by three membranes [10,11]. The plastid genome of this species is circular and comprises 143.17 kb. It contains 96 protein and RNA gene loci [12], group II and III introns, and twintrons (i.e. introns within introns) [13].

The evolutionary transition from an endosymbiont to the plastid organelle was accompanied by a loss of many genes and gene transfer from the endosymbiont genome(s) to the host genome [14]. Gene transfer from plastids and mitochondria is an ongoing process, as it has been demonstrated in animals, plants, fungi as well as protists [15,16].

Nuclear copies of plastid DNA (NUPT) can be found in a variety of species [17]. However, some species, e.g. the pelagophyte alga *Aureococcus anophagefferens* or the apicomplexan parasite *Babesia bovis*, have been found to contain no nuclear organellar DNA (norgDNA) [16]. Species containing only one plastid per cell have fewer NUPTs than those with more plastids per cell. The total length of NUPTs is higher in polyplastidic species [16,18]. There is no obvious correlation between the norgDNA content and the organellar

* Corresponding author at: Department of Biology and Ecology, Faculty of Science, University of Ostrava, Chittussiho 10, 710 00 Ostrava, Czech Republic. Fax: +420 597092382.
E-mail addresses: vesteg@fns.uniba.sk, matej.vesteg@osu.cz (M. Vesteg).

http://dx.doi.org/10.1016/j.febslet.2014.01.034
0014-5793/

genome size. However, there is a strong positive correlation between the nuclear genome size and the NUPT content [16].

3′-terminal polyadenylation of nucleus-encoded transcripts is a very well-studied process that is essential for stability and proper function of most mRNA molecules in the cytosol [19]. Less is known about polyadenylation in organelles. It has been documented in metazoan mitochondria, where it is particularly important for some transcripts by restoring their termination codons [19]. On the other hand, polyadenylation of transcripts in primary plastids has been shown to serve as a signal for degradation [20–22]. As a result, polyadenylated transcripts are much more abundant in the cytosol and mitochondria than in plastids [23].

Polyadenylation-facilitated RNA degradation in plastids in those species, where it was studied in detail, proceeds in three steps: (i) endonucleolytic cleavage, (ii) polyadenylation of fragments and (iii) exonucleolytic degradation of polyadenylated fragments [20–22,24]. The polyA tails can be homopolymeric, composed solely of adenosines, or heteropolymeric composed of all four nucleotides, but with adenosines being the most abundant ones. The enzymes performing polyadenylation are polynucleotide phosphorylase (PNPase) and several polyA polymerases (PAPs) [21,22].

We anticipated that expressed sequence tag (EST) data from *E. gracilis*, which were obtained by sequencing of cDNAs synthesized by reverse transcription using oligo(dT) priming [25–27], should primarily include sequences from nuclear mRNA. However, our analysis of the EST data revealed that they include transcripts apparently derived from genes known to reside on the plastid genome. These sequences could represent transcripts of functional nucleus-localized copies of plastid genes or polyadenylated transcripts of plastid genes. The aim of the current study was to distinguish between these two possibilities.

2. Materials and methods

2.1. Nucleic acid isolation and purification, cDNA synthesis

E. gracilis wild type strains *Z* and *bacillaris* (hereafter abbreviated as EgZ and EgB, respectively) were cultivated under constant illumination or in the dark in Hutner medium [28]. Genomic DNA was isolated as in the protocol described earlier using phenol–chloroform method [29]. RNA extraction and mRNA isolation were carried out as described previously [29,30]. cDNA synthesis was performed with oligo(dT) and random hexanucleotide primers using ImProm-II Reverse Transcription System (Promega, Madison, WI, USA).

2.2. Polymerase chain reaction

PCR was performed as described earlier [29,30] except the annealing temperature of 60 °C instead of 58 °C. In the first set of PCR reactions, the 26 bp-long SL sequence present at the 5′-end of *E. gracilis* nuclear mRNAs was used as a forward primer (SL). Reverse primers (R) were designed based on the sequence of six *E. gracilis* EST sequences with 100% identity with plastid genes. In the second set of PCR reactions, the same reverse primers were used, but forward primers (F) were derived from the 5′-end sequence of the respective ESTs. Sequences of primers are listed in Supplementary Table 1. cDNAs synthesized using both oligo(dT) and random hexanucleotide primers were used as templates in quantitative PCR reactions. Total DNAs and cDNAs were used as templates in control reactions. Control PCRs were performed using primers for the nuclear genes *Nop1* (data not shown) and *PsbO*, and for the plastid gene *rpl16* (Supplementary Table 1). The identity of the products obtained by PCR was verified by sequencing.

2.3. Quantitative PCR

Quantitative PCRs were performed using LightCycler® 480 SYBRGreen Master mix (Roche, Basel, Switzerland) and the protocol according to the manufacturer's instructions, using the same primers as in the second set of PCRs (see above). All measurements were taken in triplicates. The ratios of (i) total mRNA or rRNA (full set of RNAs for individual genes) to polyadenylated mRNA or rRNA (polyadenylated fraction of total RNAs for individual genes), and (ii) RNAs from light- compared to dark-grown cultures were calculated using standard curve method for relative quantification [31]. Hereafter, total RNA means the full complement of mRNAs for individual genes and polyadenylated RNA means the polyadenylated fraction thereof. We used the expression of nuclear gene *PsbO* to normalize the calculations, because our previous studies have revealed that its mRNA level is the same in the light- and dark-grown *E. gracilis* [30,32].

2.4. Bioinformatics analyses

The *E. gracilis* plastid genome sequence RefSeq accession number NC_001603.2 [12] was used as a query in a BLASTn search of currently available *E. gracilis* EST data [25–27].

To search for enzymes potentially involved in polyadenylation-facilitated RNA degradation pathway in euglenid plastids, PNPase and PAP sequences from *Arabidopsis thaliana* and *Chlamydomonas reinhardtii* were used as queries in tBLASTn search in *E. gracilis* EST data [25–27] and in the EST data of the marine euglenid *Eutreptiella gymnastica* (Marine Microbial Eukaryote Transcriptome Sequencing Project; http://camera.calit2.net/mmetsp/). Homologs found were screened for the presence of targeting presequences using the program TargetP (http://www.cbs.dtu.dk/services/TargetP/).

3. Results

Transcripts with 100% sequence identity with parts of six plastid genes – *atpA, ccsA, psbD, ycf13, rrn23S and rrn16S* – were found in *E. gracilis* EST data. Therefore, we tested whether these ESTs were derived from functional copies of plastid genes transferred to the nucleus, or whether representation of these sequences in the EST dataset reflected the fact that some plastid transcripts are polyadenylated. Our approach was based on the assumption that mRNAs originating by transcription of nuclear loci possess the SL sequence at their 5′-ends.

No PCR products were detected (lanes SL in Fig. 1) in PCRs using a forward primer (SL) corresponding to the conserved SL sequence and a reverse primer (R) designed based on particular EST sequences (listed in Supplementary Table 1). This suggests that primary transcripts of all studied genes are not *trans*-spliced. For control PCR reactions we used total DNA, as well as cDNA synthesized using oligo(dT) and random hexanucleotide primers as templates. All reactions yielded products of the expected sizes (Fig. 1). This confirms presence of the intact templates. The *rpl16* gene was detected in both *E. gracilis* strains using gene-specific F and R primers and cDNA synthesized using random hexanucleotide primers as the template. No PCR product was detected when oligo(dT) primed cDNA was used (lanes oligoT cDNA in Fig. 1). In all control reactions using primers for the nuclear gene *PsbO*, a product of the expected length was obtained (Fig. 1).

The most plausible interpretation of these results is that the studied genes presented in the *E. gracilis* EST database are not localized in the nucleus, but are exclusively present in the plastid and a part of their transcripts can be polyadenylated, at least to some extent. Therefore, quantitative PCR experiments were