TRANSLATION REGULATION IN CHLOROPLASTS

Seliverstov A.V.*[*], *Lyubetsky V.A.
Institute for Information Transmission Problems, RAS, Moscow, Russia
[*] Corresponding author: e-mail: slvstv@iitp.ru

Key words: translation, chloroplasts, multiple alignments

SUMMARY

Motivation: Gene expression in chloroplasts of algae and plants is regulated through binding of chloroplast mRNA by nuclear-encoded proteins. It is therefore important to determine such protein binding sites and study them from evolutionary perspective.

Results: An algorithm of finding conservative protein-RNA binding sites is designed, also see details in (Lyubetsky *et al.*, 2004). The algorithm was applied to infer these sites upstream of chloroplast genes. As a result, candidate protein-RNA binding sites were detected upstream of the *atpF*, *petB*, *clpP*, *psaA*, *psbA* and *psbB* genes in many chloroplasts of algae and plants. We suggest that some of these sites are involved in suppressing translation until the completion of splicing.

INTRODUCTION

Gene expression in chloroplasts of algae and plants is regulated by binding of chloroplast mRNA by nuclear-encoded proteins (Nickelsen, 2003). These proteins are involved in editing, translation and maintaining stability of chloroplast mRNA. Detailed analysis of regulatory sites is available from published evidence for alga *Chlamydomonas reinhardtii*, as well as some plants (Hauser *et al.*, 1996; Zerges, 2000; Nickelsen, 2003). For example, protein binding to the *psbA* 5′-untranslated region in *C. reinhardtii* results in activation of translation (Hauser *et al.*, 1996).

Many chloroplast protein-coding genes contain introns. Thus, their translation should not start immediately after transcription. However, the translation machinery of chloroplasts closely resembles that of bacteria, particularly, in the ribosome being able to immediately follow the RNA-polymerase on mRNA strand. If the ribosome arrives at the end of exon before splicing is completed, the splicing process halts. To avoid this, in some rare cases the AUG start codon is derived from ACG by editing mRNA, which prevents translation from starting immediately (Zerges, 2000). RNA editing is known for chloroplasts of higher plants and is absent, e.g., in the liverwort *Marchantia polymorpha*.

Our algorithm detected candidate protein-RNA binding sites upstream of *atpF*, *petB*, *clpP*, *psaA*, *psbA* and *psbB* genes in many chloroplasts.

We suggest that some of these sites are involved in suppressing translation until splicing is completed. This conjecture is in agreement with observation that multiple alignments of the site-containing regions upstream of these genes are highly conservative, and is also supported by experimental evidence (Hauser *et al.*, 1996).

ALGORITHM

Consider a dataset of leader regions upstream of orthologous genes and a corresponding species tree. A set of shallow phylogenetic subtrees (groups of taxa) is selected in the species tree. For each of the groups, the algorithm searches for conserved regions of fixed length *n* (which can be varied) by finding cliques in a suitable multipartite graph. The basic idea is as follows. The algorithm finds clusters of very similar sites, called signals or motifs, of a fixed length *n* for each of these phylogenetic groups. From a motif, a weight matrix 4 × *n* is generated, where the *k*th column of the matrix, $1 \leq k \leq n$, contains letter frequencies in the *k*th site position from the motif. Further, the algorithm generates clusters of weight matrices for different suitable *n* across all groups. The clusters of matrices are generated accounting for distances in the species tree between the ancestors of the initial phylogenetic groups. The algorithm of clique finding can also be used for constructing these clusters of matrices. In each matrix cluster, the matrices are replaced by the corresponding motif, thus defining sets of motifs. The described procedure can be iterated. The algorithm is described in detail in (Lyubetsky, Seliverstov, 2004).

IMPLEMENTATION AND RESULTS

Chloroplast genomes were obtained from GenBank (NCBI). The initial dataset contained 5'-untranslated intergenic regions from chloroplast genomes of algae and plants.

Occurrence of predicted sites upstream of chloroplast genes is shown in the table.

In many chloroplasts, the algorithm found long conserved binding sites containing conserved helices upstream of the genes *atpF* (ATP-synthetase subunit), *petB* (cytochrome b6), *clpP* (ATP-dependent Clp protease proteolytic subunit), *psaA* (photosystem I P700 apoprotein A1), *psbA* (photosystem II protein D1) and *psbB* (photosystem II P680 chlorophyll A apoprotein).GenBank annotation of the *psbA* gene of *Amborella trichopoda* probably misses a short N-terminal sequence, which might explain why in this case the algorithm failed to find the corresponding motif.

For the genes *atpF*, *clpP* and *petB*, there is a strong correlation between the occurrence of splicing and the existence of the predicted protein-binding sites. On the other hand, with *psaA*, *psbA* and *psbB* no such correlation is found. With *clpP*, *petB*, *psaA*, *psbA*, the sites always contain helices, but for *atpF* and *psbB* they do not.

Table 1. Occurrence of predicted sites and introns upstream of chloroplast genes *atpF*, *petB*, *clpP*, *psaA*, *psbA* and *psbB*. Notation: "+" – candidate protein binding site present; "-" – no candidate binding site; "s" – introns present; "n" – no gene homolog in the species; "&" – helices in the site; "E" – editing of start codon

Species	*atpF*	*clpP*	*petB*	*psaA*	*psbA*	*psbB*
Euglena gracilis	–s	–	-s	–s	-s	-s
Odontella sinensis	–	–	–	+&	+&	–
Guillardia theta	–	–	–	+&	+&	–
Cyanidioschyzon merolae	–	–	–	–	+&	–
Cyanidium caldarium	–	–	–	–	–	–
Porphyra purpurea	–	–	–	+&	+&	+
Gracilaria tenuistipitata	–	–	–	–	+&	–
Chlamydomonas reinhardtii	–	–	–	-s	+&s	–
Nephroselmis olivacea	–	–	–	+&	+&	+
Chaetosphaeridium globosum	–	+&s	-s	+&	+&	+
Mesostigma viride	–	–	–	+&	–	–
Anthoceros formosae	+s	+&s	+&s	+&	+&	+
Marchantia polymorpha	+s	+&s	+&s	+&	+&	+
Huperzia lucidula	+s	+&s	+&s	+&	+&	+

Species	*atpF*	*clpP*	*petB*	*psaA*	*psbA*	*psbB*
Adiantum capillus-veneris	+sE	+&s	-sE	+&	+&	+
Psilotum nudum	+s	+&s	+&s	+&	+&	+
Pinus thunbergii	+s	+&	+&s	+&	+&s	+
Amborella trichopoda	+s	+&s	+&s	+&	–	+
Arabidopsis thaliana	+s	+&s	+&s	+&	+&	+
Atropa belladonna	+s	+&s	+&s	+&	+&	+
Calycanthus floridus	+s	+&s	+&s	+&	+&	+
Cucumis sativus	+s	+&s	+&s	+&	+&	+
Epifagus virginiana	n	+&s	n	n	n	n
Lotus corniculatus	+s	+&s	+&s	+&	+&	+
Nicotiana tabacum	+s	+&s	+&s	+&	+&	+
Nymphaea alba	+s	+&s	+&s	+&	+&	+
Panax ginseng	+s	+&s	+&s	+&	+&	+
Spinacia oleracea	+s	+&s	+&s	+&	+&	+
Oryza nivara, Oryza sativa	+s	+&s	+&s	+&	+&	+
Triticum aestivum	+s	+&s	+&s	+&	+&	+
Zea mays	+s	+&s	+&s	+&	+&	+

DISCUSSION

Conserved motifs detected upstream of the *atpF*, *petB*, *clpP*, *psaA*, *psbA* and *psbB* genes are likely to be involved in translation regulation.

The conserved region upstream of *atpF* contains an AG-rich motif typical for ribosome binding sites, although being considerably longer than typical binding sites. This might be relevant to presence of introns in the gene, which suggests that translation initiates only after completion of splicing.

Upstream region of the *petB* gene does not have a typical ribosome-binding site but instead contains a conserved helix, which might suggest posttranscriptional modification of the 5′-untranslated regions or binding of a translation activator. In all plants, the *petB* gene contains introns.

Translational regulation of the *psbA* gene was experimentally observed in *Chlamydomonas reinhardtii*, where transcription is continuous, but translation is activated at light by a 47 kDa protein that forms a complex with other proteins and mRNA not interacting with mRNA directly (Hauser *et al.*, 1996). The complex is inactivated in the dark. The conserved nature of this region in plants and algae might suggest that the translation regulation machinery for gene *psbA* preceded the evolutionary emergence of introns.

Conserved regions in the 5'-untranslated regions of *clpP* and *psbA* genes were observed upstream of almost all their orthologs, even those lacking introns. Notably, conserved RNA motifs in the transcripts of *petB*, *clpP*, *psaA* and *psbA* contain helices with conserved flanks likely interacting with a protein mediator, which is typical for most regulatory systems (Seliverstov *et al.*, 2005).

Long conserved motifs were found upstream of the *psaA* and *psbB* genes, which lack introns in all species containing the motifs. On the other hand, in chloroplasts of *Adiantum*, all studied 5′-untranslaled regions are considerably diverged. Hence, the motif was not found upstream of *petB*, while it was in the five other cases. In the latter situation, however, site trees and species trees disagreed considerably at the node containing the name of the corresponding species.

Other intron-containing genes in the studied chloroplast genomes were not found to have conserved 5′-motifs, or their 5′-untranslated regions were too short or absent. Two such examples are discussed. In studied plants, the upstream regions of gene *rbcL* encoding a ribulose 1,5-bisphosphate carboxylase/oxygenase subunit contain only a short conserved motif with the consensus ARGGAGGGACYT, which core constitutes a ribosome-binding site. We have no reason to assign a regulatory role to this motif, as the *rbcL* gene in plants lacks introns. On the other hand, *rbcL* contains introns in chloroplasts of both algae *Euglena gracilis* and *Chlamydomonas reinhardtii*, and, in the latter case, it is regulated by mRNA-binding proteins (Hauser *et al*., 1996). This seeming discrepancy is not surprising, since in both algae the structure of 5′-untranslated region is completely different from that in studied plants.

A different situation is with the *ycf3* gene (photosystem I assembly protein Ycf3). It contains introns and a long 5′-untranslated regions not overlapping with other genes in plant chloroplasts, but it was not found to possess conserved motifs.

ACKNOWLEDGEMENTS

The authors are grateful to M.S. Gelfand for valuable discussion and to L.Y. Rusin for discussion and help. This study was partially supported by ISTC 2766.

REFERENCES

Hauser C.R., Gillham N.W., Boynton J.E. (1996) Translation regulation of chloroplast genes. *The J. of Biol. Chemistry*, **271**, 1486–1497.

Lyubetsky V.A., Seliverstov A.V. (2004) Note on cliques and alignments. *Information Processes*, **4**, 241–246.

Nickelsen J. (2003) Chloroplast RNA binding proteins. *Current Genet*, **43**, 392–399.

Seliverstov A.V., Putzer H., Gelfand M.S., Lyubetsky V.A. (2005) Comparative analysis of RNA regulatory elements of amino acid metabolism genes in Actinobacteria. *BMC Microbiology*, **5,** 54.

Zerges W. (2000) Translation in chloroplasts. *Biochimie*, **82**, 583–601.