Sequence homologies among mitochondrial DNA-coded URF2, URF4 and URF5

Reiko Kikuno* and Takashi Miyata*+

*Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812 and +National Institute of Genetics, Mishima, Japan

Amino acid sequences coded for by 13 different genes of mammalian mitochondrial DNA (mtDNA) including 8 unassigned open reading frames (URFs) were compared in pairs. It was found that significant homologies exist among the amino acid sequences of the three URFs (URF2, URF4 and URF5) with a probability of occurrence of less than 10^{-5}. This result strongly suggests that the 3 URFs evolved from a single ancestral gene by a series of gene duplications. A phylogenetic tree based on the alignment of the URF sequences from mammals, an insect, a fungus and protozoa revealed a very remote divergence of the 3 URFs, going back to a time before separations of animal/protozoa and animal/fungus.

Mitochondrion *URF* *Homology* *Gene duplication* *Evolution*

1. INTRODUCTION

Gene duplication is known to be an important mechanism for generating diverse functions of genes [1]. The nuclear genome of higher eukaryotes can afford many redundant copies within it, from which a wide variety of genes with different functions have evolved. In contrast, animal mitochondrial DNA (mtDNA) is very small in size, only about 16 kilobase pairs long [2–4], and the genomic organization shows a marked economy: the mtDNA tightly packs genes for 2 ribosomal RNAs, 22 tRNAs and 5 proteins as well as 8 unassigned open reading frames (URFs) and there are few or no noncoding sequences between genes [2–4]. To know whether or not gene duplication has also been an important mechanism for the evolution of these mtDNA-coded genes, we subjected the amino acid sequences coded by these genes to computer-assisted search for homology. In addition, most proteins coded by the URFs have yet to be identified, except for URFA6L which has recently been assigned as a component of ATPase subunits [5]. Sequence homologies between these URF-coded proteins together with other proteins would provide some insight into structural and evolutionary significances of the URFs. We demonstrate here that the URF2, 4 and 5 products have extensive homologies in common. This strongly suggests that they diverged from a single common ancestor by a series of gene duplications.

+ To whom correspondence should be addressed

2. MATERIALS AND METHODS

Mammalian mtDNA sequences were taken from [2–4] and *Drosophila* counterparts were from [6,7]. Two protozoa sequences were from [8,9] and a fungus sequence was from [10].

The method for calculating homology matrix between the amino acid sequences compared and that for alignment of homologous sequences were described previously [11]. Using the scoring system MDM78 by Schwartz and Dayhoff [12], the aligned sequences were subjected to statistical test by the method described [11].

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

3. RESULTS AND DISCUSSION

Amino acid sequences coded for by 13 different genes of mammalian mtDNAs including 8 URFs were compared in pairs by the method described previously [11]. Marked homologies have been observed between pairs of three URFs: URF2, URF4 and URF5. Fig.1 demonstrates the presence of homologies between (a) URF4 and URF5 products and (b) URF4 and URF2 products. In both cases long diagonal lines representing the extensive homologies have been detected. Similar homology was also observed when URF2 and URF5 products were compared. On the basis of the homology matrices, the amino acid sequences were aligned for each of the URF pairs (fig.2). The

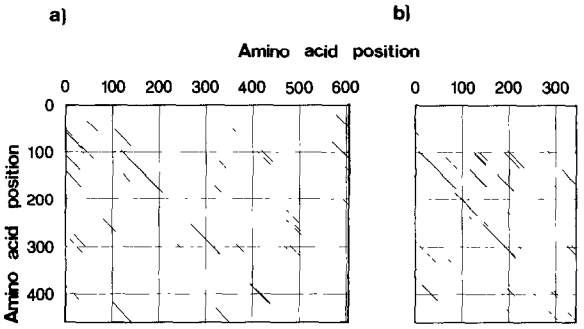

Fig.1. Homology matrix comparisons of amino acid sequence of mouse mtDNA URF4 product (ordinate) with those of (a) URF5 and (b) URF2 products (abscissa). Each diagonal line indicates a segment of 30 residues long which shows homology with a probability of occurrence of 1.3×10^{-3}.

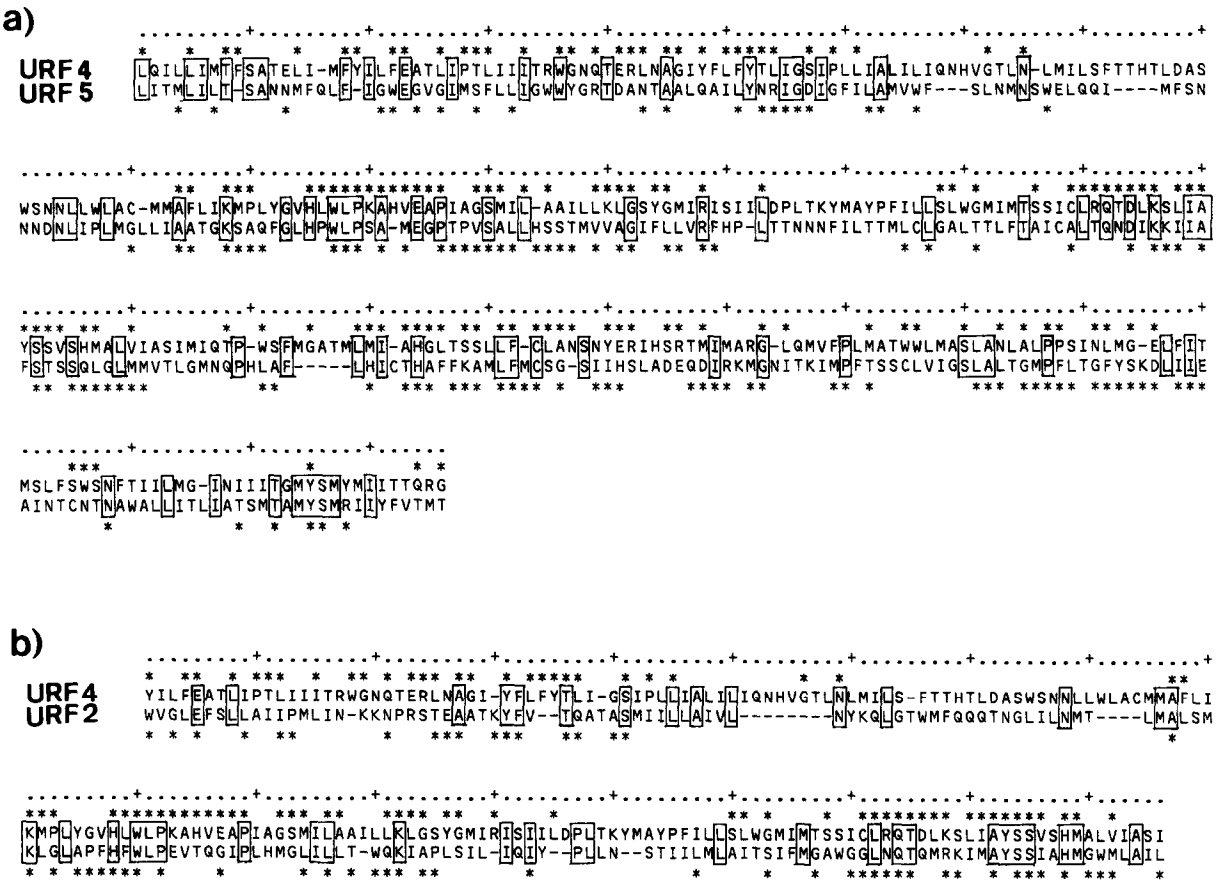

Fig.2. Alignments of amino acid sequences between (a) URF4 and URF5 products and (b) URF4 and URF2 products. Aligned regions: (a) amino acid positions 102–417 and 125–433 of URF4 and URF5, respectively, and (b) 119–301 and 30–194 of URF4 and URF2, respectively. For each of the URF products, positions that are invariant among human, bovine, mouse and *Drosophila* were shown by * above and below the alignments. Identical amino acids between different URFs were boxed. Gaps (–) were introduced to increase sequence similarity.

URF4 product shares 24% homologies with both the URF5 and URF2 products (gaps were counted as substitutions regardless of their length).

The statistical test shows that the observed homologies are highly significant; the probabilities that such sequence homologies are realized by chance are 6.6×10^{-7} and 1.0×10^{-5} for the pairs (a) URF4/URF5 and (b) URF4/URF2, respectively. In addition, each of the three URF products was analysed by a computer program that provides a moving average of homology along the length of the sequence [13]. Their profiles closely resembled each other. Hydropathy profiles [14] are also similar for the three URF products. These results strongly suggest that the three URFs evolved from a single ancestral gene by a series of gene duplications, followed by deletions and/or insertions of DNA pieces coding for non-homologous segments.

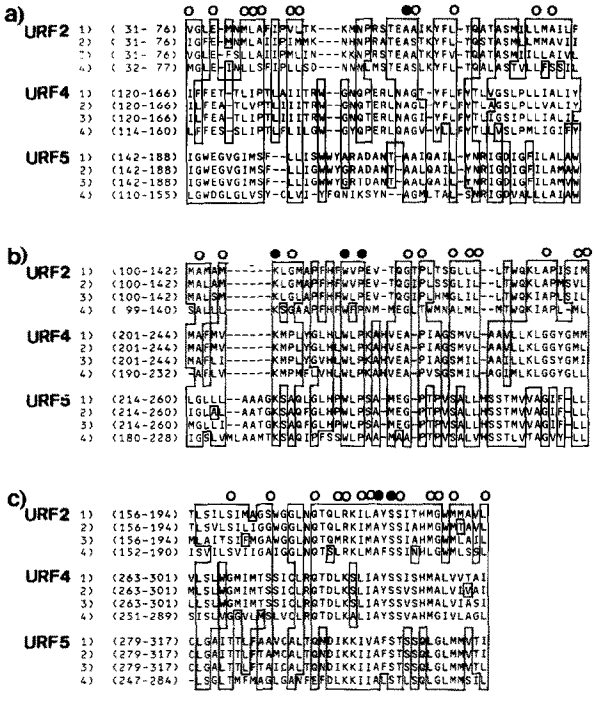

Fig.3. Alignments of the URF2, URF4 and URF5 products. Only regions that are conserved between different URFs (the positions were shown in parentheses) were aligned. (1) Human, (2) bovine, (3) mouse and (4) *Drosophila*. Positions that are occupied by identical or chemically similar amino acids [11] between different URFs were boxed for most common amino acids. (●, ○), Positions that are occupied by identical and chemically similar amino acids among all the sequences, respectively. Gaps (−) were inserted to increase sequence similarity.

The URF2, 4 and 5 products have homology in common. They were each aligned to three mammalian and *Drosophila* sequences (fig.3). A highly conserved stretch of amino acids was found in the C-terminal portion of alignment (c) of fig.3 (positions 23–28). Such highly conserved sequences might carry an important function which is common among the three URF products. Similar sequences were found in URF1 products of mammals (positions 102–107 of mouse). *Drosophila* and *A. nidulans*. We do not know whether this sequence similarity is a result of divergent evolution or convergent evolution due to similar function. Although the URF2, 4 and 5 differ in size, the homologous region may aggregate to form a subunit structure in a pseudosymmetric fashion.

On the basis of the alignments shown in fig.3 and including further homologous sequences from *T. brucei*, *L. tarentolae* and *A. nidulans*, a phylogenetic tree representing evolutionary relationships among the URF2, URF4 and URF5 was constructed by a modified matrix method [15]. According to the inferred tree topology, a series of gene duplications that generated three different genes had occurred before the times of separation of animal/protozoa and animal/fungus. It is highly likely that this set of URFs is present in fungal and protozoa mtDNAs, although some of them have not yet been identified. It may be in-

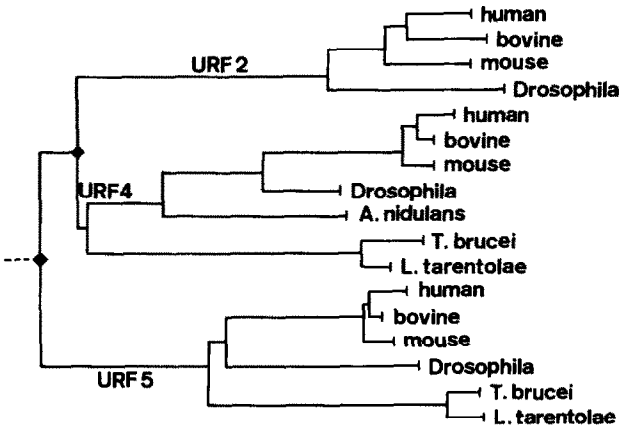

Fig.4. Phylogenetic tree representing evolutionary relationships among the URF2, URF4 and URF5. (♦) Gene duplication is expected to have occurred.

teresting to know whether or not the homologous sequences exist in prokaryotic genomes. If so, the gene duplications may be considered to have occurred at very remote times before the endosymbiosis.

ACKNOWLEDGEMENTS

We thank Drs H. Hayashida and H. Toh for discussion and Miss K. Mitsuyasu for technical assistance. This work was supported in part by grants from the Ministry of Education, Science and Culture of Japan.

REFERENCES

[1] Ohno, S. (1970) Evolution by Gene Duplication, Springer, Berlin.
[2] Anderson, S., Bankier, A.T., Barrell, B.G., De Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. (1981) Nature 290, 457–465.
[3] Bibb, M.J., Van Etten, R.A., Wright, C.T., Walberg, M.W. and Clayton, D.A. (1981) Cell 26, 167–180.
[4] Anderson, S., De Bruijn, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F. and Young, I.G. (1982) J. Mol. Biol. 4, 683–717.
[5] Lang, B.F., Ahne, F., Distler, S., Trinki, H., Kaudewitz, F. and Wolf, K. (1983) in: Nucleo-mitochondrial Interactions, Mitochondria (Schweyen, R.J., Wolf, K. and Kaudewitz, F. eds) pp.319, De Gruyter, Berlin.
[6] Clary, D.O. and Wolstenholme, D.R. (1983) Nucleic Acids Res. 11, 6859–6871.
[7] Clary, D.O., Wahleithner, J.A. and Wolstenholme, D.R. (1984) Nucleic Acids Res. 12, 3747–3762.
[8] Hensgens, L.A.M., Brakenhoff, J., De Vries, B.F., Sloof, P., Tromp, M.C., Van Boom, J.H. and Benne, R. (1984) Nucleic Acids Res. 12, 7327–7344.
[9] De la Cruz, V.F., Neckelmann, N. and Simpson, L. (1984) J. Biol. Chem. 259, 15136–15147.
[10] Terence, A., Brown, R., Wayne, D., Ray, J.A., Waring, R.B. and Scazzocchio, C. (1983) EMBO J. 2, 427–435.
[11] Toh, H., Hayashida, H. and Miyata, T. (1983) Nature 305, 827–829.
[12] Schwartz, R.M. and Dayhoff, M.O. (1978) in: Atlas of Protein Sequence and Structure (Dayhoff, M.O. ed.) vol.5, suppl.3, pp.353–358, National Biomedical Research Foundation, Washington, DC.
[13] Yamawaki-Kataoka, Y., Miyata, T. and Honjo, T. (1981) Nucleic Acids Res. 9, 1365–1381.
[14] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[15] Li, W.-H. (1981) Proc. Natl. Acad. Sci. USA 78, 1085–1089.