Mol Genet Genomics (2007) 277:631–646
DOI 10.1007/s00438-007-0214-4

ORIGINAL PAPER

Comparative survey of plastid and mitochondrial targeting properties of transcription factors in Arabidopsis and rice

Rainer Schwacke · Karsten Fischer · Bernd Ketelsen · Karin Krupinska · Kirsten Krause

Received: 8 June 2006 / Accepted: 14 January 2007 / Published online: 13 February 2007

Abstract A group of nuclear transcription factors, the Whirly proteins, were recently shown to be targeted also to chloroplasts and mitochondria. In order to find out whether other proteins might share this feature, an *in silico*-based screening of transcription factors from Arabidopsis and rice was carried out with the aim of identifying putative N-terminal chloroplast and mitochondrial targeting sequences. For this, the individual predictions of several independent programs were combined to a consensus prediction using a naïve Bayes method. This consensus prediction shows a higher specificity at a given sensitivity value than each of the single programs. In both species, transcription factors from a variety of protein families that possess putative N-terminal plastid or mitochondrial target peptides as well as nuclear localization sequences, were found. A search for homologues within members of the AP2/EREBP protein family revealed that target peptide-containing proteins are conserved among monocotyledonous and dicotyledonous species. Fusion of one of these proteins to GFP revealed, indeed, a dual targeting activity of this protein. We propose that dually targeted transcription factors might be involved in the communication between the nucleus and the organelles in plant cells. We further discuss how recent results on the physical interaction between the organelles and the nucleus could have significance for the regulation of the localization of these proteins.

Keywords AP2/EREBP proteins · Chloroplasts · Dual-targeting · Mitochondria · Nucleus · Transcription factors

Abbreviations

At	*Arabidopsis thaliana*
cTP	Chloroplast targeting peptide
GFP	Green fluorescent protein
mTP	Mitochondrial targeting peptide
NLS	Nuclear localization sequence
Os	*Oryza sativa*

Introduction

Each compartment in a plant cell contains its own specific set of proteins meant to fulfil a specific function within the metabolic range of reactions. For most reactions, the general rule 'one gene-one compartment' (Small et al. 1998) applies which implies that most enzymes are targeted exclusively to one cellular

Electronic supplementary material The online version of this article (doi:10.1007/s00438-007-0214-4) contains supplementary material, which is available to authorized users.

Communicated by B. F. Lang.

R. Schwacke · K. Fischer
Institute of Botany, University of Cologne,
Gyrhofstr. 15, 50931 Cologne, Germany

B. Ketelsen · K. Krupinska · K. Krause
Institute of Botany, Christian-Albrechts-University Kiel,
Olshausenstr. 40, 24098 Kiel, Germany

Present Address:
K. Fischer · K. Krause (✉)
Institute for Biology, University of Tromsø,
9037 Tromsø, Norway
e-mail: Kirsten.Krause@ib.uit.no

compartment. As a consequence, similar metabolic steps are often performed by isoenzymes that are presumed to have evolved by gene duplication. However, an increasing number of examples for proteins that possess either one ambiguous targeting peptide or two or more targeting signals have emerged over the last decade. Selective targeting of proteins to different cellular compartments can be important for plant development and interorganellar communications. This phenomenon has been discussed in several recent reviews and the various mechanisms of dual targeting and combinations of intracellular targets have been summarized (Small et al. 1998; Silva-Filho 2003; Karniely and Pines 2005). Among the known combinations of target compartments, the combination of mitochondria and plastids is particularly abundant in plant cells (Silva-Filho 2003). In contrast, it is striking that hardly any target combinations of nucleus/plastids or nucleus/mitochondria have been reported. The aim of this study was to determine to what extent dual localization to the nucleus and one of the other DNA-containing organelles might occur among proteins involved in the regulation of gene expression. To this end we used an *in silico* approach to screen the genomes from *Arabidopsis thaliana*, a dicotyledonous plant, and rice, a monocotyledonous plant, for transcription factors that possess the relevant plastid, mitochondrial and nuclear-targeting sequences.

In contrast to plants, the available data from yeast and mammalian cells show that here, a significant number of proteins are active in mitochondria as well as in the nucleus. Several of such dually targeted proteins are involved in tRNA-processing like the yeast Trm1, Mod5, Cca1 and Rpm2 proteins (Ellis et al. 1989; Boguta et al. 1994; Wolfe et al. 1994, 1996; Stribinskis et al. 2005) or in DNA mismatch repair as the human uracil-DNA glycosylase (Slupphaug et al. 1993). One example implicated in chromatin remodeling, transcription, splicing and translation processes is the K protein of the hnRNP complex that has been found not only in the nucleus but also in the cytoplasm and in the mitochondria (Bomsztyk et al. 2004). Other proteins found in the mitochondria and in the nucleus are involved in programmed cell death such as the apoptosis inducing factor, AIF, which has been found in mammals and in yeast (Wissing et al. 2004; Ruchalski et al. 2006).

In land plants, one of the very few examples for dually targeted nuclear/plastid proteins was described in 1997 by Luo et al., who reported on the existence of two sets of transcripts of the bifunctional carrot dihydrofolate reductase/thymidylate synthase. A longer transcript of the corresponding gene encodes a protein with an N-terminal plastid target peptide that can direct the precursor protein to the chloroplasts while a shorter transcript produced from the same gene lacks the N-terminal extension and therefore apparently codes for a nuclear version of the protein (Luo et al. 1997). In 2001, the presence of a protein similar in size and immunologically related to a nuclear DNA-binding protein, SEBF, that acts as a repressor of the potato pathogenesis-related gene *PR-10a*, has been observed in chloroplasts (Boyle and Brisson 2001). Recently, the three members of a new family of transcription factors in *Arabidopsis thaliana*, the Whirly (Why) protein family, were shown to be directed to either plastids or mitochondria in protoplasts transformed with the respective GFP fusion proteins (Krause et al. 2005). Previous reports on the Why1 protein of potato (alias p24, Desveaux et al. 2000) have described the interaction between this protein and the promoter of the nuclear pathogen response gene *PR-10a* in infected cells (Desveaux et al. 2000, 2004). Most recently, reports on two further dually targeted DNA-binding proteins with localization in the nucleus and in one of the other two DNA-containing organelles have been published (Sunderland et al. 2006; Raynaud et al. 2006). In case of the DNA ligase 1, translation initiation from a first in-frame start codon produces a protein that is exclusively targeted to mitochondria, while the use of an alternative second start codon produces a protein that is found only in the nucleus (Sunderland et al. 2006). The existence of a chloroplast-localized protein initiated at a potential third AUG that was previously proposed (Sunderland et al. 2004) could not be confirmed.

The present study demonstrates that these proteins likely are just the tip of the iceberg and that dual-targeting activity to the nucleus and the plastids or mitochondria seems to be a broader phenomenon in plant cells than currently anticipated.

Materials and methods

Sequence retrieval

Predicted putative transcription factor sequences of *Arabidopsis thaliana* were obtained from the Arabidopsis Transcription Factor Database (Davuluri et al. 2003; http://www.arabidopsis.med.ohio-state.edu/AtTFDB). The gene names follow the AGI locus identifier and the annotation is based on TAIR v.6 (http://www.arabidopsis.org). The different loci coding for putative transcription factors of rice (*Oryza sativa*) were obtained from the Rice Transcription Factor Database

(http://www.ricetfdb.bio.uni-potsdam.de). The rice genes were named according to the TIGR locus identifier and the annotation is based on TIGR v.4 (http://www.tigr.org/tdb/e2k1/osa1).

Prediction of subcellular localization

All predictions were based on a consensus prediction using a naïve Bayes method. For this, individual predictions of chloroplast and mitochondrial target peptides were performed by several publicly available web services (Table 1). These individual predictions were combined mathematically to a consensus score. In detail, two complementary hypotheses for the location of a protein in the chloroplast (and two more for the location in the mitochondrion) were tested: the hypothesis that a protein is located and the hypothesis that a protein is not located there, given a positive prediction. For each prediction program the likelihoods, i.e. the probability of a positive prediction regarding one or the other hypothesis, were evaluated by considering its prediction data for sets of plant proteins with known subcellular localization. Plant proteins for these test sets were selected from the UniProt database (Schneider et al. 2005) or the Arabidopsis Subcellular Proteomic Database (Heazlewood et al. 2005) (see supplemental files 1–3). Redundancy within the protein sets was reduced in a way that no two proteins shared greater than 40% sequence identity.

To combine the different methods, it was assumed that their predictions are independent of each other. This naïve assumption allowed us to compute the likelihood of the parameters given several prediction data simply as product of the individual likelihoods. The ratio of the posterior probabilities of both hypotheses was computed by

$$\frac{p(c|a_1,a_2,\ldots,a_n)}{p(\bar{c}|a_1,a_2,\ldots,a_n)} = \frac{p(c)\prod_{i=1}^{n} p(a_i|c)^{w_i}}{p(\bar{c})\prod_{i=1}^{n} p(a_i|\bar{c})^{w_i}}$$

where c is the location of a protein in the chloroplast or mitochondrion, respectively, (the negation of c is written $\bar{c}$) and a_1 to a_n are the individual positive predictions. Based on predictions for the whole genomes of *A. thaliana* and *O. sativa* (data not shown), the chloroplast-targeted and mitochondrion-targeted proteins were estimated to constitute 15% and 12% of all open reading frames. $p(c)$ for chloroplast-targeting was set, accordingly, to 0.15 and $p(c)$ for mitochondrion-targeting to 0.12. The weight w_i is given by the score value of the corresponding prediction program and was normalized to a value between 0 and 1. Programs without scoring (IPsort, WoLF-PSort) can be viewed as a special case of weighting where weights are restricted to either 0 or 1. The logarithm in base 2 of the ratio that resulted from this calculation was used as consensus score value.

Evaluation of the consensus prediction method

To show an improvement of this consensus method over each of the individual methods that contribute to it, the specificities of all methods were compared by applying them to the plant protein test sets described earlier (suppl. files 1–3). The specificity (computed as 1—false positives/all negatives) depends on the score value threshold (above which the prediction is positive) chosen for an individual prediction program. In general, a higher threshold generates a higher specificity but sacrifices sensitivity (computed as true positives/all positives). Therefore, the comparison of the

Table 1 Web services used to predict plastid (cTP) or mitochondrial (mTP) targeting sequences of plant transcription factors

Program	Reference	Spec. (sens.) plastid	Spec. (sens.) mitochondrion
ChloroP v1.1	Emanuelsson et al. (1999)	0.917	–
iPSort	Bannai et al. (2002)	0.917 (0.595)	0.823 (0.766)
Mitopred	Guda et al. (2004)	–	0.762
MitoProt v2	Claros and Vincens (1996)	–	0.819
PCLR v0.9	Schein et al. (2001)	0.895	–
PProwler v1.1	Bodén and Hawkins (2005)	0.959	0.945
Predotar v1	Small et al. (2004)	0.955	0.938
PredSL	Petsalaki et al. (2006)	0.939	0.849
TargetP v1	Nielsen et al. (1997); Emanuelsson et al. (2000)	0.937	0.908
WoLF-PSort	Horton et al. (2006)	0.828 (0.713)	0.811 (0.688)
Consensus	This publication	0.971	0.952

The specificity values (spec.) for a reference sensitivity value of 0.7 were evaluated for the individual prediction methods as well as for the consensus method using two plant protein test sets (see supplementary material). For prediction programs lacking a score value (iPsort, WoLF-PSort) a trimming of the threshold score value resulting in a reference sensitivity value of 0.7 was not possible, instead the sensitivity values (sens.) are shown in parentheses

specificities was based on a common reference sensitivity value. The specificity was evaluated after trimming the method score threshold to a value that results in a reference sensitivity of 0.7. This reference sensitivity was used for all further calculations.

Sequence alignments of orthologous proteins from different plant species and reconstruction of phylogenetic trees

Protein and translated EST databases were examined for sequences homologous to Arabidopsis transcription factors using the blastp and tblastn tools of the BLAST program (Altschul et al. 1990). The sequences were aligned using the Clustal X program (Thompson et al. 1997). The sequence alignments were subsequently inspected and edited by hand as recommended by Harrison and Langdale (2006) using the graphical multiple sequence alignment editor (BioEdit v.7.0.5.3) in order to obtain optimal alignment and eliminate gap-rich stretches. Nuclear localization sequences were identified with the programs PredictNLS (Cokol et al. 2000) and PSORT (Nakai and Horton 1999). Unrooted trees were prepared by the neighbor joining method (Saitou and Nei 1987) using Clustal X (v1.81) and TreeView (v1.5.2) with 1,000 replicates performed for obtaining bootstrap confidence values. The measure for the distances between sequences was percent divergence.

Localization of an At2g44940-GFP-fusion protein

The entire cDNA sequence and the sequence corresponding only to the plastid target peptide, respectively, were amplified by PCR using isolated cDNA from Arabidopsis. The PCR products were subsequently cloned, sequenced and then inserted in-frame in front of the *gfp* coding sequence using the binary gateway vector pBatTL-B-GFP2 that contains a double 35S promoter. Protoplasts from *Arabidopsis thaliana* were produced from Arabidopsis light-grown suspension culture cells according to the protocol of Negrutiu et al. (1987). The recombinant plasmids with the GFP fusion constructs were introduced into the protoplasts using PEG-mediated transformation (Negrutiu et al. 1987). Transiently transformed cells were analyzed for GFP fluorescence using a fluorescence microscope.

Results

Validation of the screening method

For most of the annotated plant transcription factors no experimental data concerning their subcellular localization are available. Analyses of these proteins are complicated by the fact that they are often present in trace amounts only. Sensitive methods like mass spectrometric analysis of compartmental proteomes are prone to artifacts because of the danger of cross-contamination from other cell compartments. Optical in vivo techniques based on the fusion with fluorescent proteins such as GFP or immunological methods are more reliable but are only available for a few selected proteins. For the task of identifying potential candidates that are targeted to one of the organelles, a prediction method of the subcellular localization that picks up as many true positives for a given compartment while keeping the number of false positives or true negatives as low as possible is highly desired. Wagner and Pfannschmidt (2006) have recently listed 48 putatively plastid-targeted transcription factors from Arabidopsis based on the prediction with the program TargetP (Nielsen et al. 1997). In contrast, we have chosen an approach where the results of several prediction programs were combined to a consensus prediction using a naïve Bayes method (see Materials and methods). In order to compare the performance of the consensus prediction to those of the individual single prediction programs that contribute to it, the specificities were calculated using sets of organellar test proteins consisting of >500 proteins from Arabidopsis and other species. For control, a test set of >600 proteins of confirmed non-organellar localization was used. We found that for both plastid and mitochondrial proteins, the consensus prediction method showed a higher specificity at a reference sensitivity of 0.7 than the single predictions which contribute to the consensus (Table 1). The vast majority of the organellar test set proteins achieved consensus score values of 10 and above (up to 21) (data not shown). When used on the experimental sets of DNA-binding SET domain proteins (Springer et al. 2003) (Table 2) and transcription factors (Tables 3, 4, 5, 6), we found again that those proteins with a confirmed localization (ATXR5, AtWhy1-3) had values of above 10. Our algorithm predicted high scores of 19.2 (AtWhy1), 17.1 (AtWhy3), 16.3 (ATXR5) and 10.6 (AtWhy2) for these proteins, respectively (Tables 2, 3, 4). Two more SET domain proteins also received high scores for plastids (At1g26760) and mitochondria (At5g06620) (Table 2), whereas the remaining 34 SET domain proteins were not indicated as being organelle-targeted by the prediction method. This is consistent with their confirmed (At1g02580, Choi et al. 2004) or presumed location according to the SUBA proteomic database (Heazlewood et al. 2005). Based on these results we decided to use 10 as cutoff value. Below this value the risk of

Table 2 Mitochondrial and plastid consensus scores for SET domain proteins

Common name	Gene locus	NLS	mTP consensus	cTP consensus
ATXR1	*At1g26760*	Yes	1.7	**18.3**
ATXR2	*At3g21820*	No	–1.5	–0.5
ATXR3	*At4g15180*	Yes	–1.5	–0.4
ATXR4	*At5g06610*	No	–1.7	–0.2
ATXR4	*At5g06620*	No	**13.7**	1.5
ATXR5*	*At5g09790*	Yes	2.1	**16.3**
ATXR6	*At5g24330*	Yes	5.7	–0.9
ATX1	*At2g31650*	Yes	–1.6	0.4
ATX2	*At1g05830*	Yes	–1.7	–0.2
ATX3	*At3g61740*	Yes	–0.1	–0.7
ATX4	*At4g27910*	Yes	6.3	0.6
SUVH1	*At5g04940*	Yes	1.0	0.6
SUVH2	*At2g33290*	Yes	–1.5	1.2
SUVH3	*At1g73100*	Yes	–1.5	1.7
SUVH4	*At5g13960*	Yes	4.4	1.8
SUVH5	*At2g35160*	Yes	–1.4	1.2
SUVH6	*At2g22740*	Yes	–1.6	1.6
SUVH7	*At1g17770*	Yes	–1.5	2.6
SUVH9	*At4g13460*	No	–1.7	7.3
SUVH10	*At2g05900*	Yes	–1.4	–1.1
SUVR1	*At1g04050*	No	–0.3	0.4
SUVR3	*At3g03750*	Yes	2.6	–0.2
SUVR4	*At3g04380*	Yes	–1.2	–0.2
SUVR5	*At2g23740*	Yes	–1.0	–0.9
SDG3	*At2g17900*	No	2.8	1.8
SDG29	*At5g53430*	Yes	1.4	–0.8
CLF	*At2g23380*	Yes	–1.6	7.6
MDH9	*At5g42400*	Yes	0.1	5.1
MRH10	*At5g43990*	Yes	2.3	–0.8
EZA1	*At4g02020*	No	–1.7	–1.1
MEA*	*At1g02580*	Yes	–1.4	–1.0
ASHH1	*At1g76710*	Yes	–1.1	–1.2
ASHH2	*At1g77300*	Yes	–1.6	0.7
ASHH3	*At2g44150*	No	–0.9	–0.4
ASHH4	*At3g59960*	Yes	0.8	0.4
ASHR2	*At2g19640*	No	–0.2	–0.1
ASHR3	*At4g30860*	Yes	–1.5	4.7

Gene loci were taken from Baumbusch et al. (2001). The scores were determined as described in Materials and methods (mTP consensus = mitochondrial score; cTP consensus = plastid score). An asterisk (*) marks the proteins for which experimental confirmation of the localization is existent. Values above 10 are printed bold

contamination by false positives was observed to increase.

Identification of putative plastid and mitochondrial transcription factors

The Arabidopsis transcription factor database currently lists 1,747 different proteins from 50 transcription factor families. A similar list containing currently 2,309 different loci grouped in 53 transcription factor families was compiled for rice by the Rice Transcription Factor Database. The protein sequences from these lists were subjected to a search for targeting sequences to plastids and mitochondria.

Among the Arabidopsis transcription factors, we identified 78 proteins that possess putative plastid targeting sequences (cTPs) and 12 proteins with a putative mitochondrial presequence. Fifty-one of the proteins with a cTP possess an additional sequence (NLS) that can target the protein to the nucleus, while 27 proteins lack such a sequence (Fig. 1). Of the 12 putative mitochondrial proteins 7 possess no additional targeting sequences while 5 contain a NLS (Fig. 1). Most of the proteins without known nuclear localization sequences have a molecular weight below 40 kDa and might thus not necessarily need a NLS for nuclear import. In rice, 80 proteins with a cTP and 23 proteins with a mitochondrial presequence possess a NLS. Furthermore, 40 proteins exclusively possess a cTP while 15 proteins have only a mitochondrial presequence (Fig. 1). In Tables 3, 4, 5, and 6 these proteins are listed according to their affiliation with the different transcription factor families.

Of the 50 Arabidopsis transcription factor families and the 53 transcription factor families of rice, 23 and 33, respectively, possess members with putative organellar presequences. These include large families with numerous members such as the C2H2 and CH3 zinc finger domain protein families or the AP2/EREBP proteins. On the other hand also small protein families like the GeBP or Whirly transcription factor families are included (Tables 3, 4, 5, and 6).

Apart from the three Whirly proteins of Arabidopsis (Krause et al. 2005, see Introduction), only three proteins from the list of identified proteins (At1g47870 alias E2FC, the GeBP protein At4g00270 and the GRAS protein At3g54220 alias Scarecrow) were so far analyzed for their subcellular localization using fluorescence-based techniques (proteins marked with asterisks in Tables 3, 4). All three were reported to be in the nucleus (Curaba et al. 2003; Heidstra et al. 2004; Koroleva et al. 2005). However, in the case of the YFP-At4g00270 fusion, the confocal images showed more than one fluorescent spot per cell. These spots were not seen with a nuclear control construct (Curaba et al. 2003) and can thus not be assigned to a specific compartment. A dual localization of this protein was, therefore, not refuted. Three transcription factors were identified by different mass spectrometric approaches but no confirmation of these by other methods exists. Only one (At4g00870) was identified as a nuclear protein (Bae et al. 2003), whereas the other two (At5g27070, At5g38560) were detected in a plasma membrane fraction (Nuhse et al. 2003).

Table 3 Characteristics of *Arabidopsis thaliana* transcription factors with putative plastid localization sequences

Family/name	Gene locus	cTP score	NLS	kDa
AP2-EREBP				
	At2g44940	20.8	Yes	32.0
	At1g77640	18.7	No	27.0
	At1g44830	15.1	No	23.0
	At1g21910	14.7	No	25.5
Ail5	*At5g57390*	14.6	No	60.3
	At3g16280	14.2	Yes	20.4
Wri1	*At3g54320*	13.8	Yes	48.0
Rap2.10	*At5g52020*	12.8	No	25.1
	At2g22200	11.3	No	29.8
	At4g31060	10.1	No	20.8
RAV2	*At1g68840*	10.0	Yes	39.5
bHLH				
AtbHLH147	*At3g17100*	17.1	Yes	25.3
AtbHLH148	*At3g06590*	15.2	Yes	24.2
AtbHLH14	*At4g00870*	15.1	Yes	47.0
AtbHLH128	*At1g05805*	11.2	Yes	39.0
AtbHLH62	*At3g07340*	10.7	Yes	50.2
bZIP				
AtbZIP33	*At2g12900*	17.2	Yes	30.0
AtbZIP31	*At2g13150*	16.5	Yes	29.7
GBF4	*At1g03970*	15.6	Yes	30.5
GBF5	*At2g18160*	13.7	Yes	19.1
AtbZIP8	*At1g68880*	12.2	Yes	16.2
AtbZIP69	*At1g06070*	12.2	Yes	47.1
GBF6	*At4g34590*	11.6	Yes	18.8
AtbZIP13	*At5g44080*	11.2	Yes	35.0
AtbZIP43	*At5g38800*	10.6	Yes	19.2
AtbZIP44	*At1g75390*	10.3	Yes	19.1
AtbZIP14	*At4g35900*	10.2	Yes	27.0
AtbZIP34	*At2g42380*	10.1	Yes	35.7
BZR				
Bzr1	*At1g75080*	13.4	Yes	36.5
Bzr2	*At1g19350*	11.5	Yes	36.5
C2C2-Dof				
Dag2	*At2g46590*	13.7	No	40.5
	At5g65590	10.4	No	34.9
C2C2-Yabby				
Yab3	*At4g00180*	16.5	Yes	26.3
Yab1	*At2g45190*	10.8	Yes	25.8
C2H2				
Knu	*At5g14010*	16.7	No	18.0
	At2g02080	16.0	Yes	55.8
	At1g14580	14.3	Yes	50.6
	At5g01310	13.8	Yes	101.4
	At2g02070	12.5	Yes	64.4
	At3g18290	12.2	Yes	141.5
	At4g02670	11.9	Yes	44.5
Zpf2	*At5g57520*	11.6	No	17.0
	At5g01860	11.1	Yes	24.1
	At5g27880	10.8	Yes	30.9
C3H				
	At3g26730	18.6	Yes	85.0
	At1g68070	15.3	No	38.3
	At5g45290	14.2	Yes	60.8
	At1g73760	14.2	Yes	40.6
	At2g39100	13.0	No	34.4
	At5g55970	12.0	No	39.0
	At2g04240	11.2	No	17.9
	At2g01735	11.2	No	40.1
	At4g23450	10.6	No	16.9
CAMTA				
AtCAMTA3	*At2g22900*	10.8	Yes	52.1
CPP				
	At4g14770	10.3	Yes	72.1
E2F-DP				
E2FA	*At2g36010*	14.6	Yes	56.4
E2FC*	*At1g47870*	11.9	No	44.5
G2-like				
	At5g29000	12.2	No	46.2
GeBP				
	At4g00610	14.2	Yes	37.0
*	*At4g00270*	10.5	Yes	34.1
GRAS				
Las	*At1g55580*	14.2	No	50.0
AtGras8	*At1g63100*	14.1	Yes	73.5
Scr*	*At3g54220*	13.2	No	71.5
Homeobox				
Wox4	*At1g46480*	11.3	No	28.7
MADS				
AGL103	*At3g18650*	19.8	Yes	43.5
AGL98	*At5g39810*	15.5	Yes	37.3
AGL81	*At5g39750*	13.2	Yes	65.2
AGL77	*At5g38740*	10.8	Yes	48.4
AGL93	*At5g26950*	10.7	No	32.8
AGL53	*At5g27070*	10.5	No	32.1
AGL89	*At5g27580*	10.3	Yes	25.6
NAC				
	At3g10480	10.0	Yes	50.4
TCP				
	At1g35560	14.1	Yes	36.0
Trihelix				
	At5g38560	17.1	Yes	72.0
Whirly				
AtWhy1*	*At1g14410*	19.2	No	29.1
AtWhy3*	*At2g02740*	17.1	No	29.7
WRKY				
AtWRKY33	*At2g38470*	13.6	No	57.1
AtWRKY20	*At4g26640*	10.6	Yes	53.6

Gene loci were taken from the AtTFDB database on the Arabidopsis Gene Regulatory Information Server (AGRIS). The cTP consensus score was determined based on the calculation described in Materials and methods. Only values of 10 and higher are shown. Asterisks (*) mark the proteins for which experimental confirmation of the localization is existent

Phylogenetic relationship of putative plastid and mitochondrial proteins of the AP2/EREBP family

The AP2/EREBP protein family is among the families with the most putative plastid or mitochondrial targeting sequences (see Tables 3, 4, 5, 6). This protein family is defined by the AP2/EREBP domain which consists of 60–70 amino acids and is involved in DNA binding (Weigel 1995). Based on the number of AP2/EREBP domains and other conserved motifs, the AP2/EREBP transcription factor family is divided into four

Table 4 Characteristics of *Arabidopsis thaliana* transcription factors with putative mitochondrial localization sequences

Family/name	Gene locus	mTP score	NLS	kDa
AP2-EREBP				
Shine3	*At5g11190*	11.2	No	21.4
Shine2	*At5g25390*	10.0	No	20.8
C2H2				
	At5g20220	11.2	Yes	46.0
C3H				
	At1g68180	12.7	No	28.8
G2-like				
	At1g79430	10.5	Yes	32.4
GeBP				
	At2g01370	13.4	Yes	29.4
HRT				
	At5g56770	11.6	No	29.3
MADS				
AGL92	*At1g31640*	11.8	Yes	21.2
AGL86	*At1g31630*	11.2	Yes	36.9
TUB				
AtTLP9	*At3g06380*	12.5	No	42.3
AtTLP7	*At1g53320*	12.3	No	42.2
Whirly				
Why2*	*At1g71260*	10.6	No	29

Gene loci and the corresponding common names were taken from the AtTFDB database on the Arabidopsis Gene Regulatory Information Server (AGRIS). The mTP consensus score was determined based on the calculation described in Materials and methods. Only values of 10 and above are shown. Asterisks (*) mark the protein for which experimental confirmation of the localization is existent

Table 5 Characteristics of *Oryza sativa* transcription factors with putative plastid localization sequences

Family	Gene locus	cTP score	NLS	kDa
ABI3VP1				
	Os01g51610	13.4	Yes	31.9
	Os07g37610	11.8	Yes	105.7
Alfin-like				
	Os01g73460	17.5	Yes	43.5
	Os06g08790	15.5	Yes	92.1
	Os06g14010	12.9	Yes	19.0
	Os06g01170	12.9	Yes	111.1
	Os06g51450	10.4	Yes	87.7
AP2-EREBP				
	Os12g03290	17.1	Yes	49.3
	Os11g03540	17.0	Yes	49.9
	Os04g46400	16.2	Yes	29.6
	Os06g11860	15.2	No	37.8
	Os01g04800	14.6	Yes	39.3
	Os05g49010	14.3	Yes	30.3
	Os04g46440	14.2	Yes	23.0
	Os07g22770	13.9	Yes	25.4
	Os07g47330	13.8	Yes	33.5
	Os08g43200	13.2	No	24.5
	Os09g35020	12.8	Yes	25.5
	Os10g41130	12.7	Yes	29.7
	Os10g25170	12.6	Yes	34.1
	Os04g46410	12.1	Yes	26.2
	Os04g32790	11.7	Yes	29.8
	Os01g73770	11.1	Yes	23.8
	Os09g25600	11.0	Yes	41.0
	Os03g12950	10.0	No	68.7
ARF				
	Os01g54990	16.0	Yes	79.6
	Os04g59430	11.5	No	57.2
AUX/IAA				
	Os01g18360	11.8	Yes	21.9
	Os07g08460	11.5	No	23.0
	Os01g48450	11.2	No	28.4
	Os11g11410	11.0	No	37.0
	Os02g49160	10.2	No	22.2
	Os05g08570	10.0	No	27.2
BES1				
	Os02g03690	10.3	Yes	80.1
bHLH				
	Os05g50900	17.1	Yes	52.8
	Os04g28280	16.4	Yes	28.1
bZIP				
	Os02g03960	14.7	Yes	16.9
	Os05g36160	13.8	Yes	25.2
	Os01g36220	12.1	Yes	19.0
	Os08g26880	12.0	Yes	20.1
	Os09g13570	10.9	Yes	17.1
	Os02g10860	10.7	Yes	27.1
	Os05g03860	10.6	Yes	16.0
C2C2-Dof				
	Os10g35300	16.1	No	24.7
	Os04g58190	12.3	Yes	21.6
	Os03g55610	10.4	Yes	36.7
C2C2-GATA				
	Os02g43150	19.0	Yes	45.0
C2C2-Yabby				
	Os10g36420	12.0	Yes	29.2

subfamilies, the ERF subfamily, the APETALA2 (AP2) subfamily, the RAV subfamily and the DREB subfamily (Sakuma et al. 2002). ERF and DREB subfamilies are both characterized by the possession of a single AP2/ERF domain and are thus often regarded as one protein family (Nakano et al. 2006; Shigyo et al. 2006).

To analyze the phylogenetic position of the putative organellar proteins among the AP2/EREBP proteins, we constructed a phylogenetic tree with all 149 AP2 domain-containing proteins of Arabidopsis (not shown). Of the twelve putative plastid proteins, nine were identified as members of the DREB subfamily (Table 7). DREB proteins are reportedly involved in drought and low temperature stress responses in plant cells (Hao et al. 2002; Sakuma et al. 2002). Two of the other putative plastid proteins are members of the AP2 subfamily and a third one belongs to the RAV subfamily, whereas both putative mitochondrial proteins belong to the ERF subfamily (Table 7).

For phylogenetic comparison of the individual putative organellar AP2 proteins from Arabidopsis and rice, a phylogenetic tree was constructed using only the sequences of the putative organellar proteins from

Table 5 continued

Family	Gene locus	cTP score	NLS	kDa
C2H2				
	Os04g08290	18.3	Yes	51.7
	Os04g02510	16.2	Yes	50.2
	Os08g20580	15.9	Yes	23.7
	Os06g07020	15.5	Yes	45.4
	Os09g19940	13.9	No	57.9
	Os02g44120	12.7	No	65.7
	Os04g46670	11.3	No	59.6
	Os09g38340	10.3	Yes	54.9
	Os05g37190	10.0	Yes	42.1
C3H				
	Os05g11860	18.3	Yes	27.1
	Os07g27950	17.5	Yes	38.9
	Os05g32350	17.1	No	59.1
	Os08g43670	16.5	No	26.6
	Os09g27380	16.0	No	20.8
	Os02g46340	15.4	No	39.0
	Os01g72480	14.2	Yes	30.4
	Os06g32720	13.5	No	31.1
	Os04g49700	13.3	Yes	38.5
	Os04g02730	13.3	No	75.6
	Os02g09060	12.1	Yes	52.6
	Os06g09310	11.5	No	33.5
	Os05g01940	11.1	Yes	41.6
	Os03g26300	10.6	No	19.0
CCAAT-Hap5				
	Os03g63530	13.9	Yes	32.9
CPP				
	Os04g09560	20.5	No	16.4
	Os05g43380	12.8	No	41.4
	Os02g17460	10.9	Yes	54.4
E2F-DP				
	Os10g30420	11.0	Yes	36.9
G2-like				
	Os03g21240	13.5	Yes	46.7
	Os03g45194	10.2	No	60.9
GeBP				
	Os08g36450	13.7	Yes	48.6
GRAS				
	Os11g03110	17.5	Yes	69.9
	Os05g31420	12.7	Yes	58.2
	Os06g01620	12.5	No	51.0
Homeobox				
	Os03g52239	17.1	No	81.9
	Os03g55990	12.9	Yes	26.7
	Os06g04870	12.7	No	32.1
	Os03g47042	12.7	No	20.7
	Os04g55590	12.3	Yes	25.8
	Os06g39906	10.3	Yes	35.1
HSF				
	Os01g39020	11.5	Yes	43.9
Jumjonji				
	Os11g36450	15.5	Yes	57.7
MYB				
	Os06g14700	15.6	Yes	19.9
	Os04g58020	13.2	No	46.3
	Os03g19630	11.9	Yes	22.2
	Os01g63160	10.5	Yes	44.3

Table 5 continued

Family	Gene locus	cTP score	NLS	kDa
MYB-related				
	Os09g03690	16.3	Yes	31.9
	Os02g10060	15.3	No	55.1
	Os04g01970	13.6	Yes	97.4
	Os06g14710	12.2	No	16.4
	Os05g10690	11.7	Yes	30.5
	Os05g37040	11.2	No	14.6
	Os01g44370	10.3	No	10.1
NAC				
	Os03g59730	12.8	Yes	53.6
Orphans				
	Os06g48610	16.7	Yes	49.3
	Os02g05470	14.9	Yes	49.8
	Os10g08970	14.4	Yes	86.4
	Os12g01080	12.3	No	16.2
	Os05g38990	11.5	No	34.9
SNF2				
	Os05g15890	17.4	Yes	97.1
	Os06g01320	14.7	Yes	226.7
TCP				
	Os04g11830	10.9	Yes	18.8
Trihelix				
	Os02g33610	17.1	Yes	97.4
	Os01g70230	12.7	Yes	31.6
	Os10g41460	12.1	No	35.7
	Os04g45750	11.1	Yes	57.5
Whirly (PBF2-like)				
	Os06g05350	15.4	No	30.1
WRKY				
	Os05g39720	18.8	No	57.1
	Os05g46020	14.5	No	23.2
	Os11g29870	13.4	Yes	25.9
	Os01g61080	12.7	No	59.3
	Os05g40060	11.6	Yes	38.1
	Os01g08710	11.2	Yes	59.7

Gene loci are based on version 4 of the TIGR Rice Pseudomolecules and Genome Annotation database (http://www.tigr.org). The cTP consensus score was determined based on the calculation described in Materials and methods. Only values of 10 and above are shown

both species (Fig. 2). The proteins were designated using the nomenclature defined by Nakano et al. (2006), where DREB proteins are represented by ERF groups I to IV and ERF proteins *in senso stricto* are represented by groups V to X. The phylogenetic tree showed that most Arabidopsis genes contain one or more closely related orthologues in rice, the only exceptions being the four Arabidopsis proteins belonging to group II of the ERF proteins (Fig. 2). No Arabidopsis orthologues could be found for any of the rice proteins belonging to groups XI to XIV which is consistent with previous observations (Nakano et al. 2006).

Table 6 Characteristics of *Oryza sativa* transcription factors with putative mitochondrial localization sequences

Family	Gene locus	mTP score	NLS	kDa
Alfin-like				
	Os09g27620	11.0	Yes	73.8
	Os11g12650	10.1	No	79.9
AP2-EREBP				
	Os08g41030	12.4	Yes	20.3
	Os12g41030	11.6	Yes	15.8
	Os10g38000	11.1	No	20.4
	Os02g55380	10.7	No	18.7
	Os06g08340	10.3	No	19.2
BES1				
	Os01g08180	11.5	Yes	17.4
C2H2				
	Os08g44830	14.9	Yes	47.8
	Os03g05480	13.5	Yes	69.5
	Os02g44130	12.0	No	35.3
C3H				
	Os10g32740	13.4	Yes	76.9
	Os05g10670	11.3	Yes	49.7
	Os05g41520	11.2	No	31.3
	Os07g06540	10.7	Yes	20.3
	Os02g06584	10.7	Yes	49.1
	Os03g04890	10.1	Yes	75.8
	Os04g56750	10.2	No	50.5
CCAAT-Hap5				
	Os12g25120	11.2	No	14.1
	Os07g36130	11.2	No	14.0
	Os08g33100	10.3	No	13.9
	Os07g36140	10.2	No	14.0
Homeobox				
	Os05g02730	10.4	No	25.9
MADS				
	Os01g23760	12.4	Yes	42.7
	Os01g18420	11.2	Yes	26.4
	Os08g02070	10.6	Yes	25.2
	Os01g68560	10.6	Yes	51.1
MYB-rel				
	Os03g13790	10.8	Yes	78.7
	Os11g08080	10.2	Yes	85.1
	Os01g43230	10.1	No	8.9
	Os05g07010	10.0	No	26.7
NAC				
	Os02g38130	14.2	Yes	43.6
	Os10g26240	11.7	Yes	19.4
Orphans				
	Os08g10780	10.2	Yes	47.1
SNF2				
	Os07g44210	10.3	Yes	80.0
Whirly (PBF2-like)				
	Os02g06370	13.0	No	25.2
WRKY				
	Os12g02440	13.7	Yes	24.7
zfHD				
	Os09g24810	12.2	Yes	11.8

Gene loci are based on version 4 of the TIGR Rice Pseudomolecules and Genome Annotation database (http://www. tigr.org). The mTP consensus score was determined based on the calculation described in Materials and methods. Only values of 10 and above are shown

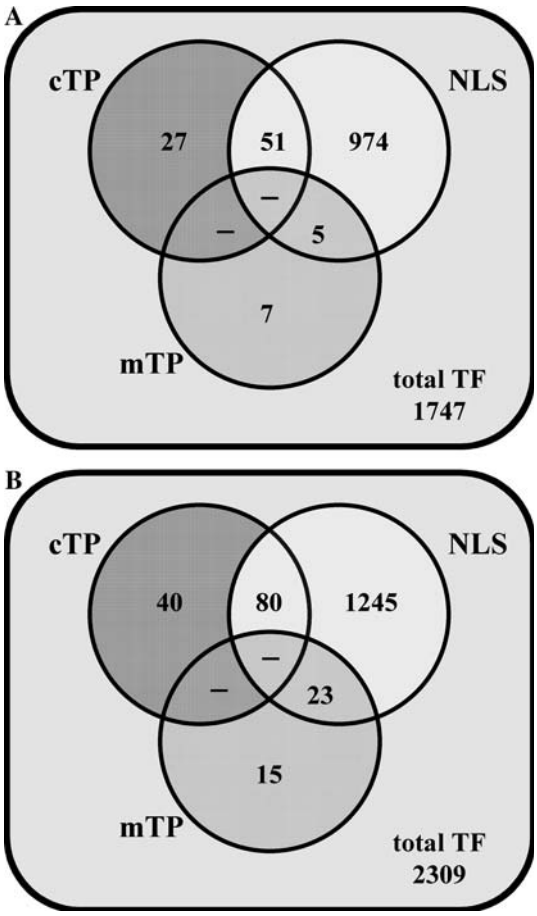

Fig. 1 Venn diagram of Arabidopsis (**a**) and rice (**b**) transcription factors possessing targeting sequences. The number of proteins with plastid (*cTP*), mitochondrial (*mTP*) and nuclear (*NLS*) localization sequences and combinations thereof are depicted

The existence of homologous pairs or groups of putative organellar proteins in Arabidopsis and rice prompted us to search for related proteins in other species. For the AP2 protein from Arabidopsis that gained the highest chloroplast score and that is encoded by the gene locus *At2g44940*, several homologous proteins from both dicotyledonous and monocotyledonous species could be identified. These include a protein from maize (ZmDBF2), one from *Triticum monococcum* (TmCbf7), one from barley (HvCbf7), a protein from *Medicago trunculata* (MtERF) and one from potato that was deduced from the fused amino acid sequences of two overlapping EST sequences (StPPCBR81) (Fig. 3). A similar number of homologues were found for the gene product of *At5g11190* that is putatively targeted to mitochondria (Fig. 3). Table 8 shows that all proteins

Table 7 Distribution of putative organellar proteins among the AP2/EREBP transcription factor family of Arabidopsis

Subfamily	Total number	Number of proteins with predicted cTP	Number of proteins with predicted mTP
ERF	65	–	2
DREB	55	9	–
AP2	18	2	–
RAV	11	1	–

cTP chloroplast target peptide, *mTP* mitochondrial presequence

from these species are strongly predicted to be targeted to either the plastids or the mitochondria.

An alignment of six sequences homologous to the *At2g44940* gene product revealed a high sequence identity within the AP2 domain and, beyond that, the existence of further domains that are highly conserved (Fig. 4). AP2 domains are characterized by several well-conserved amino acids that constitute a putative amphipathic α-helix and are generally divided into a DNA-binding and an oligomerization domain. These domains can be either adjacent to each other or separated by a few amino acids (Riechmann and Meyerowitz 1998; Liu et al. 1999). In the present case, the two parts of the AP2 domain are separated by a stretch of basic amino acids that constitute the nuclear localization sequence (Fig. 4). The N terminus of each protein, although being considerably variable, is extremely rich in hydroxylated amino acids and in alanine, leucine and arginine and thus fulfils the classical features of chloroplast-targeting sequences (Bruce 2000). Taken together, these findings indicate that this group of proteins has evolved before the monocotyledonous and dicotyledonous plants have split up.

Cellular localization of At2g44940

For the AP2 protein encoded by the *At2g44940* gene, GFP fusion constructs of the entire gene product or the putative plastid target peptide sequence were used to examine the localization of this protein. Transient

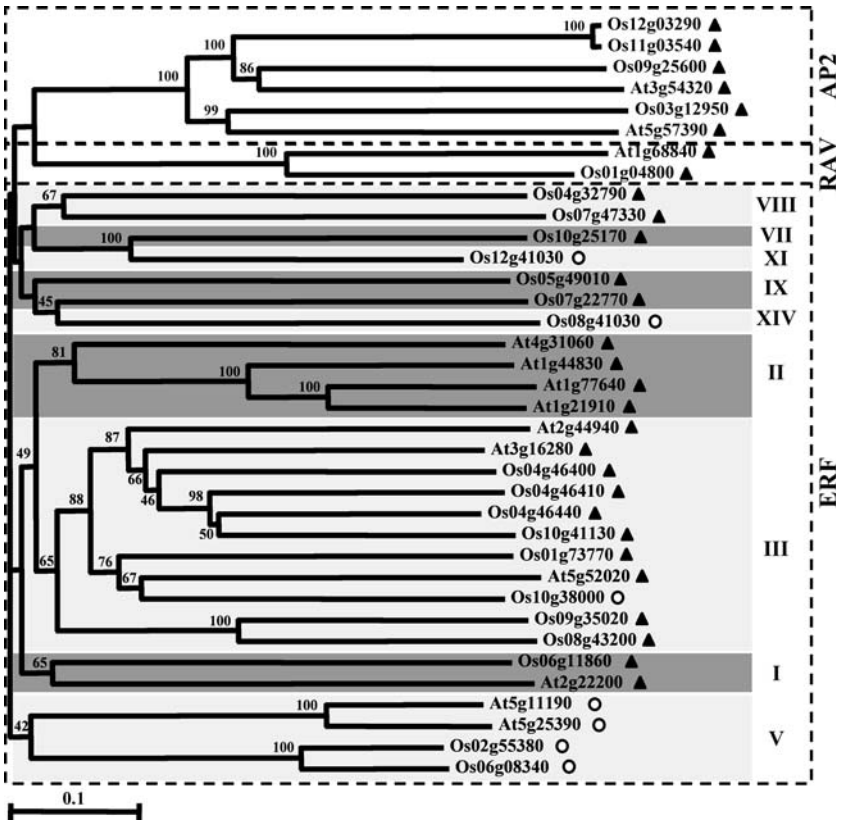

Fig. 2 Phylogenetic tree of AP2/EREBP proteins with putative mitochondrial and plastid targeting sequences. Arabidopsis and rice sequences were obtained from the public databases (see Materials and methods). Full length amino acid sequences were aligned using the programs Clustal X and BioEdit. The resulting alignment was used to construct a neighbor joining tree (Saitou and Nei 1987) with the program TreeView. *Numbers* at the nodes represent bootstrap values in percentage based on 1,000 repeats. Only nodes with bootstrap values above 40 are labeled. The *scale bar* represents the number of substitutions per site. Proteins with a high chloroplast score (*filled triangle*) and proteins with high mitochondrial score (*open circle*) are designated. Classification of proteins into subfamilies as defined by Nakano et al. (2006) is indicated

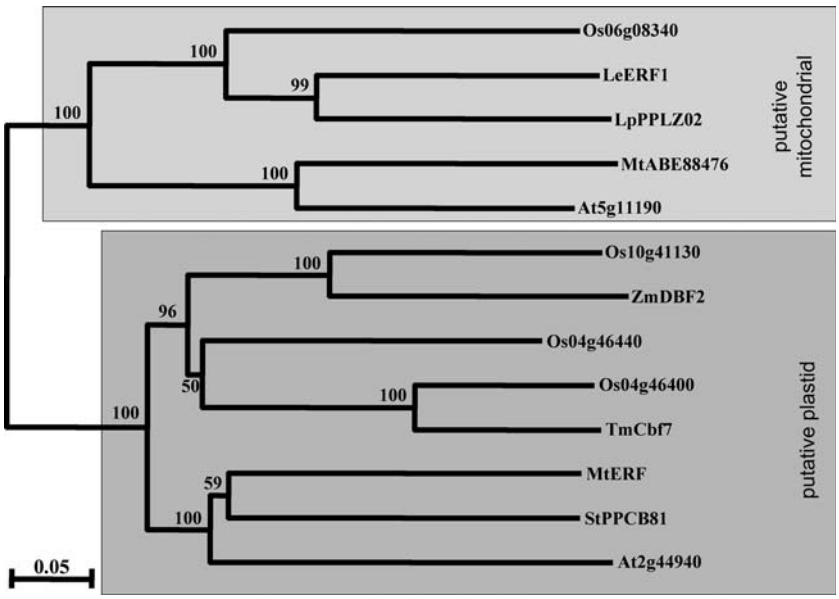

Fig. 3 Phylogenetic relationship of homologues of *At2g44940* and *At5g11190* gene products from different monocotyledonous and dicotyledonous plant species. Sequences from other plant species were obtained through BLAST searches. The alignment of full length amino acid sequences and construction of the neighbor joining tree was done as described in Fig. 2. *Numbers* at the nodes represent bootstrap values in percentage based on 1,000 repeats. The *scale bar* represents the number of substitutions per site

Table 8 Localization predictions for homologues of the *At2g44940* and *At5g11190* gene products

Species	Protein	cTP score	mTPscore	NLS
Arabidopsis thaliana	At2g44940	20.8	0	Yes
Solanum tuberosum	StPPCBR81	10.4	0	Yes
Medicago trunculata	MtERF	12.0	0	Yes
Oryza sativa	Os04g46400	16.2	0	Yes
Triticum monococcum	TmCBF7	16.2	0	Yes
Zea mays	ZmDBF2	12.5	0	Yes
Arabidopsis thaliana	At5g11190	0	11.2	No
Medicago trunculata	ABE88476	0	10.6	No
Oryza sativa	Os06g08340	0	10.3	No
Lycopersicon esculentum	LeERF1	0	12.7	No
Lupinus polyphyllus	LpPPLZ02	0.5	12.0	Yes

cTP chloroplast target peptide, *mTP* mitochondrial presequence, *NLS* nuclear localization signal

expression of these fusion proteins in protoplasts from a light-grown mesophyll cell suspension culture from *Arabidopsis thaliana* showed that the GFP fused to the entire *At2g44940* gene product is indeed targeted to both compartments (Fig. 5a). The dual localization confirmed that both of the targeting signals, i.e. the N-terminal plastid target peptide and the NLS were correctly predicted. However, we observed that most of the recombinant protein was located inside the nucleus, whereas the chloroplasts showed only weak fluorescence. We therefore fused only the putative plastid target peptide to GFP and transformed protoplasts with this construct. As expected, the GFP fluorescence coincided only with the chlorophyll autofluorescence of the chloroplasts and no nuclear signal was observed (Fig. 5b).

Discussion

Existence of proteins with sequences targeting them to the nucleus and either plastids or mitochondria

A systematic *in silico* search for dually targeted DNA-binding proteins from Arabidopsis and rice was performed by integrating the individual predictions of several prediction programs into a consensus prediction. With this approach, we identified approximately 90 transcription factors in Arabidopsis and almost twice as many transcription factors in rice that have a very high probability of possessing targeting sequences for the nucleus and at least one of the other two organelles (Fig. 1; Tables 3, 4, 5, 6). Many of the identified proteins were found to form orthologous groups and possess homologues in other plant species as well (Figs. 2, 3, 4 and data not shown). The same was observed for the SET domain proteins where all putative target-peptide containing proteins belong to the group of trx-related proteins (Table 2 as well as unpublished data).

Xiong et al. (2005) reported in a genome-wide comparative analysis between monocots and eudicots that approximately 50% of Arabidopsis and rice transcription factor genes form orthologous pairs or groups. They argue that the existence of such groups in two or more species hints at conserved functions of the proteins in monocotyledonous and dicotyledonous plants. A potential transit peptide for plastids or mitochondria has been conserved in orthologous proteins of the AP2/EREBP transcription factor family in a number of species (Figs. 2, 3, 4), suggesting that these

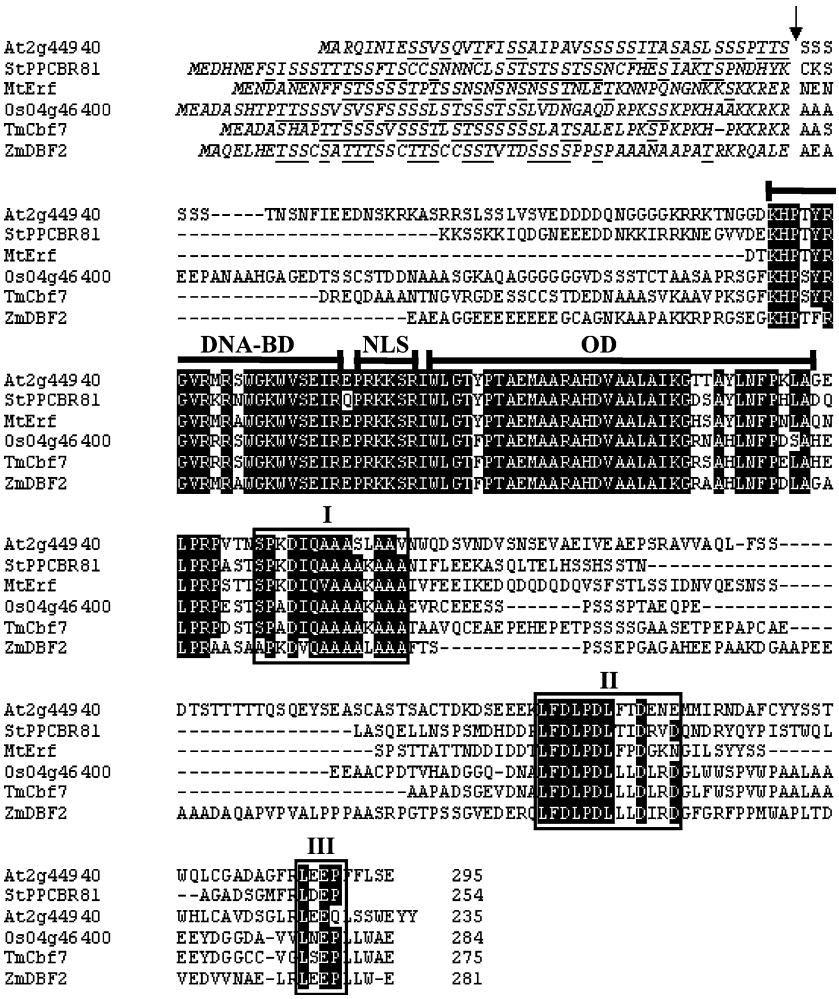

Fig. 4 Alignment of amino acid sequences of AP2 domain containing homologues of At2g44940. Amino acids that are identical in at least 5 out of 6 sequences are shown in *white against a black background*. The chloroplast target peptide (cTP) is depicted in *italic letters* and ends at the cleavage site that is marked by a *downward arrow*. The DNA-binding and oligomerization domains (DNA-BD, OD) of the AP2 motif and the nuclear localization sequence (NLS) are indicated by *contiguous lines* above the sequence. Other conserved domains are framed and designated I to III

proteins could indeed have a functional role within these organelles. Of particular interest with respect to this possible role is the fact that AP2 domain-containing proteins were recently discovered in a cyanobacterium, *Trichodesmium erythraeum* (Magnani et al. 2004; Wessler 2005). One possible interpretation of this observation is that the eukaryotic AP2 domain-containing proteins were derived originally from the algal ancestor of plastids. After multiplication, some of them could have retained a function in these organelles while many others were assigned new functions in the other DNA-containing compartments.

It is conspicuous that many putative plastid AP2-proteins belong to ERF groups II and III. These groups are characterized by additional specific C-terminal motifs. ERF group II is further subdivided into three subgroups, IIa, IIb and IIc (Nakano et al. 2006). Four putative dually targeted Arabidopsis proteins belong to the small subgroup IIb consisting of only seven members. All these proteins are characterized by the C-terminal CMII-3 motif. Interestingly, the same motif was also found in several members of the ERF group III, among them three further potentially dually targeted proteins. Whether there is a connection between the possession of this motif and a role inside the plastids cannot be resolved at this stage. Given this striking cluster of CMII-3 motif-containing proteins among the putative plastid-targeted transcription factors, it is surprising that no orthologues of these proteins were found in rice (see Fig. 2). However, two group II rice proteins achieved cTP consensus scores of 8.5 and 7.7, respectively, and therefore failed to reach our cut-off value. It cannot be precluded that these two proteins might represent plastid orthologues of the four Arabidopsis ERF group II members shown in Fig. 2.

So far, the localization of one AP2/EREBP protein from the DREB subgroup (*At2g44940*) was analyzed with fluorescent microscopical techniques. This analysis confirmed the presence and functionality of the predicted dual targeting signals in vivo (Fig. 5). Further experimental evidence will be needed to validate a

presumed function of the identified candidates in the organelles. However, in many cases, the existence of paralogues and hence the possibility of functional redundancy could complicate the interpretation of experimental results.

Potential significance of nucleus/plastid and nucleus/mitochondria dually targeted proteins

A communication between the DNA-containing compartments is essential for plant cells since most organellar enzyme complexes are composed partly of nuclear-encoded subunits and partly of organelle-encoded subunits. This communication is characterized, for example, by nuclear control over plastid gene expression and a retrograde control of nuclear genes by a plastid signal. These mechanisms were summarized in a number of recent reviews (Richly et al. 2003; Strand 2004; Beck 2005).

Transcription factors that are dually targeted might play a key role in the coordinated regulation of nuclear and organellar genes in this context. Two possible ways are feasible by which the transcription factors could coordinate the gene expression in the different compartments. Both ways have been realized in yeast or animal cells. The first possibility implies that a protein would accumulate in both compartments simultaneously, either in the same cell type or under a similar developmental context. An example from yeast is the Rpm2 protein (Stribinskis et al. 2005). Such proteins can directly influence and co-regulate the expression of nuclear-encoded as well as organelle-encoded organellar proteins. The second possibility involves a development- or environment-induced retargeting of proteins as is evidently the case with the apoptosis-inducing factor (AIF) of yeast and mammalian cells. AIF is released from the mitochondria when these get disrupted during programmed cell death and is imported into the nucleus where it fulfils an important role in the coordinate degradation of nuclear DNA (Susin et al. 1999; Cregan et al. 2002; Ruchalski et al. 2006). Other well-studied examples for an influence of environmental factors on the localization of plant proteins are the phytochromes A, B, C, D and E whose

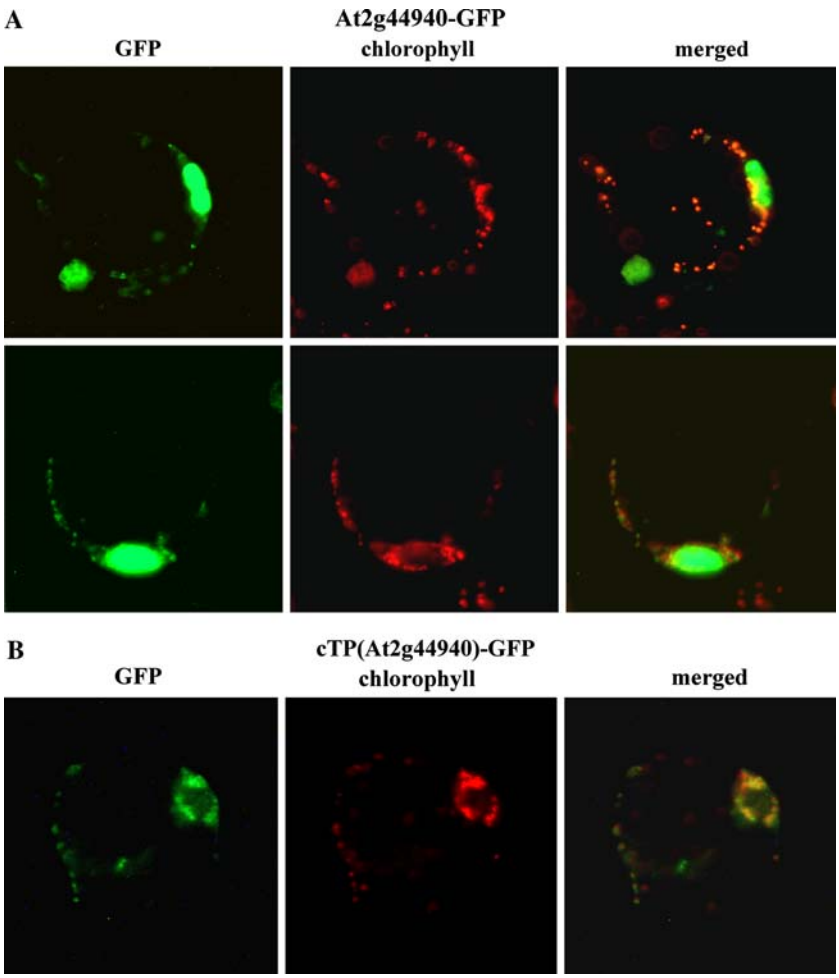

Fig. 5 Subcellular localization of *At2g44940* gene products fused to GFP in Arabidopsis protoplasts. Fluorescent microscope images of GFP fluorescence and chlorophyll autofluorescence are shown in the *left and middle images*, respectively. The *third column on the right* depicts the merged images. **a** Two individual protoplasts that express the entire At2g44940 protein fused to GFP are shown. **b** One protoplast showing expression of the chloroplast target peptide (cTP At2g44940) fused to GFP is depicted

nucleocytoplasmic partitioning is regulated by a diurnal rhythm and by light conditions (Kircher et al. 2002; Chen et al. 2005) or phototropin 1 that moves from the plasma membrane to the cytosol in response to blue light (Sakamoto and Briggs 2002).

So far, we can only speculate on whether a scenario similar to the ones mentioned also applies to plant transcription factors, since experimental data on nucleus/plastid- and nucleus/mitochondria-targeted plant proteins are extremely scarce. An interesting example is provided, however, by the dually targeted plant protein SEBF. This protein possesses a functional plastid target peptide and an RNA-binding domain reminiscent of that of heterogenous nuclear ribonucleoproteins (hnRNPs) (Boyle and Brisson 2001). The processed mature form of the protein was detected in the chloroplasts and, surprisingly, also in the nucleus, whereas the unprocessed form did not occur there (Boyle and Brisson 2001). Since no indication for a differential splicing was obtained, this raises the question whether the precursor was processed outside the chloroplast or whether the imported mature plastid protein was re-targeted to the nucleus. In line with such speculations, observations regarding the physical interaction of plastids as well as mitochondria with the nuclear envelope gain importance. Plastids seem to be attracted to the nucleus under certain circumstances and can interact with the nuclear envelope through stroma-filled tubular extensions termed stromules (Kwok and Hanson 2004). A clustering of plastids around the nucleus was, surprisingly, also seen in Arabidopsis protoplasts expressing the At2g44940 fusion protein (Fig. 5). The reason for this is unclear. A similar behavior was recently reported for mitochondria that seem to accumulate close to the nuclear envelope in leaf mesophyll cells undergoing programmed cell death (Selga et al. 2005).

In contrast, disintegration of chloroplast envelope membranes and vesicle 'blebbing' have recently been brought up as possible fates of ageing chloroplasts in senescing plant cells (Krupinska 2005). According to this scenario, plastid proteins might be released to the cytosol under these conditions. From there they could be imported into the nucleus, as is the case for some mitochondrial proteins in animal cells undergoing apoptotic cell death (e.g. AIF, see previous). A conditional re-targeting of organellar proteins could represent a novel mechanism of communication between the nucleus and the organelles, especially in situations such as pathogen attack, abiotic stresses or senescence, and would add a new dimension to our knowledge on the complex network of intercompartmental crosstalk. Indeed, a number of the proteins identified by our screen belong to families such as the DREB proteins whose association with stress responses is known. These proteins would thus be candidates for such a regulatory role.

In summary, our survey demonstrates the likely existence of more than the currently known proteins with nuclear as well as plastid or mitochondrial localization. Many of these factors belong to families that respond to external or internal stress stimuli and play a role in stress response reactions. Whether these putative dually targeted proteins are indeed part of the interorganellar communication network in plant cells and are able to affect the gene expression in two or more compartments and thereby contribute to stress response reactions will certainly be revealed in the future by a closer characterization of these proteins.

Acknowledgments The authors gratefully acknowledge Dr. Mario Brosch and Isabell Kilbienski (University of Kiel) for stimulating discussions. Prof. Martin Huelskamp (University of Cologne) is thanked for providing the Arabidopsis cell culture.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Bae MS, Cho EJ, Choi EY, Park OK (2003) Analysis of the Arabidopsis nuclear proteome and its response to cold stress. Plant J 36:652–663

Bannai H, Tamada Y, Maruyama O, Nakai K, Miyano S (2002) Extensive feature detection of N-terminal protein sorting signals. Bioinformatics 18:298–305

Baumbusch LO, Thorstensen T, Krauss V, Fischer A, Naumann K, Assalkhou R, Schulz I, Reuter G, Aalen RB (2001) The *Arabidopsis thaliana* genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. Nucleic Acids Res 29:4319–4333

Beck CF (2005) Signalling pathways from the chloroplast to the nucleus. Planta 222:743–756

Bodén M, Hawkins J (2005) Prediction of subcellular localization using sequence-biased recurrent networks. Bioinformatics 21:2279–2286

Boguta M, Hunter LA, Shen W-C, Gillman EC, Martin NC, Hopper AK (1994) Subcellular locations of MOD5 proteins: mapping of sequences sufficient for targeting to mitochondria and demonstration that mitochondrial and nuclear isoforms comingle in the cytosol. Mol Cell Biol 14:2298–2306

Bomsztyk K, Denisenko O, Ostrowski J (2004) hnRNP K: one protein multiple processes. BioEssays 26:629–638

Boyle B, Brisson N (2001) Repression of the defense gene PR-10a by the single-stranded DNA binding protein SEBF. Plant Cell 13:2525–2537

Bruce BD (2000) Chloroplast transit peptides: structure, function and evolution. Trends Cell Biol 10:440–447

Chen M, Tao Y, Lim J, Shaw A, Chory J (2005) Regulation of phytochrome B nuclear localization through light-dependent unmasking of nuclear localization signals. Curr Biol 15:637–642

Choi Y, Harada JJ, Goldberg RB, Fischer RL (2004) An invariant aspartic acid in the DNA glycosylase domain of DEMETER is necessary for transcriptional activation of the imprinted MEDEA gene. Proc Natl Acad Sci USA 101:7481–7486

Claros MG, Vincens P (1996) Computational method to predict mitochondrially imported proteins and their targeting sequences. Eur J Biochem 241:779–786

Cokol M, Nair R, Rost B (2000) Finding nuclear localization signals. EMBO Rep 1:411–415

Cregan SP, Fortin A, MacLaurin JG, Callaghan SM, Cecconi F, Yu SW, Dawson TM, Dawson VL, Park DS, Kroemer G, Slack RS (2002) Apoptosis-inducing factor is involved in the regulation of caspase-independent neuronal cell death. J Cell Biol 158:507–517

Curaba J, Herzog M, Vachon G (2003) GeBP, the first member of a new gene family in Arabidopsis, encoded a nuclear protein with DNA-binding activity and is regulated by KNAT1. Plant J 33:305–317

Davuluri RV, Sun H, Palaniswamy SK, Matthews N, Molina C, Kurtz M, Grotewold E (2003) AGRIS: Arabidopsis Gene Regulatory Information Server, an information resource of Arabidopsis cis-regulatory elements and transcription factors. BMC Bioinformatics 4:25

Desveaux D, Després C, Joyeux A, Subramaniam R, Brisson N (2000) PBF-2 is a novel single-stranded DNA binding factor implicated in PR-10a gene activation in potato. Plant Cell 12:1477–1489

Desveaux D, Subramaniam R, Despres C, Mess JN, Levesque C, Fobert PR, Dangl JL, Brisson N (2004) A ''Whirly'' transcription factor is required for salicylic acid-dependent disease resistance in Arabidopsis. Dev Cell 6:229–240

Ellis SR, Hopper AK, Martin NC (1989) Amino-terminal extension generated from an upstream AUG codon increases the efficiency of mitochondrial import of yeast N^2,N^2-dimethylguanosine-specific tRNA methyltransferases. Mol Cell Biol 9:1611–1620

Emanuelsson O, Nielsen H, von Heijne G (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci 8:978–984

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300:1005–1016

Guda C, Fahy E, Subramaniam S (2004) MITOPRED: a genome-scale method for prediction of nuclear-encoded mitochondrial proteins. Bioinformatics 20:1785–1794

Hao D, Yamasaki K, Sarai A, Ohme-Takagi M (2002) Determinants in the sequence specific binding of two plant transcription factors, DBF1 and NtERF2, to the DRE and GCC motifs. Biochemistry 41:4202–4208

Harrison DJ, Langdale JA (2006) A step by step guide to phylogeny reconstruction. Plant J 45:561–572

Heazlewood JL, Tonti-Filippini J, Verboom RE, Millar H (2005) Combining experimental and predicted datasets for determination of the subcellular location of proteins in Arabidopsis. Plant Physiol 139:598–609

Heidstra R, Welch D, Sheres B (2004) Mosaic analyses using marked activation and deletion clones to dissect Arabidopsis SCARECROW action in asymmetric cell division. Genes Dev 18:1964–1969

Horton P, Park K-J, Obayashi T, Nakai K (2006) Protein subcellular localization prediction with WoLF PSORT. In: Proceedings of the 4th annual Asia Pacific bioinformatics conference APBC06, Taipei, Taiwan, pp 39–48

Karniely S, Pines O (2005) Single translation—dual destination: mechanisms of dual protein targeting in eukaryotes. EMBO Rep 6:420–425

Kircher S, Gil P, Kozma-Bognár L, Fejes E, Speth V, Husselstein-Muller T, Bauer D, Adam E, Schäfer E, Nagy F (2002) Nucleocytoplasmic partitioning of the plant photoreceptors phytochrome A, B, C, D, and E is regulated differentially by light and exhibits a diurnal rhythm. Plant Cell 14:1541–1555

Koroleva OA, Tomlinson ML, Leader D, Shaw P, Doonan JH (2005) High-throughput protein localization in Arabidopsis using Agrobacterium-mediated transient expression of GFP–ORF-fusions. Plant J 41:162–174

Krause K, Kilbienski I, Mulisch M, Rödiger A, Schäfer A, Krupinska K (2005) DNA-binding proteins of the Whirly family in *Arabidopsis thaliana* are targeted to the organelles. FEBS Lett 579:3707–3712

Krupinska K (2005) Fate and activities of plastids during leaf senescence. In: Wise RR, Hoober JK (eds) The structure and function of plastids. Springer, Heidelberg, pp 433–449

Kwok EY, Hanson MR (2004) Plastids and stromules interact with the nucleus and cell membrane in vascular plants. Plant Cell Rep 23:188–195

Liu L, White MJ, MacRae TH (1999) Transcription factors and their genes in higher plants. Functional domains, evolution and regulation. Eur J Biochem 262:247–257

Luo M, Orsi R, Patrucco E, Pancaldi SRC (1997) Multiple transcription start sites of the carrot dihydrofolate reductase-thymidylate synthase gene, and sub-cellular localization of the bifunctional protein. Plant Mol Biol 33:709–722

Magnani E, Sjölander K, Hake S (2004) From endonucleases to transcription factors: evolution of the AP2 DNA binding domain in plants. Plant Cell 16:2265–2277

Nakai K, Horton P (1999) PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem Sci 24:34–36

Nakano T, Suzuki K, Fujimura T, Shinshi H (2006) Genome wide analysis of the ERF gene family in Arabidopsis and rice. Plant Physiol 140:411–432

Negrutiu I, Shillito RD, Potrykus I, Biasini G, Sala F (1987) Hybrid gene in the analysis of transformation conditions. I. Setting up a simple method for direct gene transfer in plant protoplasts. Plant Mol Biol 8:363–373

Nielsen H, Engelbrecht J, Brunak B, von Heijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng 10:1–6

Nuhse TS, Stensballe A, Jensen ON, Peck SC (2003) Large-scale analysis of in vivo phosphorylated membrane proteins by immobilized metal ion affinity chromatography and mass spectrometry. Mol Cell Proteomics 2:1234–1243

Petsalaki EI, Bagos PG, Litou ZI, Hamodrakas SJ (2006) PredSL: a tool for the N-terminal sequence-based prediction of protein subcellular localization. Genomics Proteomics Bioinformatics 4:48–55

Raynaud C, Sozzani R, Glab N, Domenichini S, Perennes C, Cella R, Kondorosi E, Bergouinoux C (2006) Two cell-cycle regulated SET-domain proteins interact with proliferating cell nuclear antigen (PCNA) in Arabidopsis. Plant J 47:395–407

Richly E, Dietzmann A, Biehl A, Kurth J, Laloi C, Apel K, Salamini F, Leister D (2003) Covariations in the nuclear chloroplast transcriptome reveal a regulatory master-switch. EMBO Rep 4:491–498

Riechmann JL, Meyerowitz EM (1998) The AP2/EREBP family of plant transcription factors. Biol Chem 379:633–646

Ruchalski K, Mao H, Li Z, Wang Z, Gillers S, Wang Y, Mosser DD, Gabai V, Schwartz JH, Borkan SC (2006) Distinct

hsp70 domains mediate apoptosis-inducing factor release and nuclear accumulation. J Biol Chem 281:7873–7880
Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425
Sakamoto K, Briggs WR (2002) Cellular and subcellular localization of phototropin 1. Plant Cell 14:1723–1735
Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K, Yamaguchi-Shinozaki K (2002) DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. Biochem Biophys Res Commun 290:998–1009
Schein AI, Kissinger JC, Ungar LH (2001) Chloroplast transit peptide prediction: a peek behind the black box. Nucleic Acids Res 29:82
Schneider M, Bairoch A, Wu CH, Apweiler R (2005) Plant protein annotation in the UniProt Knowledgebase. Plant Physiol 138:59–66
Selga T, Selga M, Pavila V (2005) Death of mitochondria during programmed cell death of leaf mesophyll cells. Cell Biol Int 29:1050–1056
Shigyo M, Hasabe M, Ito M (2006) Molecular evolution of the AP2 subfamily. Gene 366:256–265
Silva-Filho MC (2003) One ticket for multiple destinations: dual targeting of proteins to distinct subcellular locations. Curr Opin Plant Biol 6:589–595
Slupphaug G, Markussen F-H, Olsen LC, Aasland R, Aarsaether N, Bakke O, Krokan HE, Helland DE (1993) Nuclear and mitochondrial forms of human uracil-DNA glycosylase are encoded by the same gene. Nucleic Acids Res 21:2579–2584
Small I, Wintz H, Akashi K, Mireau M (1998) Two birds with one stone: genes that encode products targeted to two or more compartments. Plant Mol Biol 38:265–277
Small I, Peeters N, Legeai F, Lurin C (2004) Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics 4:1581–1590
Springer NM, Napoli CA, Selinger DA, Pandey R, Cone KC, Chandler VL, Kaeppler HF, Kaeppler SM (2003) Comparative analysis of SET domain proteins in maize and Arabidopsis reveals multiple duplications preceding the divergence of monocots and dicots. Plant Physiol 132:907–925
Strand A (2004) Plastid-to-nucleus signalling. Curr Opin Plant Biol 7:621–625
Stribinskis V, Heyman H-C, Ellis SR, Steffen MC, Martin NC (2005) Rpm2p, a component of yeast mitochondrial RNaseP, acts as a transcriptional activator in the nucleus. Mol Cell Biol 25:6546–6558
Sunderland PA, West CE, Waterworth WM, Bray CM (2004) Choice of a start codon in a single transcript determines DNA ligase 1 isoform production and intracellular targeting in *Arabidopsis thaliana*. Biochem Soc Trans 32:614–616
Sunderland PA, West CE, Waterworth WM, Bray CM (2006) An evolutionarily conserved translation initiation mechanism regulates nuclear or mitochondrial targeting of DNA ligase 1 in *Arabidopsis thaliana*. Plant J 47:356–367
Susin SA, Lorenzo HK, Zamzami N, Marzo I, Snow BE, Brothers GM, Mangion J, Jacotot E, Costantini P, Loeffler M, Larochette N, Goodlett DR, Aebersold R, Siderovski DP, Penninger JM, Kroemer G (1999) Molecular characterization of mitochondrial apoptosis-inducing factor. Nature 397:441–446
Thompson JD, Gibson TJ, Plewiak F, Jeanmougin F, Higgins DG (1997) The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882
Wagner R, Pfannschmidt T (2006) Eukaryotic transcription factors in plastids—bioinformatic assessment and implications for the evolution of gene expression machineries in plants. Gene 381:62–70
Weigel D (1995) The APETALA2 domain is related to a novel type of DNA binding domain. Plant Cell 7:388–389
Wessler SR (2005) Homing into the origin of the AP2 DNA binding domain. Trends Plant Sci 10:54–56
Wissing S, Ludovico P, Herker E, Büttner S, Engelhardt SM, Decker T, Link A, Proksch A, Rodrigues F, Corte-Real M, Fröhlich K-U, Manns J, Candé C, Sigrist SJ, Kroemer G, Madeo F (2004) An AIF orthologue regulates apoptosis in yeast. J Cell Biol 166:969–974
Wolfe CL, Lou Y-C, Hopper AK, Martin NC (1994) Interplay of heterogenous transcriptional start sites and translational selection of AUGs dictate the production of mitochondrial and cytosolic/nuclear tRNA nucleotidyltransferase from the same gene in yeast. J Biol Chem 269:13361–13366
Wolfe CL, Hopper AK, Martin NC (1996) Mechanisms leading to and the consequences of altering the normal distribution of ATP (CTP):tRNA nucleotidyltransferase in yeast. J Biol Chem 271:4679–4686
Xiong Y, Liu T, Tian C, Sun S, Li J, Chen M (2005) Transcription factors in rice: a genome-wide comparative analysis between monocots and eudicots. Plant Mol Biol 59:191–203