Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 2328–2332, March 1996
Evolution

Genes encoding the same three subunits of respiratory complex II are present in the mitochondrial DNA of two phylogenetically distant eukaryotes

(mitochondrion/succinate:ubiquinone oxidoreductase/succinate dehydrogenase/*Porphyra purpurea*/*Reclinomonas americana*)

GERTRAUD BURGER*‡, B. FRANZ LANG*‡, MICHAEL REITH†, AND MICHAEL W. GRAY*§

*Program in Evolutionary Biology, Canadian Institute for Advanced Research, ‡Département de biochimie, Université de Montréal, Montréal, QC H3C 3J7, Canada; †Institute for Marine Biosciences, National Research Council, Halifax, NS B3H 3Z1, Canada; and §Department of Biochemistry, Dalhousie University, Halifax, NS B3H 4H7, Canada

Communicated by Charles S. Levings III, North Carolina State University, Raleigh, NC, November 8, 1995 (received for review July 28, 1995)

ABSTRACT **Although mitochondrial DNA is known to encode a limited number (<20) of the polypeptide components of respiratory complexes I, III, IV, and V, genes for components of complex II [succinate dehydrogenase (ubiquinone); succinate:ubiquinone oxidoreductase, EC 1.3.5.1] are conspicuously lacking in mitochondrial genomes so far characterized. Here we show that the same three subunits of complex II are encoded in the mitochondrial DNA of two phylogenetically distant eukaryotes, *Porphyra purpurea* (a photosynthetic red alga) and *Reclinomonas americana* (a heterotrophic zooflagellate). These complex II genes, *sdh2*, *sdh3*, and *sdh4*, are homologs, respectively, of *Escherichia coli sdhB*, *sdhC*, and *sdhD*. In *E. coli*, *sdhB* encodes the iron-sulfur subunit of succinate dehydrogenase (SDH), whereas *sdhC* and *sdhD* specify, respectively, apocytochrome b_{558} and a hydrophobic 13-kDa polypeptide, which together anchor SDH to the inner mitochondrial membrane. Amino acid sequence similarities indicate that *sdh2*, *sdh3*, and *sdh4* were originally encoded in the protomitochondrial genome and have subsequently been transferred to the nuclear genome in most eukaryotes. The data presented here are consistent with the view that mitochondria constitute a monophyletic lineage.**

Mitochondrial DNA (mtDNA) encodes a limited number (usually <20) of the protein components of the functional mitochondrion, among which are subunits of key complexes involved in electron transport and oxidative phosphorylation—namely, complexes I (NADH:ubiquinone oxidoreductase), III (ubiquinol:cytochrome *c* oxidoreductase), IV (cytochrome *c* oxidase), and V (ATP synthase) (1, 2). In all of these cases, the mitochondrially encoded and synthesized polypeptides interact with proteins that are encoded in the nuclear genome, synthesized in the cytosol, and imported into the organelle. Given the accumulated evidence that the mitochondrial genome was acquired in evolution from a eubacteria-like endosymbiont (3–6), it is usually assumed that the low coding capacity of contemporary mitochondrial genomes reflects a massive loss of genes originally encoded in the protomitochondrial genome or their transfer to the nucleus (5, 6). The latter assumption draws support from specific examples of mitochondrial genes that have been relocated to the nucleus relatively recently in evolutionary history, including genes coding for components of the respiratory chain (7, 8).

Usually, a given mitochondrial function will be found to be encoded in the nuclear DNA in some organisms but in the mtDNA in others (5). In particular, animals and fungi have nuclear DNA counterparts for a number of genes that are encoded in the mtDNA of other eukaryotes. An example is the gene for the largest subunit (NAD11) of the NADH dehydrogenase of complex I. Until recently, only nuclear genes for NAD11 had been described; however, this gene has now been identified in the mitochondrial genomes of the oomycete *Phytophthora infestans* (9), the cellular slime mold *Dictyostelium discoideum* (10), and the amoeboid protozoon *Acanthamoeba castellanii* (11). Such examples strengthen the inference that particular nuclear genes were once encoded by, and have been acquired from, the mitochondrial genome.

In contrast to the situation with respiratory complexes I, III, IV, and V, genes for components of complex II [succinate dehydrogenase (ubiquinone); succinate:ubiquinone oxidoreductase, EC 1.3.5.1] have not been found in any animal or fungal mtDNA (1, 2, 5), and, until very recently, there was also no evidence for such genes in plant or protist mitochondrial genomes. Complex II comprises the tricarboxylic acid cycle enzyme succinate dehydrogenase [SDH; succinate:(acceptor) oxidoreductase, EC 1.3.99.1] plus polypeptides that anchor SDH to the inner mitochondrial membrane (12–14). A number of nuclear genes for components of complex II have been identified and characterized (15–30), as have the homologous genes in both eubacteria (31–38) and archaebacteria (39). In *Escherichia coli*, the genes *sdhA* (32) and *sdhB* (31) encode the two subunits of SDH (the flavin protein and the iron-sulfur protein, respectively), whereas *sdhC* and *sdhD* (32) specify apocytochrome b_{558} and a hydrophobic 13-kDa polypeptide, the two integral membrane components of complex II. These membrane-anchoring proteins play an important role in electron transfer to ubiquinone (30, 33, 34).

The apparent absence of mtDNA-encoded genes for complex II components raises questions about the evolutionary origin of the corresponding nuclear genes—i.e., were they originally present in the nuclear genome of the amitochondriate eukaryote that served as host to the eubacteria-like, protomitochondrial endosymbiont, or were they brought into the eukaryotic host by the endosymbiont and subsequently transferred to the nuclear genome? Recently, Daignan-Fornier *et al.* (28) characterized the yeast (*Saccharomyces cerevisiae*) nuclear gene *SDH3*, a homolog of *E. coli sdhC*, noting that this gene is also homologous at the amino acid level to an unidentified open reading frame, *orf137*, that had previously been found by Ohyama *et al.* (40) in the mitochondrial genome of the liverwort *Marchantia polymorpha*. Here we show that three *sdh* genes are encoded in the mtDNA of two evolutionarily distant eukaryotes, *Porphyra purpurea*, a multicellular red alga (rhodophyte) and *Reclinomonas americana*, a unicellular, early diverging bacterivore, a member of a group of protists known as the jakobid flagellates (41). Comparison of the derived amino acid sequences of these three mtDNA-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: SDH, succinate dehydrogenase; ORF, open reading frame; NCBI, National Center for Biotechnology Information.

§To whom reprint requests should be addressed.

encoded *sdh* genes with their nuclear and prokaryotic homologs allows us to draw conclusions about the evolution of nuclear *SDH* genes in eukaryotes, as well as about the origin of the mitochondrial genome itself.

MATERIALS AND METHODS

mtDNAs from *P. purpurea* and *R. americana* (ATCC 50394) have been sequenced under the auspices of the Organelle Genome Megasequencing Program (OGMP). Details of growth conditions for these organisms and isolation and cloning of mtDNA will be presented in conjunction with a description of the complete sequences of the *P. purpurea* and *R. americana* mitochondrial genomes (unpublished data). DNA sequencing, data entry, and sequence analysis were performed as described (11). Information about OGMP as well as detailed experimental protocols may be obtained from WWW site URL http://megasun.bch.umontreal.ca/. The Protist Image Data Base at this site provides additional information about *P. purpurea* and *R. americana*, including comments on their phylogenetic positions.

Searches for and retrieval of sequences were carried out using the services of the National Center for Biotechnology Information (NCBI). Multiple alignments of the deduced protein sequences were generated with CLUSTAL V and W (42) using the PAM250 matrix (43). The mitochondrial SDH2, SDH3, and SDH4 protein sequences from *P. purpurea* and *R. americana* have been deposited in the Swiss-Prot data bank (accession nos. P80477–P80482).

RESULTS AND DISCUSSION

Homologs of *E. coli sdhB*, *sdhC*, and *sdhD* (here designated *sdh2*, *sdh3*, and *sdh4*, respectively) were encountered in the course of determining the complete primary sequences of the mitochondrial genomes of *P. purpurea* and *R. americana*. In *P. purpurea* mtDNA, the three *sdh* genes are widely separated in the orientation . . . *-rps11-sdh3-trnG-* . . . *-rnl-sdh2-trnL-* . . . *-nad4-sdh4-nad3-* . . . (transcription from left to right in all cases). In marked contrast, *sdh* genes in *R. americana* mtDNA are tightly linked in the cluster . . . *-sdh3-sdh4-sdh2-nad4L-* . . . , with these genes also all transcribed in the same direction (from left to right). Physical linkage to respiratory chain genes (*nad2*, *nad4*, *nad4L*) and to the mitochondrial large subunit rRNA gene (*rnl*) indicates that these *sdh* genes reside in mtDNA rather than in contaminating nuclear or bacterial DNA. This conclusion has been verified by complete (*P. purpurea*) or nearly complete (*R. americana*) sequencing of the respective mitochondrial genomes (to be reported in detail elsewhere).

The derived amino acid sequences of the *sdh2* genes from *P. purpurea* and *R. americana* mtDNAs are shown in Fig. 1, aligned with their eubacterial (*E. coli*, *Coxiella burnetii*) and nuclear (*Ustilago maydis*, *S. cerevisiae*, *Caenorhabditis elegans*, and *Homo sapiens*) homologs. Within a well-aligned region encompassing amino acid residues 40–228 of *Reclinomonas* SDH2 (corresponding to 79% of this sequence), the mitochondrial sequences share a high degree of sequence identity with their nuclear homologs, ranging from 62% to 77%; however, they display a substantially lower level of identity with the two eubacterial representatives (51–55%) and particularly with the single known archaebacterial sequence, that of *Thermoplasma acidophilum* (38–39%). Notably, the *T. acidophilum* SDHB sequence (NCBI gi:479545) is about equally distant in these comparisons from the eubacterial, mitochondrial, and nuclear sequences. At their N termini, the mitochondrial sequences more closely approximate their eubacterial homologs in length; indeed, *Reclinomonas* and *Coxiella* SDHB begin at

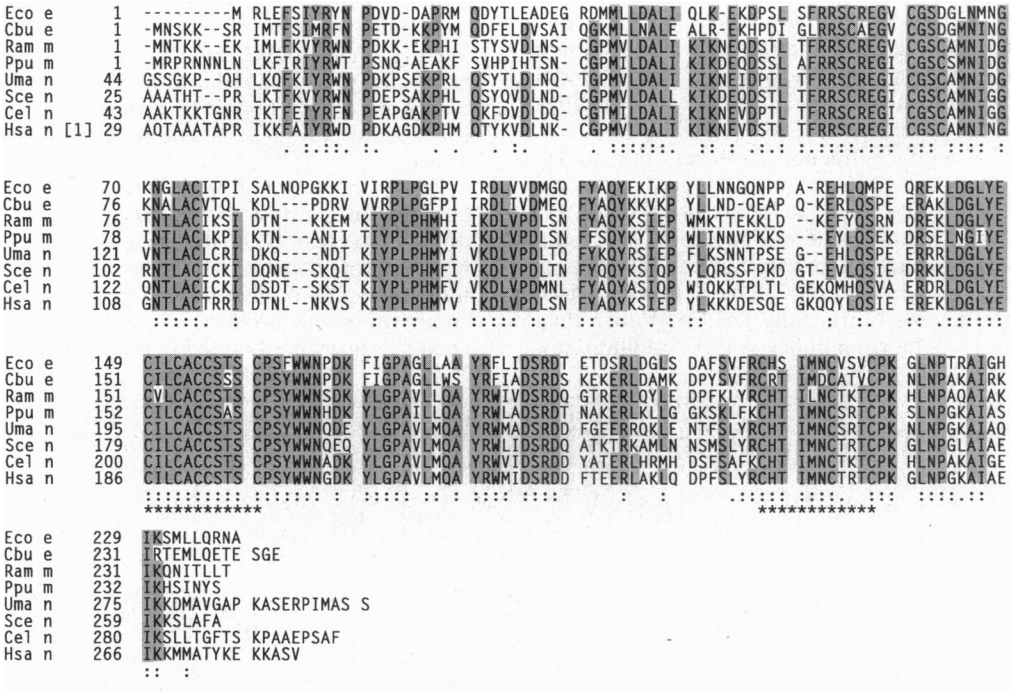

FIG. 1. Alignment of SDH2 protein sequences. Abbreviations used for organism names (with NCBI gi: identifiers for each SDH2 sequence indicated in brackets) are as follows: Eco, *E. coli* (118615); Cbu, *C. burnetii* (541241); Ram, *R. americana* (this paper); Ppu, *P. purpurea* (this paper); Uma, *U. maydis* (416903); Sce, *S. cerevisiae* (yeast) (118619); Cel, *C. elegans* (642191); Hsa, *H. sapiens* (human) (118616). The single lowercase letter indicates which genome encodes SDH2 in a particular organism (e, eubacterial; m, mitochondrial; n, nuclear). In the case of human nucleus-encoded SDH2 (Hsa n), codon 29 of its gene corresponds to the known N terminus of the mature protein (indicated by [1] in the figure). Alignment gaps are denoted by dashes. In columns where >67% of the residues are the same, these identical residues are emphasized by shading. Columns in which all of the residues are chemically similar (e.g., all aromatic or basic) are denoted by a colon, whereas those in which >67% of the residues are of the same chemical type are marked by a dot. Asterisks delineate highly conserved, cysteine-rich regions that form the iron-sulfur centers of the protein.

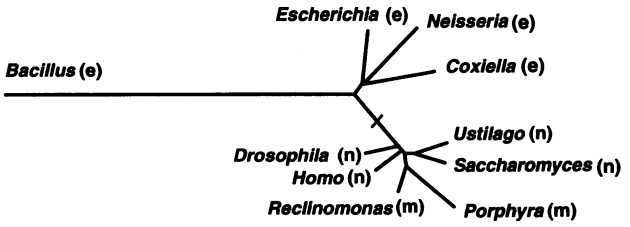

FIG. 2. Phylogenetic tree of SDH2 protein sequences. Data set comprises 222 amino acid positions within unambiguously aligned sequence blocks (alignment available upon request from B.F.L. at email address langf@bch.umontreal.ca). Data were analyzed by a distance approach (PROTDIST, Dayhoff PAM matrix option, and FITCH) (44, 45) and submitted to bootstrap analysis (45). Alternative trees (not shown) were constructed with a maximum likelihood algorithm (PROTML) (46), using an exhaustive tree optimization procedure with the jf parameters. No significant differences were observed when the distance matrices created by PROTDIST were combined with either a NEIGHBOR or FITCH algorithm. e, Eubacterial; n, nuclear; m, mitochondrial. Branch leading to the mitochondrial and nuclear SDH2 sequences (horizontal bar) is supported by a bootstrap value of 100%. NCBI gi: identifiers for all sequences are listed in Fig. 1 with the exception of *Bacillus* (*B. subtilis*; 118613), *Neisseria* (*N. gonorrhoeae*; 150368), and *Drosophila* (*D. melanogaster*; 508849).

precisely the same position and share evident primary sequence similarity within the first 20 residues. The longer N-terminal extensions of the nuclear SDHB sequences undoubtedly reflect a requirement for mitochondrial targeting of these proteins. Nevertheless, consideration of the sequence alignment and pairwise identities clearly indicates that the mitochondrial and nuclear SDHB sequences are more closely related to one another than to the two eubacterial sequences.

The phylogenetic relationships suggested by the results of pairwise comparison are further emphasized in the unrooted phylogenetic tree shown in Fig. 2, which incorporates additional eubacterial sequences (*Bacillus subtilis* and *Neisseria gonorrhoeae*). In this tree, the nuclear and mitochondrial sequences form a clade to the exclusion of the bacterial sequences. Bootstrap analysis indicates that the topology within this clade is not robust, so the relative branching order cannot be inferred with confidence; however, the branch leading to the cluster of nuclear and mitochondrial sequences is strongly supported (100%), and the same tree topology was obtained using a maximum likelihood approach (PROTML). From these results, we conclude that the nuclear *SDH2* genes originated by evolutionary transfer from a mitochondrial genome in which they were originally resident rather than directly and more recently from eubacteria. In addition, our data do not support an ancient acquisition of these genes from eubacterial or archaebacterial donors that may have contributed substantially to the initial formation of the eukaryotic nuclear genome (47).

Although the selection of eubacterial SDH2 sequences available for comparison is limited, it is notable that the three Proteobacteria (*Escherichia*, *Neisseria*, *Coxiella*) form a tight clade that is well separated from *B. subtilis* (a Gram-positive eubacterium). Moreover, as a group the proteobacterial sequences appear more closely related to the mitochondrial/nuclear clade than does the *Bacillus* sequence. This relationship is consistent with much molecular data that support the view that mitochondria originated in evolution from within the α-proteobacterial phylum of eubacteria (5, 6).

Our analysis of SDH2 amino acid similarities does not support the recent hypothesis (48) that the ancestor of the animal mitochondrial genome was a prokaryotic endosymbiont sharing specific ancestry with either *Sulfolobus* (an archaebacterium) or *Mycoplasma* (a close relative of Gram-positive eubacteria such as *B. subtilis*; see ref. 49), rather than with an α-proteobacterium. Instead, our data clearly indicate that animal SDH2 protein sequences not only form a monophyletic grouping with their mitochondrion- and nucleus-encoded counterparts from other organisms but that they are more closely related to proteobacterial SDH2 sequences than to those of Gram-positive eubacteria or archaebacteria.

An SDH3 alignment (Fig. 3) shows that these sequences display a substantially lower level of conservation than do SDH2 sequences, although sequence similarities within short N-terminal and C-terminal regions clearly indicate that these *sdh3* genes are homologs of one another. In this case, quantitative pairwise comparisons (results not shown) were restricted to regions within which relatively few deletions have to be assumed in optimizing the alignment (~80 residues total). Although the values obtained parallel the pattern observed in SDH2 comparisons, the phylogenetic conclusions are less compelling because of the lower overall level of sequence identity among SDH3 proteins. As expected for a membrane-associated protein, the inferred SDH3 sequences are considerably more hydrophobic than the SDH2 ones (54% vs. 33% hydrophobic residues, *P. purpurea*; 54% vs. 32%, *R. americana*).

Fig. 4 shows a multiple alignment of SDH4 protein sequences. In contrast to the *sdh2* and *sdh3* cases, it proved difficult to identify *P. purpurea* and *R. americana sdh4* as homologs of bacterial (*sdhD*) and yeast (*SDH4*) genes. *E. coli sdhD* encodes a protein that resembles in size and amino acid composition the product of *frdD*, the fumarate reductase anchor, suggesting that *sdhD* may also encode a membrane

```
Eco e      1                  MIRN VKKQRPVNLD LQTIRFPITA IASILHRVSG VITFVAVGIL LWLLGTSLSS PEGFEQASAI
Cbu e      1                       MNAKRPVNLD LTKFHFPPMA ILSIGHRISG FVLFLCMPLM FYLLHRATAS AESFYHLHQL
Rpr e      1                MTKIKQEI YN-KRPTSPH LTIYKPQISS TLSILHRMTG VAL---FFVV SILV------ -WWLILSKYD
Ram m      1      MISINF NFLKIKGIIN MNINRPISPH LTIYKLQITN TLSIFHRITG GVLALTLCFF ILILKMLNFH LSSYAFYSIA
Ppu m      1                     M YNINRPISPH LTIYNTQKSS LFSIWHRISG VAMFTLIASP PLFLKLATFS YKSFNILDLM
Mpo m      1                       MKINRPLSPH LTIYKPQLTS TFSIFHRISG AFLATMVLFS ILFFKIGDLS LTFYHFYQYF
Sce n [1] 51  NVASEMNTKA AIAEEQILNK QRAKRPISPH LTIYQPQLTW YLSSLHRISL VLMGLGFYLF TILFGVSGLL GLGLTTEKVS
Cel n     28  TKSEAKTPIQ KFGWEYLLKQ RSKNRPIAPH LTVYQPQLTW MLSGFHRISG CVMAGTLLVG GIGFAVLPFD FTAFVDFIRS
Bta n      1           ME RFWSKN---- TTLNRPLSPH ISIYGWSLPM AMSICHRGTG IALSAGVSLF GLSALLVPGS FESHLEFVKS
                                    .::.... :........   .:. ::.:. ..... . ........ .   ..

Eco e     65  MG-SFFVKFI MWGILT---- -ALAYHVVVG IRHMMMDFGY LEETFEAGKR SAKISFVITV VLSLLAGVLV W
Cbu e     61  LLHNGWIKLA VWIMLS---- -ATLFHLFAG IRHLAMDLGF WES-VPEGRI SAYTVFVVSF IAIVLAGVWI W
Rpr e     58  NNYLQLASCC IIKICLVAFS YSWCYHLCNG IRHLFWDIGY GFS-IKAVNI TGWCVVVCSI LLTMLLWV
Ram m     77  YTLN-QYSGF LFIAISFFLL LFIFYHLFAG LRHLVWDAGY ALE-IENVYL TGYIMLGLAF LFTLIAWIIF
Ppu m     62  LNNSSLILPW FIVIIS---- VIFLYHIING IRHFLWDSVV NVN-TESIIK DSNTLLALVF LIMLFKFIL
Mpo m     61  FFLT-FYLNW FIISLVNFTL LALCYHMSNG VRHLLWDLGF FLE-LSKVYT SGIIMLFCAA FLALLNIIRQ HWSNGQIPY
Sce n    131  NWYHQKFSKI TEWSIKGSFA YLFAIHYGGA IRHLIWDTAK ELT-LKGVYR TGYALIGFTA VLGTYLLTL
Cel n    108  WNLPCAVTAV FKYIIA---- FPIIFHTLNG IRFLGFDLAK GVNNVGQIYK SGYLVSGLSA ILALAIVFNS CQNKSNKTA
Bta n     69  LCLGPALIHT AKFALV---- FPLMYHTWNG IRHLMWDLGK GLT-ISQLHQ SGVAVLVLTV LSSVGLAAM
                .  .       ...      .. .:   : ::....:... .  .  .   .:. .. ..  .....  ..
```

FIG. 3. Alignment of SDH3 protein sequences. With the exception of Rpr (*Rickettsia prowazekii*), Mpo (*Marchantia polymorpha*), and Bta (*Bos taurus*), abbreviations for organism names are as in Fig. 1. In the case of yeast nucleus-encoded SDH3 (Sce n), codon 51 corresponds to the known N terminus of the mature protein (indicated by [1] in the figure). In columns where >50% of the residues are the same, these identical residues are emphasized by shading. Columns in which all of the residues are chemically similar are denoted by a colon, whereas those in which >50% of the residues are of the same chemical type are marked by a dot. Sequences other than Ram and Ppu were obtained from public domain data bases, with NCBI gi: identifiers as follows: Eco, 118621; Cbu, 495749; Rpr, 409934; Mpo, 786182 (ORF137); Sce, 313814; Cel, 433177; Bta, 786511.

```
Eco e.      1                          MVSNASAL GRNGVHDFIL VRATAIVLTL YIIYMVGFFA TSGELTYEVW IGFFASA---
Cbu e       1                            MVDRT SRRGYRDWFV QRITALLSGI YAVFVIVFLL VHHPISYPQW HALFSHL---
Rpr e       1                  MIYDFKA EIIKAKNSSF SKSGSHHWLL QRVTGVILAL CSFWLIYFMF TNKNNDINII MWEFKKP---
Ram m       1                          MTEKLLHFIR TKSGSMHWWL QRFLAILLAP IILYLLFDVA IYIGQQSDPT VMMFLNRIFN
Ppu m       1                                  M YKTLLAQVFF HSIAKKKLYF FWLPRLFSLL LVPGFL---- ----------
Mpo m       1                                MKT HRETLGHWLL QRITAAFLIP TILMA----- ---------- ----------
Sce n [1] 23  LTIPFLPVLP QKPGGVRGTP NDAYVPPPEN KLEGSYHWYM EKIFALSVVP LATTAMLTTG PLSTAADS-- ----------
                                                . .  .... .........    ....  .    .

Eco e      56  FTKVFTLLAL FSILIHAWIG MWQVLTDYVK PLALRLMLQL VIVVALVVYV IYGFVVVWGV
Cbu e      53  IMKIFTLIVI FSILWHAWIG MWTIFTDYVK NKPIRLALET LVCLLLVGYF VWAIEFLWIA R
Rpr e      65  FNIVILLITV TISLYHSVLG MRVVIEDYIN CHKLRNTLII IVKLFCILTI VSFVVAIFYS G
Ram m      61  HNSIFIFITS VILIWHVRGG MEVIIEDYVH GEKTRIVSIF LIRVIAIEIM EYLYKCSIIF
Ppu m      38  FDIEILFLFH PIILLHASLG LSVIIEDYIH IETIKFQYLS LIKLLLVLLI NLNILYLL
Mpo m      29  -NVSTLILLN ILLFWHIHVG IEEILADYVH HEVTRNWILI LLRVFCLMII KYVFVFFVF
Sce n     100  -------FFS VMLLGYCYME FNSCITDYIS ERVYGVWHKY AMYMLGLGSA VSLFGIYKLE TENDGVVGLV KSLWDS....
                 . . ...   ..... .. . ....:::.    ... ..  .:.:. :. . .. .. .
```

FIG. 4. Alignment of SDH4 protein sequences. Abbreviations for organism names are as in Figs. 1 and 3. In the case of yeast nucleus-encoded SDH4 (Sce n), codon 23 of its gene corresponds to the known N terminus of the mature SDH4 protein (indicated by [1] in the figure). Criteria for shading of residues and marking of columns (: or .) are as in Fig. 3. Sequences other than Ram and Ppu were obtained from public domain data bases, with NCBI gi: identifiers as follows: Eco, 118622; Cbu, 495749; Rpr, 729337; Mpo, 786182 (ORF86a); Sce, 585976.

anchor protein (32). As expected for membrane anchor subunits, hydrophobic residues predominate in *P. purpurea* and *R. americana* SDH4 (63% and 55%, respectively), and there is the potential to form three transmembrane domains (30). Biochemical studies and disruption of yeast *SDH4* have shown that the encoded protein is involved in attachment of the iron-sulfur and flavoprotein subunits of SDH to the inner mitochondrial membrane in the formation of complex II (30).

Our analyses further indicate that bacterial *sdhD*, yeast nuclear *SDH4*, and *M. polymorpha* (liverwort) mitochondrial *orf86a* code for homologous proteins (see below), a relationship that has not been recognized previously. Originally, *orf86a* was grouped as one of the 31 unassigned ORFs identified in *M. polymorpha* mtDNA, whereas bacterial *sdhD* and yeast *SDH4* have been considered to be functional analogs only. Our discovery of *orf86a*-like sequences in a number of protist mtDNAs suggested that this ORF is a conserved gene, and prompted us to perform a thorough search for similar proteins in sequence data bases. Several lines of evidence support the conclusion that bacterial *sdhD*, yeast *SDH4*, and liverwort *orf86a* are homologous.

(*i*) Two blocks of significant sequence similarity are found in the N- and C-terminal regions of all SDH4 sequences. These similarities can be detected by BLAST data base searches, especially between ORF86a and bacterial SDHD sequences. The conserved blocks become more clearly apparent in multiple alignments that include several representatives of this protein and that incorporate less highly diverged mitochondrial SDH4 sequences from protists (Fig. 4).

(*ii*) In all studied bacteria, *sdhD* is located immediately downstream of *sdhC*, the same order in which the homologous genes (*sdh4* and *sdh3*, respectively) are found in the mtDNAs of *R. americana* and the cryptomonad alga *Rhodomonas salina* (unpublished results).

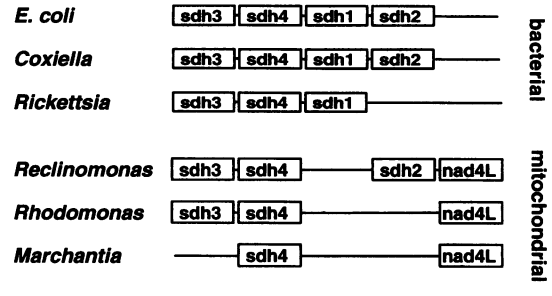

FIG. 5. Organization of *sdh* genes in bacterial and mitochondrial genomes (figure not drawn to scale). Genes (denoted by rectangles) are transcribed from left to right in all cases and are separated by intergenic spacers (horizontal lines connecting rectangles). *nad4L*, gene encoding subunit 4L of the NADH dehydrogenase of respiratory complex I.

(*iii*) In eubacteria and eukaryotes, the three largest subunits of complex II are each clearly homologous at the level of amino acid sequence among eubacteria and eukaryotes. In all cases, the fourth and smallest subunit is a hydrophobic protein, is of similar size among the compared species, and is the least conserved member within both the eubacterial and mitochondrial classes. For example, *E. coli* and *C. burnetii* SDHA share 61% sequence identity, as do the SDHB proteins of these organisms; in contrast, the SDHD proteins from these two proteobacteria display only 37% identity (38).

The selective loss of single *sdh* genes from a presumed eubacterial *sdh* gene cluster, *sdhC(3)–sdhD(4)–sdhA(1)–sdhB(2)*, is evident when comparing the order of *sdh* genes in various bacterial and mitochondrial genomes (Fig. 5). The eubacterial order *sdh3–sdh4* occurs in the mtDNA of *R. americana* and *R. salina* (unpublished results), indicating a common origin for this cluster. In addition, *nad4L* is found immediately downstream of the *sdh* gene cluster in these two cases, as well as immediately downstream of *sdh4* in liverwort (*M. polymorpha*) mtDNA, consistent with the idea of a single origin of the protomitochondrial genome.

Although *sdh* genes are dispersed in the mitochondrial genome of *P. purpurea*, it is notable that the same three genes (*sdh2*, *sdh3*, and *sdh4*) are present in the mtDNA of both *P. purpurea* and *R. americana*, two phylogenetically distant eukaryotes. Taken together with the results of comparisons of the derived SDH2 protein sequences, this observation adds to the recent molecular evidence supporting the view that mitochondria constitute a monophyletic lineage (5, 6). Although the branching position of the red algae within the eukaryotic lineage is still controversial, the available data suggest that these organisms constitute a relatively late diverging clade that branches close to the terminal eukaryotic clades that include animals, fungi, and land plants (50). In fact, recent analyses of nuclear (51) and particularly mitochondrial (ref. 52; B.F.L., unpublished results) protein-coding genes strongly suggest that red algae constitute a sister group to the green algal/land plant clade. In contrast, ultrastructural data indicate that *R. americana* and other jakobid flagellates are early diverging protists that have many features in common with amitochondriate retortamonads such as *Giardia* (41). Considering that *sdh* genes are absent from many other mtDNAs that have been completely sequenced, their presence in the mitochondrial genomes of *P. purpurea* and *R. americana* likely represents a shared character. The simplest explanation of our data is that a single, protomitochondrial genome originally contained *sdh* genes in addition to the other respiratory chain genes and protein synthesis genes commonly found in mtDNA and that these *sdh* genes were subsequently transferred to the nucleus in most but not all eukaryotes. Moreover, the selective retention of the same three *sdh* genes in the phylogenetically distant lineages leading to *Reclinomonas* on the one hand and *Porphyra* on the other suggests that multiple, independent mito-

chondrion-to-nucleus transfers of *sdh2*, *sdh3*, and *sdh4* must have occurred in the course of mitochondrial genome evolution. This also appears to have been the case for other respiratory chain genes encoded in mtDNA, such as *cox2*, *cox3*, and various *nad* genes (5).

Genes specifying components of the tricarboxylic acid cycle (of which the iron-sulfur protein encoded by *sdh2* is an example) have not previously been found in mtDNA. In view of the results reported here, it seems probable that sequencing of additional mitochondrial genomes (particularly protist ones) will reveal a wider distribution of *sdh* genes in mtDNA. In fact, sequence data obtained recently indicate the presence of the same three *sdh* homologs in the mitochondrial genome of *R. salina* (G.B. and B.F.L., unpublished results) and of both *sdh2* and *sdh3* in the mtDNA of *Cyanidium caldarium* (S. Viehmann; GenBank accession no. Z48930) and *Chondrus crispus* (53), two other red algae. The latter genome is also reported to encode a homolog of *M. polymorpha* mitochondrial *orf86a*, which we identify here as *sdh4*.

We are pleased to acknowledge the excellent technical assistance of Isabelle Plante and Diane Saint-Louis. This work was supported by grants from the Medical Research Council of Canada (SP-34) and the Canadian Genome Analysis and Technology Program (GO-12323). B.F.L. and M.W.G. are Fellows and G.B. is an Associate in the Program in Evolutionary Biology, Canadian Institute for Advanced Research (CIAR). B.F.L. and M.W.G. are grateful for salary and interaction support from CIAR. Sabbatical hospitality extended by J. Boynton and N. W. Gillham (DCMB, Duke University) to M.W.G. during the preparation of this manuscript is warmly acknowledged.

1. Tzagoloff, A. & Myers, A. M. (1986) *Annu. Rev. Biochem.* **55,** 249–285.
2. Attardi, G. & Schatz, G. (1988) *Annu. Rev. Cell Biol.* **4,** 289–333.
3. Gray, M. W. & Doolittle, W. F. (1982) *Microbiol. Rev.* **46,** 1–42.
4. Gray, M. W. (1989) *Trends Genet.* **5,** 294–299.
5. Gray, M. W. (1992) *Int. Rev. Cytol.* **141,** 233–357.
6. Gray, M. W. (1993) *Curr. Opin. Genet. Dev.* **3,** 884–890.
7. Nugent, J. M. & Palmer, J. D. (1991) *Cell* **66,** 473–481.
8. Covello, P. S. & Gray, M. W. (1992) *EMBO J.* **11,** 3815–3820.
9. Lang, B. F. & Forget, L. (1993) in *Genetic Maps*, ed. O'Brien, S. J. (Cold Spring Harbor Lab. Press, Plainview, NY), 6th ed., Book 3, pp. 3.133–3.135.
10. Cole, R. A. & Williams, K. L. (1994) *J. Mol. Evol.* **39,** 579–588.
11. Burger, G., Plante, I., Lonergan, K. M. & Gray, M. W. (1995) *J. Mol. Biol.* **245,** 522–537.
12. Hatefi, Y., Galante, Y. M., Stiggal, D. L. & Ragan, C. I. (1979) *Methods Enzymol.* **56,** 577–602.
13. Hatefi, Y. (1985) *Annu. Rev. Biochem.* **54,** 1015–1069.
14. Singer, T. P. & Johnson, M. K. (1985) *FEBS Lett.* **190,** 189–198.
15. Kita, K., Oya, H., Gennis, R. B., Ackrell, B. A. & Kasahara, M. (1990) *Biochem. Biophys. Res. Commun.* **166,** 101–108.
16. Lombardo, A., Carine, K. & Scheffler, I. E. (1990) *J. Biol. Chem.* **265,** 10419–10423.
17. Keon, J. P., White, G. A. & Hargreaves, J. A. (1991) *Curr. Genet.* **19,** 475–481.
18. Broomfield, P. L. & Hargreaves, J. A. (1992) *Curr. Genet.* **22,** 117–121.
19. Kita, K., Mizuchi, D., Wang, H., Takamiya, S., Aoki, T. & Kojima, S. (1992) *Electrophoresis* **13,** 506–511.
20. Birch-Machin, M. A., Farnsworth, L., Ackrell, B. A., Cochran, B., Jackson, S., Bindoff, L. A., Aitken, A., Diamond, A. G. & Turnbull, D. M. (1992) *J. Biol. Chem.* **267,** 11553–11558.
21. Yu, L., Wei, Y. Y., Usui, H. & Yu, C. A. (1992) *J. Biol. Chem.* **267,** 24508–24515.
22. Sulston, J., Du, Z., Thomas, K., Wilson, R., Hillier, L., Staden, R., Halloran, N., Green, P., Thierry-Mieg, J., Qiu, L., Dear, S., Coulson, A., Crazton, M., Durbin, R., Berks, M., Metzstein, M., Hawkins, T., Ainscough, R. & Waterston, R. (1992) *Nature (London)* **356,** 37–41.
23. Chapman, K. B., Solomon, S. D. & Boeke, J. D. (1992) *Gene* **118,** 131–136.
24. Robinson, K. M. & Lemire, B. D. (1992) *J. Biol. Chem.* **267,** 10101–10107.
25. Schülcke, N., Blobel, G. & Pain, D. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 8011–8015.
26. Hirawake, H., Wang, H., Kuramochi, T., Kojima, S. & Kita, K. (1994) *J. Biochem. (Tokyo)* **116,** 221–227.
27. Morris, A. A., Farnsworth, L., Ackrell, B. A., Turnbull, D. M. & Birch-Machin, M. A. (1994) *Biochim. Biophys. Acta* **1185,** 125–128.
28. Daignan-Fornier, B., Valens, M., Lemire, B. D. & Bolotin-Fukuhara, M. (1994) *J. Biol. Chem.* **269,** 15469–15472.
29. Abraham, P. R., Mulder, A., van't Riet, J. & Raué, H. A. (1994) *Mol. Gen. Genet.* **242,** 708–716.
30. Bullis, B. L. & Lemire, B. D. (1994) *J. Biol. Chem.* **269,** 6543–6549.
31. Darlison, M. G. & Guest, J. R. (1984) *Biochem. J.* **223,** 507–517.
32. Wood, D., Darlison, M. G., Wilde, R. J. & Guest, J. R. (1984) *Biochem. J.* **222,** 519–534.
33. Paudel, H. K., Yu, L. & Yu, C. A. (1991) *Biochim. Biophys. Acta* **1056,** 159–165.
34. Lee, G. Y., He, D. Y., Yu, L. & Yu, C. A. (1995) *J. Biol. Chem.* **270,** 6193–6198.
35. Magnusson, K., Philips, M. K., Guest, J. R. & Rutberg, L. (1986) *J. Bacteriol.* **166,** 1067–1071.
36. Phillips, M. K., Hederstedt, L., Hasnain, S., Rutberg, L. & Guest, J. R. (1987) *J. Bacteriol.* **169,** 864–873.
37. Aliabadi, Z., Winkler, H. H. & Wood, D. O. (1993) *Gene* **133,** 135–140.
38. Heinzen, R. A., Mo, Y. Y., Robertson, S. J. & Mallavia, L. P. (1995) *Gene* **155,** 27–34.
39. Bach, M., Reilander, H., Gartner, P., Lottspeich, F. & Michel, H. (1993) *Biochim. Biophys. Acta* **1174,** 103–107.
40. Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., Kohchi, Y. & Ohyama, K. (1992) *J. Mol. Biol.* **223,** 1–7.
41. O'Kelly, C. J. (1993) *J. Euk. Microbiol.* **40,** 627–636.
42. Higgins, D. G., Bleasby, A. J. & Fuchs, R. (1992) *Comput. Appl. Biosci.* **8,** 189–191.
43. Schwartz, R. M. & Dayhoff, M. O. (1978) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (Natl. Biomed. Res. Found., Washington, DC), Vol. 5, Suppl. 3, pp. 353–358.
44. Fitch, W. M. & Margoliash, E. (1967) *Science* **155,** 279–284.
45. Felsenstein, J. (1993) PHYLIP, Phylogeny Inference Package, Version 3.5c. (Dept. Genet., Univ. of Washington, Seattle).
46. Kishino, H., Miyata, T. & Hasegawa, M. (1990) *J. Mol. Evol.* **31,** 151–160.
47. Golding, G. B. & Gupta, R. S. (1995) *Mol. Biol. Evol.* **12,** 1–6.
48. Karlin, S. & Campbell, A. M. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 12842–12846.
49. Walker, R. T. (1983) *Yale J. Biol. Med.* **56,** 367–372.
50. Ragan, M. A. & Gutell, R. R. (1995) *Bot. J. Linn. Soc.* **118,** 81–105.
51. Zhou, Y.-H. & Ragan, M. A. (1995) *Curr. Genet.* **28,** 324–332.
52. Boyen, C., Leblanc, C., Bonnard, G., Grienenberger, J. M. & Kloareg, B. (1994) *Nucleic Acids Res.* **22,** 1400–1403.
53. Leblanc, C., Boyen, C., Richard, O., Bonnard, G., Grienenberger, J.-M. & Kloareg, B. (1995) *J. Mol. Biol.* **250,** 484–495.