1400–1403 Nucleic Acids Research, 1994, Vol. 22, No. 8

Nucleotide sequence of the *cox3* gene from *Chondrus crispus*: evidence that UGA encodes tryptophan and evolutionary implications

Catherine Boyen*, Catherine Leblanc, Géraldine Bonnard[1], Jean-Michel Grienenberger[1] and Bernard Kloareg

Centre d'Etudes d'Océanologie et de Biologie Marine, CNRS-UPR 4601, Université P. & M. Curie, BP 74, 29682 Roscoff Cedex and [1]Institut de Biologie Moléculaire des Plantes, Université Louis Pasteur, 12 rue du Général Zimmer, 67084 Strasbourg Cedex, France

Received January 24, 1994; Revised and Accepted March 15, 1994 EMBL accession no. Z29482

ABSTRACT

We present the nucleotide sequence of the gene encoding subunit 3 of cytochrome c oxidase in *Chondrus crispus*, the first report on a mitochondrial gene from a red alga. Amino acid alignment with homologous proteins shows that tryptophan is specified by UGA, as in the mitochondrial code of most organisms other than green plants. However, phylogenetic analyses of *cox3* amino acid and nucleotide sequences indicate that *C.crispus* COX3 is related to the green-plant mitochondrial lineage. No RNA editing was detected on the corresponding transcript. As the only known photosynthetic eukaryotes that both share an immediate mitochondrial ancestor with green plants and exhibit features characteristic of non-plant mitochondria, ie, a small-sized mitochondrial genome and a modified genetic code, rhodophytes may be thought of as an intermediate evolutionary link at the root of the green-plant mitochondrial lineage.

INTRODUCTION

Although the α-purple eubacterial origin of mitochondria is widely accepted (1) it is not clear whether these organelles are poly- or mono-phyletic (2). In particular higher-plant mitochondrial genomes exhibit unique characteristics, such as a large size and a marked tendency to recombination. In contrast mitochondria of ciliates, animal and fungi are characterised by a reduction of their genome size, following gene transfer into the nucleus (2). Green-plant mitochondria use the universal genetic code (3) whereas non-plant mitochondria use modified codes, with as many as seven altered codons in vertebrates (4). In global Small SubUnit (SSU) rRNA trees, green plants branch very close to the root of the mitochondrial subtree (5), a position consistent with the strong eubacterial character of plant rRNA sequences. Such a topology markedly differs from the branching position of higher plants in nuclear phylogenies and it was proposed that the rRNA genes of green-plant mitochondria may have been acquired in a separate, more recent event (2).

So far the only well-known mitochondrial genomes from photosynthetic eukaryotes other than land plants are those of the unicellular green algae, *Chlamydomonas reinhardtii* (Volvocales, 6) and *Prototheca wickerhamii* (Chlorococcales, 7) which also use the universal genetic code. Phylogenetic analysis of *P.wickerhamii* COX1 suggests a recent common ancestor for the mitochondria of Chlorococcales and those of land plants (7). In contrast, *C.reinhardtii* branches separately from land plants in SSU rRNA mitochondrial phylogenies (5) whereas it belongs to the green plant lineage in nuclear SSU trees (8). We report here on the sequence of the mitochondrial gene *cox3* encoding subunit 3 of cytochrome c oxidase from the red alga *C.crispus*. Rhodophytes are photosynthetic eukaryotes with 'primitive' plastids. In nuclear phylogenies they emerge as a distinct lineage, contemporary with the other eukaryotes lineages (8–10).

METHODS

Identification of a cox3 DNA clone

Total DNA was prepared from *C.crispus* as described by Apt and Grossman (11). Nuclear and organellar (i.e., plastidial plus mitochondrial) DNAs were separated using a Hoechst 33258-CsCl density gradient (Boyen *et al.*; submitted for publication.). In Southern hybridisation of organellar DNA, a single 3.7-kb *Eco*RI fragment annealed with the probe, consisting of a wheat mitochondrial *cox3* cDNA (12). Organellar DNA was digested to completion by *Eco*RI, a shot-gun library was prepared in pBluescript SK and a recombinant clone containing the *cox3* gene on a 3.7-kb insert was identified using the wheat probe. No other *cox3* related sequence could be detected elsewhere on the organellar or nuclear fractions.

cDNA synthesis and sequencing

In order to check whether editing could affect *C.crispus cox3* transcript, total RNA (11) was treated with RNase-free DNase

*To whom correspondence should be addressed

I and first-strand cDNAs were synthesised (12). Two synthetic oligonucleotides O_1 and O_2 (see below), corresponding respectively to the 5′ and the 3′ non coding regions of the putative *cox3*, were used to specifically amplify the first-strand cDNA. Amplified DNA with the expected size (970-bp) was cloned into the *Bam*HI−*Eco*RI sites of pBluescript KS vector and seven clones were sequenced. The presence of contaminant DNA in the RNA used for the RT-PCR reaction was ruled out as no signal was detected by Southern hybridisation in a control experiment omitting the RT step. The following oligonucleotides were used for PCR and cDNA sequencing:

O_1=CGATTGGATCCGCAAATTATAGC
O_2=GTTAAGCGAATTCGACTAAGATTAAGGC
O_3=CTTAAAACTGATTGATTACG

The underlined nucleotides were modifed in order to create *Bam*HI and *Eco*RI sites.

DNA sequencing and computer analysis

The *Eco*RI clone containing the *cox3* gene was used to create series of overlapping deletions by Exonuclease III and Mung Bean digestions (Stratagene). The nucleotide sequences of the various ExoIII deletion clones were determined by the dideoxy chain termination method (13) using a sequencing kit (Pharmacia), universal primers and, when appropriate, synthetic primers deduced from the gene sequence.

Amino acid sequences were aligned using the GCG sequence analysis software package (14). Gaps were introduced for optimal alignment, yielding 238 informative sites for comparison. Complete alignments are available upon request. Phylogenetic trees based on COX3 amino acid sequences were established with the neighbour-joining method (15) applied to a distance matrix calculated with 'protdist' (PHYLIP 3.5c, Felsenstein). Amino acid substitution was based on the Kimura two-parameters model (16). Bootstrap analysis (PHYLIP 3.5c, seqboot program, 17) was used to evaluate the solidity of branching pattern (using 100 resamplings of the data set).

RESULTS AND DISCUSSION

The *cox3* gene from *C.crispus* was sequenced from a fragment belonging to a 25.9-kb circular molecule of organellar DNA, in which a number of other mitochondrial genes such as *cox1*

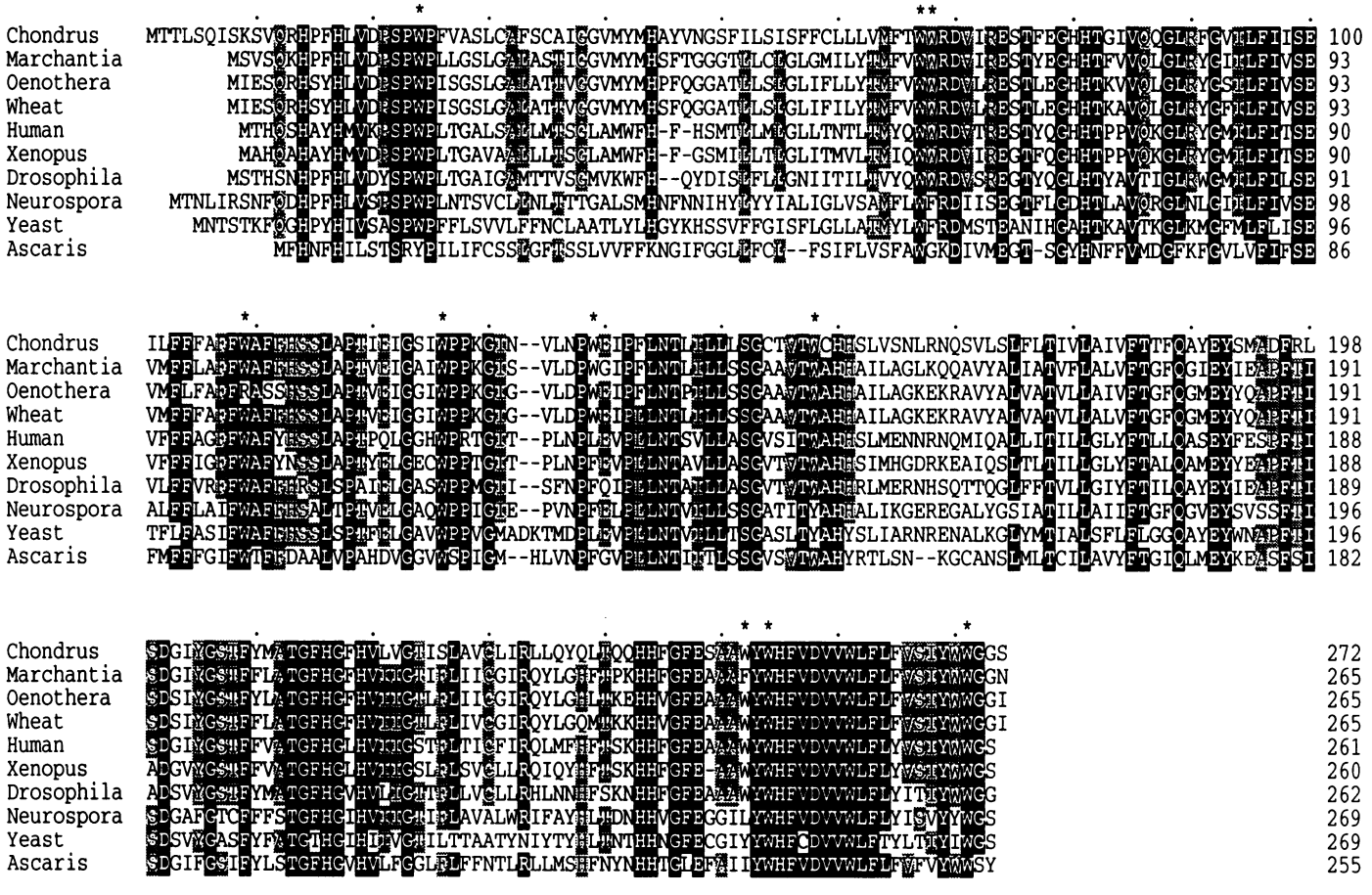

Figure 1. Alignment of COX3 protein sequences from *C.crispus*, and representatives of green plants, fungi and animals. The UGG and UGA codons in *C.crispus cox3* were translated as trp (W). Stars indicate Trp-encoding UGA codon in *Chondrus crispus*. When necessary specific codes were used to translate the gene sequences from other taxa. The wheat COX3 protein sequence was deduced from the cDNA (12). The C terminal part (from amino acid 155) of *Oenothera* protein was deduced from the cDNA (24). For the remaining part of *Oenothera* COX3, other editing sites can be predicted, such as the arg codon at position 102. Black and dark grey boxes indicate identical amino acids in at least 9 and 7 out of 10 sequences, respectively. Chondrus, *Chondrus crispus* (Z29482); Marchantia, *Marchantia polymorpha* (M68929); Oenothera, *Oenothera berteriana* (X04764); Wheat, *Triticum aestivum* (X52539); Human, *Homo sapiens* (J01415); Xenopus, *Xenopus laevis* (M10217); Drosophila, *Drosophila yakuba* (X03240); Neurospora, *Neurospora crassa* (V00668); Yeast, *Schizosaccharomyces pombe* (X16868); Ascaris, *Ascaris suum* (P24879). Parentheses refer to GenBank or EMBL accession numbers.

and *cob* were identified (Boyen *et al.*; submitted for publication). The *cox3* transcripts were detected in total RNA extracts. In the protein sequence deduced from the nucleotide sequence of cox3, the UGA codon, a termination codon in the universal code, was detected 9 times. These 9 positions correspond to highly conserved tryptophan sites in *cox3* genes from other organisms (Figure 1). In addition one UGA codon (aa 137 of *C.crispus)* corresponds to a tryptophan site found only in land plants (12). The standard codon for trp, UGG, was detected twice, at positions 259 and 268, which also are highly conserved tryptophan residues. In higher-plant mitochondria the codon corresponding to position 259 is RNA edited from CGG to UGG, leading to the conserved trp (12). To assess the possibility that the UGA codons in *C.crispus cox3* might be modified by mRNA editing, *cox3* cDNA was synthesised from total RNA, amplified and cloned. None of the seven clones that were sequenced was different from the genomic DNA sequence, indicating that the *cox3* transcript is not edited and that UGA does specify tryptophan in *C.crispus cox3* gene. Comparison with the amino-acid sequences of COX3 from various eukaryotes (Figure 1) does not suggest any other codon reassignment in the *C.crispus* gene. Change of UGA from stop to trp is thought of as the first modification in the genetic code and, according to the codon capture theory (4), is considered to be correlated to A/T pressure. The *cox3* gene of *C.crispus* exhibits a strong bias for A or T (77%) in the usage of codons. Both the *cox2−cox3* spacer (176 bp) and the *cox3−cob* intergenic region (696 bp) also have a high AT content (80%), indicating an overall evolutionary constraint towards the predominance of A/T in the mitochondrial genome of *C.crispus*.

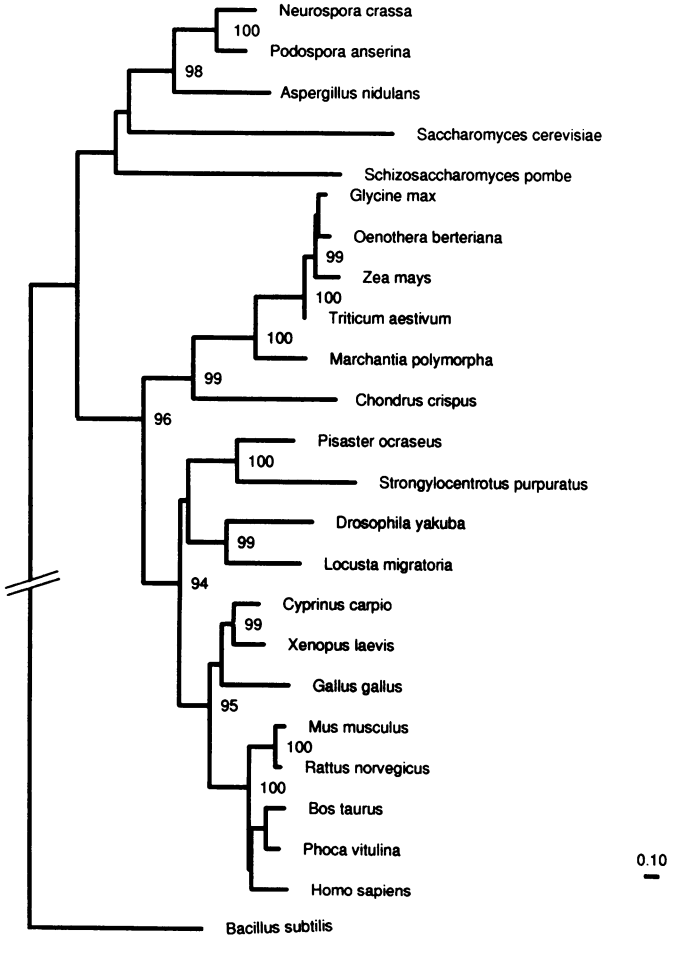

Figure 2. Phylogenetic analysis of COX3 amino acid sequences. The tree was built using the neighbour-joining method (15) applied to a categories distance matrix (PHYLIP 3.5c, Protdist program, Felsenstein unpublished). Only the significant bootstrapping values (higher than 93) are included in the tree. The horizontal bar represents 0.1 substitution per nucleotide and branch lengths are drawn to scale. *Bacillus subtilis* was chosen as the outgroup. Complete alignments are available upon request. References of the *cox3* sequences not given in Figure 1 are: *Zea mays* (X12728), *Podospora anserina* (X14734), *Aspergillus nidulans* (X06960), *Saccharomyces cerevisiae* (J01478), *Glycine max* (X15131), *Pisaster ocraseus* (X55514), *Strongylocentrotus purpuratus* (X12631), *Locusta migratoria* (X13975), *Cyprinus carpio* (X17006), *Gallus gallus* (X52392), *Mus musculus* (V00711), *Rattus norvegicus* (V01574), *Bos taurus* (V00654), *Phoca vitulina* (X63726), and *Bacillus subtilis* (X54140).

Distance matrix neighbour-joining analysis at the amino-acid level of various COX3 proteins (Figure 2) shows that the red algal COX3 is more closely related to those of green plants than to those of fungi or metazoa.The branching position of higher plants is more similar to what is observed in nuclear phylogenies (10,19) than in SSU rRNA mitochondrial phylogenies (1,2), where angiosperms are positioned very close to the root of the tree. The global topology of the tree, with plants and animals branching together but separately from higher fungi, is not congruent with nuclear trees. The bootstrap value for the clade of higher fungi, however is low and no confidence can be given to its position. In contrast, bootstrap analysis indicates that the branching of *C.crispus* at the root of the green-plant mitochondrial lineage can be inferred with high confidence in this tree. Similar neighbour-joining topologies (data not shown) were obtained when nucleotides alignments from the same 26 species were analysed by the weighed pathway method (18) or the Kimura dissimilarity matrix method (16) (performed for the first and second positions of codons). However, bootstrapping analysis (100 replicates) indicated a poor specific branching of *C.crispus* in the nucleotide trees.

In conclusion, the COX3 tree suggests that red-algal mitochondria and higher-plant mitochondria share a common evolutionary history. Yet, contrarily to green plants, *C.crispus* mitochondria display a marked AT bias and do not use the universal genetic code. In addition, unlike those of green plants, the mitochondrial genome of *C.crispus* is rather small (25.9 kb), a characteristic that can be considered as the result of a long evolution process. Interestingly, oomycetes, a phylum related to heterokonts at the nuclear level (19,20), also harbour *cox* genes that present extensive sequence similarities with those of green plants and possess a relatively small mitochondrial genome (c.a. 40 kb), but they use the universal genetic code (21,22,23). The mitochondrial genome of*Pylaiella littoralis*, a brown alga also uses the universal genetic code (S.Loiseaux-de Goër, personnal communication). Overall the mitochondria of red algae appear to be phylogenetically related with those of green plants and oomycetes but they share common features with non-plant mitochondria. We therefore suggest that red algal mitochondria have emerged early from the green-plant lineage and were not involved in the secondary process that shaped the mitochondria of extant land plants (2).

ACKNOWLEDGEMENTS

We thank Claude Leroux and Olivier Collin for their valuable assistance in data processing and construction of phylogenetic trees. We are grateful to Susan Loiseaux-de Goër for critical

reading of the manuscript. This work is supported by the G.D.R. 1002 ('Biology, biochemistry and genetics of marine macroalgae').

REFERENCES

1. Yang,D., Oyaizu,Y., Oyaizu,H., Olsen,G.J. and Woese,C.R. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4443–4447.
2. Gray,M.W. (1992) In Wolstenholme,D.R. and Jeon,K.W. (eds), Mitochondrial Genomes, Academic Press, San Diego, pp. 233–357.
3. Gualberto,J.M., Lamattina,L., Bonnard,G., Weil,J.H. and Grienenberger,J.M. (1989) *Nature* **341**, 660–662.
4. Osawa,S., Jukes,T.H., Watanabe,K. and Muto,A. (1992) *Microbiological Reviews* **56**, 229–264.
5. Cedergren,R., Gray,M.W., Abel,Y. and Sankoff,D. (1988) *J. Mol. Evol.* **28**, 98–112.
6. Kuck,U. and Neuhaus,H. (1986) *Appl. Microbiol. Biotechnol.* **23**, 462–469.
7. Wolff,G., Burger,G., Lang,B.F. and Kück,U. (1993) *Nucleic Acid Res.* **21**, 719–726.
8. Douglas,S.E., Murphy,C.A., Spencer,D.F. and Gray,M.W. (1991) *Nature* **350**, 148–151.
9. Perasso,R., Baroin,A., Qu,L.H., Bachellerie,J.P. and Adoutte,A. (1989) *Nature* **39**, 142–144.
10. Hendricks,L., de Baere,R., Van de Peer,Y., Neefs,J., Goris,A. and de Wachter,R. (1991) *J. Mol. Evol.* **32**, 167–177.
11. Apt,K.E. and Grossman,A.R. (1993) *Plant Mol. Biol.* **21**, 27–38.
12. Gualberto,J.M., Weil,J.H. and Grienenberger,J.M. (1990) *Nucleic Acid Res.* **18**, 3771–3776.
13. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
14. Devereux,J., Heaberli,P. and Smithies,O. (1984) *Nucleic Acids Res.* **12**, 387–395.
15. Saitou,N. and Nei,M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
16. Kimura,M. (1980) *J. Mol. Evol.* **16**, 111–120.
17. Felsenstein,J. (1985) *Evolution* **39**, 783–791.
18. Li,W.H., Wu,C.I. and Luo,C.C. (1985) *Mol. Biol. Evol.* **2**, 150–174.
19. Bhattacharya,D., Stickel,S.K. and Sogin,M.L. (1991) *J. Mol. Evol.* **33**, 525–536.
20. Bhattacharya,D., Medlin,L., Wainright,P.O., Ariztia,E.V., Bibeau,C., Stickel,S.K. and Sogin,M.L. (1992) *Evolution* **46**, 1801–1817.
21. Förster,H., Kinscherf,T.G., Leong,S.A. and Maxwell,D.P. (1987) *Curr. Genet* **12**, 215–218.
22. Karlovsky,P. and Fartmann,B. (1992) *J. Mol. Evol.* **34**, 254–258.
23. Sachay,D.J., Hudspeth, D.S.S. and Nadler Sa Hudspeth,M.E.S. (1993) *Experimental Phycology* **17**, 7–23.
24. Hiesel,R., Wissinger,B., Schuster,W. and Brennicke,A. (1989) *Science* **246**, 1632–1634.