A Study Using Sequence Comparison to Investigate the Molecular Evolution of Mitochondrial tRNA Genes

Jun-Ichi Sagara
jun@bi.a.u-tokyo.ac.jp
Shugo Nakamura
shugo@bi.a.u-tokyo.ac.jp
Mitsunori Ikeguchi
ike@bi.a.u-tokyo.ac.jp
Kentaro Shimizu
shimizu@bi.a.u-tokyo.ac.jp

Department of Biotechnology, The University of Tokyo
1-1-1 Yayoi, Bunkyo-ku, Tokyo, 113-8657 Japan

1 Introduction

In investigating the origins of life and molecular evolution, it is very important to study structure and functions that are common to distantly related species and to analyze the corresponding gene sequences, which are called "molecular fossils." We therefore developed a computational method [1] based on principal component analysis (PCA) and multidimensional scaling analysis (MDS), and in the present study have used it to detect bases characterizing specific sequences mitochondrial transfer RNA. Our methods first classify the sequence in a genomic database into groups by PCA of multiple sequence alignment: Gene sequences are represented as vectors in a generalized sequence space, and groups of similar sequences are revealed when these vectors are projected onto a lower-dimensional sequence space. The distribution of bases is then compared with the distribution of sequences by using MDS, in which the bases of each sequence are projected individually onto the same sequence space. This makes it possible to identify bases characteristic of each group. Applying this method to the sequences of all the mitochondrial tRNA genes, we expect to detect not only bases that are always conserved but also bases that are often conserved.

2 Method

Our method is based on the one used by the Casari *et al.* [2] to predict functional residues in protein families. We have extended it to the analysis of tRNA gene sequences and have used it to identify the groups of bases specific to particular species by applying its basic procedure recursively.

The procedure of our method is as follows:

1. Apply the PCA to the entire sequences of tRNA genes. Each sequence is plotted on a two-dimensional plane called the sequence space.
2. Apply the MDS. Bases of each sequence are projected individually onto the same two-dimensional plane in order to trace the principal components back to individual bases and positions that characterize individual groups.
3. Apply the above two classification and comparison steps recursively. This enables us to classify the groups into subgroups and makes the results of the classification clearer.

In evaluating our method, we used full-length mitochondrial tRNA gene sequences deposited in the aligned sequence database European Bioinformatics Institute (EBI) Data library [3].

3 Results

We detected many bases characteristic of all the mitochondrial tRNA gene sequences of various species. Most of the characteristic bases are in stem regions, and most of the characteristic bases that are not

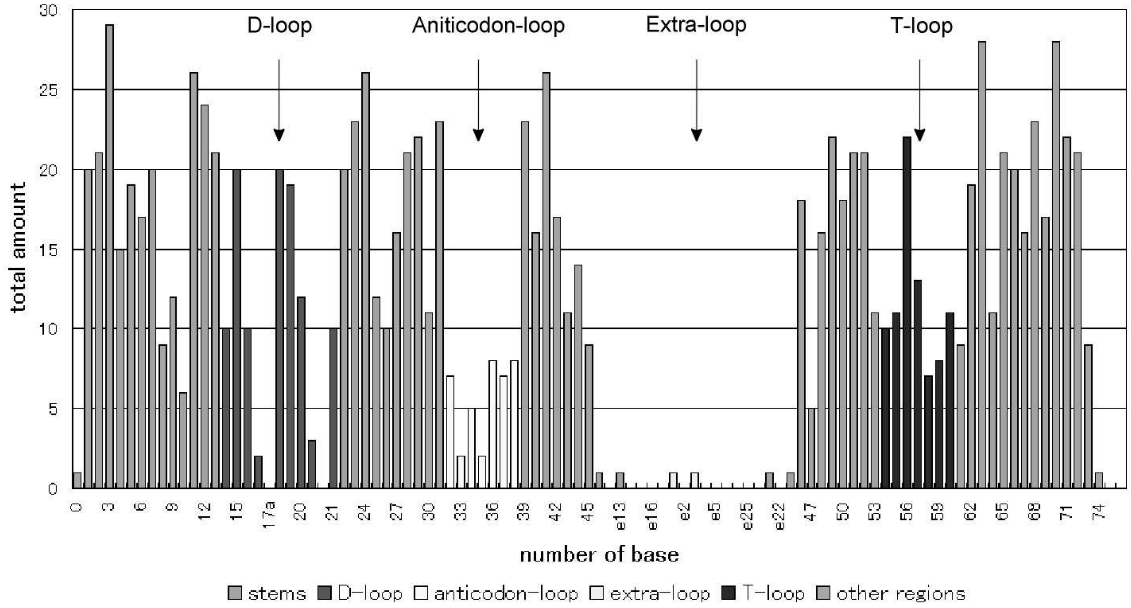

Figure 1: Histogram of characteristic bases in mitochondrial tRNA genes.

Table 1: Number of bases in each region.

	stems	D-loop	anticodon-loop	extra-loop	T-loop	other regions
number	873	106	39	2	82	48
fraction	0.759	0.092	0.034	0.002	0.071	0.042

in stem regions are in T and D domains, which are elbow regions of tRNAs (Fig. 1 and Table 1). The results suggest that the characteristic bases in these stems and domains have a role of preserving the L-shape structure of each tRNA.

References

[1] Sagara, J.-I., Shimizu, S., Kawabata, T., Nakamura, S., Ikeguchi, M., and Shimizu, K., The use of sequence comparison to detect ‘identities’ in tRNA genes, *Nucleic Acids Res.*, 26:1974–1979, 1998.

[2] Casari, G., Sander, C. and Valencia, A., A method to predict functional residues in proteins, *Natl. Struc. Biol.*, 2:171–178, 1995.

[3] Sprintzl, M., Steegborn, C., Hübel, F., and Steinberg, S., Compilation of tRNA sequences and sequences of tRNA genes, *Nucleic Acids Res.*, 24:68–72, 1996.