

Screening Human EST Database for Identification of Candidate Genes in Respiratory Chain Deficiency

Agnès Rötig, Isabelle Valnot, Claude Mugnier, Pierre Rustin, and Arnold Munnich

INSERM U393, Hôpital Necker-Enfants Malades, 149 rue de Sèvres, 75015 Paris, France

Received January 10, 2000

Disorders of mitochondrial oxidative phosphorylation (OXPHOS) are now recognized as major causes of human metabolic diseases and several mutations of mitochondrial and nuclear genes encoding respiratory chain components have been reported. Interestingly, mutations of nuclear genes involved in mitochondrial respiratory chain assembly, protein trafficking, and iron metabolism are also known to alter oxidative phosphorylation. While several hundred of these genes have been described in yeast, only a few nuclear genes have been hitherto identified in humans. Yeast gene databases present therefore an invaluable tool for identification of human homologues that should be regarded as candidate genes in OXPHOS diseases. In an attempt to identify the human counterparts of yeast genes, we developed a systematic comparison of yeast protein sequences to the GenBank dbEST database. Starting from 340 yeast protein sequences as templates, we searched the human dbEST counterparts using the BLAST similarity searching program and identified 102 groups of human EST likely to represent orthologues of yeast genes because of significant homology. This collection of human genes possibly related to mitochondrial OXPHOS may help identify nuclear genes responsible of mitochondrial disorders. © 2000 Academic Press

Key Words: human EST; mitochondrial diseases; candidate genes; yeast; oxidative phosphorylation.

Mitochondrial oxidative phosphorylation (OXPHOS) disorders represent an increasingly important group of metabolic diseases in humans. Mutations were originally reported in mitochondrial DNA (Mitomap, <http://www.gen.emory.edu/mitomap.html>) but nuclear gene mutations have been re-

cently shown to account for mitochondrial disorders as well. Indeed, mutations have been described in nuclear genes encoding respiratory chain complex subunits (1–5) and occasionally mitochondrial chaperon proteins (paraplegin, 6), respiratory chain assembly (SURF1, 7,8), traffic proteins (Tim8, 9), and iron metabolism (frataxin, 10–12). It appears therefore that a broad variety of genes and functions can be involved in mitochondrial respiratory chain dysfunction. To date however, only a few human genes involved in respiratory chain assembly, protein trafficking, or mtDNA maintenance have been identified.

The search for the molecular bases of human OXPHOS diseases can theoretically take advantage of the vast resource represented by yeast genetics. Indeed, a large number of yeast nuclear genes have been reported to cause a *petite* phenotype when mutated, demonstrating their roles in mitochondrial functions. In addition, the systematic sequencing of the yeast genome has helped to identify a large number of ORFs, products of which are predictably targeted to the mitochondria. Homologies of protein sequences and protein functions, between yeast and human have been used to identify and to clone several human genes (13,14). In addition, yeast mutants represent a valuable tool to understand the pathogenicity of gene mutations, as recently shown in Friedreich ataxia (10,11).

Similarly, identification and mapping of human genes by cross-species comparisons have been successfully achieved by using sequence homology between *Drosophila* genes and human ESTs (15). Combining data derived from yeast mitochondrial genetics and the screening of human EST databases,

we report here 102 new candidate genes for human OXPHOS diseases.

METHODS

We searched yeast mitochondria-related protein sequences using a "text-string" option (e.g., mitochondri*) in the yeast protein database (YPD, <http://www.proteome.com/databases/YPD/index.html>), the MitBASE pilot database (<http://www3.ebi.ac.uk/Research/mitbase/mitbiog.pl>), and the MITOP database (<http://websvr.mips.biochem.mpg.de/proj/medgen/mitop/>). All these yeast protein sequences were then systematically tested as templates to screen the human GenBank dbEST database using the BLAST similarity searching program (TBLASTN option: protein sequence versus six frames nucleotide translated database sequences). The human ESTs identified were then linked to their Unigene entries (<http://www.ncbi.nlm.nih.gov/UniGene/index.html>), allowing us to define their position with respect to genetic markers. The chromosomal mapping was then established by consulting the OMIM database (<http://www3.ncbi.nlm.nih.gov/80/Omim>) or the GenAtlas database (<http://www.citi2.fr/GENATLAS/>). Only ESTs with a significant homology (e value $< e^{-5}$) were considered. As for orphan ESTs, i.e., ESTs not recorded in Unigene database, only those with an even higher homology (e values $< e^{-10}$) were retained, so as to eliminate chimeric and unspecific sequences. Prediction of mitochondrial targeting sequences in proteins was performed using the MitoProt v 1.0.1 (available in the MITOP database) software program.

RESULTS

Selection of Yeast Mitochondrial Proteins

We selected 340 yeast protein sequences as "probes" to identify human orthologue genes. These yeast protein sequences originating from YPD, MitBASE, and/or MITOP databases are nuclearly encoded proteins currently known to be involved in biogenesis or assembly of respiratory chain or in mitochondrial genome expression. We also used as yeast protein probes the amino acid sequences deduced from ORFs identified during the systematic yeast genome sequencing and presenting striking similarities with known mitochondrial proteins or with a characteristic mitochondrial targeting sequence. We did not retain, however, (i) structural

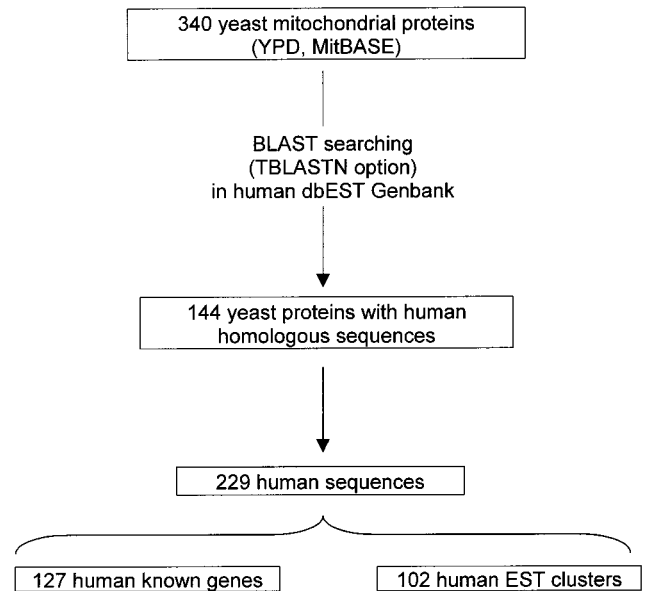


FIG. 1. Schematic diagram of the method used for human sequences retrieval through homology with yeast proteins.

proteins of the respiratory chain, as their human counterparts are already known and the corresponding genes are mapped; and (ii) proteins for which human orthologues are known to be involved in mitochondrial disorders, namely, YFH1, Tim8, and SHY1 (frataxin, DDP1, SURF1 in human) in Friedreich ataxia, dystonia–deafness syndrome, and Leigh syndrome, respectively.

Putative Human Homologues to Yeast Mitochondrial Proteins

The BLAST similarity searching program allowed us to identify human homologues for 144 of the 340 selected yeast proteins or 42.3% (Fig. 1). Some of the genes identified by this approach are the known human counterparts of yeast proteins. For example, BLAST searching using yeast adenylate kinase (ADK2, Table 3) gave the maximum score for the human adenylate kinase (AK) gene. The same is true for a number of other genes as yeast NUC1 and human endoG or yeast SUV3 and human mtRNA helicase. Taken together, 127 of the selected human sequences corresponded to known genes and their homology with the yeast counterparts reflected either functional similarity or specificity relative to mitochondrial functions. For example, the human homologues of yeast mitochondrial tRNA synthases are cytoplasmic and mitochondrial human tRNA synthases but BLAST searching using yeast mito-

TABLE 1
Human Genes and ESTs Retrieved by BLAST Homology with Yeast Carrier Proteins

YEAST Prot name	Human homolog genes				Human homolog ESTs			
	Gene symbol	Genbank accession number	Chromosomal mapping	e value	Genbank accession number	Unigene entry	Chromosomal mapping	e value
AAC1	ANT2 ANT1 ANT3	M57424 J04982 J03592	Xq24-q26 4q35 Xp22.32	5e-33 2e-31 2e-29	AI184139 AI148032 AA632137	Hs.129968 Hs.143579 Hs.181767		4e-24 1e-19 1e-17
AAC2	ANT2 ANT3 ANT1	M57424 J03592 J04982	Xq24-q26 Xp22.32 4q35	1e-41 8e-34 1e-28	AI184139 AI148032	Hs.129968 Hs.143579		2e-24 9e-22
AAC3	ANT3 ANT2 ANT1	J03592 M57424 J04982	Xp22.32 Xq24-q26 4q35	2e-41 4e-40 2e-32	AI184139 AI148032	Hs.129968 Hs.143579		2e-24 6e-21
ACR1	CTP Aralar 1 UCP2	U25147 Y14494 U94592	22q11.2-pter 11q13	2e-27 5e-08 8e-08	AA632137 AL049246 AA312705 AI377665	Hs.181767 Hs.42484 Hs.30835 Hs.183374	3q23-q24 3p14.3-3p12.3	9e-11 1e-09 6e-08 5e-07
ARG11	Aralar 1	Y14494		1e-05	AI792160			1e-10
DIC1	UCP2 oxoglu carr BMCp1	U82819 X66114 AF078544	11q13 17p13 Xq22.3-q26	3e-19 5e-17 2e-09	AI081980 AI372504	Hs.232240 Hs.237924	17q21-q22	2e-11 1e-08
FLX1					AI431745 AI200893 AI792141	Hs.42484 Hs.183047 Hs.143560	3q23-q24 1p36.32-1p36.31	6e-12 8e-10 3e-09
MIR1	PHC CTP	X60036 U25147	12q22-q23 22q11.2-pter	1e-35 1e-09				
MRS3					AI310713 AA431276 AA312705 AA632137	Hs.34401 Hs.7481 Hs.30835 Hs.181767	8p22-p21.2 1p22.1-q21 3p14.3-3p12.3	4e-22 3e-21 2e-09 7e-09
MRS4					AA234031 AI310713	Hs.7481 Hs.34401	10q24-10q23.3 8p22-8p21.2 1p22.1-q21	2e-24 4e-20
PET8					AI123566	Hs.30835	3p14.3-p12.3	5e-24
POR1	VDAC3 VDAC1 VDAC2	U90943 L06132 L06328	Xq13.1-Xq21.2	2e-19 2e-08 4e-06				
POR2	VDAC3	U90943		2e-12				
RIM2	CTP Aralar 1 ANT2 PHC	U25147 Y14494 M57424 X60036	22q11.2-pter Xq24-q26	6e-09 7e-08 7e-08 4e-07	AI431745 AA757281	Hs.42484 Hs.183047	3q23-q24 1p36.32-1p36.31	9e-30 9e-21
YBR291C	CTP UCP2	U25147 U82819	22q11.2-pter 11q13	6e-35 3e-10	AI241147 H77579	Hs.181767 Hs.7994	3p22.2-p14.2	2e-08 8e-07
YDL119C					T07400	Hs.7994	3p22.2-p14.2	2e-16
YEL006W	ANT2 ANT3	M57424 J03592	Xq24-q26 Xp22.32	3e-11 4e-10	AL049246 AI200893	Hs.42484 Hs.183047	3q23-q24 1p36.32-36.31	1e-16 3e-14
YER053C	PHC	X60036	12q22-q23	1e-38				
YFR045W	CTP UCP2	U25147 U94592	22q11.2-pter 11q13	4e-08 6e-08	AI241147	Hs.181767		2e-07
YGR096W					AT272900 AA397670	Hs.181767 Hs.16786		4e-14 2e-11
YGR257C	carrier prot	AC003083		6e-10	AA315826 AA312705	Hs.237924 Hs.30835	17q21.1 3p14.3-p12.3	3e-12 2e-07
YHR002W	ANT3 HS2OXOC	J03592 M31659	Xp22.32 10q22.1-q22.2	2e-10 3e-10	AA632137 AA809275	Hs.181767 Hs.172544		3e-25 2e-13
YIL006W	UCP2	U82819	11q13	1e-09	AL049246 AI200893	Hs.42484 Hs.183047	3q23-q24 1p36.3	4e-20 6e-12
YKL120W	BMCp1 UCP2 oxoglu carr	AF078544 U82819 X66114	11q13 17p13	1e-12 7e-12 1e-09	R27834	Hs.118918	17p13.3-17p13	6e-12
YMC1	Carnitine carrier	Y10319	3p21.31	5e-09	AI377665	Hs.183374		2e-09
YMC2	UCP2 Carnitine carrier	U82819 Y10319	11q13 3p21.31	6e-09 8e-09	AI377665	Hs.183374		2e-09
YMR166C	UCP2 Aralar 1	U82819 Y14494	11q13	3e-09 8e-08	AA312705 AA234031 T07400 AA315826	Hs.30835 Hs.7481 Hs.7994 Hs.237924	3p14.3-12.3 10q23.3-q24.1 3p22.2-p14.2 17q21.1	2e-13 8e-12 6e-10 2e-09
YNL083W					AW023500			5e-37
YOR100C	Aralar 1	Y14494		7e-07	AI377665 AA315826	Hs.183374 Hs.237924	17q21.1	2e-09 5e-07
YOR222W	CTP ANT2 ANT3 Aralar 1	U25147 J03591 J03592 Y14494	22q11.2-pter Xq24-q26 Xp22.32	5e-14 2e-10 1e-7 2e-08	AW167726			1e-15
YPL134C	CTP Aralar 1	U25147 Y14494	22q11.2-pter	3e-14 7e-09	AA632137	Hs.181767		4e-11
YPR011C	ANT2 ANT3	L78810 J03592	Xq24-q26 Xp22.32	3e-15 4e-14	AA632137	Hs.181767		2e-22
YPR021C	Aralar 1 AC002540 CTP	Y14494 AC002540 U25147	7q21-q22 22q11.2-pter	6e-36 5e-18 8e-14	H29831 AI989992	Hs.32508 Hs.170834	11p15.5-pter	2e-13 7e-13
YPR128C					AA918972	Hs.164280	Xp22.32	5e-06

Note. Unigene entries according to 12/29/99 update.

TABLE 2
Human Genes and ESTs Retrieved by BLAST Homology with Yeast Proteins
Involved in Mitochondrial Translation

Ribosomal proteins								
YEAST	Human homolog genes				Human homolog ESTs			
Prot name	Protein symbol	Genbank accession number	Chromosomal mapping	e value	Genbank accession number	Unigene entry	Chromosomal mapping	e value
DEG1					AW020963 AA127469			1e-68 4e-24
IFM1	MTIF2	L34600		3e-19				
MEF1					AI659020 AA010761 KIAA0031	Hs.41066 Hs.149489 Hs.151787	3q25.1-q25.2	7e-29 6e-28 4e-13
MRF1	MRF1	AF072934		1e-23	AA405163	Hs.97950	6q25.1-q25.2	6e-22
MRP17	MRP17	X79865	17q25.1-q25.3					
MRPL10					W00599	Hs.18349	8q11.22-8q11.23	5e-09
MRPL17					AI571752	Hs.14018	15q25-15q26.1	1e-07
MRPL19					AI188527	Hs.19077	11q13.3-q13.5	2e-26
MRPL2					W69555	Hs.7736	17q21.1-17q23	2e-11
MRP20	Ribosomal prot L23-like	U26596	11p15.5	8e-06				
MRPL23					AA430750	Hs.43946	8q22.3-8q24.13	7e-19
MRPL8					AA010799	Hs.10026	11p15.5-p15.2	5e-08
MRPL9	60S ribosomal protein L3	X06323	3q21.2-q23	2e-25				
MRPS5					N69846	Hs.7807	2p11.2-p11.1	4e-11
QSR1	Ribosomal protein L10	M73791	Xq28-qter	7e-68				
RML2	ribosomal protein L8	Z28407	8q24.3-qter	7e-15	AI185677	Hs.55041		7e-17
TUF1	mt elongation factor Tu	L38995		2e-45				
YCR024C	aspartyl-tRNA synthetase alpha-2	J05032	2q21-2q22.3	1e-09	AI697717	Hs.15502	11q14.3-q21	5e-18
YHR075C					AA234494	Hs.63304	11q13-q21	5e-14
YPL013C					AA456261	Hs.180312	10q22.2-10q22.3	1e-08

tRNA synthetase								
YEAST	Human homolog genes				Human homolog ESTs			
Prot name	Protein symbol	Genbank accession number	Chromosomal mapping	e value	genbank accession number	Unigene entry	Chromosomal mapping	e value
ISM1	IARS	D28473	9q21	1e-11	AA258087 AA987767	Hs.7558 Hs.55609	1q41-1q42	4e-21 2e-18
MSD1	KARS	D32053	16q23-q24	2e-10	AI829784	Hs.247086		9e-30
MSF1					AI338405	Hs.57969	6p25-p22	3e-33
MSK1	KARS	D32053	16q23-q24	1e-26				
MSR1	Arginyl-tRNA synthetase	S80343		1e-14	AI016336	Hs.15395	6q13	7e-41
MSS1					R13025			3e-10
MST1					W90183 AI766284 AI609524	Hs.210733 Hs.181426 Hs.184012		1e-29 6e-18 2e-17
MSY1	PGP 9.5	X04741	4p14	2e-25	AI989806	Hs.50441		3e-11
MSW1					AA227572 AA579040	Hs.168487	1q21	2e-13 2e-09
MTO1					AA121349	Hs.33979	6p11.2-6q12	4e-17
VAS1	valyl-tRNA synthetase	M98326	6p21.3	6e-67				

Note. Unigene entries according to 12/29/99 update.

chondrial transport proteins identified human mitochondrial transport proteins. Mitochondrial proteins could be distinguished from their nuclear counterparts by using the location site prediction program (Mitoprot). However, one should be aware that the identification of a mitochondrial targeting sequence could be overlooked occasionally by such programs, as was the case for frataxin (16,17).

BLAST searching also identified a large number of EST clusters. These ESTs corresponded to hitherto unknown human genes, homology being based on either functional identity or specificity with respect to mitochondrial functions. Tables 1 to 7 present these ESTs as Unigene entries, bringing together ESTs as clusters. The 102 EST clusters identified here represent putative human counterparts of yeast genes.

TABLE 3
Human Genes and ESTs Retrieved by BLAST Homology with Yeast Proteins
Involved in mtDNA and mtRNA Metabolism

YEAST		Human homolog genes			Human homolog ESTs			
Prot name	Protein symbol	Genbank accession number	Chromosomal mapping	e value	Genbank accession number	Unigene entry	Chromosomal mapping	e value
ADK2	adenylate kinase 2A adenylate kinase 3	U84371	1p34	5e-37	AA007279	Hs.43436		9e-20
		X60673	9p24-p13	6e-22				
ai5α	TRPC1	X89066	3q22-q24	6e-73				
AZF1	znfp195 Zn FINGER PROTEIN 42 Zinc finger protein OZF	AF003540	11p15	6e-23	AI637719 AA291928 AI041519	Hs.241523 Hs.22934 Hs.132130	18p11 3q25.1-q27	2e-24
		U69645	10cen-q11.2	2e-22				9e-24
		S50223		2e-22				2e-23
		X70394	19q13	2e-22				
CPF1	USF2	X90824	19q13	4e-06				
HAP2	CCAAT-binding protein	M59079	6p21.3	5e-12				
HAP3	NFYB DR1	J06145	12q22-q23	5e-28	AW136187	Hs.247228		7e-10
		M97388	1p22.1	2e-09				
HAP5	HITF2A HIS1	S74703	1p36.1-32 14q32-qter	6e-27	AI188727	Hs.19980	2p13	1e-09
		AB021179	17q21.1	9e-09				
MGM1	KIAA0820 dynamin-like dynamin 1	AB020627	1q24	7e-21				
		AF061795	12p11.2-q13.3	3e-18				
		L07807	9q34	2e-10				
MMD1	p14.5	X95384	8q22	4e-17				
MOD5					AA356092			5e-11
MSH1	hMSH2 MSH6 MSH5	U03911	2p22-p21	5e-30				
		U73737	2p16	3e-22				
		AF048986	6p21.3	1e-13				
MSS116	KIAA0801 EIF4A1 RNA helicase	AB018344	5q23-q31	1e-30	AI452653	Hs.244638		5e-17
		D13748	17p13	2e-18				
		X98743	2	7e-17				
NFS1	Nifs	AF097025	20q11-q12	4e-55	AA911739	Hs.21421	2q37.1-2q37.3	4e-15
NTG1	NTH1	U81285	16p13.2-p13.3	7e-36				
NUC1	endo G	X79444	9q34.1	8e-30				
PIF1					AI652391	Hs.112160	15q21	3e-14
PUS4					AA282195	Hs.88678		4e-14
RIM2	CTP Aralar 1 ANT2 PHC	U25147	22q11.2-pter	6e-09	AI431745 AA757281	Hs.42484 Hs.183047	3q23-q24 1p36.32-1p36.31	9e-30
		Y14494		7e-08				9e-21
		M57424	Xq24-q26	7e-08				
		X60036		4e-07				
RPO41	RNA pol	U75370	17q21.1-q23	3e-66				
SUV3	SUV3	AF042169	10q22	2e-16				
TRM1	tRNA methyltransf	AC005546	19p13	1e-14				
UNG1	UNG	X15653	12q23.3-q24.2	6e-47				

Note. Unigene entries according to 12/29/99 update.

For one-half of the yeast proteins (196/340), no human homologous sequences could be identified. As most ESTs have been obtained by sequencing the 3' portion of mRNAs, the 5' portions of large mRNAs are often lacking. Consequently, our identification strategy may be inefficient if the C-terminal regions of yeast and human counterparts were presenting low levels of homology. This feature could also be ascribed to either interspecies specificities or loss of particular mitochondrial functions during evolution. In keeping with this, disappearance of introns from human mitochondrial DNA (mtDNA) has presumably led to the loss of genes involved in mRNA splicing. It is therefore not surprising that no yeast genes involved in mitochondrial intron splicing (CBP2, MRS1, MRS2, or PET54) had human orthologues. A

third explanation may stem from a low degree of sequence conservation between species. This feature is known for mitochondrial ribosomal proteins previously shown to be highly divergent between *Escherichia coli*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, and human (18). This feature probably accounts also for the low rate (31/70) of human homologues for yeast mitochondrial translation proteins (Table 2). The absence of human orthologues of yeast proteins in the EST database may also be due to the low transcription levels of the corresponding genes, as an EST database is qualitatively but also quantitatively representative of genome transcription.

While systematic database screening represents a powerful tool for identifying putative human mito-

TABLE 4
Human Genes and ESTs Retrieved by BLAST Homology with Yeast Proteins
Involved in Protein Import

YEAST	Human homolog genes				Human homolog ESTs				
	Prot name	Protein symbol	Genbank accession number	Chromosomal mapping	e value	Genbank accession number	Unigene entry	Chromosomal mapping	e value
CPR3	cyclophilin A cyclophilin D cyclophilin F	Y00052 D63861 M80254	7p13 4q31.3 10q22.3	1e-55 2e-50 3e-45					
HSP10	HSPe1	U07550	14q23.1	3e-16					
MAS5	HSJ2 DNJ3 hsp40 hsp40-3 DNJ3/CPR3	L08069 Y13350 X62421 AF088982 AF011793	9p13-p12 19p13	1e-30 4e-24 4e-23 3e-19 7e-13	AA639658	Hs.247662	3q25.1-q28	2e-18	
MDJ1	hsp40-3 TID1 hsp40	AF088982 AF061749 X62421	16p13.3 19p13	2e-09 7e-09 1e-08					
MGE1					AA887226 AI431866	Hs.151903 Hs.164198	4p16.3-p16.1	1e-18 8e-13	
MSP1	PSMC1 Nuclear VCP- like PSMC2 PSMC4 PSMC6 p60 katanin	L02426 U68140 D11094 AF020736 D78275 AF056022	19p13.3 1q41-q42.2 7q21-q22 19q13.1-q13.4 12q15 6q24q25.2	4e-27 1e-24 1e-22 7e-19 1e-18 3e-18	AI796091 AA418426 AA976246 AA743793 T35813 AI809648	Hs.100861 Hs.234839 Hs.79387 Hs.88276 Hs.5555 Hs.131972	10q23-q24 16q22.1 17q23-q25 15q15-15q21.2	2e-34 4e-22 2e-18 2e-18 7e-17 3e-16	
SSC1	mortalin-2 HSPA5 HSHSC70 HSP70-1	L15189 M19645 Y00371 M59828	5q31.1 2q36 9q33-q34.1 11q23.3-q24.2 6p22-p21.1	2e-68 7e-51 2e-52 1e-33	AI191499	Hs.169668		3e-25	
TIM10					AI309333	Hs.109571	11p11.2-13.3	1e-12	
TIM17	TIM17 Tim17b	X97544 AF034790	1q31.3-q32	1e-31 1e-23					
TIM22					AI422894	Hs.87595	17p13.3	8e-10	
TIM23	TIM23	AF030162	10cen-q11.2	2e-07					
TIM9					AI743840	Hs.32456		1e-10	
TIM13					AA449553	Hs.76086		3e-10	
TOM20	TOM20 homolog	D13641	1q42	5e-09					
TOM70	IEF SSP 3521 KIAA0719 hTOM34p	M86752 AB018262 U58970	11p11.2-11q13 3cen-q13.1 20q12-q13.11	2e-07 1e-06 1e-06					
TOM72	IEF SSP 3521	M86752	11p11.2-11q13	1e-11	AA843594	Hs.99134		4e-07	
YDJ1	DNAJ-like 2 DNJ3/CPR3 MSJ-1 HSDNAJ hsp40-3	L08069 Y13350 AB014888 X62421 AF088982	9p13-p12 2q11.2 19p13.2-19p13.1 9	1e-30 4e-24 4e-24 4e-23 2e-19	AA599885 T78510 C17028	Hs.257313 Hs.154662 Hs.6790	3q25.1-q28	1e-18 2e-13 2e-13	
YLR168C					AA442417 AI052748	Hs.3945 Hs.9601	20q13 5q35.3	2e-21 5e-13	

Note. Unigene entries according to 12/29/99 update.

chondrial cDNAs, it has its own limitation. Indeed, the large number of yeast mitochondrial proteins originally selected (340) required strict criteria (*e* value) for reliable retrieval of human ESTs. As a result, ESTs with moderate score homologies and not belonging to an Unigene EST cluster may have been overlooked, even though it might represent the true human counterpart of a yeast protein.

Retrieving Novel Mitochondrial Protein Encoding Human Genes

The putative human genes and the EST clusters identified by BLAST searching are presented according to their mitochondrial functions: carrier protein (Table 1), translation (Table 2), mtDNA and mRNA metabolism (Table 3), protein import (Table

4), protein processing (Table 5), respiratory chain biogenesis (Table 6), and various OXPHOS related functions (Table 7). These tables also present the chromosomal mapping of the genes or EST clusters when available in Unigene database.

Comparison of the *e* value for a human gene or EST cluster homologous to a given yeast protein indicates that some EST clusters may present higher scores than genes. This may suggest that the EST clusters indeed represent the true human orthologue of the yeast protein used as probe while the retrieved gene(s) might correspond to the member(s) of the same family. Thus, RIM2, a member of the mitochondrial carrier family which is essential for mtDNA metabolism presents a significant homology (Tables 1 and 3) with known human genes encoding

TABLE 5
Human Genes and ESTs Retrieved by BLAST Homology with Yeast Proteins
Involved in Protein Processing

YEAST	Human homolog genes				Human homolog ESTs			
	Protein symbol	Genbank accession number	Chromosomal mapping	e value	Genbank accession number	Unigene entry	Chromosomal mapping	e value
AAP1	aminopeptidase	Y07701	17q21.1	2e-41	AA747425	Hs.167544		5e-29
HMT1	HMT1 PRMT2	Y10807 U80213	19q13 21q22.3	2e-37 1e-23	AA213882	Hs.191486		8e-25
IMP1					AA977197	Hs.62669		9e-16
IMP2					T67154 AA977197	Hs.13058 Hs.62669	7q22.3-q31.1	6e-23 2e-14
MAS1	core I peptidase β	D26485 AF054182	3p21 7q21.3-q22.1	2e-36 5e-32	D21064	Hs.75353	9q34.1	4e-08
MAS2	core I peptidase β core II	D26485 AF054182 J04973	3p21 7q21.3-q22.1 3p22.1-p21.33	2e-12 9e-11 2e-10	D21064	Hs.75353	9q34.1	5e-15
MAS5	HSJ2 DNJ3 hsp40 hsp40-3	L08069 Y13350 X62421 AF088982	9p13-p12 2q11.2 19p13	1e-30 4e-24 4e-23 3e-19	AA639658	Hs.247662	3q25.1-q28	2e-18
OCT1	MIPEP	U80034	13q12		AI952756	Hs.224836		1e-20
PTM1	Lon-like	U02389	8q22.2-q22.3	2e-52	AA683395 AA195153	Hs.200461 Hs.47305		1e-38 2e-20
SEC11	signal peptidase	AF090315	8p11.1-q11.21	1e-23	AA989369	Hs.68644		4e-31
SSQ1	GRP78 hsp 70 mortalin-2 HSPA2 hsc70	M19645 M11717 L15189 L26336 Y00371	9q33-q34.1 6p22-p21.3 5q31.1 2q36 14q24.1 11q23.3-q24.2	7e-51 9e-51 5e-45 4e-42 2e-43	AI191499	Hs.169668		9e-23
YME1					AA418426 AI791698 AA743793	Hs.234839 Hs.131823 Hs.88276	16q22.1	1e-20 3e-18 2e-14
YTA10	paraplegin- like paraplegin	Y18314 Y16610	18p11 16q24.3	5e-38 9e-31				
YTA12	paraplegin- like paraplegin	Y18314 Y16610	18p11 16q24.3	5e-38 9e-31				

Note. Unigene entries according to 12/29/99 update.

mitochondrial carriers, i.e., citrate carrier (e value = $6e-09$) or adenylate transporter ANT2 (e value = $7e-08$). However, two EST clusters present a much higher homology (Hs.42484, e value = $9e-30$ and Hs.183047, e value = $9e-21$) suggesting that these two EST clusters are more likely to represent parts of the human orthologue of yeast RIM2. The significance of high scores is also illustrated by the example of RNA polymerase or endonuclease G (Table 3), which shows a high homology between yeast and human (e values $3e-66$ and $8e-30$, respectively).

Conversely, a given EST cluster may well correspond to different yeast proteins belonging to the same family. The Unigene entry Hs.42484 shows significant homology with three different yeast mitochondrial carrier proteins but only one of them shows a high homology score, namely, RIM2 ($9e-30$), again suggesting that this EST cluster may be the human counterpart of yeast RIM2. Thus, while families of genes or proteins can be retrieved using this

approach, a careful examination of the homology scores is required to determine which of these genes or EST clusters is more likely expected to be the human counterpart. In other cases, the BLAST searching only allowed us to retrieve human sequences with a moderate, although significant, homology. In such cases, only additional experimental approaches will permit identification of which sequence is the actual human counterpart of the considered protein.

Yeast mtDNA and mtRNA metabolism involves 30 known proteins. Among them, 20 present at least one highly homologous sequence in human with e values between $6e-73$ and $1e-14$ (Table 3) suggesting that these are the actual human orthologues of the yeast proteins. Some of them are known to encode mitochondrial proteins, namely, endonuclease G (endoG), mitochondrial tRNA polymerase (RNA pol), mitochondrial RNA helicase (SUV3), and uracil-DNA glycosylase (UNG). Interestingly, two of them,

TABLE 6
Human Genes and ESTs Retrieved by BLAST Homology with Yeast Proteins
Involved in Respiratory Chain Biogenesis

YEAST		Human homolog genes			Human homolog ESTs			
Prot name	Protein symbol	Genbank accession number	Chromosomal mapping	e value	Genbank accession number	Unigene entry	Chromosomal mapping	e value
ABC1	BAT3	M33521	6p21.3	5e-36	AF052122	Hs.130712		3e-12
BCS1	h-bcs1	AF026849	2q33	1e-29				
COX11	COX11	AF044321	17q21.31-17q23	9e-41				
COX15	COX15	AF044323	10q24	1e-21				
COX17	COX17	L77701	13q14.3	0.060				
CYT2	holocytc-synthetase	U36787	Xp22.2-22.3	3e-29				
FAD1					U79241	Hs.118666	1q21-1q23	2e-21
ISA1					AF038186	Hs.177776		4e-26
					AI146474	Hs.63913		3e-08
					AI474489	Hs.129544		6e-06
ISU1	NifU-like	U47101	12q23-q24	7e-47				
ISU2	NifU-like	U47101	12q23-q24	5e-48				
JAC1					AI796137	Hs.130430		1e-12
MSS116	KTAA0801	AB018344	5q23-q31	1e-30	AI452653	Hs.244638		5e-17
	EIF4A1	D13748	17p13	2e-18				
	RNA helicase	X98743	2	7e-17				
MSS1					R13025	Hs.130712		3e-10
NFU1	HIRIP5	AJ132584	3p24.2-p21.33 2p13	3e-23				
OXA1	OXA1Hs	X80695	14q11.2	4e-14				
PET112	PET112	AF026851	4q27-q28	9e-23				
SCO1	SCO2 homolog	AL021683	22q13	2e-33				
	SCO1 homolog	AF026852	17p12-p13	2e-27				
SCO2	SCO2 homolog	AL021683	22q13	2e-33				
	SCO1 homolog	AF026852	17p12-p13	2e-28				
YAH1	ferredoxin	M34788	11q22	1e-14				

Note. Unigene entries according to 12/29/99 update.

namely, p14.5 (homologous to yeast MMD1) and NTH1 endonuclease III (homologous to yeast NTG1), were not known to have a mitochondrial function. However, their high scores for mitochondrial targeting (0.88 and 0.95 determined by Mito-prot) suggest that both genes may actually encode mitochondrial proteins.

CONCLUSION

Experimental cloning of human genes based on cross-species comparisons has proved to be a tedious approach largely due to nucleotide sequence divergences between species. The continuously expanding EST databases and the powerful computer pro-

TABLE 7
Human Genes and ESTs Retrieved by BLAST Homology with Yeast Proteins
Involved in Various OXPHOS-Related Functions

Heme biosynthesis								
YEAST		Human homolog genes			Human homolog EST			
Prot name	Protein symbol	Genbank accession number	Chromosomal mapping	e value	Genbank accession number	Unigene entry	Chromosomal mapping	e value
HEM1	ALAS1	X56351	3p21.1	2e-53				
	ALAS2	X56352	Xp11.21 5	1e-39				
	GAT	Z97630	22q12-13	7e-16				
HEM14	PPO	U26446	1q22	1e-05				
HEM15	ferrochelat	D00726	18q21.2-q21.3	3e-63				
HEM4	URO3	J03824	10q25.2-q26.3	7e-04				

YEAST		Human homolog genes			Human homolog EST			
Prot name	Protein symbol	Genbank accession number	Chromosomal mapping	e value	Genbank accession number	Unigene entry	Chromosomal mapping	e value
COT1					AI467909	Hs.188417		3e-15
					AA625460	Hs.55610		9e-08
					AI332677	Hs.129445	5q11.2-q12	9e-08
NHX1	NHE-6	AF030409	Xq25-q26	2e-12	AA279477	Hs.154353	Xp11.3-p11.23	2e-13
PHB1	Prohibitin	S85655	17q21	8e-55	AI936946	Hs.121973		9e-31
	hBAP	U72511	12p13	7e-50				
PHB2	hBAP	U72511	12p13	6e-53				
	Prohibitin	S85655	17q21	4e-49				

Note. Unigene entries according to 12/29/99 update.

grams now available represent a useful tool for rapid identification of new genes using sequence homologies between species. We have combined the considerable knowledge on yeast genetics and human EST database to rapidly search for nuclear genes involved in mitochondrial function. This allowed us to identify 102 human EST clusters corresponding to hitherto unknown genes likely to encode mitochondrial proteins. Future experimental approaches are required to describe the complete sequences of these genes and to determine the exact functions of the encoded proteins.

To our knowledge, none of the chromosomal localizations of the ESTs presented here correspond to loci previously linked to mitochondrial disorders. As far as autosomal dominant mtDNA deletions are concerned, three loci have been mapped to chromosomes 10q, 3p, and 4q respectively (19–21). Unfortunately, none of the genes or EST clusters retrieved here can be regarded as candidate genes for these conditions. Yet the present list will hopefully help gene identification in mitochondrial disorders by pinpointing candidate genes at a given locus linked to a specific disease.

REFERENCES

1. Bourgeron T, Rustin P, Chretien D, Birch-Machin M, Bourgeois M, Viegas-Pequignot E, Munnich A, Rötig A. Mutation of a nuclear succinate dehydrogenase gene results in mitochondrial respiratory chain deficiency. *Nature Genet* **11**:144–149, 1995.
2. Loeffen J, Smeitink J, Triepels R, Smeets R, Schuelke M, Sengers R, Trijbels F, Hamel B, Mullaart R, van den Heuvel L. The first nuclear-encoded complex I mutation in a patient with Leigh syndrome. *Am J Hum Genet* **63**:1598–1608, 1998.
3. van den Heuvel L, Ruitenbeek W, Smeets R, Gelman-Kohan Z, Elpeleg O, Loeffen J, Trijbels F, Mariman E, de Bruijn D, Smeitink J. Demonstration of a new pathogenic mutation in human complex I deficiency: A 5-bp duplication in the nuclear gene encoding the 18-kD (AQDQ) subunit. *Am J Hum Genet* **62**:262–268, 1998.
4. Schuelke M, Smeitink J, Mariman E, Loeffen J, Plecko B, Trijbels F, Stockler-Ipsiroglu S, van den Heuvel L. Mutant NDUFV1 subunit of mitochondrial complex I causes leukodystrophy and myoclonic epilepsy. *Nature Genet* **21**:260–261, 1999.
5. Triepels RH, van den Heuvel LP, Loeffen JL, Buskens CA, Smeets RJ, Rubio-Gozalbo ME, Budde SM, Mariman EC, Wijburg FA, Barth PG, Trijbels JM, Smeitink JA. Leigh syndrome associated with a mutation in the NDUFS7 (PSST) nuclear encoded subunit of complex I. *Ann Neurol* **45**:787–790, 1999.
6. Casari G, De Fusco M, Ciarmatori S, Zeviani M, Mora M, Fernandez P, De Michele G, Filla A, Coccozza S, Marconi R, Durr A, Fontaine B, Ballabio A. Spastic paraplegia and OXPHOS impairment caused by mutations in paraplegin, a nuclear-encoded mitochondrial metalloprotease. *Cell* **93**:973–983, 1998.
7. Zhu Z, Yao J, Johns T, Fu K, De Bie I, Macmillan C, Cuthbert AP, Newbold RF, Wang J, Chevrette M, Brown GK, Brown RM, Shoubbridge EA. SURF1, encoding a factor involved in the biogenesis of cytochrome c oxidase, is mutated in Leigh syndrome. *Nature Genet* **20**:337–343, 1998.
8. Tiranti V, Hoertnagel K, Carrozzo R, Galimberti C, Munaro M, Granatiero M, Zelante L, Gasparini P, Marzella R, Rocchi M, Bayona-Bafaluy MP, Enriquez JA, Uziel G, Bertini E, Dionisi-Vici C, Franco B, Meitinger T, Zeviani M. Mutations of SURF-1 in Leigh disease associated with cytochrome c oxidase deficiency. *Am J Hum Genet* **63**:1609–1621, 1998.
9. Koehler CM, Leuenberger D, Merchant S, Renold A, Junne T, Schatz G. Human deafness dystonia syndrome is a mitochondrial disease. *Proc Natl Acad Sci USA* **96**:2141–2146, 1999.
10. Babcock M, de Silva D, Oaks R, Davis-Kaplan S, Jiralerspong S, Montermini L, Pandolfo M, Kaplan J. Regulation of mitochondrial iron accumulation by Yfh1p, a putative homolog of frataxin. *Science* **276**:1709–1712, 1997.
11. Foury F, Cazzalini O. Deletion of the yeast homologue of the human gene associated with Friedreich's ataxia elicits iron accumulation in mitochondria. *FEBS Lett* **411**:373–377, 1997.
12. Rötig A, de Lonlay P, Chretien D, Foury F, Koenig M, Sidi D, Munnich A, Rustin P. Aconitase and mitochondrial iron-sulphur protein deficiency in Friedreich ataxia. *Nature Genet* **17**:215–217, 1997.
13. Tugendreich S, Boguski MS, Seldin MS, Hieter P. Linking yeast genetics to mammalian genomes: Identification and mapping of the human homolog of CDC27 via the expressed sequence tag (EST) database. *Proc Natl Acad Sci USA* **90**:10031–10035, 1993.
14. Bassett DE Jr, Boguski MS, Hieter P. Yeast genes and human disease. *Nature* **379**:589–590, 1996.
15. Banfi S, Borsani G, Rossi E, Bernard L, Guffanti A, Rubboli F, Marchitelli A, Giglio S, Coluccia E, Zollo M, Zuffardi O, Ballabio A. Identification and mapping of human cDNAs homologous to *Drosophila* mutant genes through EST database searching. *Nature Genet* **13**:167–174, 1996.
16. Campuzano V, Montermini L, Molto MD, Pianese L, Cossée M, Cavalcanti F, Monros E, Rodius F, Duclos F, Monticelli A, et al. Friedreich's ataxia: Autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* **271**:1423–1427, 1996.
17. Gibson TJ, Koonin EV, Musco G, Pastore A, Bork P. Friedreich's ataxia protein: Phylogenetic evidence for mitochondrial dysfunction. *Trends Neurosci* **19**:465–468, 1996.
18. Grivell LA, Artal-Sanz M, Hakkaart G, de Jong L, Nijtmans LG, van Oosterum K, Siep M, van der Spek H. Mitochondrial assembly in yeast. *FEBS Lett* **452**:57–60, 1999.

19. Suomalainen A, Kaukonen J, Amati P, Timonen R, Haltia M, Weissenbach J, Zeviani M, Somer H, Peltonen L. An autosomal locus predisposing to deletions of mitochondrial DNA. *Nature Genet* **9**:146–151, 1995.
20. Kaukonen JA, Amati P, Suomalainen A, Rötig A, Piscaglia MG, Salvi F, Weissenbach J, Fratta G, Comi G, Peltonen L, Zeviani M. An autosomal locus predisposing to multiple deletions of mtDNA on chromosome 3p. *Am J Hum Genet* **58**:763–769, 1996.
21. Kaukonen J, Zeviani M, Comi GP, Piscaglia MG, Peltonen L, Suomalainen A. A third locus predisposing to multiple deletions of mtDNA in autosomal dominant progressive external ophthalmoplegia. *Am J Hum Genet* **65**:256–261, 1999.