Acta Biochimica Polonica

Vol. 46 No. 1/1999

155-162

QUARTERLY

A human putative Suv3-like RNA helicase is conserved between Rhodobacter and all eukaryotes*

Aleksandra Dmochowska[1,2], Katarzyna Kalita[1], Michał Krawczyk[1], Paweł Golik[1], Katarzyna Mroczek[1], Jaga Lazowska[3], Piotr P. Stępień[1,2], Ewa Bartnik[1,2]

[1]*Department of Genetics, University of Warsaw,* [2]*Institute of Biochemistry and Biophysics, Polish Academy of Sciences, A. Pawińskiego 5a, 02-106 Warsaw, Poland, and* [3]*Centre de Genetique Moleculaire du CNRS, F-91118 Gif sur Yvette, France*

Received: 2 February, 1999

Key words: human, *S. cerevisiae*, RNA helicase, SUPV3L1, *SUV3*

We have cloned and sequenced a cDNA of the human homologue of the *Saccharomyces cerevisiae* Suv3 putative RNA helicase which is indispensable for mitochondrial function in yeast. The human Suv-3-like protein has a typical mitochondrial leader sequence. Northern blot data and analysis of ESTs in the data banks indicate that this human gene (SUPV3L1) is expressed in practically all tissues, though at different levels. Sequence homology analysis has shown a strong conservation of the protein in a number of eukaryotic organisms – plants, mammals and fungi, but no close homologues exist in bacteria with the exception of the purple bacterium *Rhodobacter sphaeroides*. This gene is thus ubiquitously present in all eukaryotic organisms.

Studying the structure and function of human genes is a major challenge for molecular biology. In many cases finding homologues to a given human gene in other organisms may offer help in this research. For example the complete genome of the yeast *Saccharomyces cerevisiae* has been sequenced and due to the ease of performing genetic and molecular analysis in yeast, it has become increasingly popular to study the homologues of human genes in this organism, especially genes associated with human diseases [1]. Spectacular success had been obtained analyzing the Friederich's ataxia (frataxin) [2] and Werner-syndrome gene homologues in *S. cerevisiae* [3].

In the last decade diseases affecting mitochondrial function have become an increasingly popular area of research [4]. Numerous mutations in mitochondrial DNA responsible

*This work was supported by Grant 6P04A02612 from the State Committee for Scientific Research.
✉Correspondence to: E. Bartnik, fax: (48 22) 658 4636; e-mail: ebartnik@ibb.waw.pl
Abbreviations: aa, amino acid.

for these diseases have been discovered. However, many diseases which affect mitochondrial function are due to mutations in the nuclear DNA; and so far very few of them have been mapped and little is known about the genes which are involved. On the other hand, there are literally hundreds of nuclear genes in baker's yeast *S. cerevisiae* in which mutations affect mitochondria – mainly by causing respiratory incompetence in these unicellular organisms. As mitochondrial function is very similar in all eukaryotes, we were interested in looking for human homologues of *S. cerevisiae* genes involved in mitochondrial functions.

For some years we have been analyzing a number of *S. cerevisiae* genes, essentially involved in the processing and stability of RNA in mitochondria. The inactivation of one such gene, *SUV3*, encoding a putative RNA helicase, causes respiratory incompetence in *S. cerevisiae* [5]. The product of this gene is part of a three-protein complex responsible for processing and turnover of mitochondrial RNA. It has been postulated that this complex, called mtExo, has 3′→5′ exoribonuclease activity [6, 7]. A partial human cDNA with homology to *SUV3* was found during a systematic analysis of EST sequences [8]. In the present work we have isolated a cDNA clone of the human *SUV3* homologue and examined its sequence, expression in several human tissues and also analyzed the homology of the encoded protein to other similar sequences in the data base.

MATERIALS AND METHODS

A cDNA clone derived from EST analysis [8] was purchased from Genome Systems Inc. (St. Louis, MO, U.S.A.). A human cDNA library constructed from HeLa D98-H2 cells in plasmid pFL61 [9] was kindly provided by Dr. M. Minet. Northern blots of polyA$^+$ mRNA were obtained from Clontech (U.S.A.).

All molecular biology methods were standard, as described in [10]. Sequencing was performed using a Thermo Sequenase™ cycle sequencing kit (Amersham) according to the manufacturer's instructions.

The computer programmes used for sequence analysis were the optimized BLAST 2.0 program suite [11]. The algorithms used were BLASTP and TBLASTN. The protein sequences were aligned with the CLUSTAL W program [12] using the BLOSUM series scoring matrices [13] and the PAM250 matrix [14]. The BESTFIT program from the GCG package (Wisconsin Package Version 9.1, Genetics Computer Group (GCG), Madison, WI, U.S.A.) has been used for pairwise alignments. The Z-scores were calculated from the BESTFIT alignments as described by Landes *et al.* [15].

RESULTS AND DISCUSSION

A cDNA clone containing part of the human SUV3 homologue derived from EST analysis [8] was sequenced. The encoded protein was found to correspond to part of the C-terminal region of Suv3p [5] (Acc. no. P32580) and the *Caenorhabditis elegans* ORF C08F8.2 (Acc. no. 1321758), identified through the systematic sequencing of the *C. elegans* genome [16] on the basis of homology to the yeast Suv3p. The partial sequence was used to design primers for screening a human HeLa cDNA library in order to find a full length clone. Pools of *E. coli* carrying recombinant plasmids were screened for the presence the cDNA for the Suv3p homologue by PCR. Initially plasmid p17 was isolated, containing a 2.2 kb insert. The sequence of the human cDNA (2136 base pairs) was established. Comparison with *S. cerevisiae* and *C. elegans* sequences indicated that at least 200–300 bp of the 5′ end of the cDNA were missing, therefore the same library was analyzed again. Plasmid pKK was thus obtained, consisting of a 2.5 kb insert

within the vector pFL61. The sequence of the 5′ end of the cDNA was determined revealing the presence of an AUG codon at the beginning of an open reading frame coding for 786 amino acids. The human SUPV3L1 coding sequence has been deposited in GenBank under accession No. AF042169.

Comparison of the human SUPV3L1 protein with *S. cerevisiae* Suv3p revealed (Fig. 1) that all major blocks of sequences characteristic for helicases found in the yeast protein [5] are present in the human sequence. Analysis of the potential mitochondrial leader sequence by the method of the helical wheel indicates that the human SUPV3L1 fits this model better than the yeast *SUV3* gene product. On the basis of the PSORT (http://psort.ibb.ac.jp/) program the mitochondrial location of this gene product was predicted with the probability of 76%, which is high, even higher than for the yeast Suv3 protein (52%) which is known to be mitochondrial.

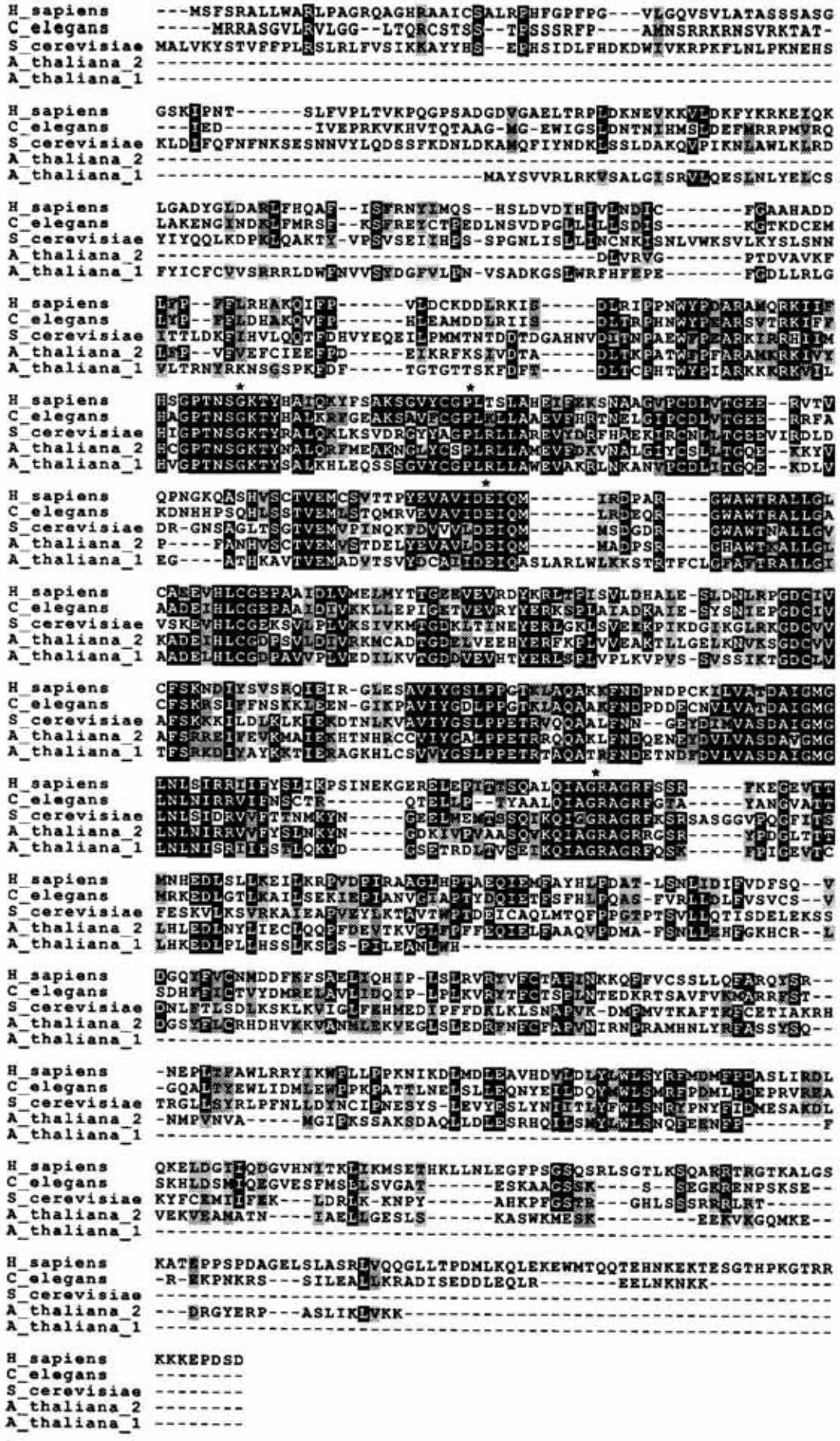
We have analyzed the expression of this gene by performing hybridization to Northern blots of polyA$^+$ mRNA using the cloned pKK cDNA as a probe. As can be seen (Fig. 2) the SUPV3L1 mRNA is expressed in all analyzed tissues, and the levels relative to actin vary several-fold. The lowest observed expression is in the lung, relatively high expression in relation to actin is seen in the liver, pancreas and kidney. The relative expression of the gene in skeletal muscle and the heart is probably underestimated, due to high expression of the actin standard in these tissues.

The predicted amino-acid sequence of the human SUPV3L1 protein (Acc. no. 2801555) has been used screen the public sequence databases (NCBI site).

Among the sequence homologues identified we have found as expected, the yeast Suv3 protein with the Z-score of 30, a value that is considered highly significant, and the *C. elegans* ORF C08F8.2. The Z-score for human and *C. elegans* SUV3 proteins is 78, making the *C. elegans* protein the nearest known orthologue of human SUV3.

Two other sequences showing significant homology to the human protein come from *Arabidopsis thaliana* and are both derived from the systematic sequencing of its genome. The first one is a protein encoded by a gene located on chromosome IV (Acc. no. 2244836) and is shown in Fig. 1 as A_thaliana_1. The sequence is only 442 amino acids long and lacks the C-terminal portion found in other Suv3 proteins (analysis of the DNA sequence in that region excludes the possibility of sequence truncation due to sequencing and/or annotating error). Within the highly conserved DEIQ box region two insertions, 6 and 4 aa long, are present. The second *A. thaliana* sequence is localized in the segment of DNA sequence of the P1 clone MYH19 from chromosome V (Acc. no. AB010077). The reading frame coding for the putative Suv3p orthologue seems to be incomplete, as no ATG start codon is present. The alignment (sequence A_thaliana_2 in Fig. 1) indicates that the N-terminal part of the protein is missing, but the C-terminal portion is similar in length and sequence to other Suv3 proteins identified. The whole MYH19 DNA sequence has already been analyzed [17], but the frame coding for the Suv3p orthologue has not been annotated as a potential gene, probably due to fact that the 5′ region of the gene has not yet been sequenced. The open reading frame codes for 553 amino acids with a stronger homology to human Suv3p than that found for the first *A. thaliana* sequence (Z-score 47 as opposed to 39). It is, however, unclear, whether this sequence is a functional gene, or a pseudogene. The latter seems more probable, as the sequence contains a continuous open reading frame, without any introns, which is uncommon for functional genes.

A maize sequence was also found (X15406) and it is a classical pseudogene, composed of several short ORF fragments interrupted by several frameshifts. It is interesting that it is localized close to another pseudogene, namely the glyceraldehyde-3-phosphate dehydrogenase 1 (GPA1) pseudogene [18].


H_sapiens
C_elegans
S_cerevisiae
A_thaliana_2
A_thaliana_1

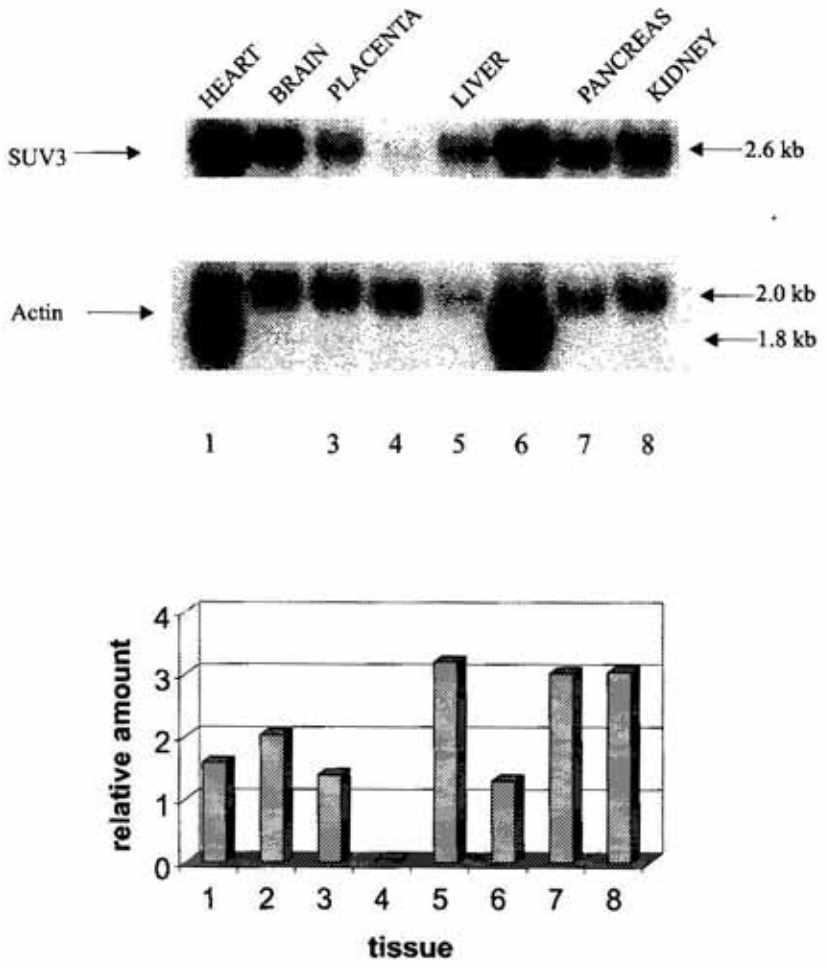


Figure 2. Expression pattern of the human SUPV3L1 gene in adult human tissues.

A multiple tissue blot (Clontech) containing polyA$^+$ mRNAs from the indicated tissues was subjected to a Northern blot analysis using either the ^{32}P-labelled pKK plasmid probe (suv) or the β-actin cDNA probe (actin). Molecular size markers are indicated at the right of the blots. The relative intensities of hybridized signals quantitated using a phosphoimager are shown in the bottom panel.

We have also found numerous short sequence segments, either sequence tags (EST and STS sequences) or putative pseudogene proteins that show significant sequence homology to the human SUPV3L1. They are presented in Table 1 and were found for numerous human and mouse tissues, as well as for some *Drosophila melanogaster* tissues. The mammalian EST sequences (human, all apparently derived from the coding sequence that we have identified, and mouse and rat) come from a wide variety of tissues and sources, indicating that the SUPV3L1 gene is expressed in most cells to some degree. This is in agreement with our Nothern blots, showing the presence of the SUPV3L1 mRNA in almost all tissues analysed.

As can be clearly seen in Fig. 1 the sequences of Suv3-like proteins share considerable homology among each other, even

Figure 1. Alignment of human, *C. elegans*, *A. thaliana* and *S. cerevisiae* Suv3-like protein sequences.

The alignment has been shaded with the BOXSHADE program. Black denotes amino-acid identity, grey – homology. * – sequences characteristic for DEAD box helicases.

Table 1. List of analyzed EST sequences

Species	Acc. no.	Source	Length
Homo sapiens	AA600836	breast tumor	600 bp
	AA046407	uterus	514 bp
	AA837797	tonsils/ B cells	496 bp
	H49424	brain	465 bp
	AA910260	kidney	413 bp
	AA654266	prostate	388 bp
	H50384	brain	407 bp
	R18471	infant brain	439 bp
	T68413	liver	516 bp
	AA878814	liver	340 bp
	H40173	brain	297 bp
	H88208	cochlea	420 bp
	AA334218	embryo	268 bp
	R41323	brain	294 bp
	AA379177	skin tumor	239 bp
	AA338026	endometrial tumor	370 bp
	H40108	brain	412 bp
	T68479	liver	272 bp
	H88209	cochlea	241 bp
	AA056167	uterus	145 bp
	AA903289	soft tissue	160 bp
Mus musculus	AA116572	thymus	657 bp
	AA726682	skin	535 bp
	AA269371	fetus	526 bp
	W63890	embryo	449 bp
	AA986335	kidney	646 bp
	AA989832	kidney	430 bp
	AA763608	mammary gland	268 bp
	AA175743	spleen	352 bp
	W91752	thymus	292 bp
	C80184	blastocyst	511 bp
	AA119015	thymus	684 bp
Rattus sp.	AA799741	heart	553 bp
	H32333	PC12 cells	272 bp
Drosophila melanogaster	AA696188	ovary	761 bp
	AA941012	embryo	674 bp
	AA698735	head	418 bp

though they come from distantly related organisms. They all contain the ATPase motif A GXXXXGKT (as GPTNSGKTY), the ATPase motif B in the form of DEIQ (the DE is always very highly conserved in RNA helicases and many other proteins) and a variant of the YIHRIGRXXR box (QIA/GGRAGR), a putative RNA binding domain which is unique to RNA helicases, and all lack another characteristic RNA helicase sequence, SAT [19]. Other long blocks of homology, VAT/SDAI/VGMGLNL being the longest, seem to be characteristic for this group of proteins. It should be underlined that the homology among this group is obviously much higher than the homology to other RNA and DNA helicases of the so-called "DEAD-box" family. The closest other helicases identified by BLAST (like the DNA2 helicase, which interestingly localizes on chromosome 10 in the neighborhood of SUPV3L1 in the human genome [20, 21] or the Ski2 helicase of yeast) share only short stretches of homology in the helicase signature regions and the Z-scores for their alignments with any of the Suv3-like proteins are close to 0 (data not shown). It seems therefore that the Suv3-like proteins form a distinct and highly conserved family, common to all eukaryotic organisms.

Another sequence which is very interesting from the evolutionary point of view has been identified by BLAST in the purple bacterium *Rhodobacter sphaeroides*. This gene, mgpS (Acc. no. Z50182) is involved in the transcriptional regulation of two operons [22] and shows significant homology to SUPV3L1 over a portion of its sequence (Fig. 3). The homology is limited to a few stretches, but there are thirteen consecutive identical amino acids in a domain which appears to be characteristic for this group of proteins (LVATDAIGMGLNL) and eight consecutive identical amino acids in the GRXXR domain (QIAGRAGR). The DEIQ variant of the DEAD box which is found in all Suv3-like proteins is also conserved. mgpS Is the only Suv3p homologue found in Prokaryotes, as the search of the

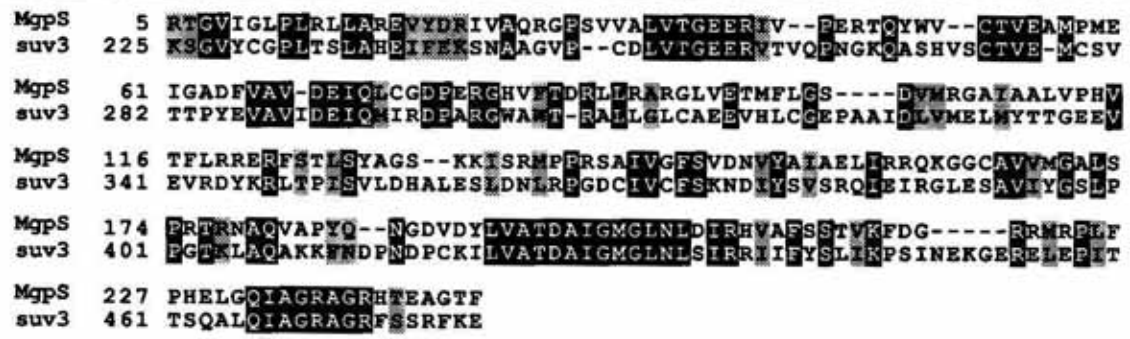


Figure 3. Alignment of partial human SUV3 and a homologous *Rhodobacter sphaeroides* sequences.

The alignment has been shaded with the BOXSHADE program. Black denotes amino-acid identity, grey – homology.

known bacterial and archaebacterial sequences, including the fully sequenced genomes, has given no other potential Suv3-like protein. DEAD-box helicases are common in bacteria, but outside the short "signature" sequence stretches show no homology to Suv3. It seems therefore that the Suv3 proteins are particular to Eukaryotes, and are highly conserved among all the eukaryotic taxa. It interesting to note that α-purple bacteria, like *R. sphaeroides* are believed to be the direct evolutionary ancestors of the mitochondria of modern Eukaryotes.

A.D. would like to thank Dr. M. Minet for her encouragement and interest.

REFERENCES

1. Bassett, D.E., Jr., Boguski, M.S. & Hieter, P. (1997) Yeast genes and human disease. *Nature* **379,** 589–590.

2. Koutnikova, H., Campuzano, V., Foury, F., Dolle, P., Cazzalini, O. & Kooening, M. (1997) Studies of human, mouse and yeast homologues indicate a mitochondrial function for frataxin. *Natue Genetics* **16,** 345–351.

3. Sinclair, D.A., Mills, K. & Guarente, L. (1997) Accelerated aging and nucleolar fragmentation in yeast sgs1 mutants. *Science* **277,** 1313–1316.

4. Wallace, D.C. (1995) Mitochondrial DNA Variation in Human Evolution. Degenerative Disease and Aging. *Am. J. Hum. Genet.* **57,** 201–223.

5. Stepien, P.P., Margossian, S.P., Landsman, D. & Butow, R. (1992) The yeast nuclear gene suv3 affecting mitochondrial post-transcriptional processes encodes a putative ATP-dependent RNA helicase. *Proc. Natl. Acad. Sci. U.S.A.* **89,** 6813–6817.

6. Dmochowska, A., Golik, P. & Stepien, P.P. (1995) The novel nuclear gene DSS-1 of *Saccharomyces cerevisiae* is necessary for mitochondrial biogenesis. *Curr. Genet.* **28,** 108–112.

7. Margossian, S.P., Li, H., Zassenhaus, H.P. & Butow, R.A. (1996) The DexH box protein Suv3p is a component of a yeast mitochondrial 3'-to-5' exoribonuclease that suppresses group I intron toxicity. *Cell* **84,** 199–209.

8. Adams, M.D. *et al.* (1995) Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* **377** (Suppl.) 3–174.

9. Minet, M. & Lacroute, F. (1990) Cloning and sequencing of human cDNA coding for multifunctional polypeptide of the purine pathway by complementation of the ade2-101 mutant in *Saccharomyces cerevisiae*. *Curr. Genet.* **18,** 287–291.

10. Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory Press.

11. Altschul, S.F., Madden, T.F., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402.

12. Thompson, J.D., Higgins, D.H. & Gibson, T.J. (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673–4680.

13. Henikoff, S. & Henikoff, J.G. (1993) Performance evaluation of amino acid substitution matrices. *Proteins* **17**, 49–61.

14. Dayhoff, M.O. Barker, W.C. & Hunt, L.T. (1983) Establishing homologies in protein sequences. *Methods Enzymol.* **91,** 524–545.

15. Landes, C., Henault, A. & Risler, J.-L. (1992) A comparison of several similarity indices used in the classification of protein sequences: A multivariate analysis. *Nucleic Acids Res.* **20**, 3631–3637.

16. Wilson, R. *et al.* (1994) 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans. Nature* **368,** 32–38.

17. Sato, S., Kaneko, T., Kotani, H., Nakamura, Y., Asamizu, E., Miyajima, N. & Tabata, S. (1998) Structural analysis of *Arabidopsis thaliana* Chromosome 5. IV. Sequence features of the regions of 1,456,315 bp covered by nineteen physically assigned P1 and TAC clones. *DNA Res.* **5**, 1–54.

18. Quigley, F., Brinkmann, H., Martin, W.F. & Cerff, R. (1989) Strong functional GC pressure in a light-regulated maize gene encoding subunit GAPA of chloroplast glyceraldehyde-3-phosphate dehydrogenase: Implications for the evolution of GAPA pseudogenes. *J. Mol. Evol.* **29**, 412–421.

19. Lueking, A., Stahl, U. & Schmidt, U. (1998) The protein family of RNA helicases. *Crit. Rev. Biochem. Mol. Biol.* **33**, 259–266.

20. Eki, T., Okumura, K., Shiratori, A., Abe, M., Nogami, M., Taguchi, H., Shibata, T., Murakami, Y. & Hanaoka, F. (1996) Assignment of the closest human homologue of the yeast Dna2 helicase gene to chromosome band 10q21.3-q22.1. *Genomics* **37,** 408–410.

21. Dmochowska, A., Stankiewicz, P., Golik, P., Stepien, P.P., Bocian, E., Hansmann, I. & Bartnik, E. (1998) Assignment of SUPV3L1 to human chromosome band 10q22.1 by *in situ* hybridization. *Cytogen. Cell Genet.* **83**, 84–85.

22. Sabaty, M. & Kaplan, S. (1996) mgpS, a complex regulatory locus involved in the transcriptional control of the puc and puf operons in *Rhodobacter sphaeroides* 2.4.1. *J. Bacteriol.* **178**, 35–45.