Planta (2002) 216: 107–119
DOI 10.1007/s00425-002-0890-6

ORIGINAL ARTICLE

**H. Üner Kolukisaoglu · Lucien Bovet · Markus Klein
Thomas Eggmann · Markus Geisler · Dierk Wanke
Enrico Martinoia · Burkhard Schulz**

Family business: the multidrug-resistance related protein (MRP) ABC transporter genes in *Arabidopsis thaliana*

Received: 2 May 2002 / Accepted: 9 August 2002

Abstract Despite the completion of the sequencing of the entire genome of *Arabidopsis thaliana* (L.) Heynh., the exact determination of each single gene and its function remains an open question. This is especially true for multigene families. An approach that combines analysis of genomic structure, expression data and functional genomics to ascertain the role of the members of the multidrug-resistance-related protein (*MRP*) gene family, a subfamily of the ATP-binding cassette (ABC) transporters from *Arabidopsis* is presented. We used cDNA sequencing and alignment-based re-annotation of genomic sequences to define the exact genic structure of all known *AtMRP* genes. Analysis of promoter regions suggested different induction conditions even for closely related genes. Expression analysis for the entire gene family confirmed these assumptions. Phylogenetic analysis and determination of segmental duplication in the regions of *AtMRP* genes revealed that the evolution of the extraordinarily high number of ABC transporter genes in plants cannot solely be explained by polyploidisation during the evolution of the *Arabidopsis* genome. Interestingly *MRP* genes from *Oryza sativa* L. (rice; *OsMRP*) show very similar genomic structures to those from *Arabidopsis*. Screening of large populations of T-DNA-mutagenised lines of *A. thaliana* resulted in the isolation of AtMRP insertion mutants. This work opens the way for the defined analysis of a multigene family of important membrane transporters whose broad variety of functions expands their traditional role as cellular detoxifiers.

Keywords ABC transporter · *Arabidopsis* · Functional genomics · Gene family evolution · MRP genes

Abbreviations aa: amino acid · ABC: ATP-binding cassette · dag: days after germination · EST: expressed sequence tag · MDR: multi-drug resistance protein · MRP: multi-drug-resistance-related protein · MSD: membrane-spanning domain · NBD: nucleotide-binding domain · nt: nucleotide · PGP: p-glycoprotein · SA: salicylic acid

Dedicated to Nikolaus Amrhein, Zürich, on the occasion of his 60th birthday.

H.Ü. Kolukisaoglu · D. Wanke · B. Schulz (✉)
Universität zu Köln, Botanik II (AG Flügge),
Max-Delbrück-Laboratorium in der Max-Planck-Gesellschaft,
Carl-von-Linné-Weg 10, 50829 Köln, Germany

L. Bovet · M. Klein · T. Eggmann · M. Geisler · E. Martinoia
Université de Neuchâtel, Physiologie Végétale,
Rue Emile-Argand 13, 2007 Neuchâtel, Switzerland

Present address: H.Ü. Kolukisaoglu
Universität Rostock, Fachbereich Biowissenschaften,
Abteilung Pflanzenphysiologie,
Albert-Einstein-Str. 3, 18051 Rostock, Germany

Present address: D. Wanke
Max-Planck-Institut für Züchtungsforschung,
Abt. Hahlbrock, Carl von Linne Weg 10,
50829 Köln, Germany

Present address: B. Schulz
Universität Tübingen,
ZMBP-Pflanzenphysiologie, Auf der Morgenstelle 5,
72076 Tübingen, Germany
e-mail: burkhard.schulz@zmbp.uni-tuebingen.de
Fax: +49-7071-295135

Introduction

The first complete genomic sequence of a flowering plant provides invaluable insight into the inventory, composition and organisation of higher-plant genes. One striking result of the analysed data was the significantly higher proportion of gene families in the *Arabidopsis* genome than in prokaryotic and other eukaryotic genomes. It was assumed that this form of gene multiplication is mostly due to duplication of chromosome segments (Arabidopsis Genome Initiative 2000; Blanc et al. 2000). The functional definition of each gene of a gene family is the challenge of *Arabidopsis* post-genomics.

Although the functional analysis of ABC transporters is a field of increasing research interest in recent years,

little is known about their biological function in plants. Additionally, there is an increasing number of reports supporting the assumption that the function of these proteins is not limited to the so-called phase-III detoxification (Martinoia et al. 1993; Kreuz et al. 1996), but that they are also involved in ion transport regulation, auxin distribution, peroxisomal fatty acid transport, Fe incorporation in proteins or protochlorophyllide transport (Gaedeke et al. 2001; Kushnir et al. 2001; Møller et al. 2001; Zolman et al. 2001; Martinoia et al. 2002). Dealing with the genomics of such a large gene family as the ABC transporters creates several difficulties that have to be overcome. The large number of ABC transporter genes have to be ordered in a reasonable way into subgroups that facilitate the functional analysis of these groups. A subfamily of ABC transporters can be defined as a group of paralogous genes that exhibit the same exon–intron pattern. In some cases, such as the *AtMRP* genes, even more than one subgroup can be found on the basis of splicing patterns. A critical analysis of the data compiled in the public databases such as the MIPS database (http://mips.gsf.de/proj/thal/) has to be done, because in many cases annotation based on gene prediction programs produces conflicting results. Therefore, the more intimate knowledge of a gene family and its biology can strongly support the genomics work. Direct use of data from the public databases without re-annotation might give a first impression of a gene family like the ABC transporters (Sánchez-Fernández et al. 2001). However, a detailed analysis of the combination of predicted genomic data and expressed sequences of genes provides a more complete view of this gene family. Such an analysis might also help to facilitate the elucidation of gene function. Several aspects of genome analysis such as (i) investigation of the physical structure of genes by comparing cDNA/expressed sequence tag (EST) sequences and genomic sequences to determine splicing patterns, putative start and stop sites of a gene; (ii) analysis of gene origin in a genome by chromosomal segmental or gene duplication, as well as promoter and regulator sequence analysis; and (iii) isolation of mutants for the investigated genes, provide a great amount of data which support the functional analysis of an entire gene family.

A structural characteristic of all ABC transporters is the combination of integral membrane-spanning domains (MSDs) with cytoplasmic nucleotide-binding domains (NBDs). ABC proteins that consist of only one pair of these functional domains are called ''half-size'' ABC transporters, which can form a ''full-size'' ABC transporter through dimerisation of two independent polypeptides. ''Full-size'' ABC transporter proteins consist of a combination of two MSDs and NBDs each. The arrangement of these domains as MSD-NBD-MSD-NBD is characteristic of the two subgroups MDR/PGP (multidrug-resistance protein/p-glycoprotein) and MRP (multidrug-resistance-related protein). The mirror-image structure NBD-MSD-NBD-MSD is the hallmark of the PDR (pleiotropic drug resistance) and the ABC1-like transporters, as well as the PMP70 ABC transporters (Martinoia et al. 2002). A total of 53 genes for full-size ABC transporters have been found in the nearly completed *Arabidopsis* genome sequence (Arabidopsis Genome Initiative 2000). Twenty-two of these belong to the AtMDR subfamily, which contains the majority of all full-size ABC transporter genes. The closely related MRP gene family consists of 14 members, whereas for PDR-like ABC transporter genes 15 members could be found. The remaining two full-size ABC transporter genes, *AtABC1* and *AtPXA1*/*Ped3p*/*COMATOSE* (Zolman et al. 2001; Footitt et al. 2002; Hayashi et al. 2002), are single-copy genes without any paralogs in the *Arabidopsis* genome.

Here we describe a comprehensive analysis of the *AtMRP* gene family, which is still the best characterised group of ABC transporters in plants to date. However, most genes of this gene family are only poorly characterised. To present a comprehensive impression of this gene family, several methods of analysis have been combined. This diverse array of procedures will help us to ask revealing questions about the specific functional role of each and every member of this large gene family of *Arabidopsis* membrane transporters.

Materials and methods

Computational analysis of AtMRP sequences and their genomic loci

The *AtMRP* genes were identified by comparison of the *AtMRP1*–*AtMRP5* sequences (Lu et al. 1997, 1998; Tommasini et al. 1997; Sánchez-Fernández et al. 1998; Gaedeke et al. 2001) with entries in the GenBank database (http://www.ncbi.nlm.nih.gov/BLAST) and the *Arabidopsis* MIPS database (http://mips.gsf.de/proj/thal/db/index.html) using BLASTP and TBLASTN algorithms (Altschul et al. 1990, 1997). Putative protein sequences were derived from annotation of the sequences and aligned with Clustal X (Thompson et al. 1997). Only sequences meeting the following two criteria were included in the *AtMRP* gene family: they had to contain two Walker A and B boxes, and ABC motifs at conserved positions with preceding transmembrane domains (Martinoia et al. 2002). In addition, the homology of *AtMRP* sequences had to be highest to other characterised members of this gene family and not to other members of the ABC proteins. Sequence annotations differing between the databases or aberrant sequences were either re-annotated, based on alignment or identical intron position. In questionable cases, either the corresponding EST sequences were compared using BLASTN in the GenBank EST database or cDNA sequences were isolated and sequenced. Multiple alignments were processed with MEGALIGN (DNASTAR) and corrected manually to determine the homology scores.

To construct a phylogenetic tree with the program package PHYLIP version 3.6 alpha (http://evolution.genetics.washington.edu/phylip.html) the AtMRP amino acid sequences were aligned together with chosen representatives of MRPs and PGPs from other species. Therefore, a PAM matrix (Dayhoff 1979) was calculated in PROTDIST and the program NEIGHBOR was employed to construct a neighbour-joining tree (in both programs default settings were used). In a successive bootstrap analysis (100 replicates) the topology of the tree was confirmed.

To analyse chromosomal segment duplications of loci bearing *AtMRP* genes, their map positions in the *Arabidopsis* genome were determined based on the sequence information of the BAC or P1 clones that harbor the genes. Comparison of the surroundings of

AtMRP loci with each other allowed duplicated segments in the *Arabidopsis* genome to be found. To find other homologous loci (lacking *AtMRP* genes), BLASTN searches in the non-redundant *Arabidopsis* GenBank database with coding sequences flanking the *AtMRP* genes were performed. If at least three sequences in the proximity (< 50 kb) of a particular *AtMRP* showed significant homology (BLAST cutoff 10^{-10}) to sequences of another genomic contig of similar size, both loci were analysed in more detail. For this purpose genomic contigs of at least 100 kb were compared using BLAST 2 sequences (http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html). Two loci were regarded as duplicated when at least eight different genes appeared in the considered region with nucleotide identities of more than 75% and mostly identical orientation.

Transcript levels of *AtMRP* genes

To analyse the expression levels of *AtMRP* genes in different organs and developmental stages, reverse transcription–polymerase chain reaction (RT–PCR) experiments were performed. To achieve specific amplification of individual genes in the RT–PCR reaction one pair of PCR primers was designed for each *AtMRP* gene. The primers were denoted as follows: gene name (*MRP1* to *MRP14*)-S (sense) or (*MRP1* to *MRP14*)-AS (antisense). Each primer pair spanning at least one intron was tested with genomic DNA and cDNA as template. PCR products were sequenced to ensure identity with the template sequence. The pairs of primers used to amplify the different *AtMRP* genes by RT–PCR were:

- *AtMRP1-S*, ccgcagaaatcctcttggtcttgatg; *AtMRP1-AS*, gtgaatcat caccgttagcttctctgg;
- *AtMRP2-S*, ccgcagaaatcctcttggtcttgatg; *AtMRP2-AS*, ccttgtaag tggtgtgagtcatctttgg;
- AtMRP3-S, ccactgcttctgttgacactg; *AtMRP3-AS*, gaggtgtactcag ccacaagc;
- *AtMRP4-S*, ctggaacggtgtcaactcaaggatgttg; *AtMRP4-AS*, attccg gcagatcggagagcgtactctt;
- *AtMRP-5S*, cacttggacgagcattactga; *AtMRP5-AS*, tcttctaatag ccgtgcagga;
- *AtMRP6-S*, ggtcagagacaattggtgtgc; *AtMRP6-AS*, accttggtctagg agcaggac;
- *AtMRP-7S*, aactggtgtgtcttggacgag ; *AtMRP7-AS*, tcttgaatccgaa cttgctgt;
- *AtMRP-8S*, ctctgcaagggagctgattaggatca; *AtMRP8-AS*, tggagc tatgtagcccttgggaataa;
- *AtMRP9-S*, gccactgcttctgttgattct; *AtMRP9-AS*, gagccggcaaagtg attagat;
- *AtMRP10-S*, cgagcgatgtcaacttaagga; *AtMRP10-AS*, gcaaacaacg actgtctctcc;
- *AtMRP11-S*, tcgctcacagattgaatacca; *AtMRP11-AS*, ccaccttgactc attccattc;
- *AtMRP12-S*, attcgcgaggaattcaagtct; *AtMRP12-AS*, cacccacac tcattccattct;
- *AtMRP13-S*, tcagagccctttctgtttca; *AtMRP13-AS*, tcgtgttgtggag tagggaag;
- *AtMRP14-S*, ccacggcatcgatagataatg; *AtMRP14-AS*, gccgagttg-taatgagaccaa.

As an internal control, the genes coding for the ribosomal protein S16 (At5g18380) and actin2 (At3g18780) were chosen. PCR primers for these genes were:

- S16-S, ggcgactcaaccagctactga; S16-AS, cggtaactcttctggtaacga;
- actin2-S, tggaatccacgagacaaccta;actin2-AS, ttctgtgaacgattcctg-gac.

The RT–PCRs were performed in a final volume of 25 µl containing the following mixture: 1× PCR buffer, 0.2 mM dNTPs, 1 µM of both 5′ and 3′ primers, 1 U Taq DNA polymerase (Promega, Catalys, Wallisellen, Switzerland) and adjusted amounts of first-strand cDNA. Total RNA was purified from plants using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) and stored at –80 °C after quantification by spectrophotometry. After DNAse treatment (DNase, RQ1, RNase free; Promega, Catalys), first-strand cDNA was prepared using M-MLV reverse transcriptase, RNaseH minus, point mutant (Promega, Catalys) as indicated by the manufacturer and stored at –20 °C. First-strand cDNA was diluted 10- to 100-fold before being used as template in the PCR. After a 3-min denaturation step at 94 °C, 35 PCR cycles (94 °C for 45 s, 58 °C for 45 s and 72 °C for 1 min) were performed.

Semi-quantitative RT–PCR

The cDNA synthesis reactions were performed exactly as indicated above. For PCR amplification, 0.5 MBq α-[^{33}P]ATP (110 TBq/mmol) was added to the PCR mixture using the following set of dNTPs: 0.07 mM dATP, 0.2 mM dGTP, 0.2 mM dTTP and 0.2 mM dCTP. The number of PCR cycles was variable for both the control gene (S16) and the *AtMRP* genes. The linearity of amplification was verified by taking aliquots during the PCR after the elongation step of the PCR (72 °C). The radioactive PCR products were then separated by gel electrophoresis in 2% agarose gels. After incubation in 0.25 M HCl for 10 min and two washes in 2× SSC (0.03 M Na-citrate, 0.3 M NaCl, pH 7.4), the PCR products were transferred onto nitrocellulose membranes (Porablot NY plus; Macherey and Nagel, Düren, Germany) via vacuum-transfer for 2 h. The membranes were then subjected to overnight autoradiography (BIOMAX-MR; Kodak). The incorporation of radioactive nucleotides was quantified using a phosphoimager (BioRad, Rheinach, Switzerland).

Tissue culture of *Arabidopsis* seedlings

Seeds were surface-sterilised by immersion in a saturated aqueous solution of calcium hypochlorite [$Ca(OCl)_2$] for 10 min. After a short centrifugation (3,000 *g*), seeds were first washed with 70% ethanol and then four times with sterile distilled water. The seeds were incubated overnight at 4 °C in water, before placing them on 0.6% (w/v) bactoagar plates containing the following nutrient solution (1 l): 0.099 g *myo*-inositol, 2.5 mM Mes–KOH (pH 5.7), 10 g sucrose, 0.132 g $(NH_4)_2HPO_4$ and 0.163 g Hoaglands salts (H-2395; Sigma). After 7 days in the phytotron (25 °C, 16 h light and 70% humidity), seedlings were harvested and frozen in liquid nitrogen for RNA extraction.

Results

Annotation of *AtMRP* genes

To employ the information from completely sequenced eukaryotic genomes one has to ensure that the annotation of the genes of interest is correct. As annotation of the gene sequences of *Arabidopsis* is still in progress, we decided to verify the annotation of all database entries from the GenBank and MIPS databases for *MRP* genes and also to search for not yet annotated *MRP* genes in the genomic sequence of *A. thaliana* generated by the *Arabidopsis* Genome Initiative (2000; see also *Materials and methods*). Using this approach, we found 16 sequences with significant homologies to known *MRP* genes in the *Arabidopsis* genome.

As the next step of this analysis we determined the exact gene structure – showing exon and intron regions – of the putative *AtMRP* genes. This is mandatory to verify if all these sequences actually belong to the

AtMRP gene family or to other groups within the ABC transporter superfamily. The identification of pseudogenes would also be possible this way. For this purpose we included all available ESTs, cDNA sequences and protein sequence entries in our analysis. In cases where no such information was available, comparison- and alignment-based annotations were performed. For those genes where all these mentioned processes led to unsatisfactory or uncertain results the cDNA of the corresponding genes was isolated and sequenced.

As seen in Table 1, out of the 16 putative *AtMRP* gene sequences one is missing: in the case of At1g71330 it could be shown that the genomic sequence has a high homology to *AtMRP3*, but the annotated gene encodes just for one membrane-spanning region (data not shown). This sequence seems to be a remnant of an ancient gene duplication from an ancestral gene of *AtMRP3* and cannot be accounted as an *MRP* gene due to its aberrant structure.

For another gene, *AtMRP15*, the result of the annotation process was a matter of debate. Although the genomic sequence was nearly identical to that of *AtMRP9* (>99% nucleotide identity) the predicted amino acid (aa) sequence differed by more than 400 aa. This difference is due to three point mutations and an insertion of 230 bp in the N-terminus of AtMRP15 leading to the truncation of almost the complete first MSD. This structural aberration made it questionable whether AtMRP15 is a functional member of the *MRP* gene family in *Arabidopsis*. This assumption was also supported by expression analysis (see below), leading to the exclusion of this gene sequence from further considerations.

Structure of *AtMRP* genes

The predicted aa sequences of all found *AtMRP* genes lay within a length of 1,146–1,623 aa (Table 1) and showed the topology of full-length ABC transporters of the MDR/MRP class (MSD-NBD-MSD-NBD). Also, the three conserved sequence motifs (Walker A and B boxes and the ABC transporter consensus motif) were found in all nucleotide-binding domains of the AtMRPs (data not shown). In particular cases, differences from the consensus sequences occurred, which have also been described for other *Arabidopsis* ABC proteins (Davis and Coleman 2000; Martinoia et al. 2002).

All 14 *AtMRP* genes described in Table 1 share aa identity of 27.8% up to 90.2% (Table 2). The degree of identity is even higher if only the C-terminal part of the genes is considered. The relationships between the *AtMRP* genes are also supported by phylogenetic analysis. A phylogenetic tree of these genes, together with several other ABC genes from yeast and man, shows that the *AtMRP* genes form a common cluster together with their homologs from other species (Fig. 1). These data let us subdivide the *AtMRP* genes into at least two

Table 1. List of MRP genes in *Arabidopsis thaliana*. All characterised or putative *AtMRP* genes are listed in the first column. The second column shows the chromosomal location. Column 3 lists the protein entry codes of the MIPS *Arabidopsis thaliana* database (MATDB; http://mips.gsf.de/proj/thal/db). Accession numbers for the BAC clones (bacterial artificial chromosomes) and the available cDNA entries in GenBank are listed in the fourth and fifth columns. The lengths of the predicted proteins are given as the number of amino acids (*AA*). The length of the coding sequence (*CDS*), as the genomic sequence from translation start to stop codon of the open reading frame, is listed in base-pairs (column 7); the number of exons per gene is given in column 8. To determine the coding sequence of the *AtMRP* genes, BLAST searches with the annotated CDS were made using all GenBank ESTs in TAIR (http://arabidopsis.org/BLAST). When necessary, cDNA sequences of *AtMRP* genes were isolated and sequenced to confirm the annotation. *nd* Not determined

Gene	Locus	MIPS-Code	BAC-Acc.	Gene-Acc.	AA	CDS (genomic)	Exons	Reference	Remarks
AtMRP1	I/T4K22	At1g30400	AC025295	AF008124	1,622	9,559	28	Lu et al. 1997	Intron in 5′-UTR
AtMRP2	II/T29F13	At2g34660	AC003096	AF020288	1,623	9,121	26	Lu et al. 1998	–
AtMRP3	III/MJG19	At3g13080	AP000375	U92650	1,515	5,232	8	Tommasini et al. 1997	–
AtMRP4	II/F17A22	At2g47800	AC005309	AF243509	1,516	5,440	11	Sánchez-Fernández et al. 1998	–
AtMRP5	I/F20D22	At1g04120	AC002411	Y11250	1,514	5,549	11	Gaedeke et al. 2001	Intron in 5′-UTR
AtMRP6	III/MJG19	At3g13090	AP000375	AY052368	1,466	5,158	9	–	Referred as MRP8
AtMRP7	III/MJG19	At3g13100	AP000375	AJ507129	1,493	5,315	9	–	–
AtMRP8	III/MXL8	At3g21250	AB023045	AJ507057	1,294	4,823	11	–	–
AtMRP9	III/T2O9	At3g60160	AL138658	–	1,470	5,180	10	–	–
AtMRP10	III/F17J16	At3g59140	AL163527	–	1,454	5,183	11	–	–
AtMRP11	I/T4K22	At1g30420	AC025295	–	1,488	7,501	26	–	–
AtMRP12	I/T4K22	At1g30410	AC025295	–	1,488	7,661	26	–	–
AtMRP13	II/T5E7	At2g07680	AC006225	–	1,146	6,254	25	–	–
AtMRP14	III/F26K9	At3g62700	AL162651	–	1,539	5,300	9	–	–
AtMRP15	III/T27I15	At3g60970	AL358732	–	nd	nd	nd	–	Pseudogene

Table 2. Homology between AtMRPs. Homology scores of amino acid sequences are given for AtMRP1 to AtMRP14 as percent identity of amino acids. Homologies of genes at homeologous loci are indicated in *italics* and homologies between clustered genes are indicated in *bold*. AtMRP15 has been excluded from this comparison for reasons explained in the text

	At-MRP2	At-MRP3	At-MRP4	At-MRP5	At-MRP6	At-MRP7	At-MRP8	At-MRP9	At-MRP10	At-MRP11	At-MRP12	At-MRP13	At-MRP14
AtMRP1	*87.0*	31.5	33.3	33.1	33.7	32.2	32.7	34.4	33.2	**67.9**	**70.1**	30.8	31.4
AtMRP2	–	31.5	33.1	33.1	33.9	32.7	32.2	34.5	33.6	68.6	70.6	30.6	31.2
AtMRP3		–	40.8	49.5	**65.7**	**66.2**	40.0	49.1	40.9	31.1	30.6	29.7	39.9
AtMRP4			–	40.7	41.7	40.0	37.5	40.0	37.6	32.4	32.1	27.8	*82.0*
AtMRP5				–	48.4	47.4	40.8	50.7	41.5	32.9	32.7	29.3	40.3
AtMRP6					–	**81.8**	41.4	50.3	43.5	33.9	33.8	32.4	39.8
AtMRP7						–	40.0	48.7	42.1	33.4	33.0	31.1	38.1
AtMRP8							–	41.8	45.4	32.7	32.2	40.6	36.2
AtMRP9								–	42.4	33.4	33.3	31.7	37.7
AtMRP10									–	33.0	32.6	31.0	35.9
AtMRP11										–	**90.8**	30.6	30.4
AtMRP12											–	31.0	30.4
AtMRP13												–	26.3

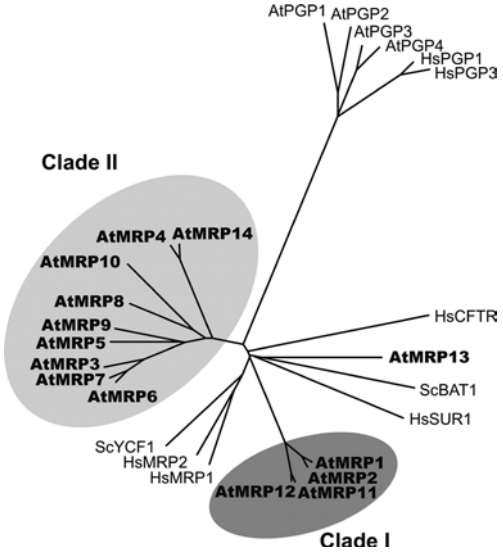

Fig. 1. Phylogenetic tree of the *AtMRP* gene family and other ABC transporter genes from *Arabidopsis thaliana*, yeast (*Saccharomyces cerevisiae*) and humans (*Homo sapiens*). Accession numbers for AtMRPs are given in Table 1. Yeast ABC transporters ScYCF1 (P39109), ScBAT (P32386), and human ABC transporters HsMRP1 (P33527), HsMRP2 (NP000383), HsSUR1 (Q09428), HsCFTR (M28668), HsPGP1 (P08183), and HsPGP3 (P21439) were used for comparisons. In addition, selected ABC transporters from the PGP/MDR gene subfamily from *Arabidopsis* were included in the comparison: AtPGP1 (At2g36910), AtPGP2 (At4g25960), AtPGP3 (At4g01820), AtPGP4 (At2g47000)

different clades. Members of clade I cluster with their non-plant homologs, whereas clade II seems to form a distinct plant-specific lineage. *AtMRP13* branches at a point in this tree clearly distinct from both mentioned clades. Its aberrant intron–exon structure, as well as the lack of sequence homology to any other *AtMRP* gene, makes it impossible to group this gene within the two *AtMRP* clades.

The classification into different clades or subfamilies is also reflected by the different exon–intron structures of the *AtMRP* genes (Fig. 2): *AtMRP1*, *2*, *11*, and *12*, which show a high aa sequence identity to each other (67.9–90.8%; Table 2), form a distinct subgroup of genes with a particular exon–intron structure. Here we find 26 or 27 introns distributed across the entire coding sequence. Only in *AtMRP1* was an additional intron found in the 5′ UTR, which is missing in the same region of the most related *AtMRP2* gene. Compared with *AtMRP11*/*12* the coding sequences of *AtMRP1* and *2* differ by one additional intron in the C-terminal part, leading to the extension of the predicted aa sequence of AtMRP1 and 2 by more than 100 aa.

Members of the *AtMRP* genes clustered in clade II show a consistent gene structure with eight introns spread over the region between the two nucleotide-binding motifs. In contrast to all other *AtMRP* genes the N-termini of these genes are entirely intronless. With respect to the gene structure, *AtMRP13* is clearly an exception within the *AtMRP* genes. We found no N-terminal extension and a truncated first membrane-spanning region. The positions of nearly all 24 introns found in this gene were unique for *AtMRP* genes (Fig. 2). That *AtMRP13* is a single-copy gene without close relatives in *Arabidopsis* supports the impression that it is an exceptional member of the *AtMRP* gene family.

It is noteworthy that the number, positions and phases of introns are highly conserved within the mentioned clades: as can be seen in Fig. 2, only 2 out of 26 or 28 introns within the coding sequence of *AtMRP* genes of clade I are at variable positions. One of the last introns of AtMRP2 is missing and there is an additional intron in *AtMRP1* and *AtMRP2* in the outermost 3′-end of the coding sequence. The situation is similar for the 10 introns of *AtMRP* genes of clade II. Only the introns 1 and 7 are missing in particular genes. Additionally, in *AtMRP3* the introns 6 and 9 are also not present. Apart from the described differences, all genomic sequences encoding *AtMRP* genes possess introns identical in number, position and phase.

Preliminary results of the analysis of the rice genome, which has been completely sequenced recently (Goff et al.

2002; Yu et al. 2002), revealed a set of *MRP* genes similar to those of *Arabidopsis*. Members of all described *AtMRP* clades were also present in the grass genome. Most interestingly, again only one homolog of *AtMRP13* is encoded in the rice genome, but there seems to be also just one representative of clade-I *MRP* genes. Both rice orthologues of *AtMRP1* and *13* are nearly identical in their genomic structure to their *Arabidopsis* counterparts. For instance AtMRP1 and its rice orthologue possess 27 introns in their coding sequences, of which 25 are identical in phase and position. Similar identities to *AtMRP* genes have also been observed for all other *MRP* genes in *Oryza*.

Expression of *AtMRP* genes

The transcript levels of *AtMRP* genes in different organs of *A. thaliana* have been examined using RT–PCR. Gene-specific primers were designed in order to amplify a cDNA fragment near the 3′-end of each gene. The RT–PCR products were sequenced to verify that only the transcripts of the corresponding gene were amplified. This approach circumvented the problem of cross-hybridisation with highly homologous gene sequences as in Northern blot or cDNA array experiments.

RT–PCR experiments (see *Material and methods*) with first-strand cDNAs from different organs as template showed that almost the entire set of *AtMRP* genes is expressed in leaves, roots, flowers, stems and siliques (Fig.3). Only in siliques could transcripts of *AtMRP2* and *AtMRP10* not be detected after ethidium bromide staining of the gel. Consequently, we performed an additional RT–PCR experiment in the presence of α-[^{33}P]ATP using the same cDNA from siliques as template. In this experiment, bands corresponding to *AtMRP2* and *AtMRP10* were detected on the autoradiograph, demonstrating low transcript levels of these genes in siliques (data not shown). RT–PCR analyses were also performed using RNAs isolated from seedlings 1–7 days after germination. Expression of all 14 *AtMRP* genes was detected throughout this time period (Fig. 4a). However, 24 h after plating seeds on agar plates, expression of *AtMRP8* and *AtMRP11* appeared to be weak or missing completely. Semi-quantitative RT–PCR using α-[^{33}P]ATP (Fig. 4b) revealed that *AtMRP8* transcripts were detectable in germinating seeds 1 day after plating, whereas

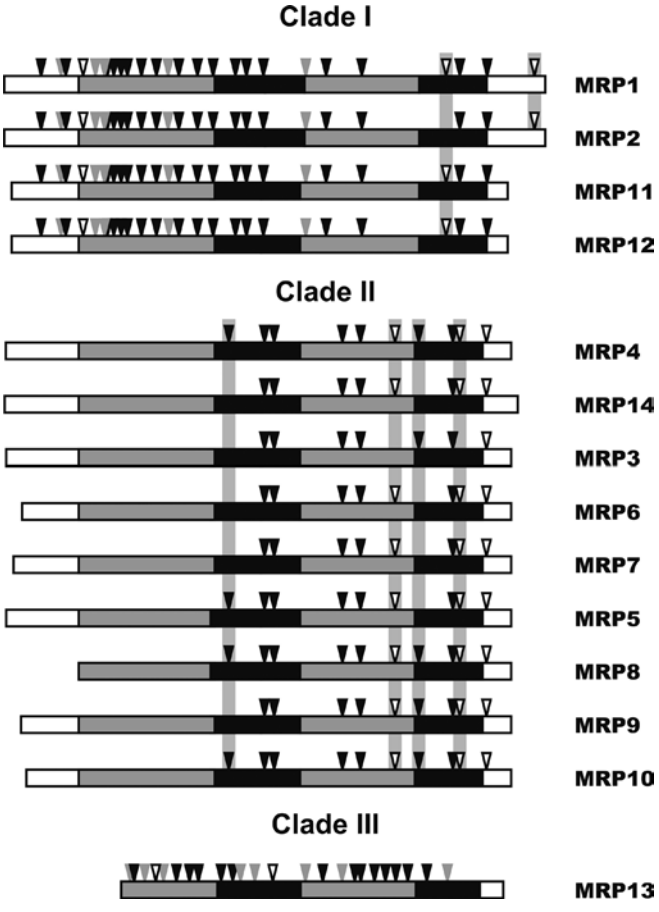

Fig. 2. Intron–exon structure of *AtMRP* genes. Each *AtMRP* gene has two transmembrane-spanning domains (MSD1 and MSD2; *grey bars*) and two nucleotide-binding domains (NBD1 and NBD2; *black bars*). Intron positions are indicated by *triangles*. The patterns of the triangles indicate the phase of introns: phase-0 introns (intron located between two codons) are drawn in *black*, phase-1 introns (intron located between first and second nucleotide of a codon) are drawn in *white*, and phase-2 introns (intron located between second and third nucleotide of a codon) are drawn in *grey*. Introns that are variable within one clade are highlighted by *grey boxes*

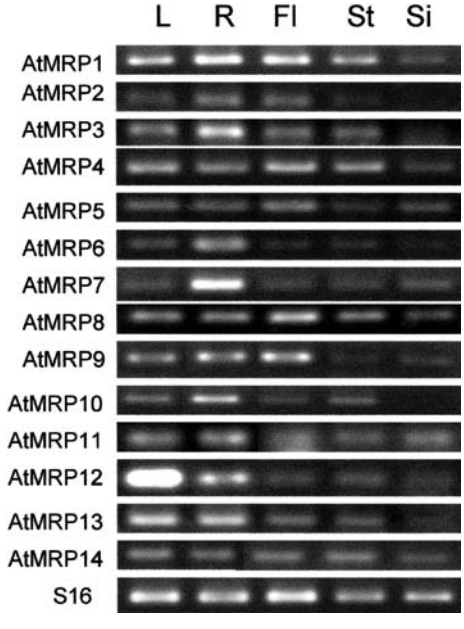

Fig. 3. Expression of *AtMRP* genes in different organs. Different organs (*L* leaves, *R* roots, *Fl* flowers, *St* stems, *Si* siliques) from *Arabidopsis thaliana* (ecotype Columbia) were collected and frozen in liquid nitrogen. RNA extraction was performed using the RNeasy plant mini Kit (Qiagen). Non-quantitative two-step RT–PCRs were performed using cDNA templates prepared as indicated in *Material and methods*. After 35 PCR cycles, amplified DNA 3′-end fragments were stained with ethidium bromide

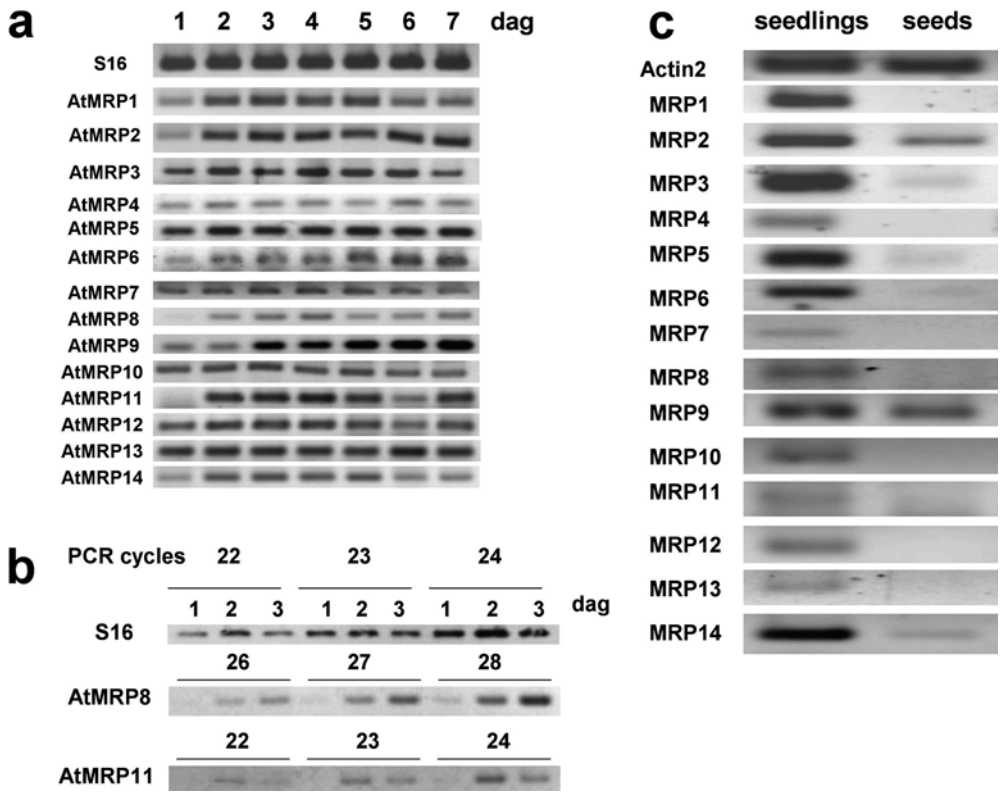

Fig. 4a–c. Expression of *AtMRP* genes in *A. thaliana* seedlings and seeds. Seedlings at different developmental stages grown for 1 week on agar-solidified medium were harvested from 1 day after germination (dag) to 7 dag for semiquantitative RT–PCR (**a**). Seedlings as well as seeds were frozen in liquid nitrogen and stored at –80 °C before RNA extraction. Non-quantitative two-step RT–PCRs were performed as indicated in Fig. 3. To show the accumulation of mRNA for *AtMRP8* and *AtMRP11* during the first 3 dag, semi-quantitative RT–PCRs were performed on cDNA templates at 1, 2 and 3 dag using primers that amplify *S16*, *AtMRP8* and *AtMRP11* (**b**). At the indicated amplification cycles, aliquots of 6 µl were collected from the PCR tubes after the 72 °C elongation step and run on an agarose gel. After transfer of the DNA onto nitrocellulose membranes, the PCR products were revealed by autoradiography after hybridisation with gene-specific probes. Differences in mRNA accumulation for all *AtMRP* genes in seedlings and seeds are shown in (**c**)

AtMRP11 was barely detectable at this stage. Expression of *AtMRP* was also analysed in dry seeds (Fig. 4c). Transcripts were detected for *AtMRP2*, *AtMRP3*, *AtMRP5*, *AtMRP9* and *AtMRP14*, suggesting involvement of these transporters (i) in detoxification processes of the seeds, (ii) in the late steps of grain filling or (iii) in very early processes of mobilisation. Although RT–PCR analyses of expression in organs and during germination were performed under non-quantitative conditions (Figs. 3, 4), in some cases elevated transcript levels of a particular gene could be observed: the expression of *AtMRP12* seemed to be strongly induced in leaves, whereas *AtMRP3*, *AtMRP6* and *AtMRP7* showed elevated levels of expression in roots. Although most of the *AtMRP* genes exhibited unaltered expression levels during germination, *AtMRP1*, *2*, *8*, *11* and *14* showed a strong induction at the second day after germination. In contrast, *AtMRP9* transcripts increased at day 3 after germination.

Identification of *AtMRP15* as a pseudogene

Another question was also answered by the application of RT–PCR: *AtMRP9* and *AtMRP15* are highly homologous. However, *AtMRP15* is likely to be a pseudogene because a deletion of 1 bp at position 513 and two insertions of 1 bp each at positions 765 and 789 lead to a premature stop after 549 bp. Therefore, only a truncated MRP peptide of 183 aa could be encoded by the *AtMRP15* gene. In addition, a 229-bp-long insertion at position 943 disrupts the *MRP15* gene sequence. To obtain experimental evidence for this assumption, we performed RT–PCR experiments with primers specific for *AtMRP15*. These experiments revealed no detectable *AtMRP15* expression with cDNAs from any of the tested organs (Fig. 5). Because of the high rate of sequence identity between the two genes (99%) only the sense primers were specific either for *AtMRP9* or *15*, whereas the antisense primer matched both of them. Although the gene-specific primers differed only in their first two 3′ bases their specificity could be validated by PCR on genomic DNA as template and subsequent analysis of restriction polymorphisms (Fig. 5). None of our experiments can exclude the possibility that *AtMRP15* is expressed under certain, but as yet untested, conditions. The fact that all other *AtMRP* genes are constitutively expressed, at least at a low level, in all organs analysed supports the assumption that this gene

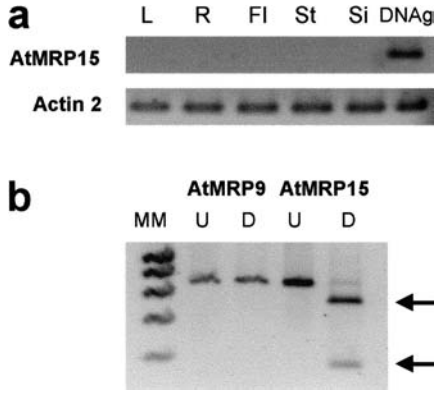

Fig. 5a, b. Expression of *AtMRP15* in different organs. **a** Using the same procedure described in Fig. 3, no *AtMRP15* transcripts were detected in the different organs tested. *L* leaves, *R* roots, *Fl* flowers, *St* stems, *Si* siliques, *DNAg* genomic DNA. **b** Using genomic DNA as template, PCR products of similar size (758 bp) were found for *AtMRP9* and *AtMRP15* when using the primers *AtMRP9-S*: cttgggattattcctcaagatcc, *AtMRP15-S*: cttgggattattcctcaagataa, and *AtMRP9,15-AS*: gagccggcaaagtgattagat (*lanes U*). Both PCR-amplified DNA fragments were then digested with *Hin*cII and separated on agarose gels (*lanes D*). The digested DNA fragments exhibited sizes predicted from genomic data (568 and 190 bp)

is rather a pseudogene than a transcribed member of the *AtMRP* gene family.

Analysis of promoter sequences and up-regulation of *AtMRP* genes

As little is known about the in planta function of AtMRPs, the presence of known *cis*-regulatory elements within a 1.5-kb promoter sequence of *AtMRP1* to *14* was examined by performing a PLACE database search (http://www.dna.affrc.go.jp/htdocs/PLACE; Higo et al. 1999). Only motifs that were at least hexamers with a minimum GC content of at least 2 GC nucleotides (nt) per hexamer were included in this analysis. We found 20 motifs that met these criteria. Most motifs of this subset were known to drive tissue-specific expression. As this specificity could not be sufficiently tested by RT–PCR experiments, only seven well-studied motifs were taken into consideration.

The conserved binding site TTGAC(C/T) for WRKY transcription factors, the W-Box (Eulgem et al. 2000; Maeo et al. 2001), was found in all but one of the promoter sequences (*AtMRP9*; Fig. 6). The TTGAC pentamer is invariant in all reports of WRKY protein/W-box interactions and appears 49 times on both strands in all *AtMRP* promoters. The second most abundant motif TGTCTC of the auxin-responsive element is the binding site for ARF1 (Ulmasov et al. 1997a, b), which was found 18 times in 8 *MRP* promoters. Also, uncommon motifs such as the As1-like elements TGACGT (13 times in 8 promoters) were found (Liu et al. 1994). Further motifs are the G-box CACGTG, the ABA-response element ACGTGTC, two jasmonic acid (JA)-response elements ACCGCC and two GCC-boxes (GCCGCC) (Liu et al. 1997; Menke et al. 1999; Ohme-Tagaki et al. 2000; Uno et al. 2000; Rushton et al. 2002).

Despite the fact that none of the motifs was significantly enriched in the *MRP* promoters according to the statistically expected frequency per 1 kb promoter sequence published by Maleck et al. (2000), we still expected at least one of the most abundant motifs to be functional. As an example, we focused on the hexameric W-boxes (26 copies in 14 *MRP* promoters), which are well characterised for their role in salicylic acid (SA)- and pathogen-induced defense pathways (Eulgem et al. 2000; Maleck et al. 2000).

Treatments of plants with SA induced the expression of *AtMRP4*, *AtMRP6* and *AtMRP7* (Fig. 7). Therefore,

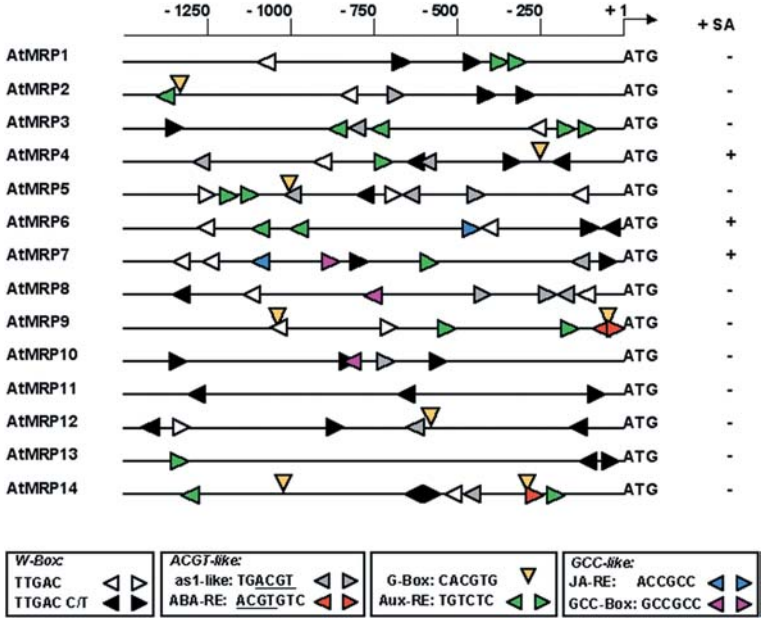

Fig. 6. Analysis of *AtMRP* genes, showing the presence of specific promoter elements that are well-described in the literature. The directions of the arrowheads indicate the orientation of the promoter elements. The palindromic sequence of the G-box element (CACGTG) is indicated as a *downward-pointing arrowhead*. Inducibility of the respective genes with SA is indicated (+; see also Fig. 7)

we examined the promoter sequences using MEME/MAST (http://meme.sdsc.edu) to identify motifs common in these three regulatory regions. The consensus sequences found with the algorithm did not contain patterns that were unique to the three genes examined. Thus, semi-quantitative RT–PCRs were performed to confirm the induction of *AtMRP4*, *6* and *7* by SA. Different PCR cycles for sampling aliquots showed that SA treatment was responsible for the induction of at least three of the *AtMRP* genes (Fig. 7).

Distribution of *AtMRP* genes in the *Arabidopsis* genome

Using the approaches outlined by several researchers (Blanc et al. 2000; Paterson et al. 2000; Vision et al. 2000) to analyse the *Arabidopsis* genome revealed that a large portion of the genome (58–64%) was arranged in large chromosomal segmental duplications. As sequence comparisons and analysis of the exon–intron structure had already shown that most *AtMRP* genes have closely related counterparts, it was obvious that the occurrence of *AtMRP* gene pairs originating through duplications of large chromosomal segments should be investigated.

The results of this analysis showed that 12 of the 15 *AtMRP* genes are located in duplicated chromosomal segments (Fig. 8a). Only for *AtMRP9*, *AtMRP13* and *AtMRP15* could no duplicated chromosomal segments be found. The gene pair *AtMRP9*/*AtMRP15* seems to have evolved through a genomic duplication event of a different type (see below). *AtMRP13* is located in the direct neighbourhood of the integration site of mitochondrial genome segments in chromosome 2, for which no duplication is known. In four of the six cases analysed the duplicated segments no longer harboured any *AtMRP* gene.

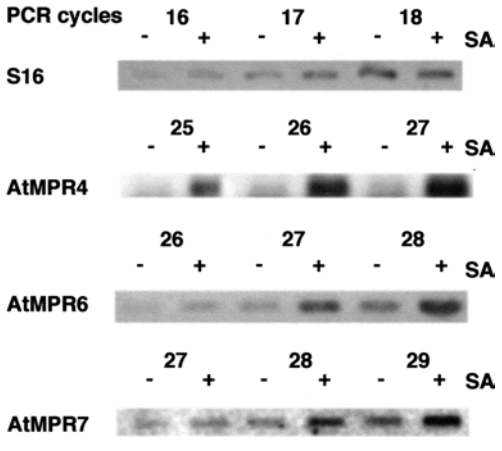

Fig. 7. Effect of salicylic acid (SA) on the expression of the *AtMRP* genes. *A. thaliana* seedlings were treated with 1 mM SA for 24 h before RNA extraction. Two-step qualitative RT-PCRs (35 PCR cycles) were performed with the corresponding cDNA templates. As the genes coding for *AtMRP 4*, *6* and *7* were induced by SA, semi-quantitative RT–PCRs were run for *AtMRP4*, *6* and *7*, as well as for S16 at PCR cycles indicated. Under these conditions the PCR was in the exponential phase of amplification

As shown in Fig. 8a, only *AtMRP4*/*AtMRP14* and *AtMRP1*/*AtMRP2* are located on duplicated chromosomal segments, which also harbour duplications of the corresponding *AtMRP* genes (Fig. 8b). In no other cases (Fig. 8b) were *AtMRP* genes found at the homologous loci. This means that in these cases either the AtMRP genes were lost during or after the duplication event or that they were transposed to their actual position. Although there is a lack of highly homologous gene pairs like *AtMRP4*/*AtMRP14* and *AtMRP1*/*AtMRP2* for the other cases mentioned (see also Table 2), this argument cannot rule out the possibility of transposition of an *AtMRP* followed by the loss of one of the copies.

Regarding the duplication of *AtMRP9*/*AtMRP15*, it is noteworthy that both genes show a high degree of homology to each other, even on the nucleotide level (95%). Moreover, they are not located on duplicated chromosomal segments, but both genes are located on chromosome 3, just about 100 kb apart from each other. Most probably they originated via transposition, which is supported by sequence identity of 93% in the 858 nt upstream of the start ATG codon and 98% identity in the 1,055 nt downstream of the stop codon of *AtMRP9* and *AtMRP15*, respectively. In addition to the remarkably high degree of sequence conservation in non-coding regions, highly identical transposable retroelements can be found just 3 kb upstream of both genes. This event must be rather recent because of the extraordinarily high degree of nucleotide identity between *AtMRP9* and *AtMRP15*.

The relative position of some *AtMRP* genes to each other might also have consequences for their functional relationship. In the cases of *AtMRP1*/*AtMRP12*/*AtMRP11* and *AtMRP3*/*AtMRP6*/*AtMRP7*, the three genes are each arranged as tandem duplications (Fig. 8a, b). In each cluster a pair of highly homologous genes could be found accompanied by an *AtMRP* with significantly lower homology. This observation gives rise to the assumption that in both clusters two independent duplication events took place. This model is supported by the duplication event leading to *AtMRP1* and *AtMRP2*. Although *AtMRP1* is found in close proximity to *AtMRP12* and *AtMRP11* on chromosome 1 its identity is highest toward *AtMRP2* on the duplicated chromosomal segment on chromosome 2 (Table 2, Fig. 8). This means that the ancestors of *AtMRP1* and *AtMRP11*/*12* existed before this segment of the *Arabidopsis* genome was duplicated, leading to the generation of an additional paralog – the ancestor of *AtMRP2*. The tandem duplication that resulted in the split divergence of *AtMRP11* and *AtMRP12* should have taken place during or after this event. This assumption is also supported by phylogenetic analysis showing that *AtMRP1* and *AtMRP2* are found on one branch whereas *AtMRP11* and *AtMRP12* are found on another branch of the tree (Fig. 1).

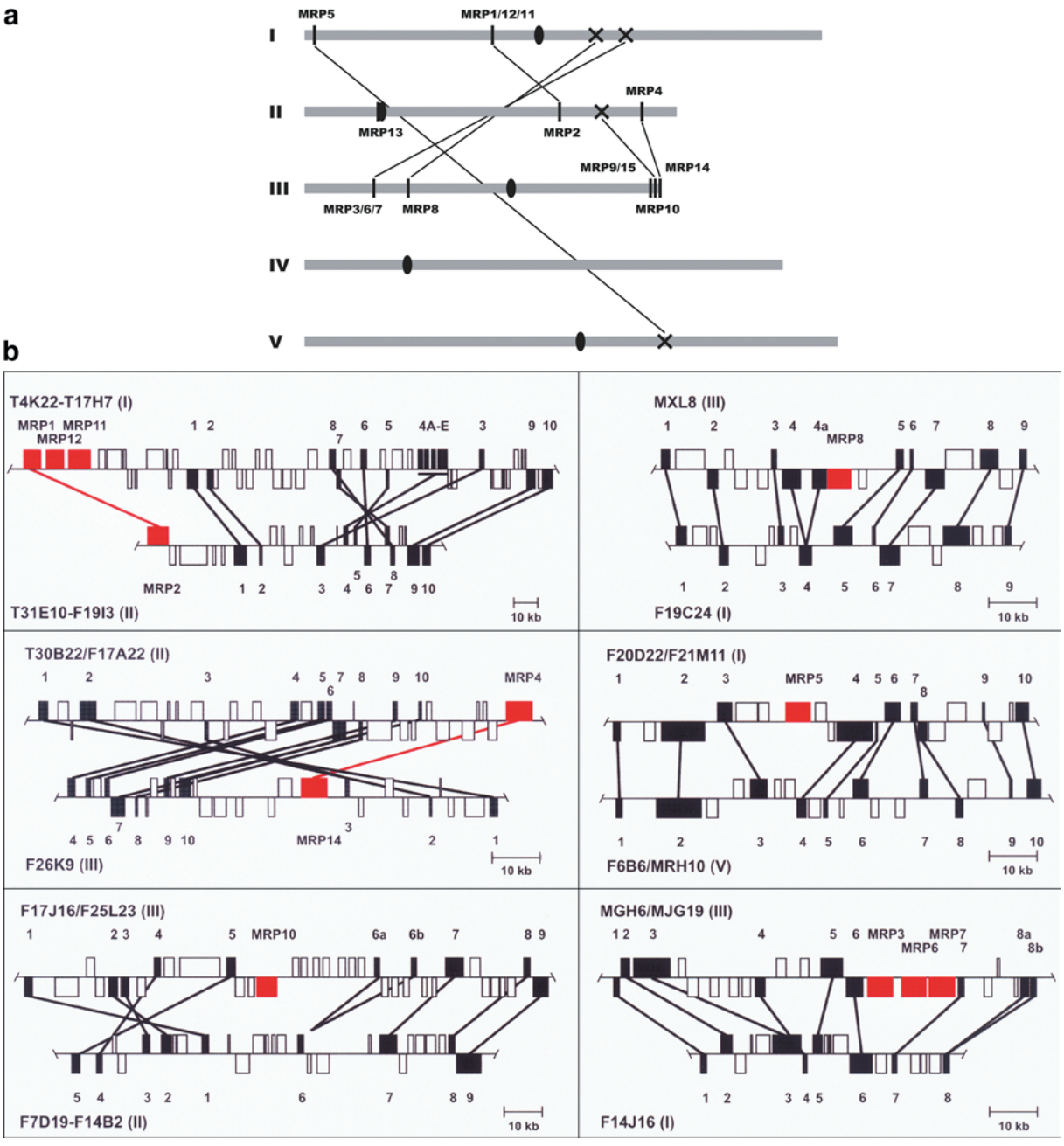

Fig. 8. **a** Schematic representation of all described segmental chromosomal duplications containing *Mrp* genes throughout the *Arabidopsis* genome. *Mrp* genes harboring loci are indicated as *vertical bars* on the five chromosomes of *Arabidopsis*. *Crosses* mark loci homeologous to *Mrp*-bearing regions, where this gene is missing. Duplicated regions are connected with lines to each other. Centromeric regions are marked with *ellipses*. **b** *AtMRP* genes (*red boxes*) and their neighbouring genes (*black* and *white boxes*) are shown in comparison to their homeologous loci on different chromosomes. Chromosomes and BAC clones of the loci are given at the upper or lower left edges of the corresponding figures. Genes encoded by the upper strand are drawn as boxes sitting on top of the chromosomes, whereas genes encoded by the lower strand are drawn below the chromosomes. Homologous genes on homologous loci (*black boxes*) are connected by dotted lines and can be identified by the same gene numbers, given either above or below the genes. The same numbers with different letters are given for genes that arose by tandem duplications

Discussion

Gene structure and annotation of *AtMRP* genes

The *AtMRP* family represents a group of genes whose coding sequences and genomic structures show a high degree of conservation within the described clades. According to these features two subgroups of *AtMRP* genes were defined. Each clade consists of genes with nearly identical exon–intron distribution. The relationship of these genes is also supported by the identity scores between the genes within each subclade and the positioning of the particular sequences in phylogenetic trees.

Conservation of position and phase of introns helped to determine the correct annotation of particular genes, in some cases differing from database entries. According to such considerations the first exon of *AtMRP11* was re-annotated. This finding was validated by cDNA sequencing. Another example of alignment-based re-annotation has been shown for *AtMRP15* (see above): according to the annotation in GenBank and MIPS this gene consists of an AtMRP-like sequence with a truncation of its N-terminal part. Comparison of nucleotide sequences of *AtMRP9* and *AtMRP15* showed that both sequences are nearly identical. However, the *AtMRP15* sequence contains several single-bp deletions and insertions in the part homologous to the first exon of *AtMRP9* as well as a bigger insertion of 229 bp, which clearly disrupts the coding sequence. The resulting assumption that this gene is a pseudogene rather than a truncated version of a functional *AtMRP* was validated by RT–PCR experiments.

More examples of incorrect annotations and misinterpretations of the genome data were reported recently (Martinoia et al. 2002). This review about full-size ABC transporters in *Arabidopsis* showed that several genes were annotated differently by MATDB and GenBank. Either different coding sequences were given for the same gene or particular genes were interpreted as pseudogenes in one database but not in the other. There were also cases of misinterpretation of the genomic data throughout all databases. For *AtPDR2* the annotation suggested two separate genes.

In all cases mentioned, comparison of annotated aa sequences with alignments of validated sequences of homologs helped to improve the annotation. These examples show that alignment-based re-annotation of genes can be a powerful tool to complement the methods applied in gene annotation of completely sequenced genomes. Programs like GENSCAN, GRAIL and NETPLANTGENE, which have been employed extensively to annotate the *Arabidopsis* genome, are based on algorithms using canonical motifs in genomic sequences. All these programs share the characteristic to neglect structural similarities within individual gene families. Experimental approaches to elucidate correct gene structures like EST or cognate cDNA sequencing lead in nearly all cases to the correction of at least one exon of a gene (Brendel and Zhu 2002). But they often struggle with different problems, which are connected with genome complexity, and also the cost of such approaches has so far hampered the sequencing of all cognate cDNAs from *Arabidopsis*.

Evolution of the AtMRPs

Another result of our analysis was the subdivision of the *AtMRP* gene family into three major groups, as has been described above. One remarkable finding was the separation of this gene family into a clade (clade I) with similarities to its human homologs and another one that seems to be more plant-specific (clade II). Furthermore the characteristic exon–intron structures of the particular subgroups also imply a polyphyletic origin of this gene family. This assumption is also supported by the findings about *MRP* genes in the rice genome: as has already been described above the three *MRP* clades found for *Arabidopsis* were also found in rice. The existence of all three types of gene in grasses indicates an early separation of these groups from each other and therefore an independent evolution of each clade, at least during angiosperm evolution.

The classification of particular *AtMRP* genes into different subgroups is also supported by the evolution of their genomic locations. The occurrence of chromosomal segmental duplication in *Arabidopsis*, which can only be analysed for entirely sequenced genomes, was a rather unexpected finding (Blanc et al. 2000; Paterson et al. 2000; Vision et al. 2000). However, this fact might help to explain the evolution and phylogeny of a multigene family for which the AtMRPs serve as a role model. Our results have shown, that the genes belonging to clade I seem to have evolved from a single locus (see above). The genesis of these genes also reflects the general processes leading to the formation of the whole gene family. In the beginning of the evolution of these *AtMRP1/2/11/12* genes, one ancestral gene underwent tandem duplication that resulted in *AtMRP1* and *12*. During the polyploidisation of the ancestral *Arabidopsis* genome the copy of *AtMRP12* must have disappeared which left *AtMRP2* alone. In the same time range another tandem duplication took place at the first locus leading to the formation of *AtMRP11*.

Although the scenario described above still leaves open questions about the succession of some of its steps, it is the most likely process according to the available data. Also the development of the remaining members of the *AtMRP* gene family showed striking parallels to it: only in one additional case has the polyploidisation led to the duplication of the *MRP* gene (see Fig. 8). In all other cases the *MRP* genes are missing from the duplicated loci. The homology among *AtMRP3*, *6* and *7* is another indication of stepwise tandem duplication as described above. Unfortunately the orthologous locus does not harbour any *MRP* genes at all, so a direct comparison between the development of the *AtMRP1*- and *AtMRP3*-bearing loci and their duplicated regions is not possible.

Functions of *AtMRP* genes

The classification of different groups within the AtMRPs is not only interesting from the evolutionary point of view, but might also have implications for their functional characterisation. Although each *AtMRP* that has been cloned and heterologously expressed in yeast shows at least a weak glutathione-conjugate transport activity and at least two AtMRPs transport glucuronide conjugates (Liu et al. 2001; Gaedeke et al. 2001), information

about their functions in planta is still scarce. Mutant analysis indicates that additional roles can be found for MRPs in plants. A T-DNA knock-out mutant of *AtMRP5* (atmrp5-1) exhibits a clear phenotype, resulting in reduced root growth and development of lateral roots, changes in auxin distribution along the root as well as changes in stomatal opening (Gaedeke et al. 2001).

A more subtle phenotypic analysis of other *AtMRP* mutants might result in more obvious phenotypic differences. However, one should keep in mind that the lack of phenotypic alteration in particular mutants might be masked by closely related isoforms sharing identical functions. The number of tandem duplications and the difference in numbers of *MRP1* homologs between *Arabidopsis* and rice indicate a redundant set of particular members of this gene family in *Arabidopsis*. This raises the question about the proper functional analysis of these genes in *Arabidopsis*. Insertion mutagenesis will provide clear indications about the role of single genes in this model plant, but it will probably not lead to the functional analysis of genes duplicated in tandem. To solve this problem reverse genetic screens of physically mutagenised lines might be the appropriate way (Li et al. 2001).

Expression analysis of the MRP gene family showed that the expression of three *MRP* genes was induced upon SA treatment. Although we can now distinguish these genes from the other 11 *MRP* genes by their response towards a simple stimulus, the basis for this could not be traced to either of the known *cis*-elements found by database analysis. Previous work demonstrated that various, isolated single *cis*-acting elements are able to trigger gene expression specifically after stimulation by biotic and abiotic stresses (Rushton et al. 2002). Recent identification of potato genes induced during colonisation by *Phytophthora infestans* detected induced transcripts of one WRKY transcription factor, as well as one MRP-type ABC transporter (Beyer et al. 2001) that shares 60% identity to AtMRP3. Furthermore, earlier expression studies for *AtMRP1* to *4* using treatments with herbicides, xenobiotics and compounds inducing oxidative stress also showed a differential induction of *AtMRP* genes (Tommasini et al. 1997; Sánchez-Fernández et al. 1998). These findings also support our experimental approach. Taken together, database analysis of *cis* elements in promoter regions is helpful as a targeted approach to propose possible treatments that might affect the expression patterns of not only genes of the *MRP* family. This approach should be understood as a powerful strategy to gather necessary information on the gene expression and functional analysis of members of a gene family.

Acknowledgements The authors acknowledge support provided to E.M. by the Swiss National Foundation, to M.G. and M.K. by the Humboldt Stiftung, to B.S. by the Deutsche Forschungsgemeinschaft, the EC (LATIN, BIOTEC 4) and the Ministerium für Schule, Wissenschaft und Forschung des Landes NRW. B.S. thanks Prof. U.I. Flügge for continuous support. L.B. thanks Marlyse Meylan-Bettex for technical assistance. H. Üner Kolukisaoglu and Lucien Bovien contributed equally to this work.

References

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

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSIBLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Beyer K, Binder A, Boller T, Collinge M (2001) Identification of potato genes induced during colonization by *Phytophthora infestans*. Mol Plant Pathol 2:125–134

Blanc G, Barakat A, Guyot R, Cooke R, Delseny M (2000) Extensive duplication and reshuffling in the *Arabidopsis* genome. Plant Cell 12:1093–1101

Brendel V, Zhu W (2002) Computational modeling of gene structure in *Arabidopsis thaliana*. Plant Mol Biol 48:49–58

Davis TGE, Coleman JOD (2000) The *Arabidopsis thaliana* ATP-binding cassette proteins: an emerging superfamily. Plant Cell Environ 23:431–443

Dayhoff MO (1979) Atlas of protein sequence and structure, vol 5, suppl 3, 1978. National Biomedical Research Foundation, Washington, DC

Eulgem T, Rushton PJ, Robatzek S, Somssich IE (2000) The WRKY superfamily of plant transcription factors. Trends Plant Sci 5:199–206

Footitt S, Slocombe SP, Larner V, Kurup S, Wu Y, Larson T, Graham I, Baker A, Holdsworth M (2002) Control of germination and lipid mobilisation by COMATOSE, the *Arabidopsis* homologue of human ALDP. EMBO J 21:2912–2922

Gaedeke N, Klein M, Kolukisaoglu U, Forestier C, Müller A, Ansorge M, Becker D, Mamnun Y, Kuchler K, Schulz B, Müller-Roeber B, Martinoia E (2001) The *Arabidopsis thaliana* ABC transporter *At*MRP5 controls root development and stomata movement. EMBO J 20:1875–1887

Goff SA, Ricke D, Lan TH et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92–100

Hayashi M, Nito K, Takei-Hoshi R, Yagi M, Kondo M, Suenaga A, Yamaya T, Nishimura M (2002) Ped3p is a peroxisomal ATP-binding cassette transporter that might supply substrates for fatty acid beta-oxidation. Plant Cell Physiol 43:1–11

Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant *cis*-acting regulatory DNA elements (PLACE) database. Nucleic Acids Res 27:297–300

Kreuz K, Tommasini R, Martinoia E (1996) Old enzymes for a new job. Herbicide detoxification in plants. Plant Physiol 111:349–353

Kushnir S, Babiychuk E, Storozhenko S, Davey MW, Papenbrock J, De Ricke R, Engler G, Stephan UW, Lange H, Kispal G, Lill R, Van Montagu M (2001) A mutation of the mitochondrial ABC transporter Sta1 leads to dwarfism and chlorosis in the *Arabidopsis* mutant *starik*. Plant Cell 13:89–100

Li X, Song Y, Century K, Straight S, Ronald P, Dong X, Lassner M, Zhang Y (2001) A fast neutron deletion mutagenesis-based reverse genetics system for plants. Plant J 27:235–242

Liu G, Sanchez-Fernandez R, Li Z-S, Rea PA (2001) Enhanced multispecificity of *Arabidopsis* vacuolar multidrug resistance-associated protein-type ATP-binding cassette transporter, AtMRP2. J Biol Chem 12:8648–8656

Liu ZB, Ulmasov T, Shi X, Hagen G, Guilfoyle TJ. (1994) Soybean GH3 promoter contains multiple auxin-inducible elements. Plant Cell 6:645–657

Liu ZB, Hagen G, Guilfoyle TJ (1997) A G-box-binding protein from soybean binds to the E1 auxin-response element in the soybean GH3 promoter and contains a proline-rich repression domain. Plant Physiol 115:397–407

Lu Y-P, Li Z-S, Rea PA (1997) *AtMRP1* gene of *Arabidopsis* encodes a glutathione *S*-conjugate pump: isolation and functional definition of a plant ATP-binding cassette transporter gene. Proc Natl Acad Sci USA 94:8243–8248

Lu Y-P, Li Z-S, Drozdowicz YM, Hörtensteiner S, Martinoia E, Rea PA (1998) AtMRP2, an *Arabidopsis* ATP binding cassette transporter able to transport glutathione *S*-conjugates and chlorophyll catabolites: functional comparisons with AtMRP1. Plant Cell 10:267–282

Maeo K, Hayashi S, Kojima-Suzuki H, Morikami A, Nakamura K (2001) Role of conserved residues of the WRKY domain in the DNA-binding of tobacco WRKY family proteins. Biosci Biotechnol Biochem 65:2428–2436

Maleck K, Levine A, Eulgem T, Morgan A, Schmid J, Lawton KA, Dangl JL, Dietrich RA (2000) The transcriptome of *Arabidopsis thaliana* during systemic acquired resistance. Nat Genet 26:403–410

Martinoia E, Grill E, Tommasini R, Kreuz K, Amrhein N (1993) An ATP-dependent glutathione S-conjugate ''export'' pump in the vacuolar membrane of plants. Nature 364:247–249

Martinoia E, Klein M, Geisler M, Bovet L, Forestier C, Kolukisaoglu U, Müller-Röber B, Schulz B (2002) Multifunctionality of plant ABC transporters – more than just detoxifiers. Planta 214:345–355

Menke FLH, Champion A, Kijne JW, Memelink J (1999) A novel jasmonate- and elicitor-responsive element in the periwinkle secondary metabolite biosynthetic gene Str interacts with a jasmonate- and elicitor-inducible AP2-domain transcription factor, ORCA2. EMBO J 18:4455–4463

Møller SG, Kunkel T, Chua NH (2001) A plastidic ABC protein involved in intercompartmental communication of light signaling. Genes Dev 15:90–103

Ohme-Tagaki M, Suzuki K, Shinshi H (2000) Regulation of ethylene-induced transcription of defense genes. Plant Cell Physiol 41:1187–1192

Paterson AH, Bowers JE, Burow MD, Draye X, Elsik CG, Jiang CX, Katsar CS, Lan TH, Lin YR, Ming R, Wright RJ (2000) Comparative genomics of plant chromosomes. Plant Cell 12:1523–1539

Rushton PJ, Reinstädler A, Lipka V, Lippok B, Somssich IE (2002) Synthetic plant promoters containing defined regulatory elements provide novel insights into pathogen- and wound-induced signaling. Plant Cell 14:749–762

Sánchez-Fernández R, Ardiles-Diaz W, Van Montagu M, Inze D, May MJ (1998) Cloning and expression analyses of AtMRP4, a novel MRP-like gene from *Arabidopsis thaliana*. Mol Gen Genet 258:655–662

Sánchez-Fernández R, Davies TGE, Coleman J, Rea PA (2001) The *Arabidopsis thaliana* ABC protein superfamily, a complete inventory. J Biol Chem 276:30231–30244

Thompson JD, Gibson TJ, Plewniak 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 24:4876–4882

Tommasini R, Vogt E, Schmid J, Fromentau M, Amrhein N, Martinoia E (1997) Differential expression of genes coding for ABC transporters after treatment of *Arabidopsis thaliana* with xenobiotics. FEBS Lett 411:206–10

Ulmasov T, Hagen G, Guilfoyle TJ (1997a) ARF1, a transcription factor that binds to auxin responsive elements. Science 276:1865–1868

Ulmasov T, Murfett J, Hagen G, Guilfoyle TJ (1997b) Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. Plant Cell 9:1963–1971

Uno Y, Furihata T, Yoshida R, Shinozaki K, Yamaguchi-Shinozaki K (2000) *Arabidopsis* basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. Proc Natl Acad Sci USA 97:11632–11637

Vision TJ, Brown DG, Tanksley SD (2000) The origins of genomic duplications in *Arabidopsis*. Science 290:2114–2117

Yu J, Hu S, Wang J et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296: 79–92

Zolman BK, Silva ID, Bartel B (2001) The *Arabidopsis* pxa1 mutant is defective in an ATP-binding cassette transporter-like protein required for peroxisomal fatty acid β-oxidation. Plant Physiol 127:1266–1278