ACTA BIOLOGICA CRACOVIENSIA Series Botanica 45/1: 119–124, 2003

THE INVOLVEMENT OF ORGANELLES IN PLANT SEXUAL REPRODUCTION: A POST-GENOMIC APPROACH

RICHARD BERTHOMÉ, NICOLE FROGER, SOPHIE HIARD, HERVÉ BALASSE, ALFRED MARTIN-CANADELL, AND FRANÇOISE BUDAR

Station de Génétique et d'Amélioration des Plantes, INRA, route de Saint Cyr, 78026 Versailles, France

Received July 7, 2002; revision accepted November 20, 2002

To study mitochondria and plastid involvement in plant development, particularly in sexual reproduction, we made use of the *Arabidospis thaliana* T-DNA insertion collection developed in Versailles. Mutants affected in the nuclear genes that encode proteins predicted to be targeted to organelles were identified using two complementary strategies. In the first (forward genetics), mutants chosen for their sterile or gametophytic lethal phenotype were screened for T-DNA insertion in genes encoding mitochondrial or plastid proteins after systematic sequencing of the Flanking Sequence Tag (FST). The second (reverse genetics) enabled us to identify other mutants using the following tools: systematic *A. thaliana* proteome analysis, bioinformatics software to predict the sub-cellular localization of putative proteins, and the FST sequencing program FlagDB. Preliminary results for the first set of 82 putative mutants are presented and discussed.

Key words: Reproduction, sterility, mitochondria, plastids, protein targeting, T-DNA insertion mutants.

INTRODUCTION

Plant cells possess three kinds of organelles: nuclei, plastids and mitochondria. Their functions are essential to the cell and they have their own genomes. Despite this, more than 90% of the proteins used for biogenesis and the functions of plastids and mitochondria are nuclear-encoded, translated in the cytoplasm, and addressed to the organelles by targeting information encrypted in the protein sequence. These proteins are estimated to represent 10% of the total amount of nuclear-encoded proteins (Abdallah et al., 2000).

In the yeast *Saccharomyces cerevisae*, mutants affected in nuclear genes encoding mitochondrial proteins have been very useful in studying the processes controlling mitochondria biogenesis and functions (Contamine and Picard, 2000). This has been made possible by the fact that mutant yeasts have a facultative aerobic phenotype. In higher plants, however, such mutants are much more difficult to obtain because of the lethality of respiratory dysfunction at the sporophytic or gametophytic stage. However, some mutant plants affected in the nuclear genes that encode mitochondrial proteins have been described (Brangeon et al., 2000; Liu et al., 2001; Skinner et al., 2001). Cytoplasmic male sterility (CMS) is another well-documented phenotype characterized by mitochondrial or chloroplastic DNA recombination leading to the appearance of new organellar open reading frames (ORFs). The peptide products of these new ORFs act either at the RNA or at the protein level within organelles, inducing male sterility or alteration of floral development. In general, nuclear restorer genes that suppress CMS specifically modify expression of the mitochondrial CMS-associated regions, but not other mitochondrial genes (for review: Budar and Pelletier, 2001). However, a limited number of nuclear genes that restore fertility have been identified up to now (Cui et al., 1996). These examples underline the major significance of interactions between organ-

PL ISSN 0001-5296

elles and the nucleus in certain aspects of plant sexual reproduction. To gain insight into the mechanisms involved in these processes, we studied the role of organelles in this specific stage of plant development. For this purpose, we used bioinformatics tools and the *Arabidospis thaliana* T-DNA insertion collection from Versailles (Bechtold et al., 1993) in two complementary strategies of forward and reverse genetics. Here we describe the developed strategies, and present preliminary results.

MATERIALS AND METHODS

PLANT MATERIAL

Mutant lines of *A. thaliana* (L.) ecotype Wassilewskaja (Ws) used in our approaches were isolated from the Versailles collection of T-DNA insertion mutants generated by Bechtold et al. (1993). Mutants were screened for their sterile or gametophyte lethal phenotypes by members of the *Arabidopsis* reproduction group in our institute (Bonhomme et al., 1998; Mercier et al., 2001).

IDENTIFICATION OF T-DNA INSERTION MUTANTS

In the forward genetics approach (Fig. 1a), mutants chosen for their sterile or gametophytic lethal phenotype were screened for the localization of the T-DNA insertion as follows: systematic sequencing of the Flanking Sequence Tag (FST) (Balzergue et al., 2001), followed by a standard nucleotide-nucleotide blastn search of the nonredundant nucleotide databases (GenBank, EMBL, DDBJ, PDB sequences) (Madden et al., 1996; www.ncbi.nlm.nih.gov/BLAST/). From these insertion sites we identified putative protein sequences, which were then analyzed with four software packages. These packages were recently developed to predict the cellular localization of unknown or unannotated proteins, to gain some indication of their function. They recognize the N-terminal presequences of classically targeted precursor proteins. We used TargetP (V1.01), known in the literature (Emanuelsson et al., 2000) to be the most reliable analyzer of the N-terminal sorting signal, and which predicts targeting of protein sequences to chloroplasts, mitochondria and the secretory system (www.cbs.dtu.dk/services/TargetP/). To maximize the number of proteins putatively targeted to organelles, no specificity cutoff was used in submitting the protein sequences to TargetP. Predotar (V 0.5), constructed by I. Small (URGV, INRA; www.inra.fr/Internet/Produits/Predotar/) is a very powerful package that distinguishes and efficiently discriminates mitochondrial and plastidic targeting sequences. We also used MitoProt, predicting mitochondrial targeting and presequence cleavage (Claros et al., 1996; www.mips.biochem.mpg. de/cgi-bin/proj/medgen/mitofilter), and IPSORT, which is a new version of PSORT, with prediction accuracies very close to TargetP (Bannai et al., 2002; www.HypothesisCreator.net/iPSORT/). To maximize the accuracy of the predictions, we selected only mutants with the T-DNA insertion in genes encoding proteins predicted to be localized in organelles with a score above 0.5 in three out of the four packages.

The second approach (reverse genetics) (Fig. 1b) is based on the analysis of the full genome sequence of *Arabidopsis* that was recently published and annotated (The Arabidopsis Genome Initiative, 2000). In collaboration with Small (URGV, INRA), on the basis of the predicted *A. thaliana* proteome and using the latest version of the prediction targeting software PREDOTAR (V1), we established a list of genes encoding proteins putatively addressed to mitochondria. We used this new version of PREDOTAR because it is much more stringent than the previous one (V 0.5) and even more stringent than TargetP. Moreover, in contrast to the forward genetic approach, we chose a targeted probability cutoff above 0.5 to limit the number of proteins selected. We looked for mutant plants in which one of these genes was disrupted by a T-DNA insertion using the FlagDB program (Samson et al., 2002).

RESULTS

SEARCH FOR STERILE OR GAMETOPHYTIC MUTANTS: FORWARD GENETICS (FIG. 1a)

The T-DNA mutant collection available in Versailles contains 55,500 independent lines; of these, 30,719 have been individually observed for reduction of fertility. In compiling the results obtained from these fertility screens (Bonhomme et al., 1998; Mercier et al., 2001), about 1.9% (575) of the total number of examined lines showed altered fertility or a gametophytic viability defect. Using the Flanking Sequence of the T-DNA insertion sequencing program FlagDB, genomic sequences of nuclear genes were identified for 540 mutants. We performed bioinfomatic analyses on the corresponding encoded proteins using software packages for predicting targeting to organelles. A total 38 mutants were finally selected (6.8% of the total number of mutants). Of these, 18 correspond to genes encoding proteins potentially addressed to plastids,

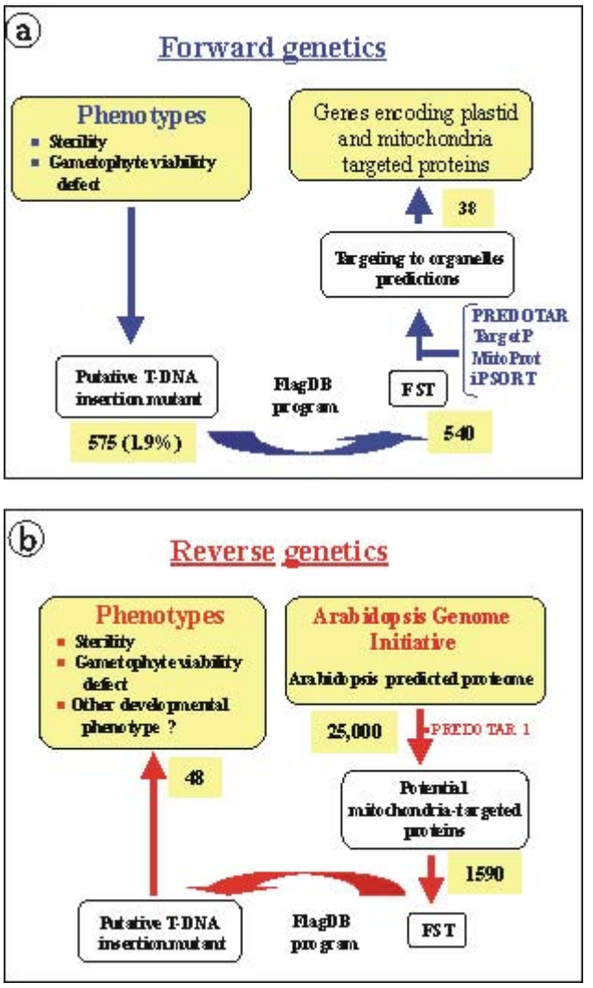

Fig. 1. Two complementary strategies (**a, b**) used to identify reproduction mutants corresponding to genes encoding proteins potentially addressed to organelles.

and 20 to genes encoding proteins potentially addressed to mitochondria (Fig. 2).

IDENTIFICATION OF GENES ENCODING MITOCHONDRIAL PROTEINS IN *ARABIDOPSIS THALIANA* NUCLEAR GENOME: REVERSE GENETICS (FIG. 1b)

Analysis of the predicted *A. thaliana* proteome with PREDOTAR (V 1) enabled us to identify 1590 genes encoding proteins potentially exported to mitochondria. Using the FlagDB program, a T-DNA insertion was found in 46 of these genes. Of these 46 lines, 2 correspond to mutants already detected with the forward genetics strategy, and the other 44 are considered to be new potential mutants (Fig. 2).

Altogether, using these two complementary strategies, 82 putative mutants were selected. Figure 2 summarizes these results and is organized according to the metabolic pathways in which the corresponding encoded proteins are predicted to be involved. Analyses are currently in progress to check the linkage between T-DNA insertion and phenotype. Encoded proteins corresponding to verified mutants will be tested for their real targeting to organelles by biolistics. Functional analyses will then be undertaken to clarify the roles of the identified proteins in plant development and sexual reproduction. Several promising mutants have already been isolated, allowing us to validate our approach.

DISCUSSION

In their study of the *Arabidopsis* mitochondrial proteome using two-dimensional gel separation of *Arabidopsis* cell culture mitochondrial proteins, Millar et al. (2001) excised and identified 170 proteins. Analysis of their targeting using bioinformatics software to predict their localization revealed significant variation in the predictions, and also a lack of targeting prediction for several of them known to be mitochondrial. These results obtained with Mitoprot, PSORT and TargetP suggest that at present one program alone cannot provide a complete, high-confidence analysis of the nuclear-encoded components of plant mitochondrial and plastid proteomes. Numerous examples of targeting to plant mitochondria by a means other than N-terminal extensions, and the limited use of characterized plant mitochondrial proteins in the construction of these software packages, might exacerbate this problem (Sjoling and Glaser, 1998). In view of these drawbacks, we decided to systematically verify the real targeting of the predicted proteins to organelles. For this purpose and for each mutant selected, transient expression assays on *Nicotiana benthamiana* leaves will be done using GFP fusion constructs.

With our reverse genetics approach we obtained only 46 FST out of 1590 genes potentially encoding proteins addressed to mitochondria. Forward genetics would also have been expected to detect more sterile mutants having a T-DNA insertion in a gene encoding a protein putatively targeted to organelles.

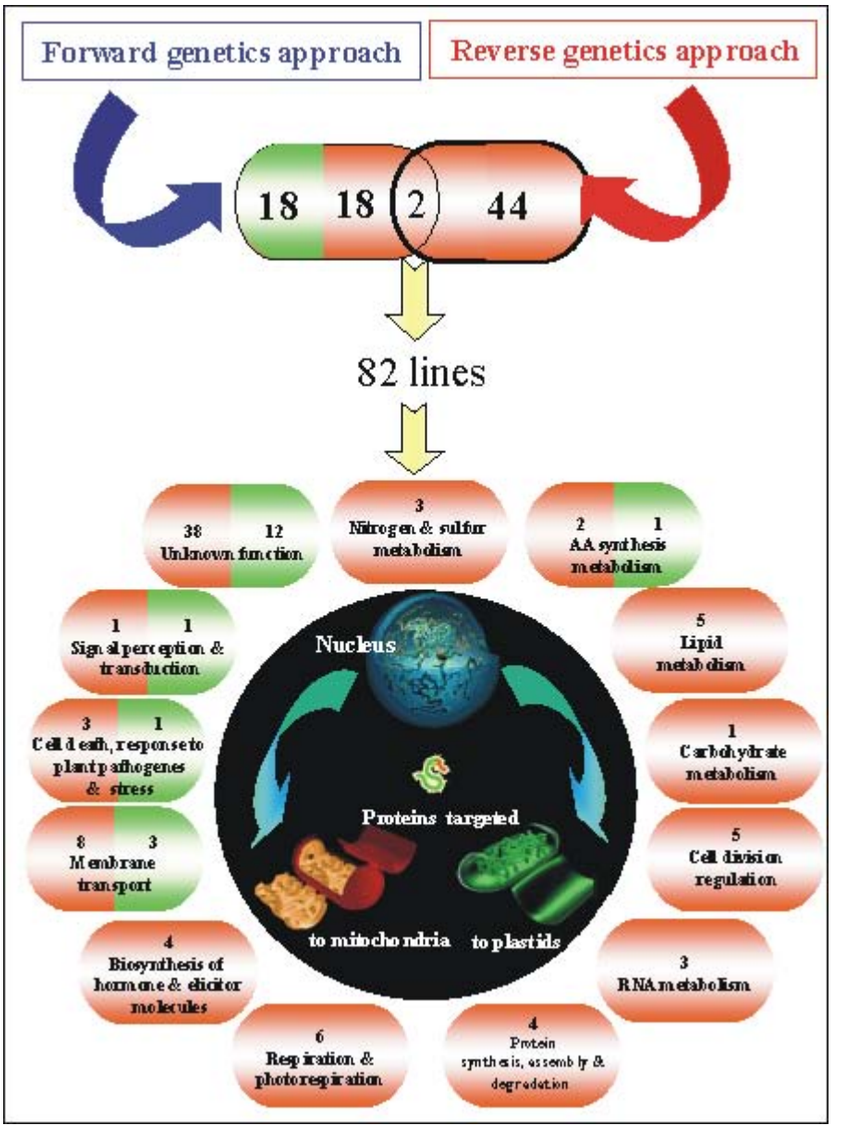

Fig. 2. Preliminary results obtained from our post-genomic approach. The 82 mutants selected are distributed according to the putative function of the proteins corresponding to the disrupted genes. Red – protein predicted to be targeted to mitochondria; green – protein predicted to be localized within proplastids.

These results might be explained by the crucial role of mitochondria in plant life. Indeed, we should expect that a disruption of a gene encoding an essential organellar protein will be lethal either at both the gametophytic male and female stages (forward genetics approach) or at the sporophytic stage, thus explaining the low number of FST lines identified. Bonhomme et al. (1998) put forward several hypotheses to explain the limited number of gametophytic defective mutants they identified. One possibility is that a gene expressed in gametophytes might be preferentially duplicated and thus could not be selected using a gene disruption strategy. As gametophytes are tightly connected to sporophytic tissues, metabolic supplementation could counteract the mutation effect.

Only two mutants were identified by both strategies. This result could be explained by the

greater stringency of the new PREDOTAR version compared to the oldest one used in the forward genetics approach and to TargetP (Small, personal communication). Supporting this explanation, we found that when we resubmitted the protein sequences detected in our forward genetics approach to PREDOTAR I, the major part of them were rejected from our screen (data not shown).

Altogether we detected 82 putative mutants corresponding to genes encoding proteins potentially targeted to plastids and mitochondria. If we consider the predicted function of the products involved (Fig. 2), we observe that the mutants detected by both strategies are distributed in each of the thirteen pathways and that an important part of the putative proteins encoded are of unknown function. In previous studies on sorghum CMS, Chen et al. (1995) showed the possible involvement of a deletion in the *rpoC2* plastid gene. Because of the ambiguity of the targeting predictions obtained with the software, in our forward approach we decided to enlarge our screening to mitochondria and plastids potentially encoding proteins, even though no examples implying proplastids in CMS have been described thus far. The aim was to maximize our chance of detecting interesting mutants affecting sexual reproduction. In this way we obtained 18 putative mutants encoding proteins that might be involved in the following pathways: membrane transport, protein synthesis, signal perception-transduction, and programmed cell death.

So far the preliminary results presented here are encouraging evidence that our approach can help identify and analyze original mutants in plant development and reproduction.

ACKNOWLEDGEMENTS

We thank Dr. I. Small for his contribution to the bioinformatics analyses, and all members of the *Arabidopsis* reproduction group for screening the data and for stimulating discussions.

REFERENCES

ABDALLAH F, SALAMINI F, and LEISTER D. 2000. A prediction of the size and evolutionary origin of the proteome of chloroplasts of *Arabidopsis*. *Trends in Plant Science* 5: 141–142.

BALZERGUE S, DUBREUCQ B, CHAUVIN S, LE-CLAINCHE I, LE BOULAIRE F, DE ROSE R, SAMSON F, BIAUDET V, LECHARNY A, CRUAUD C, WEISSENBACH J, CABOCHE M, and LEPINIEC L. 2001. Improved PCR-walking for large-scale isolation of plant T-DNA borders. *Biotechniques* 30: 496–498, 502, 504.

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.

BECHTOLD N, ELLIS J, and PELLETIER G. 1993. In planta, *Agrobacterium* mediated gene transfer by integration of adult *Arabidopsis thaliana* plants. *Comptes Rendus de l' Academie des Sciences Paris* 316: 1194–1199.

BONHOMME S, HORLOW C, VEZON D, DE LAISSARDIRE S, GUYON A, FÉRAULT M, MARCHAND M, BECHTOLD N, and PELLETIER G. 1998. T-DNA mediated disruption of essential gametophytic genes in *Arabidopsis* is unexpectedly rare and cannot be inferred from segregation distortion alone. *Molecular and General Genetics* 260: 444–452.

BRANGEON J, SABAR M, GUTIERRES S, COMBETTES B, BOVE J, GENDY C, CHETRIT P, DES FRANCS-SMALL CC, PLA M, VEDEL F, and DE PAEPE R. 2000. Defective splicing of the first nad4 intron is associated with lack of several complex I subunits in the *Nicotiana sylvestris* NMS1 nuclear mutant. *Plant Journal* 21: 269–280.

BUDAR F, and PELLETIER G. 2001. Male sterility in plants: occurrence, determinism, significance and use. *Comptes Rendus de l' Academie des Sciences Serie III-Sciences de la Vie-Life Sciences* 324: 543–550.

CHEN Z, SCHERTZ KF, MULLET JE, DUBELL A, and HART GE. 1995. Characterization and expression of the *rpoC2* in CMS and fertile lines of sorghum. *Plant Molecular Biology* 28: 799–809.

CLAROS MG, and VINCENS P. 1996. Computational method to predict mitochondrially imported proteins and their targeting sequences. *European Journal of Biochemistry* 241: 779–786.

CONTAMINE V, and PICARD M. 2000. Maintenance and integrity of the mitochondrial genome: a plethora of nuclear genes in the budding yeast. *Microbiology and Molecular Biology Reviews* 64: 281–315.

CUI X, WISE RP, and CHNABLE PS. 1996. The *rf2* nuclear restorer gene of male-sterile T-cytoplasm maize. *Science* 272: 1334–1336.

EMANUELSSON O, NIELSEN H, BRUNAK S, and VON HEIJNE G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology* 300: 1005–1016.

LIU F, CUI X, HORNER HT, WEINER H, and SCHNABLE PS. 2001. Mitochondrial aldehyde dehydrogenase activity is required for male fertility in maize. *Plant Cell* 13: 1063–1078.

MADDEN TL, TATUSOV RL, and ZHANG J. 1996. Applications of network BLAST server. *Methods in Enzymology* 266: 131–141.

MERCIER R, GRELON M, VEZON D, HORLOW C, and PELLETIER G. 2001. How to characterize meiotic functions in plants. *Biochimie* 83: 1023–1028.

MILLAR AH, SWEETLOVE LJ, GIEGÉ P, and LIEVER CJ. 2001. Analysis of the *Arabidopsis* mitochondrial proteome. *Plant Physiology* 127: 1711–1727.

SAMSON F, BRUNAUD V, BALZERGUE S, DUBREUCQ B, LEPINIEC L, PELLETIER G, CABOCHE M, and LECHARNY A. 2002. FLAGdb/FST: a database of mapped flanking insertion sites (FSTs) of *Arabidopsis thaliana* T-DNA transformants. *Nucleic Acids Research* 30: 94–97.

SJOLING S, and GLASER E. 1998. Mitochondrial targeting peptides in plants. *Trend in Plant Science* 3: 136–140.

SKINNER DJ, BAKER SC, MEISTER RJ, BROADHVEST J, SCHNEITZ K, and GASSER CS. 2001. The *Arabidopsis* HUELLENLOS gene, which is essential for normal ovule development, encodes a mitochondrial ribosomal protein. *Plant Cell* 13: 2719–2730.

THE ARABIDOPSIS GENOME INITIATIVE. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.