

Sugarcane genes related to mitochondrial function

Ghislaine V. Fonseca¹, José Humberto M. Tambor,^{1,2} Marina P. Nobrega¹,
Rafael Santos¹ and Francisco G. Nobrega^{1*}

Abstract

Mitochondria function as metabolic powerhouses by generating energy through oxidative phosphorylation and have become the focus of renewed interest due to progress in understanding the subtleties of their biogenesis and the discovery of the important roles which these organelles play in senescence, cell death and the assembly of iron-sulfur (Fe/S) centers. Using proteins from the yeast *Saccharomyces cerevisiae*, *Homo sapiens* and *Arabidopsis thaliana* we searched the sugarcane expressed sequence tag (SUCEST) database for the presence of expressed sequence tags (ESTs) with similarity to nuclear genes related to mitochondrial functions. Starting with 869 protein sequences, we searched for sugarcane EST counterparts to these proteins using the basic local alignment search tool TBLASTN similarity searching program run against 260,781 sugarcane ESTs contained in 81,223 clusters. We were able to recover 367 clusters likely to represent sugarcane orthologues of the corresponding genes from *S. cerevisiae*, *H. sapiens* and *A. thaliana* with E-value $\leq 10^{-10}$. Gene products belonging to all functional categories related to mitochondrial functions were found and this allowed us to produce an overview of the nuclear genes required for sugarcane mitochondrial biogenesis and function as well as providing a starting point for detailed analysis of sugarcane gene structure and physiology.

INTRODUCTION

Mitochondria play a central role in the life of eukaryotic organisms because of their role in respiration, a fundamental process for most nucleated cells (Saraste, 1999). The plant cell is a particularly complex environment, with individual genomes enclosed in the nucleus, the mitochondrion and the chloroplast. These genomes interact by means of specific gene products to ensure proper physiologic homeostasis and the biogenesis of the organelles themselves. The mitochondrion genome codes for a limited set of its own proteins and is dependent on nuclear gene products for the large majority of its constituent proteins and enzymes. The unicellular eukaryote *Saccharomyces cerevisiae* has a pivotal position in the study of mitochondrial function due to the ease with which its mitochondrial and nuclear genomes can be manipulated (Tzagoloff and Dieckmann, 1990).

It has become clear that mitochondria are essential cellular organelles for most eukaryotes not only because they produce ATP by aerobic catabolism. Research on *S. cerevisiae* has been instrumental in demonstrating that even cells without mitochondrial DNA (mtDNA) have an organelle that performs essential functions (Schatz, 1995) and there is growing evidence that mitochondria have a central position in regulating the response to stress and programmed cell death (apoptosis) in eukaryotic cells (Susin *et al.*, 1998). These new insights place mitochondria and their

metabolic and biosynthetic activities at the focus of renewed interest.

Saccharomyces cerevisiae was the first eukaryote to have its genome completely sequenced (Goffeau, 1996) thus allowing the overall evaluation of the genes involved in mitochondrial functions in this organism. Due to the importance of mitochondrial dysfunction in genetic diseases of humans, the molecular analysis of mitochondrial functions in complex eukaryotes is most advanced in *Homo sapiens* (Wallace, 1999) and about 260 gene products have been linked to mitochondrial functions in humans (MITOP database, www.mips.biochem.mpg.de/proj/medgen/mitop). A different approach by Rötig *et al.* (2000) has identified in humans, 229 potential orthologues of *S. cerevisiae* nuclear mitochondrial genes, 102 of which were expressed sequence tag (EST) matches. *Arabidopsis thaliana* was the first plant to have its genome completely sequenced (Anon, 2000) providing a rich source of genetic material for comparative studies.

In this paper we present the results of a search of the Sugarcane EST (SUCEST) database for candidate orthologues of nuclear encoded mitochondrial genes. The (SUCEST) project is part of the Organization for Nucleotide Sequencing and Analysis (ONSA) network created and supported by the São Paulo State Research Foundation (*Fundação de Amparo à Pesquisa do Estado de São Paulo* (FAPESP)).

¹Instituto de Pesquisa e Desenvolvimento (IP&D) Universidade do Vale do Paraíba (UNIVAP), 12244-000 São José dos Campos, SP, Brazil.

²Microbiology Graduate Program, ICB/USP, São Paulo, SP, Brazil.

Send correspondence to Francisco G. Nobrega. E-mail: fgdnobre@univap.br.

Human and plant (*Arabidopsis*) annotated gene sequences were also used to expand our search for mitochondria related genes in sugarcane because *S. cerevisiae* lacks a number of important mitochondrial proteins (e.g. those present in the proton-translocating NADH-dehydrogenase-ubiquinone) and activities that are present in more complex eukaryotes.

Our research extracted from the SUCEST database a set of 367 clusters representing potential orthologues of nuclear sugarcane genes coding for gene products linked to mitochondrial function.

MATERIAL AND METHODS

Using the search tools available for data mining at the SUCEST site (<http://sucest.lad.ic.unicamp.br/en/>) we launched an EST-based strategy for gene discovery centered on mitochondrial functions.

A basic local alignment search tool TBLASTN search (protein sequence against 6-frame nucleotide translated database sequences) was carried out against 81,223 consensi from the fragment assembly program (phrap, Altschul *et al.*, 1997) clustering of 260,781 ESTs (reference: Telles and da Silva, 2001) contained in the SUCEST database which were sequenced mostly from the 5' end and derived from 37 cDNA libraries constructed from distinct sugarcane tissues grown under different conditions (Vettore *et al.*, 2001). The EST reads were generated by a consortium of laboratories and deposited and managed at the bioinformatics laboratory at the university of Campinas (Telles *et al.*, 2001). Our search included all mitochondrially located gene products using the information available in the *S. cerevisiae* Proteome, Inc. database (Costanzo *et al.*, 2001; www.proteome.com/databases/index.html).

Some transcription factors linked to the expression of genes related to mitochondria were also searched, although their site of action is the nucleus. Matches/Hits with an E-value of $\leq 10^{-10}$ against any protein were recorded, the annotation usually referring to well-characterized mitochondrial sequences from *S. cerevisiae* or the corresponding sequences from *Arabidopsis thaliana* or *Homo sapiens*. The *Arabidopsis thaliana* protein sequences were retrieved using the relevant categories listed at the Functional Catalogue available at the MIPS site (PEDANT database - <http://pedant.mips.biochem.mpg.de/at>) while the *Homo sapiens* sequences were retrieved from the MITOP database (Scharfe *et al.*, 2000; www.mips.biochem.mpg.de/medgen/mitop) which lists the annotated human genes linked to mitochondrial functions.

RESULTS

Table I shows the origins of the 869 gene products derived from the reference organisms and known organellar locations used in the search along with the number of SUCEST matches with E-value $\leq 10^{-10}$ and the number of full-length clones recovered. We selected 459 *S. cerevisiae*

Table I - Summary of SUCEST matches for all gene products related to mitochondrial function.

Organisms: cellular location of gene product	Gene products	Total SUCEST clusters E-value $\leq 10^{-10}$	Full-length clusters
<i>Saccharomyces cerevisiae</i> : inner membrane	149	86	41
<i>S. cerevisiae</i> : inter-membrane space	11	8	4
<i>S. cerevisiae</i> : matrix	51	44	16
<i>S. cerevisiae</i> : outer membrane	27	11	4
<i>S. cerevisiae</i> : mitochondria and other sites	221	123	39
Total for <i>S. cerevisiae</i>	459	272	104
<i>Arabidopsis thaliana</i>	149	146	55
<i>Homo sapiens</i>	261	179	77
Grand total ^a	869	597	236

^aTotal for *S. cerevisiae* + *A. thaliana* + *H. sapiens*.

gene products assigned to mitochondria, including some important transcription factors and cytoplasmic proteins known to be relevant to the function of the organelles, of which 272 showed significant similarity (E-value $\leq 10^{-10}$) to SUCEST clusters with 104 of these clusters probably representing full-length cDNA clones.

In calculating the full-length clones the criteria were that there should be no more than 30 amino acids out of alignment at the N-terminal end of the driver-sequence and an excess of nucleotides at the 5' end of the SUCEST cluster to accommodate the divergent N-terminal sequence. The 149 *Arabidopsis thaliana* gene products used in the search resulted in 146 matches with E-value $\leq 10^{-10}$ and 55 apparently full-length cDNA clones while the 261 human sequences resulted in 179 SUCEST matches (E-value $\leq 10^{-10}$) and 77 apparently full-length cDNA clones.

The data shown in Table I was sorted and compared to generate a non-redundant set (Table II) of 367 unique clusters out of the 597 SUCEST matches. Table II shows the number of SUCEST clusters with similarity to single proteins or two or more proteins from the different organisms used for the search. Seventy clusters were similar to proteins from two organisms and 19 clusters were similar to sequences from all three organisms. The complete list of clusters is extensive and is presented in two supplementary tables (Table I-S and Table II-S: available at the SUCEST site (<http://sucest.lad.ic.unicamp.br/en/>) which contain the 367 SUCEST clusters identified and the single or multiple matches to the different proteins used in the search. These tables list the sugarcane cluster, the reference organism(s) protein or gene name, protein size and function, score and E-value of the match. Clusters are labeled with the category number from Table III and probable full-length clusters are indicated by an asterisk. In Table I-S we list 130 clusters

Table II - Number of non-redundant SUCEST clusters with matches to one or more protein sequences used as queries in the searches.

Organism	SUCEST clusters
Single matches ^a	
<i>Saccharomyces cerevisiae</i>	140
<i>Arabidopsis thaliana</i>	44
<i>Homo sapiens</i>	53
Two or more matches ^b	
<i>S. cerevisiae</i>	10
<i>Arabidopsis thaliana</i>	16
<i>Homo sapiens</i>	15
<i>S. cerevisiae</i> + <i>Arabidopsis thaliana</i>	22
<i>S. cerevisiae</i> + <i>Homo sapiens</i>	40
<i>Arabidopsis thaliana</i> + <i>Homo sapiens</i>	8
<i>S. cerevisiae</i> + <i>Arabidopsis thaliana</i> + <i>Homo sapiens</i>	19
Total	367

^aIn this set each cluster corresponds to just one protein from one of the reference organisms.

^bIn this set each cluster corresponds to two or more proteins from the indicated reference organism(s).

that have matches with two or more proteins used in the search from the same or different organisms. Table II-S lists 237 clusters with hits to single proteins from *S. cerevisiae*, *H. Sapiens* and *A. thaliana*.

In order to have an overview of the cellular roles and functions covered by this set of putative mitochondria-related genes expressed in sugarcane we grouped the annotated SUCEST sequences in a more convenient way for analysis (Table III) by classifying them in functional categories based on cellular processes or enzyme complexes. It can be seen that we were able to obtain putative orthologues for all functional categories of relevance to mitochondrial functions.

DISCUSSION

As mentioned in the introduction, mitochondrial functions go beyond energy metabolism and the generation of ATP by oxidative phosphorylation. The biogenesis of this cellular organelle is dependent upon the cooperation of two genomes: the mitochondrial genome encoding a relatively small number of gene products and the nuclear genome coding, in *S. cerevisiae*, for over 400 proteins that are needed for the function of this organelle (Tzagoloff and Dieckmann, 1990). Genome-wide surveys have suggested that in plants and vertebrates 10% of all proteins coded could be targeted to the mitochondria (Emanuelsson, 2000). In Table III we have grouped genes linked to amino acid metabolism, the Krebs cycle and other functions (not included in any specific category) under the heading 'metabolism', which contains 76 SUCEST clusters and among them, 27 putative full-length clones. Another category with

many representatives is the mitochondrial carrier family of proteins (MCF) and other membrane transporters. Here we found, for example, candidates for transporters of citrate, phosphate, carnitine, ornithine and ATP/ADP.

Solute transport systems are essential for communication between cells and the environment and in organelle homeostasis. This large protein family is eukaryotic-specific, and supposedly arose after the appearance of mitochondria in the eukaryotic lineage, although its members are exclusively localized in the mitochondria (El Moulaj *et al.*, 1997). Plant mitochondrial carriers are attracting interest (Laloi, 1999) and our search found many members of this family (Table III). The uncoupling proteins found in plants and animals are a subfamily of mitochondrial carriers (Ricquier and Bouillaud, 2000) with a proposed role in cell defense. One of the endogenous damage-inducing mechanisms that have been confirmed by a variety of different approaches are the cellular and organellar effects of reactive oxygen radicals generated as byproducts of respiration. The study of energy-dissipating systems like the uncoupling proteins in plants and animals and the presence of alternative oxidases in plants indicate that cells also use these

Table III - Summary of SUCEST clusters from different functional categories recovered from the TBLASTN similarity searches of all mitochondrial-related proteins from *Saccharomyces cerevisiae*, *Arabidopsis thaliana* and *Homo sapiens* against cluster consensi (phrap).

Functional categories	SUCEST clusters with an E-value $\leq 10^{-10}$	Full-length clusters
1 Metabolism	76	27
2 Mitochondrial carriers and transporters	36	19
3 Fe/S assembly	10	4
4 Mitochondrial dynamics	5	1
5 Protein dynamics	22	9
6 Protein import	10	4
7 ATP synthase	8	4
8 NADH - cytochrome C reductase	20	10
9 Succinate dehydrogenase	3	2
10 Cytochrome oxidase	8	4
11 General respiratory ability	8	2
12 Lipid metabolism	32	13
13 Ubiquinone synthesis	4	1
14 Heme synthesis	4	1
15 RNA metabolism and transcription	10	1
16 DNA dynamics	6	1
17 Protein synthesis	45	13
18 Free radical defense, stress and apoptosis	14	6
19 Unclassified	46	23
Total	367	145

systems along with superoxide dismutase and peroxidases to reduce the damage caused by reactive oxygen species (Kowaltowski *et al.*, 1999). Our data indicates the presence of an alternative oxidase and uncoupling protein in sugarcane.

A small but growing number of essential *S. cerevisiae* gene products have been located in mitochondria, going against the expectation that the only result of inactivating a mitochondrion-related gene (either nuclear or mitochondrial) would be the inability to metabolize non-fermentable carbon sources. This small set of genes code for proteins involved in the mitochondrial machinery of protein import (Neupert, 1997). Our group (Manzella *et al.*, 1998; Barros and Nobrega, 1999) have detected two new mitochondrial proteins (Arh1p and Yah1p) in *S. cerevisiae* which are es-

sential for growth under all the conditions tested, and which seem to be part of an electron transfer chain (ETC) distinct from the well-studied ATP-generating electron transport chain (ETC). Experiments with the *YAH1* gene by Lange *et al.* (2000) have shown that this essential gene is one of a number of genes linked to Fe/S cluster formation, a complete outline of this new function having been given by Lill and Kispal (2000). We found 10 potential orthologues of genes linked to the assembly of Fe/S clusters in the SUCEST database. These prosthetic groups are regarded as important structural elements for many proteins and are redox centers for mitochondrial (Figure 1) and cytoplasmic enzymes (Beinert *et al.*, 1997). In *S. cerevisiae*, the export of assembled Fe/S centers is mediated by the ATP-binding cassette transporter Atm1p (Kispal *et al.*, 1999) and sugar-

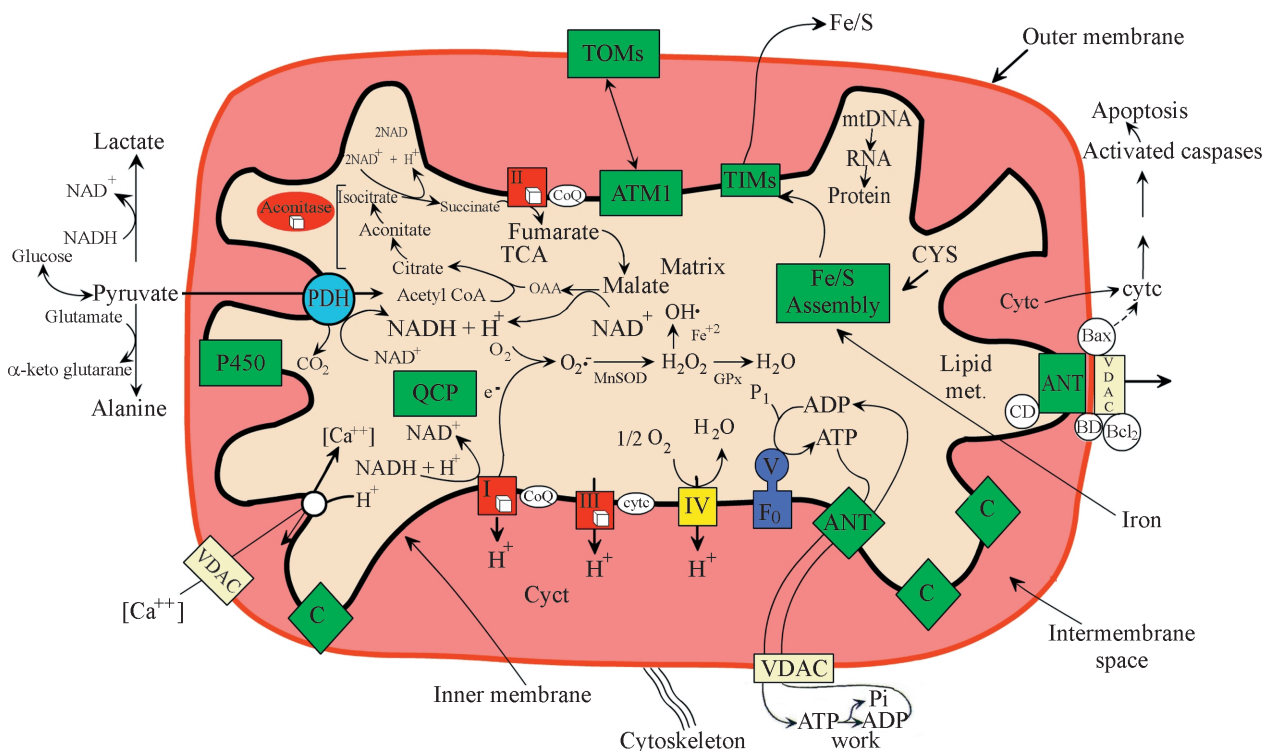


Figure 1 - Mitochondrial functions are represented in simplified form in this figure (modified from Wallace, 1999). Substrate entry and metabolism (TCA cycle and lipid metabolism) generate reducing equivalents that drive vectorial proton pumping by three respiratory complexes (I, III and IV) of the inner membrane involved in oxidative phosphorylation. Purified complex I in plants has over 30 subunits and includes as prosthetic groups flavin mononucleotide and six iron-sulfur (Fe/S) centers (white cubes) which transfer electrons from NADH to ubiquinone. Complex II carries a flavin dinucleotide, a cytochrome b and three Fe/S centers which transfer electrons from succinate to ubiquinone. Complex III (ubiquinol:cytochrome c oxidoreductase) has as catalytic core cytochrome b, cytochrome c_1 and the Fe/S Rieske protein. Cytochrome c oxidase (complex IV) carries cytochromes $a + a_3$ and copper atoms (CuA and CuB) as prosthetic groups. Energy is conserved by the action of ATP synthase or complex V. In all these complexes a set of additional polypeptides co-purifies with the enzyme and are usually essential for activity, although not in a direct (catalytic) manner. Pyruvate enters mitochondria via the pyruvate dehydrogenase complex (PDH). Small molecules (e.g. Ca^{2+}) enter through the outer membrane via porins or voltage-dependent anion channels (VDAC), a protein that together with the adenine nucleotide translocator (ANT) and apoptosis factors (Bax, cyclophilin D or CD) create the pores that initiate cytochrome c release and programmed cell death. A large number of mitochondrial carrier family proteins allow transport of specific solutes in and out of the organelle (C). P450 cytochromes are linked to the membrane and carry out detoxification and steroid modification. Assembly of the inner membrane enzymes and import of mitochondrial proteins depend upon the supramolecular complexes of translocators of the outer (TOM) and inner membrane (TIM). Degradation of proteins for turnover and quality control during organelle assembly is dependent on quality control proteases (QCP). Iron and cysteine are processed by a set of enzymes that carry out Fe/S center assembly for mitochondrial proteins and assembled centers are exported for use outside the organelle through the ATM1 transporter. Damaging oxygen radicals generated by electron transport, and increased when transport is blocked, are inactivated by mitochondrial superoxide dismutase (MnSOD) and other enzymes and are also reduced by the action of cyanide resistant alternative oxidase and plant uncoupling proteins. Mitochondrial connections with the dynamic cytoskeleton have been shown to be important for organelle sorting, shape and stability.

cane has a potential homologue to this protein. Other key components of the pathway seem to exist in sugarcane *e.g.* *ISAI*, *NFSI*, *NFUI* and *YAH1*. This mitochondrial function has been inherited from the bacterial endosymbiont from which mitochondria originated, as suggested by similarities between the corresponding bacterial and eukaryotic proteins. Interestingly, mammalian homologues of the components of the electron transfer chain (*ARH1* and *YAH1*) are well known for donating electrons to a mitochondrial cytochrome P450 in steroidogenic tissues, the first and fundamental step in the synthesis of all steroid hormones. Analogous reactions are to be expected in plant mitochondria and it is interesting to note that the SUCEST carries eight P450 cytochrome-like genes.

The power of molecular genetics has produced important new developments in understanding the molecular machinery behind the inheritance, shape and distribution of the mitochondrial network (Yaffe, 1999), aspects of mitochondrial physiology which are essential for normal cell proliferation. The mitochondrial network has important connections with the cytoskeleton (Figure 1) and disruption of some mitochondrion-related genes in *S. cerevisiae* results in aggregation or defective segregation as well as instability of the mitochondrial genome. Our search found five genes in this class (mitochondrial dynamics, Table III).

The machinery for protein import in mitochondria has been very intensively studied in *S. cerevisiae* (Neupert, 1997) and is being studied in plants as well (Glaser *et al.*, 1998). Among this class one finds many essential genes in *S. cerevisiae* and our search detected ten putative transporters: seven belong to the translocase of the inner membrane (TIM complex) and three to the translocase of the outer membrane (TOM complex). The two supra-molecular assemblies cooperate to import (Figure 1) nuclear encoded proteins synthesized in the cytoplasm into the mitochondrion.

The inner membrane electron translocating complexes and ATP synthase are central to the ability of mitochondria to generate abundant energy by complete oxidation of substrates and because of this the assembly of these supra-molecular complexes (Saraste, 1999) has been much studied. The active purified respiratory enzyme complexes have, besides the catalytic subunits like cytochromes and Fe/S center-bearing subunits (Figure 1), accessory proteins that are needed for the activity of the enzyme. In addition it has been found (Tzagoloff and Dieckmann, 1990) that there are additional proteins that are needed for the proper functional assembly of these heteromeric membrane enzymes but these proteins are not part of the isolated biochemically active enzymes. These extra-enzyme activity factors are thought to be 'assembly facilitators' or specialized chaperones but the molecular mechanism of their action is usually unknown, although their inactivation generates clear respiratory deficient phenotypes (Nobrega *et al.*, 1992; Souza *et al.*, 2000). This class is mostly dispersed among

NADH-cytochrome c reductase, cytochrome oxidase and general respiratory ability categories (Table III).

Our search found 20 candidate clusters in the combined NADH – cytochrome c oxidoreductase activity (complex I and complex III). Complex I, well studied in mammals has 43 subunits and is the largest of all the complexes, and the same complex in plants is also very large and similar to the mammalian enzyme (Rasmusson *et al.*, 1998). Complex III receives electrons from ubiquinone and is reduced by either complex I or complex II. Complex III has been very well studied with its tri-dimensional structure having been described by Xia *et al.* (1997). This complex has cytochrome b, the Rieske Fe/S protein and cytochrome c₁ as active components. Eight additional subunits surround the catalytic subunits and the so-called 'core proteins' contact the matrix space and are, in plants, identical to subunits of the mitochondrial processing peptidases that catalyze the specific cleavage of pre-sequence peptides from a large number of imported proteins (Glaser and Dessi, 1999).

Succinate:ubiquinone reductase or complex II is also a component of the Krebs cycle and is linked to the inner membrane by a b-type cytochrome. In addition it has flavin-adenine dinucleotide and several Fe/S centers and does not translocate protons across the membrane (Hägerhäll, 1997). We found three putative clusters in sugarcane (Table III).

Cytochrome c oxidase, the terminal enzyme of the respiratory chain, has also been extensively investigated. The mammalian enzyme has been crystallized (Tsukihara, 1996) and consists of 10 nuclear-encoded subunits and three mitochondrial-DNA encoded proteins. We detected 8 putative orthologues of this enzyme in sugarcane, with three being structural subunits and the others representing heme *a* synthesis enzymes and assembly facilitators.

ATP synthase is a complex and fascinating molecular motor. The detailed structure of the bacterial enzyme has been studied by crystallography and its function as a thirteen subunit reversible molecular rotary motor has been confirmed by Wang and Oster (1998). We found eight putative components of this complex in sugarcane (Table III).

Another important area of development is the identification of mitochondrial defects linked to genetic and degenerative diseases, *e.g.* aging and cancer (Wallace, 1999), this having led to the identification of many human homologues of genes linked to the function of, for example, cytochrome oxidase (Wallace, 1999) and other complexes. The human genes were instrumental in finding the putative orthologues of 53 sugarcane genes. Among them we found eleven nuclear subunits of complex I or NADH dehydrogenase (ubiquinone) and nine proteins related to complex III or ubiquinol:cytochrome c reductase (category 8, Table III). In plants antisense suppression of complex I activity through repression of the 55 kDa subunit results in healthy plants. However the reduced respiratory ability is insufficient for normal pollen development (Rasmusson *et*

al., 1998) pointing to the feasibility of complex I control of male fertility.

In category 11 (general respiratory ability) we have genes that are needed for respiratory competence but so far have not been linked to a specific complex. Other categories in Table III are lipid metabolism with 32 clusters and protein synthesis with 45 clusters. Protein synthesis components are synthesized at high levels and as a consequence are usually well represented in the cDNA libraries. Ribosomal proteins are an abundant class followed by the amino acid activating enzymes. A number of clusters are putative orthologues of genes linked to cell defense and programmed cell death (category 18, Table III). We left 48 clusters with activities listed as unclassified.

Our survey was based on 260,781 individual cDNA clones from 37 libraries. Rötig *et al.* (2000) used 340 *S. cerevisiae* mitochondrial proteins to scan the human genome for genes and ESTs and found 229 matches (E-value $\leq 10^{-5}$), of which 102 were EST clusters. The human genome database at the time had more than 2 million EST sequences and the total number of unique exons identified was over 600,000. Consequently our results, based on a smaller database, can be considered particularly successful. Our set of 367 clusters contains an estimated 145 full-length clones that are candidates for complete sequencing and homologous or heterologous functional studies (Hamel *et al.*, 1997). The E-value $\leq 10^{-10}$ threshold we adopted was chosen to maximize the likelihood that matches were significant and we also collected information on matches up to E-value $\leq 10^{-5}$ and found that the increase in the number of additional clusters found did not warrant a change in the threshold value. For example, with the 459 *S. cerevisiae* proteins used we obtained just 28 hits with E-values between 10^{-5} and 10^{-10} .

We are aware of the importance of comparing the sequences found with known counterparts from other organisms, and so far we have analyzed (results not shown) for five full-length cDNAs (the sugarcane putative orthologues of *BIO2*, *COX11*, *CYB2*, *YAH1* and *ODC2*) and have produced high quality complete sequences. The alignments were done with two or three other proteins and confirm the validity of an E-value based approach to collect a representative set of mitochondria-related nuclear genes. We hope to expand this more detailed analysis for a larger set of proteins and to use the alignments for construction of phylogenies able to better identify gene families and evolutionary relationships. We also repeated our analysis (using the 149 inner membrane *S. cerevisiae* driver-sequences and the 261 sequences from *Homo sapiens*) by doing a TBLASTN search against the CAP3 cluster consensi available at the SUCEST site. Using this approach, we recovered from the *S. cerevisiae* set 89 matches as opposed to 86 using phrap, while from the human set we recovered 181 matches compared with 179 phrap matches. So, for our purposes, it appears that the two clustering methods (phrap or

CAP3) yielded almost identical results, although there was an important change in the number of apparent full-length clones. The CAP3 assembly reduced the number of putative full-length clones: for the *S. cerevisiae* set CAP3 assembly produced 38 putative full-length clones as against 41 for phrap assembly, while for the human set CAP3 assembly produced 52 putative full-length clones compared to 77 with phrap assembly. The survey reported in this paper is, to our knowledge, the first genomic-wide EST study of mitochondria-related plant genes and will be a starting point for future and more focused studies on the molecular biology of mitochondrial physiology.

RESUMO

A mitocôndria funciona como uma usina geradora metabólica por meio da fosforilação oxidativa e tem sido alvo de um renovado interesse devido aos progressos no entendimento de sua biogênese e na descrição de novos papéis ligados à senescência, morte celular e montagem dos centros Fe/S. Uma análise global dos genes de planta ligados à esta organela é agora possível. A base de dados do projeto SUCEST foi examinada para detecção de ESTs com similaridade a genes nucleares relacionados às funções mitocondriais usando-se proteínas de *Saccharomyces cerevisiae*, *Homo sapiens* e *Arabidopsis thaliana*. Foram utilizadas 869 seqüências de proteínas para varrer o banco de ESTs do projeto SUCEST por meio do programa de busca de similaridade TBLASTN, sendo examinados 81.223 agrupamentos. Encontramos 367 agrupamentos com E-value $\leq 10^{-10}$ que representam os prováveis ortólogos em cana-de-açúcar dos genes correspondentes humanos, de levedura e de *Arabidopsis*. Encontramos produtos gênicos relacionados a todas as categorias funcionais ligados à atividade mitocondrial de maneira que este estudo serve de ponto de partida para a identificação dos genes de cana-de-açúcar envolvidos na biogênese e função da organela e para o estudo da estrutura e fisiologia destes genes.

ACKNOWLEDGMENTS

We thank the Brazilian agencies FAPESP (99/02859-0; 99/02844-2) and CNPq (301070/80) for support. The assistance of Edilson Gomes de Faria was invaluable. We also thank Jefferson Viude, Priscila M.S. Correa and Julio C. Moreira for technical help. Support was also provided at times by Alexandre R. Lamm, Raquel F. Guedes, Tatiana M. Pereira and Elaine Crespim. JHMT is recipient of a FAPESP fellowship.

REFERENCES

- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W. and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389-3402.
- Anon. (2000). The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796-815.

- Barros, M.H. and Nobrega, F.G.** (1999). *YAH1* of *Saccharomyces cerevisiae*: a new essential gene that codes for a protein homologous to human adrenodoxin. *Gene* 233:197-203.
- Beinert, H., Holm, R. H. and Munck, B.** (1997). Iron-sulfur clusters: Nature's modular, multipurpose structures. *Science* 277: 653-659.
- Costanzo, M.C., Crawford, M.E., Hirschman, J.E., Kranz, J.E., Olsen, P., Robertson, L.S., Skrzypek, M.S., Braun, B.R., Hopkins, K.L., Kondu, P., Lengieza, C., Lew-Smith, J.E., Tillberg, M. and Garrels, J.I.** (2001). YPD, PombePD, and WormPD: model organism volumes of the BioKnowledge library, an integrated resource for protein information. *Nucleic Acids Res.* 29: 75-79.
- Emanuelsson, O., Nielsen, H., Brunak, S. and von Heijne, G.** (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300: 1005-1016.
- El Moulidj, B., Duyckaerts, C., Lamotte-Brasseur, J., and Sluse, F.E.** (1997). Phylogenetic classification of the mitochondrial carrier family of *Saccharomyces cerevisiae*. *Yeast* 13: 573-581.
- Glaser E. and Dessi P.** (1999). Integration of the mitochondrial-processing peptidase into the cytochrome bc(1) complex in plants. *J. Bioenerg. Biomembr.* 31: 259-274.
- Glaser, E., Sjolung, S., Tanudji, M. and Whelan, J.** (1998). Mitochondrial protein import in plants - Signals, sorting targeting, processing and regulation. *Plant Mol. Biol.* 38:311-338.
- Goffeau, A.** (1996). Life with 6000 genes. *Science* 274: 562-567.
- Hägerhäll, C.** (1997). Succinate: quinone oxidoreductases. Variations on a conserved theme. *Biochim. Biophys. Acta.* 1320: 107-141.
- Hamel, P., Sakamoto, W., Wintz, H. and Dujardin, G.** (1997). Functional complementation of an oxal-yeast mutation identifies an *Arabidopsis thaliana* cDNA involved in the assembly of respiratory complexes. *Plant J.* 12: 1319-1327.
- Kispal, G., Csere, P., Prohl, C. and Lill, R.** (1999). The mitochondrial proteins Atm1p and Nfs1p are essential for biogenesis of cytosolic Fe/S proteins. *EMBO J.* 18: 3981-3989.
- Kowaltowski A.J. and Vercesi, A.E.** (1999). Mitochondrial damage induced by conditions of oxidative stress. *Free Rad. Biol.Med.* 26: 463-471.
- Laloi, M.** (1999). Plant mitochondrial carriers: an overview. *Cell. Molec. Life Sci.* 56: 918-944.
- Lange, H., Kaut, A., Kispal, G. and Lill, R.** (2000). A mitochondrial ferredoxin is essential for biogenesis of cellular iron-sulfur proteins. *Proc Natl Acad Sci USA* 97: 1050-1055.
- Lill, R. and Kispal, G.** (2000). Maturation of cellular Fe-S proteins: an essential function of mitochondria. *Trends Biochem. Sci.* 25: 352-356.
- Manzella, L., Barros, M.H. and Nobrega, F.G.** (1998). *ARH1* of *Saccharomyces cerevisiae*: a new essential gene that codes for a protein homologous to the human adrenodoxin reductase. *Yeast* 14: 839-846.
- Neupert, W.** (1997). Protein import into mitochondria *Annu. Rev. Biochem.* 66: 863-917.
- Nobrega, F.G., Nobrega, M.P. and Tzagoloff, A.** (1992). BCS1, a novel gene required for the expression of functional Rieske iron-sulfur protein in *Saccharomyces cerevisiae*. *EMBO J.* 11: 3821-3829.
- Ricquier, D. and Bouillaud, F.** (2000). The uncoupling protein homologues: UCP1, UCP2, UCP3, StUCP and AtUCP. *Biochem. J.* 345: 161-179.
- Rötig, A., Valnot, I., Mugnier, C., Rustin, P. and Munnich, A.** (2000). Screening HumanEST database for identification of candidate genes in respiratory chain deficiency. *Molec. Genet. Metab.* 69: 223-232.
- Rasmusson, A.C., Heiser, V., Zabaleta, E., Brennicke, A. and Grohmann, L.** (1998). Physiological, biochemical and molecular aspects of mitochondrial complex I in plants. *Biochim. Biophys. Acta* 1364: 101-111.
- Saraste, M.** (1999). Oxidative phosphorylation at the *fin de siècle*. *Science* 283: 1488-1492.
- Scharfe, C., Zaccaria, P., Hoertnagel, K., Jaksch, M., Klopstock, T., Dembowski, M., Lill R., Prokisch H., Gerbitz K.D., Neupert W., Mewes H.W. and Meitinger T.** (2000). MITOP, the mitochondrial proteome database: 2000 update. *Nucleic Acids Res.* 28: 155-158.
- Schatz, G.** (1995). Mitochondria: beyond oxidative phosphorylation. *Biochim. Biophys. Acta* 1271: 123-126.
- Souza, R.L., Green-Willms, N.S., Fox, T.D., Tzagoloff, A. and Nobrega, F.G.** (2000). Cloning and characterization of *COX18*, a *Saccharomyces cerevisiae* *PET* gene required for the assembly of cytochrome oxidase. *J. Biol. Chem.* 275: 14898-14902.
- Susin, A.A., Zamzami, N. and Kroemer, G.** (1998). Mitochondria as regulators of apoptosis: doubt no more. *Biochim. Biophys. Acta* 1366: 151-165.
- Telles, G. P., Braga, M. D. V., Dias, Z., Lin, T-L., Quitzau, J. A. A., da Silva, F. R. and Meidanis, J.** (2001). Bioinformatics of the sugarcane EST project. *Genetics and Mol. Biol.* 24 (1-4): 9-15.
- Telles, G. P. and da Silva, F. R.** (2001). Trimming and clustering sugarcane ESTs. *Genetics and Mol. Biol.* 24 (1-4): 17-23.
- Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R. and Yoshikawa, S.** (1996). The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 Å. *Science* 272: 1136-1144.
- Tzagoloff, A. and Dieckmann, C.L.** (1990). *PET* genes of *Saccharomyces cerevisiae*. *Microbiol. Rev.* 54: 211-225.
- Vettore, A. L., da Silva, F. R., Kemper, E. L., and Arruda, P.** (2001). The libraries that made SUCEST. *Genetics and Mol. Biol.* 24 (1-4): 1-7.
- Wallace, D.C.** (1999). Mitochondrial diseases in man and mouse. *Science* 283: 1482-1488.
- Wang, H. and Oster, G.** (1998). Energy transduction in the F1 motor of ATP synthase. *Nature* 396: 279-282.
- Xia, D., Yu, C.A., Kim, H., Xia, J.Z., Kachurin, A.M., Zhang, L., Yu, L. and Deisenhofer, J.** (1997). Crystal structure of the cytochrome bc1 complex from bovine heart mitochondria. *Science* 277: 60-66.
- Yaffe, M.P.** (1999). The machinery of mitochondrial inheritance and behavior. *Science* 283: 1493-1497.