

AMPDB: the Arabidopsis Mitochondrial Protein Database

Joshua L. Heazlewood and A. Harvey Millar*

Plant Molecular Biology Group, School of Biomedical and Chemical Sciences, The University of Western Australia, Crawley 6009, WA, Australia

Received August 12, 2004; Revised and Accepted September 30, 2004

ABSTRACT

The Arabidopsis Mitochondrial Protein Database is an Internet-accessible relational database containing information on the predicted and experimentally confirmed protein complement of mitochondria from the model plant *Arabidopsis thaliana* (<http://www.ampdb.bcs.uwa.edu.au/>). The database was formed using the total non-redundant nuclear and organelle encoded sets of protein sequences and allows relational searching of published proteomic analyses of Arabidopsis mitochondrial samples, a set of predictions from six independent subcellular-targeting prediction programs, and orthology predictions based on pairwise comparison of the Arabidopsis protein set with known yeast and human mitochondrial proteins and with the proteome of *Rickettsia*. A variety of precomputed physical-biochemical parameters are also searchable as well as a more detailed breakdown of mass spectral data produced from our proteomic analysis of Arabidopsis mitochondria. It contains hyperlinks to other Arabidopsis genomic resources (MIPS, TIGR and TAIR), which provide rapid access to changing gene models as well as hyperlinks to T-DNA insertion resources, Massively Parallel Signature Sequencing (MPSS) and Genome Tiling Array data and a variety of other Arabidopsis online resources. It also incorporates basic analysis tools built into the query structure such as a BLAST facility and tools for protein sequence alignments for convenient analysis of queried results.

INTRODUCTION

Mitochondria carry out a wide variety of biochemical processes within the eukaryotic cell. Their primary role is the

oxidation of organic acids via the tricarboxylic acid cycle and the synthesis of ATP. However, they also undertake transcription and translation, actively import proteins and metabolites from the cytosol, influence programmed cell death, and respond to cellular signals such as oxidative stress. In plants, mitochondria also perform many important secondary functions such as synthesis of nucleotides, amino acids, lipids and vitamins, participate in photorespiration and export organic acid intermediates for cellular biosynthesis [reviewed in (1)] (<http://www.aspb.org/publications/arabidopsis/>). These diverse range of roles, further underscore the need to fully understand how the mitochondrion functions in plants. Currently, much of the detailed knowledge is limited to a handful of biochemical pathways; whereas our understanding of mitochondrial biogenesis or how mitochondrial activity is perceived by the nucleus is less developed.

A major limitation in comprehending the full function of the mitochondria has been a lack of precise information about the protein components involved. Although a handful of mitochondrial proteins are encoded in its small organelle genome, the majority of proteins found in mitochondria are encoded in the nucleus and targeted to the organelle via signal sequences that often reside in the N-terminal portion of the protein. The subcellular location of a protein can, in theory, be predicted through commonly identified characteristics in these signal sequences. Several bioinformatics methods have been developed to predict the subcellular localization of a protein using both defined characteristics and machine-learning techniques [reviewed in (2)]. This has led to a number of publicly available programs that can be used to predict protein subcellular localization. MitoProt II (3), PSORT (4) and iPSORT (5) utilize predefined parameters of signal peptides to predict subcellular localization. The programs TargetP (6), Predotar (7) and SubLoc (8) utilize machine-learning techniques for subcellular predictions. A basic assessment of the mitochondrial localization capabilities of these prediction programs using the Arabidopsis nuclear proteome found that each prediction program over-predicts localization to this organelle (9). In contrast, an approach using a novel combination of these targeting prediction programs and comparative

*To whom correspondence should be addressed. Tel: +61 8 6648 7245; Fax: +61 8 6648 1148; Email: hmillar@cyllene.uwa.edu.au

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use permissions, please contact journals.permissions@oupjournals.org.

genomics analysed in 10 different eukaryotic genomes proposed much smaller putative mitochondrial proteomes (10). Phylogenetic profiling has also been utilized providing estimations of 660 nuclear-encoded mitochondrial proteins in *Caenorhabditis elegans* and an estimated yeast nuclear-encoded set of 630 mitochondrial proteins (11). The accuracy of these various prediction sets is currently unclear due to a lack of substantial experimental data for comparison, although recent technical advances have started to make an impact in this area. A proteome-scale epitope tagging analysis of yeast to determine protein localization estimated that 14% of the yeast proteome (~850 proteins) is potentially localized to the mitochondrion (12). Most recently, by using subcellular fractionation and proteomic techniques, large mitochondrial proteomes from human (13,14), yeast (15), mouse (16) and Arabidopsis (9) have identified 680, 750, 591 and 416 proteins, respectively. The capacity to significantly expand these sets using current technologies remains to be seen.

On the one hand, predicted sets will fail to find proteins with cryptic targeting mechanisms (false negatives) and may also incorrectly predict proteins to have targeting sequences (false positives). On the other hand, experimental sets have difficulties in identifying low abundance proteins and proteins only present during specific developmental stages or in specific cell types (false negatives). Moreover, contaminants in mitochondrial preparations can be inaccurately annotated as mitochondrial in origin (false positives). Relational analysis of the predicted and experimental sets is a valuable tool in the ongoing process of completely determining the mitochondrial proteome of different eukaryotic lineages.

DATABASE STRUCTURE, DATA SOURCES AND IMPLEMENTATION

The Arabidopsis Mitochondrial Protein Database (AMPDB) was constructed using the MySQL database server and interfaces through a custom-designed web interface utilizing Hypertext Preprocessor (PHP) forms. The database is housed on a Sun Fire v880 Server running Solaris 9 (Sun Microsystems). The non-redundant nuclear protein dataset utilized to populate the database was obtained from The Institute for Genomic Research (TIGR) (17) contained in the file ATH1.pep (release 5) comprising 28 952 non-redundant proteins. Arabidopsis mitochondrial [117 open reading frames (ORFs)] and chloroplast (87 ORFs) sets were obtained from GenBankTM. The AMPDB contains a total of 29 156 proteins. Experimentally localized mitochondrial proteins identified through proteomic means were obtained from 17 public resources and represent 496 unique identifications (Table 1). Information pertaining to more detailed proteomic information regarding analysis scores for mass spectral matches, number of matched peptides and number of experiments that are generated from ongoing experimentally produced data is also available (9). A set of 168 non-redundant Arabidopsis proteins with annotations indicating a mitochondrial localization and with significant homology matches to the TIGR set (ATH1) were obtained from SWISS-PROT (February 2004) (18) and incorporated into the database in an attempt to encompass proteins identified prior to current

Table 1. List of references, type of analyses and the total number of redundantly identified proteins experimentally determined and searchable in the AMPDB

Study	Proteins	Reference
2D-PAGE Proteome	83	(36)
2D-PAGE Proteome	44	(37)
TOM20	5	(38)
Oxidative stress	31	(39)
2D-PAGE Proteome	17	H. Eubel and H.P. Braun (unpublished data) ^a
3D-PAGE complexes	23	(40)
Complex I	32	(41)
F ₁ F ₀ ATP synthase	13	(42)
Carrier proteins	8	(43)
TOM complex	8	(44)
Ascorbate/glutathione cycle	5	(45)
Supercomplexes	19	(46)
Diagonal-PAGE	13	(47)
LC/MS/MS	416	(9)
Protein import components	17	(48)
Amino acid metabolism	15	(49)
TOC64	1	(50)
Mito membranes	114	(51)
Complex II+IV	26	(52)

^aLocated at <http://www.gartenbau.uni-hannover.de/genetik/AMPP/>

proteomic analyses. Primary attributes for proteins were produced using in-house scripts calculating molecular weight, grand average of hydropathicity (GRAVY) (19) and isoelectric point (pI) (20,21). Estimations of expressed sequence tag (EST) numbers for each chromosomal locus were obtained from The Arabidopsis Information Resource (TAIR) (22). Functional assignments were made using annotations associated with each protein entry and through homology-based comparisons with the SWISS-PROT protein database using BLAST (23). Predictions of subcellular localization were undertaken using TargetP v1.01 (6) with no cut-off set and 'Plant' option selected; Predotar v1.03 (7) with 'plant sequences' selected; MitoProt II v1.0a4 (3) with a DFM cut-off between 0.7 and 1.0; iPSORT (5) with 'Plant Protein' option selected; SubLoc (8) using the eukaryotic analysis component; and MITOPRED (24) using the Arabidopsis set (<http://mitopred.sdsc.edu>) with an 85% confidence cut-off. Targeting predictions were carried out on the TIGR set outlined above. Arabidopsis mitochondrial sequence orthologs to yeast, humans and Rickettsia were determined using the program INPARANOID for two-way best pairwise matches (25). The yeast mitochondrial protein comprises 552 proteins (26); the human mitochondrial protein set (615 proteins) was obtained as Supplementary Material from Taylor *et al.* (13); while the *Rickettsia prowazekii* Madrid E protein set comprising 834 entries was downloaded from the Comprehensive Microbial Resource at TIGR (27). Hyperlinks to other databases and resources for each protein details page (flatfile) comprise TAIR, TIGR, MIPS *Arabidopsis thaliana* Database (MAtdB) (28), *Arabidopsis thaliana* Plant Genome Database (AtGDB), Massively Parallel Signature Sequencing (MPSS) (29), The Plant Specific Database (30), ARAMEMNON (31), *Arabidopsis thaliana* Insertion Database (32), Salk Insertion Sequence Database (33) and Arabidopsis Tiling Array Transcriptome (34). An overview of the database is shown in Figure 1.

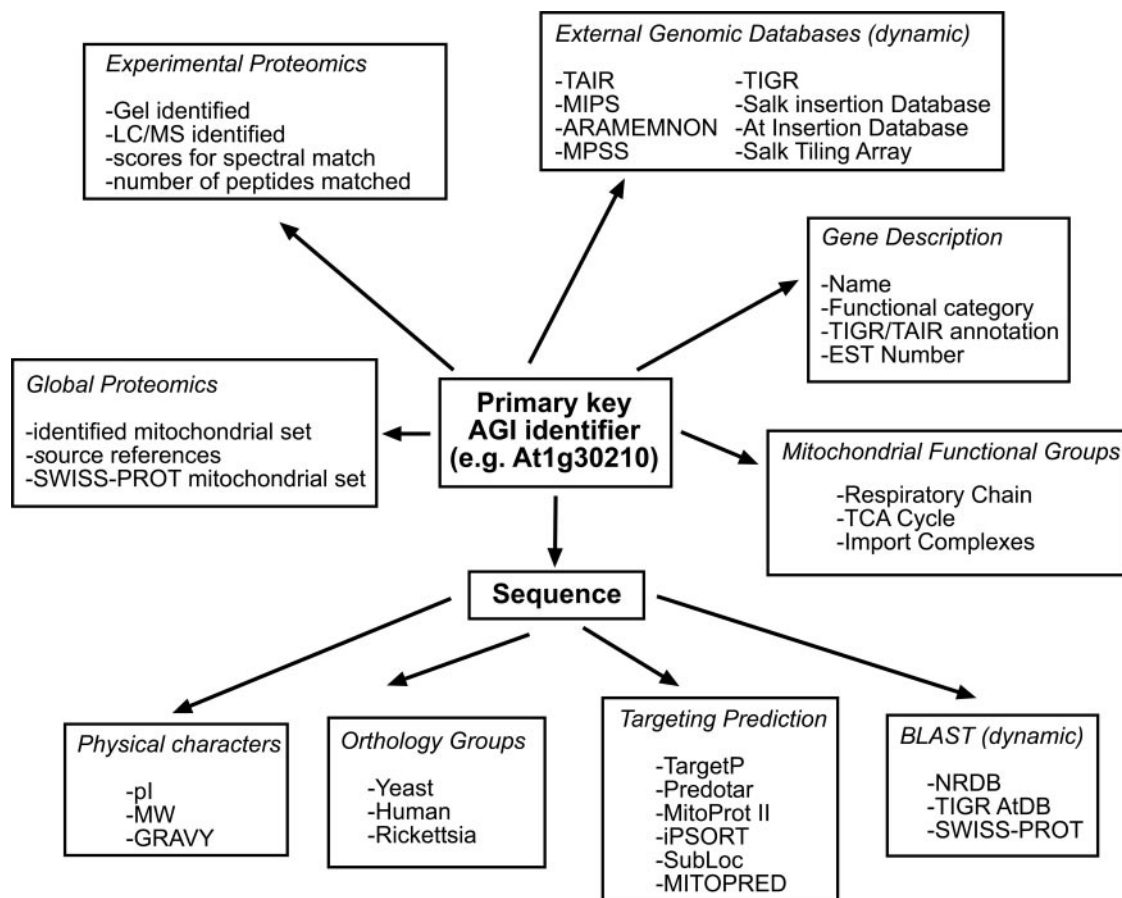


Figure 1. Overview of the data stored in the AMPD. Arrows indicate the relationship between the data types (AGI: Arabidopsis Genome Initiative).

THE ARABIDOPSIS MITOCHONDRIAL PROTEIN DATABASE INTERFACE

Based on a theoretical set of 29 156 non-redundant proteins, AMPDB uses the SQL language structure for data queries. The database houses protein sequences, a variety of physical-chemical data derived from these protein sequences, descriptions and names of genes. Through a custom web interface, AMPDB can be relationally queried with a series of mitochondrial-specific characteristics such as targeting presequence predictions, and orthology to yeast, human mitochondrial proteins and the mitochondrial progenitor Rickettsia. Also searchable are a variety of basic characteristics such as Arabidopsis gene loci number (AGI identifier), protein description, name, MW, pI, GRAVY, number of ESTs, chromosome location, broad functional category or specialized functional groups associated with recognized mitochondrial functions, metabolic pathways and protein complex component lists (Figure 2A). This interface enables complex data queries with no knowledge of SQL required. Querying of the database has been implemented via custom-built PHP forms. The query interface has been designed for ease of use with the primary interface consisting of pull down menus and simple text boxes. Query results are shown as a table of matches, each row containing protein AGI identifier and protein attribute data requested on the front page (Figure 2B). Multiple matches in the result table can be compared by ticking boxes to undertake sequence

alignment by CLUSTALW (35) (Figure 2C). Each protein match in the result table can be further viewed in a flatfile format (Figure 2D) where further information and hyperlinks to the well-established Arabidopsis databases, e.g. TIGR, TAIR and MIPS can be found.

HOW TO USE THE ARABIDOPSIS MITOCHONDRIAL PROTEIN DATABASE

In the query window (<http://www.ampdb.bcs.uwa.edu.au/>), any number of characteristics or sets can be selected to define an inclusive set for a relational query. This allows a large range of different types of searches to be conducted. Below are some examples of query classes that show how this database can help with particular questions.

Comparison of prediction programs

Targeting prediction programs are mostly used to determine the likelihood that a specific protein sequence is targeted to a particular subcellular structure. Usually only one is used or at best, two programs are used, one at a time. However, as this database contains pre-analysed predictions for all nuclear-encoded protein sequences from Arabidopsis, the set of mitochondrial proteins predicted by multiple targeting prediction programs can be rapidly compared. For example, 3182

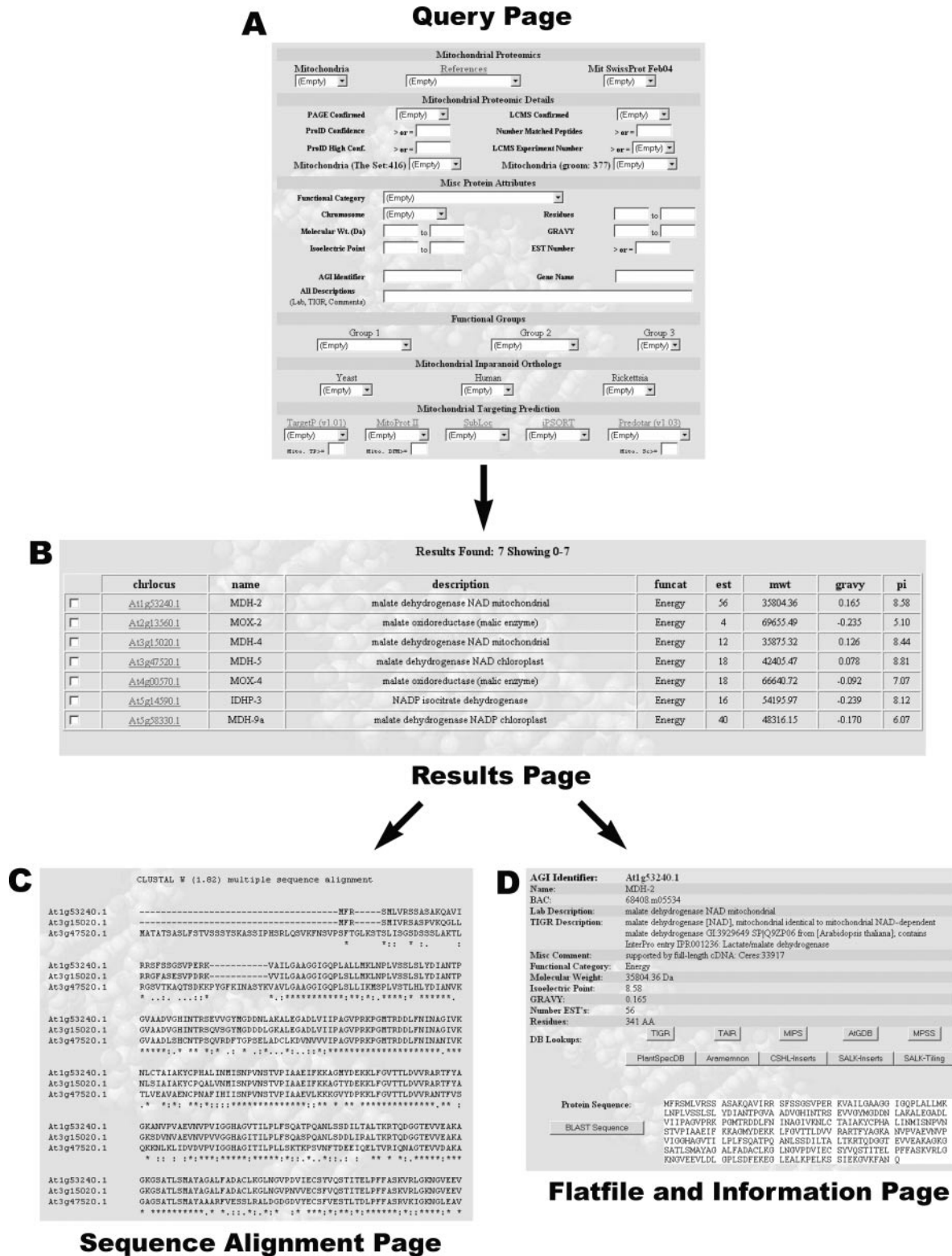


Figure 2. An example of an AMPD browser query. Links from the query choice (A) to a result page (B), and a branch to a sequence alignment of result matches (C) and to a details page (flatfile) (D) with hyperlinks to a variety of related resources.

proteins are predicted to be mitochondrial by TargetP, but this set is reduced to 1176 if MitoProt II and iPSORT are also queried to form an inclusive consensus set across all three predictors. Further, this type of search can be used to assess

the range of targeting predictions for the members of a gene family, to identify the most likely member with a particular subcellular location. For example, only one of the five monodehydroascorbate reductases (by description search) is

consistently identified as containing an organelle targeting presequence by multiple predictors.

Protein or protein families of interest

For an interest in proteins that use malate as a substrate, you could select proteins with a description including the word 'malate', which yields 42 entries from the entire Arabidopsis protein set (29 156). If you then select both malate and the current mitochondrial proteome set, the selection results in a set of proteins that we have experimentally identified in mitochondrial samples, a total of 7 of the 42 entries (Figure 2A and B).

Small, basic proteins found in mitochondria

For an interest in the physical characteristics of proteins, a query can be undertaken to assess small, basic proteins that have been experimentally found in Arabidopsis mitochondria. For pI between 9 and 14 and molecular mass between 1 and 15 000 Da, there are 44 proteins found in mitochondria using this search.

Conserved mitochondrial proteins across eukaryotic lineages

Owing to the evolutionary history of mitochondria among eukaryotes, some mitochondrial proteins and functions are highly conserved between organisms, while others are not. Searching with experimentally determined Arabidopsis mitochondrial proteins and their orthologs in yeast and humans, a set of 107 proteins are found. Not surprisingly, this set contains many proteins with core mitochondrial functions in metabolism. Notably, by adding yeast and human orthologs to the 'malate' search noted above, a total of 4 of the 42 entries that mention malate were found experimentally and are known orthologs of yeast and human mitochondrial proteins.

In this manner, very complex queries can be made to investigate many different questions ranging from expression levels (using ESTs or MPSS) to physical location on chromosomes (using AGI identifiers and chromosome selection) to physical characteristics of proteins (GRAVY, pI, MW) to orthology sets with other eukaryotic mitochondrial sets (using yeast, human and Rickettsia orthologs), and prediction and/or experimental verification of mitochondrial location. Such queries have been used to derive much of the comparative data recently highlighted by Heazlewood *et al.* (9).

FUTURE PROSPECTS

Incorporation of further experimental datasets is planned to enhance the value of this database in raising the level of higher order data analysis for mitochondrial research in plants. Incorporation of microarray datasets will allow the expression profile of genes encoding mitochondrial pathways, complexes and functional sets to be easily compared. Further, new prediction programs will be added, and as available, the results of high-throughput green fluorescent protein tagging experiments and protein-protein interaction studies will be incorporated. This will provide a relational search environment bringing together large datasets with a focus on questions pertaining to mitochondrial structure and function.

ACKNOWLEDGEMENTS

We are grateful to the contributions and comments from Julian Tonti-Filippini and A/Prof. James Whelan (Biochemistry Department, University of Western Australia) and for the computational support from Lindsay Campbell and the team at Sun Microsystems (Perth, Australia). This work is supported through grants provided by the Australian Research Council Discovery Program, an ARC QEII Research Fellowship to A.H.M. and a UWA Postdoctoral Research Fellowship to J.L.H.

REFERENCES

1. Millar, A.H., Day, D.A. and Whelan, J. (2004) Mitochondrial biogenesis and function in Arabidopsis. In Somerville, C.R. and Meyerowitz, E.M. (eds), *The Arabidopsis Book*. American Society of Plant Biologists, Rockville, MD.
2. Schneider, G. and Fechner, U. (2004) Advances in the prediction of protein targeting signals. *Proteomics*, **4**, 1571–1580.
3. Claros, M.G. and Vincens, P. (1996) Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur. J. Biochem.*, **241**, 779–786.
4. Nakai, K. and Horton, P. (1999) PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.*, **24**, 34–36.
5. Bannai, H., Tamada, Y., Maruyama, O., Nakai, K. and Miyano, S. (2002) Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics*, **18**, 298–305.
6. 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.
7. Small, I., Peeters, N., Legeai, F. and Lurin, C. (2004) Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics*, **4**, 1581–1590.
8. Hua, S. and Sun, Z. (2001) Support vector machine approach for protein subcellular localization prediction. *Bioinformatics*, **17**, 721–728.
9. Heazlewood, J.L., Tonti-Filippini, J.S., Gout, A.M., Day, D.A., Whelan, J. and Millar, A.H. (2004) Experimental analysis of the Arabidopsis mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins. *Plant Cell*, **16**, 241–256.
10. Richly, E., Chinnery, P.F. and Leister, D. (2003) Evolutionary diversification of mitochondrial proteomes: implications for human disease. *Trends Genet.*, **19**, 356–362.
11. Marcotte, E.M., Xenarios, I., van der Bliek, A.M. and Eisenberg, D. (2000) Localizing proteins in the cell from their phylogenetic profiles. *Proc. Natl Acad. Sci. USA*, **97**, 12115–12120.
12. Kumar, A., Agarwal, S., Heyman, J.A., Matson, S., Heidtman, M., Piccirillo, S., Umansky, L., Drawid, A., Jansen, R., Liu, Y. *et al.* (2002) Subcellular localization of the yeast proteome. *Genes Dev.*, **16**, 707–719.
13. Taylor, S.W., Fahy, E., Zhang, B., Glenn, G.M., Warnock, D.E., Wiley, S., Murphy, A.N., Gaucher, S.P., Capaldi, R.A., Gibson, B.W. *et al.* (2003) Characterization of the human heart mitochondrial proteome. *Nat. Biotechnol.*, **21**, 281–286.
14. Gaucher, S.P., Taylor, S.W., Fahy, E., Zhang, B., Warnock, D.E., Ghosh, S.S. and Gibson, B.W. (2004) Expanded coverage of the human heart mitochondrial proteome using multidimensional liquid chromatography coupled with tandem mass spectrometry. *J. Proteome Res.*, **3**, 495–505.
15. Sickmann, A., Reinders, J., Wagner, Y., Joppich, C., Zahedi, R., Meyer, H.E., Schonfisch, B., Perschil, I., Chacinska, A., Guiard, B. *et al.* (2003) The proteome of *Saccharomyces cerevisiae* mitochondria. *Proc. Natl Acad. Sci. USA*, **100**, 13207–13212.
16. Mootha, V.K., Bunkenborg, J., Olsen, J.V., Hjerrild, M., Wisniewski, J.R., Stahl, E., Bolouri, M.S., Ray, H.N., Sihag, S., Kamal, M. *et al.* (2003) Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell*, **115**, 629–640.
17. Wortman, J.R., Haas, B.J., Hannick, L.I., Smith, R.K., Jr, Maiti, R., Ronning, C.M., Chan, A.P., Yu, C., Ayele, M., Whitelaw, C.A. *et al.* (2003) Annotation of the Arabidopsis genome. *Plant Physiol.*, **132**, 461–468.

18. Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I. *et al.* (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.*, **31**, 365–370.
19. Kyte, J. and Doolittle, R. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, **157**, 105–132.
20. Bjellqvist, B., Basse, B., Olsen, E. and Celis, J.E. (1994) Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. *Electrophoresis*, **15**, 529–539.
21. Bjellqvist, B., Hughes, G.J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J.C., Frutiger, S. and Hochstrasser, D. (1993) The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis*, **14**, 1023–1031.
22. Rhee, S.Y., Beavis, W., Berardini, T.Z., Chen, G., Dixon, D., Doyle, A., Garcia-Hernandez, M., Huala, E., Lander, G., Montoya, M. *et al.* (2003) The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. *Nucleic Acids Res.*, **31**, 224–228.
23. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tools. *J. Mol. Biol.*, **215**, 403–410.
24. Guda, C., Guda, P., Fahy, E. and Subramaniam, S. (2004) MITOPRED: a web server for the prediction of mitochondrial proteins. *Nucleic Acids Res.*, **32** (Web Server issue), W372–W374.
25. Remm, M., Storm, C.E.V. and Sonnhammer, E.L.L. (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.*, **314**, 1041–1052.
26. Schon, E.A. (2001) Gene products present in mitochondria of yeast and animal cells. In Wilson, L. and Matsudaira, P. (eds), *Mitochondria*. Academic Press, San Diego, CA, Vol. 65, pp. 463–482.
27. Peterson, J.D., Umayam, L.A., Dickinson, T., Hickey, E.K. and White, O. (2001) The Comprehensive Microbial Resource. *Nucleic Acids Res.*, **29**, 123–125.
28. Schoof, H., Zaccaria, P., Gundlach, H., Lemcke, K., Rudd, S., Kolesov, G., Arnold, R., Mewes, H.W. and Mayer, K.F. (2002) MIPS *Arabidopsis thaliana* Database (MATDB): an integrated biological knowledge resource based on the first complete plant genome. *Nucleic Acids Res.*, **30**, 91–93.
29. Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M. *et al.* (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.*, **18**, 630–634.
30. Gutiérrez, R.A., Larson, M.D. and Wilkerson, C. (2004) The plant-specific database. Classification of Arabidopsis proteins based on their phylogenetic profile. *Plant Physiol.*, **135**, 1888–1892.
31. Schwacke, R., Schneider, A., van der Graaff, E., Fischer, K., Catoni, E., Desimone, M., Frommer, W.B., Flugge, U.I. and Kunze, R. (2003) ARAMEMNON, a novel database for Arabidopsis integral membrane proteins. *Plant Physiol.*, **131**, 16–26.
32. Pan, X., Liu, H., Clarke, J., Jones, J., Bevan, M. and Stein, L. (2003) ATIDB: *Arabidopsis thaliana* insertion database. *Nucleic Acids Res.*, **31**, 1245–1251.
33. Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R. *et al.* (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science*, **301**, 653–657.
34. Yamada, K., Lim, J., Dale, J.M., Chen, H., Shinn, P., Palm, C.J., Southwick, A.M., Wu, H.C., Kim, C., Nguyen, M. *et al.* (2003) Empirical analysis of transcriptional activity in the Arabidopsis genome. *Science*, **302**, 842–846.
35. Thompson, J.D., Higgins, D.G. and 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.
36. Millar, A.H., Sweetlove, L.J., Giege, P. and Leaver, C.J. (2001) Analysis of the Arabidopsis mitochondrial proteome. *Plant Physiol.*, **127**, 1711–1727.
37. Kruff, V., Eubel, H., Jansch, L., Werhahn, W. and Braun, H.P. (2001) Proteomic approach to identify novel mitochondrial proteins in Arabidopsis. *Plant Physiol.*, **127**, 1694–1710.
38. Werhahn, W., Niemeyer, A., Jansch, L., Kruff, V.V., Schmitz, U.K. and Braun, H.P. (2001) Purification and characterization of the preprotein translocase of the outer mitochondrial membrane from Arabidopsis. Identification of multiple forms of TOM20. *Plant Physiol.*, **125**, 943–954.
39. Sweetlove, L.J., Heazlewood, J.L., Herald, V., Holtzapffel, R., Day, D.A., Leaver, C.J. and Millar, A.H. (2002) The impact of oxidative stress on Arabidopsis mitochondria. *Plant J.*, **32**, 891–904.
40. Werhahn, W. and Braun, H.P. (2002) Biochemical dissection of the mitochondrial proteome from *Arabidopsis thaliana* by three-dimensional gel electrophoresis. *Electrophoresis*, **23**, 640–646.
41. Heazlewood, J.L., Howell, K.A. and Millar, A.H. (2003) Mitochondrial complex I from Arabidopsis and rice: orthologs of mammalian and fungal components coupled with plant-specific subunits. *Biochim. Biophys. Acta*, **1604**, 159–169.
42. Heazlewood, J.L., Whelan, J. and Millar, A.H. (2003) The products of the mitochondrial *orf25* and *orfB* genes are F₀ components in the plant F₁F₀ ATP synthase. *FEBS Lett.*, **540**, 201–205.
43. Millar, A.H. and Heazlewood, J.L. (2003) Genomic and proteomic analysis of mitochondrial carrier proteins in Arabidopsis. *Plant Physiol.*, **131**, 443–453.
44. Werhahn, W., Jansch, L. and Braun, H.P. (2003) Identification of novel subunits of the TOM complex from *Arabidopsis thaliana*. *Plant Physiol. Biochem.*, **41**, 407–416.
45. Chew, O., Whelan, J. and Millar, A.H. (2003) Molecular definition of the ascorbate-glutathione cycle in Arabidopsis mitochondria reveals dual targeting of antioxidant defenses in plants. *J. Biol. Chem.*, **278**, 46869–46877.
46. Eubel, H., Jansch, L. and Braun, H.P. (2003) New insights into the respiratory chain of plant mitochondria. Supercomplexes and a unique composition of Complex II. *Plant Physiol.*, **133**, 274–286.
47. Herald, V.L., Heazlewood, J.L., Day, D.A. and Millar, A.H. (2003) Proteomic identification of divalent metal cation binding proteins in plant mitochondria. *FEBS Lett.*, **537**, 96–100.
48. Lister, R., Chew, O., Lee, M.N., Heazlewood, J.L., Clifton, R., Parker, K.L., Millar, A.H. and Whelan, J. (2004) A transcriptomic and proteomic characterization of the Arabidopsis mitochondrial protein import apparatus and its response to mitochondrial dysfunction. *Plant Physiol.*, **134**, 777–789.
49. Taylor, N.L., Heazlewood, J.L., Day, D.A. and Millar, A.H. (2004) Lipoic acid-dependent oxidative catabolism of α -keto acids in mitochondria provides evidence for branched chain amino acid catabolism in Arabidopsis. *Plant Physiol.*, **134**, 838–848.
50. Chew, O., Lister, R., Qbadou, S., Heazlewood, J.L., Soll, J., Schleiff, E., Millar, A.H. and Whelan, J. (2004) A plant outer mitochondrial membrane protein with high amino acid sequence identity to a chloroplast protein import receptor. *FEBS Lett.*, **557**, 109–114.
51. Brugière, S., Kowalski, S., Ferro, M., Seigneurin-Berny, D., Miras, S., Salvi, D., Ravanel, S., D'Hérin, P., Garin, J., Bourguignon, J. *et al.* (2004) The hydrophobic proteome of mitochondrial membranes from Arabidopsis cell suspensions. *Phytochemistry*, **65**, 1693–1707.
52. Millar, A.H., Eubel, H., Jansch, L., Kruff, K., Heazlewood, J.L. and Braun, H.P. (2004) Mitochondrial cytochrome *c* oxidase and succinate dehydrogenase complexes contain plant-specific subunits. *Plant Mol. Biol.*, **56**, 77–89.