Plant Molecular Biology **39:** 1003–1012, 1999.

The arabinose kinase, *ARA1*, gene of *Arabidopsis* is a novel member of the galactose kinase gene family

Sarah Sherson[1,4], Isabelle Gy[2], Jonathan Medd[1], Renate Schmidt[3,5], Caroline Dean[3], Martin Kreis[2], Alain Lecharny[2] and Christopher Cobbett[1,*]

[1]*Department of Genetics, University of Melbourne, Parkville, 3052, Australia (*author for correspondence);* [2]*Laboratoire de Biologie du Développement des Plantes, Institut de Biotechnologie des Plantes, Bâtiment 630, Université de Paris-Sud, CNRS-ERS 569, 91405 Orsay Cedex, France;* [3]*Department of Molecular Genetics, John Innes Centre, Norwich Research Park, Colney, Norwich, NR4 7UH, UK; present addresses:* [4]*Institute of Cell and Molecular Biology, The King's Buildings, The University of Edinburgh, Mayfield Road, Edinburgh, EH9 3JH, UK;* [5]*Max-Delbrueck-Laboratory in the MPG, Carl-von-Linne-Weg 10, 50829 Cologne, Germany*

Received 16 June 1998; accepted in revised form 27 November 1998

Key words: Arabidopsis, sugar salvage pathways, galactose kinase, arabinose kinase

Abstract

The arabinose-sensitive *ara1-1* mutant of *Arabidopsis* is deficient in arabinose kinase activity. A candidate for the *ARA1* gene, *ISA1*, has been previously identified through the *Arabidopsis* genome sequencing initiative. Here we demonstrate that (1) the *ARA1* gene coincides with *ISA1* in a positional cloning strategy; (2) there are mutations in the *ISA1* gene in both the *ara1-1* mutant and an intragenic suppressor mutant; and (3) the *ara1-1* and suppressor mutant phenotypes can be complemented by the expression of the *ISA1* cDNA in transgenic plants. Together these observations confirm that *ISA1* is the *ARA1* gene. *ARA1* is a member of the galactose kinase family of genes and represents a new substrate specificity among this and other families of sugar kinases. A second gene with similarities to members of the galactose kinase gene family has been identified in the EST database. A 1.8 kb cDNA contained an open reading-frame predicted to encode a 496 amino acid polypeptide. The *GAL1* cDNA was expressed in a *galK* mutant of *Escherichia coli* and *in vitro* assays of extracts of the strain expressing *GAL1* confirmed that the cDNA encodes a galactose kinase activity. Both *GAL1* and *ARA1* cross-hybridise at low stringency to other sequences suggesting the presence of additional members of the galactose kinase gene family.

Introduction

Plant cell walls contain numerous polysaccharides which consist of a wide range of different sugar residues. An analysis of *Arabidopsis*, for example, identified the sugars glucose, rhamnose, galactose, xylose, arabinose and galacturonic and glucuronic acids as the major sugar constituents of the cell wall [27]. In plants, NDP sugars, particularly UDP and GDP sugars, are the immediate precursors of polysaccharide biosynthesis. Most UDP sugars, for example, can be synthesised *de novo* from UDP-glucose via a series of interconversion reactions [8]. In addition, there are, for many sugars, 'salvage' pathways which can convert free sugars into NDP sugars via a phosphorylated intermediate by the sequential action of a sugar kinase and a nucleotide triphosphate pyrophosphorylase. Salvage pathways for L-fucose, L-arabinose, D-galacturonic D-glucuronic acid, and D-galactose have been identified in plants [8]. The evidence to support the existence of salvage pathways ranges from the ability of plant cells or tissues to incorporate exogenous labelled sugar into polysaccharides [7, 8, 21] and the identification of salvage pathway enzyme activities, particularly the sugar kinases, *in vitro* [3, 17, 18].

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number AF024623 (*GAL1*).

For both D-galactose and L-arabinose, for example, sugar kinase activities, which phosphorylate the sugar at the C1 position, have been identified *in vitro* [3].

In plants the primary role of the sugar salvage pathways is presumed to be in the recycling of sugars derived from the turnover of macromolecules such as polysaccharides, galactolipids and proteoglycans, which have sugar components. Numerous bacterial, fungal and mammalian sugar kinase genes have been isolated. Comparisons of amino acid sequences coupled with biochemical data defining functionally important regions of the enzymes have led to the identification of three evolutionarily distinct gene families: the hexokinase, ribokinase and galactokinase families [2]. In plants genes encoding hexokinases which play a central role in carbohydrate metabolism have been identified [11]. Recently, Kaplan *et al.* [12] described a galactose kinase gene from *Arabidopsis*. However, genes encoding sugar kinases with other sugar specificities have not been identified in plants.

We have previously identified an L-arabinose-sensitive mutant, *ara1-1*, of *Arabidopsis* [7]. This mutant has only 10% of the wild-type levels of arabinose kinase activity and exhibits reduced metabolism of exogenous arabinose presumably because of the block in the arabinose salvage pathway. In view of these data it appeared likely that *ARA1* is the structural gene for arabinose kinase. In the presence of high levels of exogenous arabinose, the sensitive phenotype probably results from the intracellular accumulation of arabinose due to the inability of the mutant to metabolise this sugar. Here we describe the isolation of the *ARA1* gene through a map-based cloning strategy.

Materials and methods

Genetic mapping of the ara1-1 *mutation*

Sixty-three F_3 families derived from a cross between the *ara1-1* mutant (Columbia ecotype) and the L*er* ecotype were analysed for the arabinose-sensitive phenotype as previously described [7] and for their genotype with respect to the RFLP markers m326, m226, m557, m600 and m272 [4]. Recombinants between the *ARA1* and *FCA*, or *ARA1* and *COP9*, loci were identified by germinating F_2 progeny from a cross between the *ara1-1* mutant and either the late-flowering mutant *fca-1* (Ler ecotype, NASC stock number NW52) or a plant heterozygous for the seedling lethal *cop9-1* mutation (Ws ecotype, NASC stock number NW6262), respectively, on medium containing 30 mM L-arabinose as described previously [5]. Only *ARA1/ARA1* homozygous individuals survive and were scored for the wild-type *FCA* phenotype (eight recombinants from 2404 *ARA1/ARA1* F_2 progeny; $r = 0.17\%$) or *COP9* phenotype (26 recombinants from a total of 4390 F_2 progeny; $r = 1.2\%$), respectively, to identify recombinant individuals.

Isolation of clones of the chromosome walk

The chromosome walk was initiated using the yUP3F7R and EW20B3L YAC end clones. Clones were identified by high-stringency hybridisation to a genomic library of the Columbia ecotype in the cosmid vector pOCA18 [19] in most cases. Where clones could not be identified from this library a Columbia λGEM11 library (obtained from C. Somerville) or, in one case, a Landsberg *erecta* λEMBL4 library (obtained from D. Smyth) was used. Overlapping clones were identified by hybridisation using restriction fragments and were confirmed as overlapping by restriction mapping with the enzymes *Eco*RI, *Bam*HI and *Cla*I. The direction of the walk was confirmed by the hybridisation of some clones to YAC EW20B3.

Nucleotide sequencing of the ara1-1 *and* ara1-1 *sup1 alleles*

The entire wild-type, *ara1-1* and *ara1-1* sup1 alleles were amplified in a ca. 8 kb fragment of DNA using genomic DNA and the Elongase Enzyme Mix (Gibco-BRL) according to the manufacturer's instructions using primers with the sequences: 5′-TAAACTCTCTTGCTGCTGCTGC-3′ and 5′-TACTACCCATCTTTCCACTTGC-3′. The amplified products were purified by gel electophoresis using the GeneClean Kit (BIO101). Nucleotide sequencing reactions were performed by using the double-stranded templates and the dye-terminator cycle-sequencing AmpliTaq kit (ABI) with a series of primers specific for the amplified product spaced at 300–400 bp intervals. Sequence products were resolved on a 373 DNA sequencer (ABI). Any uncertainties in the sequence were resolved by using primers specific for the opposite strand. The two differences observed between the wild type and mutant sequences were determined from both strands in two independent PCR products.

Transformation of Arabidopsis

The *ISA1* cDNA was ligated into pBI121 (Clontech) downstream of the CaMV 35S promoter and transformed into *Agrobacterium* strain GV3101/pMP90 by electroporation. The construct was transformed into mutant plants by the vacuum infiltration procedure [1].

Isolation and nucleotide sequence analysis of cDNA clones

The λPRL2 cDNA library (Stock CD4-7) which is constructed in the λZip-Lox vector was obtained from the Arabidopsis Biological Resource Center, Columbus, Ohio. The inserts of the positively hybridising clones were recovered in the pZL1 plasmid within the λZip-Lox vector by an *in vivo* excision protocol described by the suppliers of the vector (Gibco-BRL). Nucleotide sequencing of both strands of the cDNA was performed as described above from various subclones ligated into the Bluescript SK+ vector (Stratagene). Sequence analysis and comparisons were done through the Australian National Genomic Information Service, Sydney.

Expression of the GAL1 *cDNA in* E. coli

The 1.8 kb cDNA was excised from the vector in which the clone was isolated, pZL1, by using the restriction enzymes *Sma*I and *Xba*I and religated between the *Sma*I and *Xba*I sites of the Bluescript KS+ vector. This construct, pKS+(*GAL1*), was confirmed by restriction enzyme digests and nucleotide sequence determination of the 5′ fusion junction. The *E. coli galK* mutant CA190 (*galK35*, *relA1*, *spoT1*) was obtained from the *E. coli* Genetic Stock Center, Yale University, New Haven, CT. The assay of galactose kinase activity was based on assays described by Sherman and Adler [25] and Robrish and Thompson [22].

Southern analysis

Arabidopsis genomic DNA was isolated using a modification of the method of Dellaporta *et al.* [6]. Southern transfer was essentially as described by Sambrook *et al.* [23]. For low-stringency hybridisation, Southern filters were pre-hybridised at 37 °C (30% v/v formamide, 5× SSPE, 5× Denhardt's solution, 1% w/v dried milk powder, 0.1% w/v SDS). The filters were hybridised (30% formamide, 5× SSPE, 3× Denhardt's solution, 0.1% w/v SDS, 9% w/v dextran sulphate, 100 μg/ml salmon sperm DNA) at 37 °C with a DNA probe (see above) and washed in 2× SSPE (30 min) and 2× SSPE/0.1% SDS (30 min) at 25 °C. The conditions for high-stringency hybridisation were similar except that the hybridisation solutions contained 50% v/v formamide and the washes, including 0.1× SSPE (15 min), were done at 65 °C.

Results

Mapping the ARA1 *locus*

The *ara1-1* arabinose-sensitive mutant is deficient in arabinose kinase activity and *ARA1* is likely to be the structural gene for arabinose kinase. We aimed to isolate the *ARA1* gene through a map-based cloning strategy. Previous mapping of *ARA1* had indicated a position on chromosome 4 between the markers *bp* and *cer2* [5]. To more accurately map *ARA1*, RFLP markers were used as described in Materials and methods. The results showed that *ARA1* lies between the markers m326 and m557 and is close to m226 (Figure 1A). No recombinants between *ARA1* and m226 were observed.

The *FCA* locus maps between m326 and m226 [16]. To position *ARA1* with respect to *FCA*, recombinants between these two loci were isolated as described in Materials and methods. To determine the relative order of *FCA* and *ARA1* with respect to flanking RFLP markers the genotype of each of eight *ARA1-FCA* recombinant lines was determined for markers m326, m226 and m557. Using these data, combined with the RFLP mapping of *ARA1* and the observation that *FCA* mapped between m326 and m226 [16], the marker order could be unequivocally determined as m326-*ARA1-FCA*-m226-m557 (Figure 1A). In addition, 26 recombinants between the *COP9* and *ARA1* loci were isolated. Using flanking RFLP markers the order of the three loci was shown to be *COP9-ARA1-FCA*.

Positional cloning of the ARA1 *locus*

To further define the position of *ARA1* and to isolate the *ARA1* gene, a chromosome walk extending over 200 kb was conducted in this region of the genome. End-clones derived from previously mapped YAC clones, EW20B3 and yUP3F7 [16], were used as starting points for the walk (Figure 1B). The positions and derivations of some of the clones isolated are illustrated in Figure 1C. Each clone identified one or more RFLPs between the Columbia and Landsberg

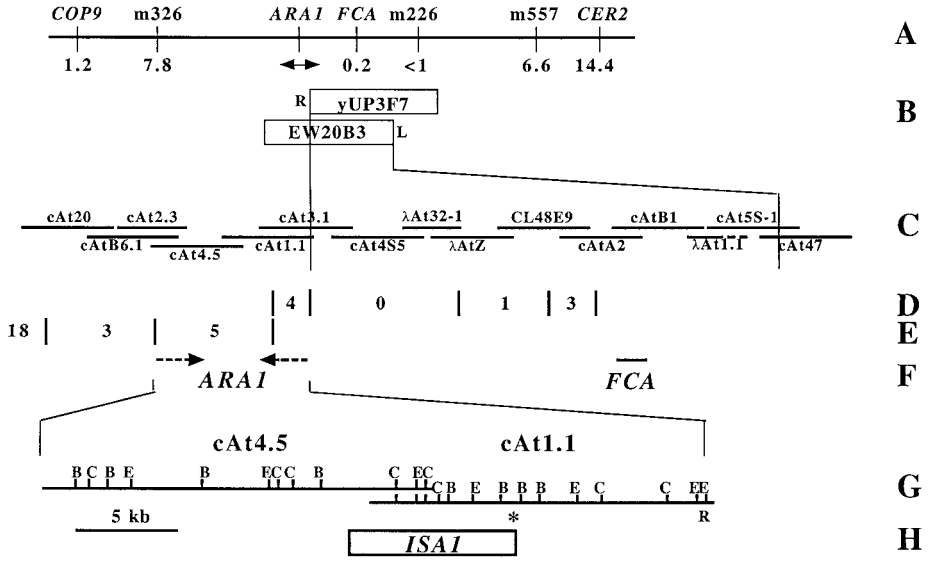

Figure 1. Physical mapping of the *ARA1* locus. A. The genetic distances in cM between *ARA1* and other markers (*COP9*, *FCA*, *CER2* and RFLP markers) determined in this or previous [5] studies are shown for each marker (error values have not been indicated). The markers are shown in their known order on the chromosome. The map is not to scale. The relative map distances of m326 and *COP9* from *ARA1* are discussed further in the text. B. Two YAC clones in the *ARA1-FCA* region [24]. End-clones derived from these YACs are indicated by R and L. C. The clones of the chromosome walk include cosmid (c-prefix and CL48E9) and lambda (λ-prefix) clones. All clones are from the Columbia ecotype with the exception of CL48E9 and λAt1.1 which are from Landsberg. λAt1.1 is a chimeric clone as indicated by the dashed line. The positions of clones in the walk which hybridise to the YAC end-clones are indicated by connecting lines. D, E. The positions of recombination events between the *ARA1* and *FCA* (D) and *COP9* and *ARA1* (E) relative to RFLPs detected by various clones. The numbers of recombination events within a particular interval defined by RFLPs detected by particular clones are indicated. The precise boundaries of each interval are uncertain because the exact positions of polymorphic restriction sites have not been determined. F. The region containing the *ARA1* locus and the known location of *FCA* [16] are indicated. G. The restriction maps of clones cAt4.5 and cAt1.1 which span the *ARA1* locus: B, *Bam*HI; C, *Cla*I; E, *Eco*RI. A 5 kb scale is indicated. The four recombination events in the corresponding interval shown in (D) lie between RFLPs detected by the 0.8 kb *Bam*HI fragment (*) and the yUP3F7R clone indicated by R. H. The position of the candidate for *ARA1*, *ISA1*, is indicated.

(for *FCA*) or Ws (for *COP9*) ecotypes. These were used to determine the genotype of each of the *ARA1-FCA* and *COP9-ARA1* recombinants for each RFLP, thus allowing the position of each recombination event to be determined (Figure 1D, E). Seven of the eight *ARA1-FCA* recombination events were clustered in two distinct regions as illustrated in Figure 1E. From these data we anticipated that *FCA* would lie to the right of, and close to, clone CL48E9. This expectation has been confirmed for *FCA* which is located within clone cAtA2 [16]. Similarly, *ARA1* was expected to lie to the left of the 0.8 kb *Bam*HI fragment of cAt1.1 and to the right of three different RFLPs detected by clone cAt4.5 (Figure 1D, E, G). This placed the gene within the adjacent, overlapping clones cAt4.5 and cAt1.1, in an interval of about 20 kb (Figure 1F, G, H).

Identification of a candidate for the ARA1 *gene*

We have determined the nucleotide sequence of a 14 kb segment within cAt4.5 and cAt1.1 as part of the ESSA genome sequencing program. Within this region we have previously identified a gene, *ISA1*, with similarities to the galactose kinase family of sugar kinases [9]. The position of *ISA1* within cAt4.5 and cAt1.1 is indicated in Figure 1. Neither clone contains the entire gene. The 3′ end of *ISA1* lies within the 0.8 kb *Bam*HI fragment of cAt1.1 and is immediately adjacent to the leftmost cluster of recombination events between *ARA1* and *FCA*. Given that *ARA1* is believed to encode an arabinose kinase and in view of the map position of *ARA1* determined above, this gene was a likely candidate for *ARA1*. The nucleotide sequence of a 3.36 kb cDNA corresponding to this gene has been determined [9] (accession number Y14404) and contains within it an open reading frame which would encode a protein of 988 amino acids. The C-terminal half of this protein contains regions which are similar to the seven conserved signature sequence blocks (A to G) used by the BlockSearch algorithm [10] to define the galactose kinase gene family (Figure 2). The N-terminal 450 amino acid sequence of the

ISA1 protein is similar to a hypothetical protein from a *Synechocystis* species which does not also contain a region similar to galactose kinases [9]. The function of this domain of the protein is unknown.

Anticipating that *ISA1* is indeed the *ARA1* gene, we determined the nucleotide sequence of *ISA1* from the *ara1-1* mutant. The sequence derived from the *ara1-1* mutant was identical to the sequence derived from the cAt4.5 and cAt1.1 clones as part of the ESSA project except for 2 bp. One of these lay within intron 22 of the gene while the second was in exon 21. Although cAt4.5 and cAt1.1 and the *ara1-1* mutant are derived from the Columbia ecotype it is possible that there exist polymorphisms between different laboratory stocks of this ecotype. To investigate this the nucleotide sequences of the regions containing these two differences were also determined for our own Columbia wild type from which the *ara1-1* mutant is derived. In this case the wild-type sequence of intron 22 was identical to that determined for the *ara1-1* mutant, while the sequence of exon 21 was identical to the sequence determined from the clones and different from that of *ara1-1*. This latter mutation is a G-to-A transition which would result in a substitution of the wild-type glutamate at position 655 in the predicted protein product by an oppositely charged lysine residue (Figure 2). This residue is immediately adjacent to Block C of the galactose kinase signature sequences and is conserved in 8 of 11 galactose kinase sequences obtained from the databases. In the remaining three examples and in the *Arabidopsis* galactose kinase there is a branched-chain amino acid at this position. It is likely that this is the *ara1-1* mutation, reinforcing the expectation that *ISA1* is *ARA1*.

Suppressors of the *ara1-1* arabinose-sensitive phenotype have been isolated previously [5]. The suppressor mutants, while resistant to arabinose, can be distinguished from the wild type by a low capacity to incorporate exogenous labelled ^{3}H-L-arabinose into ethanol-insoluble polysaccharide material using a semi-quantitative assay. These second-site mutations were closely linked to the *ara1-1* mutation and in one case the suppressor mutant, sup1, had no detectable arabinose kinase activity [5], suggesting the suppressor might lie within the *ara1-1* mutant allele. To confirm this we have determined the sequence of the *ARA1* gene from the sup1 mutant. In addition to the original *ara1-1* mutation there was a second mutation (Figure 2) which would result in premature termination of translation and consequently a complete loss of kinase activity. This observation reinforces the hypothesis that the *ARA1* gene encodes the arabinose kinase activity. It is not yet clear why this mutation suppresses the sensitive phenotype although we have speculated that the ARA1 gene product may also play a role in arabinose transport [5].

Complementation of the ara1-1 *and* ara1-1 *sup1 mutants with the* ISA1 *cDNA*

To test for complementation of the *ara1-1* mutant phenotype by *ISA1* the cDNA was cloned downstream of the CaMV 35S promoter in the binary vector pBI121 to form pBI-*ISA1*. The T-DNA regions of both pBI121 and pBI-*ISA1* were transformed into the *ara1-1* mutant by using the vacuum infiltration technique as described in Materials and methods. Kanamycin-resistant primary transformants were selected and the progeny of six independent primary transformants (A to F) for each of the constructs in the *ara1-1* mutant were scored for resistance to kanamycin (Table 1) and for sensitivity to both 3 mM (data not shown) and 10 mM L-arabinose (Table 1). Segregation of kanamycin resistance and sensitivity was observed for all the transformants and indicated a single segregating transgenic locus in most (10/12) cases. The remaining two lines exhibited more complex segregation patterns. The progeny of the pBI121 transformants of *ara1-1* were uniformly sensitive to 10 mM L-arabinose and were indistinguishable from the *ara1-1* parent strain. In contrast, the progeny of each of the six pBI-*ISA1* transformants of *ara1-1* were segregating for individuals with increased resistance to arabinose compared with the *ara1-1* parent and the control transformants (Table 1). To further test the linkage between the transgene and complementation of the *ara1-1* arabinose-sensitive phenotype, 20 progeny of the pBI-*ISA1* transformants A and C were grown for seed and re-tested for the kanamycin and arabinose resistance phenotypes. For transformant A, three families were uniformly kanamycin- and arabinose-sensitive, twelve segregated for both resistance and sensitivity, while five were uniformly resistant for both phenotypes. For transformant C these classes were represented by 4, 13 and 3 families, respectively. Thus, co-segregation of the two was observed in every case.

In addition, pBI121 and pBI-*ISA1* were transformed into the *ara1-1* sup1 mutant. Segregation of kanamycin resistance among the progeny of kanamycin-resistant transformants indicated the presence of one or, in some cases, two transgenic loci (Table 2). As expected all progeny were arabinose-

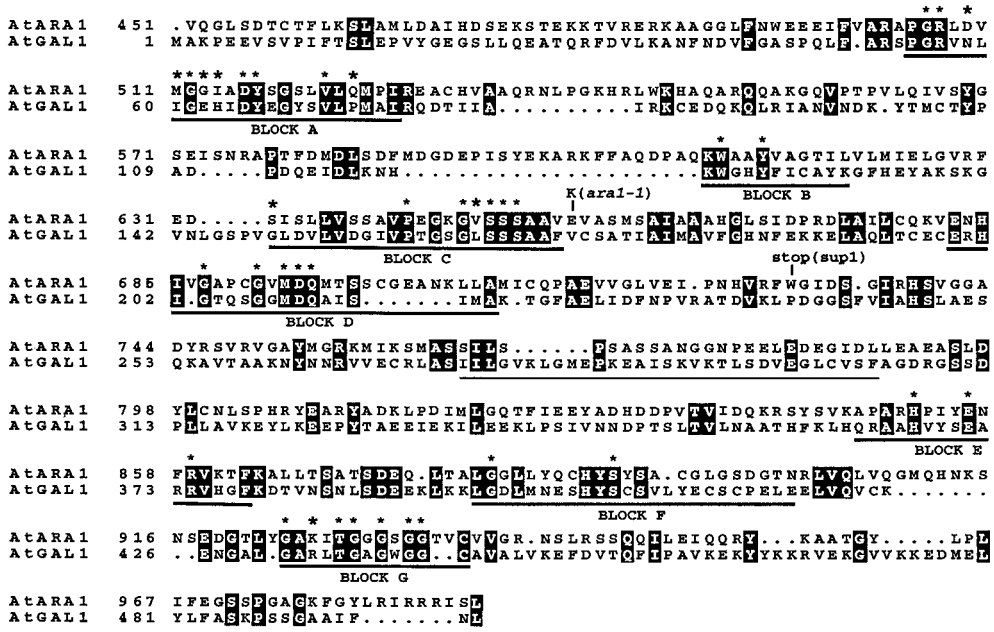

Figure 2. An alignment of the predicted amino acid sequences of the C-terminal region of the ARA1 gene product and of the GAL1, galactose kinase, protein. Gaps indicated by (.) have been introduced to maximise the alignment. Identical amino acids are highlighted. The blocks identified by BlockSearch as signature sequences for galactose kinases are heavily underlined and marked A to G. The positions of 37 amino acids within these blocks which are absolutely conserved between *GAL1* and 11 other galactose kinase sequences are indicated by asterisks. Nine of these residues, five of which are within Block A, differ in the ARA1 sequence (large asterisks). The positions of the E to K amino acid substitution identified in the *ara1-1* mutant and the additional nonsense mutation in the intragenic suppressor, sup1, are indicated above the ARA1 sequence. A region of the predicted GAL1 amino acid sequence which differs from the previously published sequence [12] is indicated (light underline).

Table 1. Complementation of the *ara1-1* mutant with the *ISA1* cDNA.

Strain		Kanamycin				L-arabinose				
		R	S	χ^2 (P)	number of loci	R	I	S	χ^2 (P)	number of loci
wild type		0	53			88	0	0		
ara1-1		0	99			0	0	123		
ara1-1 pBI121	-A	97	29	0.26 (>0.5)	1	0	0	161		
ara1-1 pBI121	-B	110	34	0.15 (>0.7)	1	0	0	136		
ara1-1 pBI121	-C	99	31	0.09 (>0.7)	1	0	0	143		
ara1-1 pBI121	-D	159	4		>2	0	0	130		
ara1-1 pBI121	-E	121	37	0.21 (>0.5)	1	0	0	135		
ara1-1 pBI121	-F	98	29	0.32 (>0.5)	1	0	0	163		
ara1-1 pBI-*ISA1*	-A	64	26	0.73 (>0.3)	1	17	0	67	1.02 (>0.3)	1
ara1-1 pBI-*ISA1*	-B	151	4		>2	32	0	92		
ara1-1 pBI-*ISA1*	-C	82	24	0.31 (>0.5)	1	33	0	83	0.74 (>0.3)	1
ara1-1 pBI-*ISA1*	-D	104	27	1.35 (>0.2)	1	0	0	99		
ara1-1 pBI-*ISA1*	-E	109	39	0.41 (>0.7)	1	30	64	41	2.16 (>0.3)	1
ara1-1 pBI-*ISA1*	-F	42	14	0.0 (>0.9)	1	36	0	19	2.67 (>0.01)	1

Progeny of kanamycin-resistant transformants were germinated on medium containing kanamycin (50 μg/ml) or L-arabinose (10 mM). In the presence of kanamycin the ratio of resistant (R) to sensitive (S) individuals was used to estimate the number of segregating transgenic loci. In the presence of L-arabinose the phenotype was compared with wild-type (R) and *ara1-1* (S) controls. Among progeny of the *ara1-1* pBI-*ISA1*-E transformant, seedlings with an intermediate (I) resistance phenotype were observed. Where the segregation of the sensitive and resistant (and intermediate) phenotypes approximated a 1:3, 3:1 or 1:2:1 ratio a χ^2 value based on the expected ratios was calculated as shown.

resistant, as was the *ara1-1* sup1 mutant parent. To test for complementation, progeny from each transformant were assayed for ^{3}H-L-arabinose metabolism. Wild-type and *ara1-1* sup1 mutant controls incorporate greater than 10% and less than 1%, respectively, of exogenous ^{3}H-L-arabinose. All progeny of pBI121 control transformants showed mutant levels of ^{3}H-L-arabinose incorporation. In contrast, the progeny of five of the six pBI-*ISA1* transformants segregated for both wild-type and mutant levels of ^{3}H-L-arabinose incorporation (Table 2). Thus the *ARA1* cDNA is able to complement both the arabinose-sensitive *ara1-1* phenotype and the phenotype of low ^{3}H-L-arabinose incorporation of the *ara1-1* sup1 intragenic suppressor mutant.

Identification of a putative Arabidopsis *galactose kinase cDNA clone*

For comparison with the *ARA1* gene we have also characterised a galactose kinase gene from *Arabidopsis*. A cDNA clone, clone 46A8T7 (NCBI ID 35035; GenBank ID T14052), exhibiting sequence similarity to various yeast and human galactose kinases was identified in the EST database. This clone was used to identify a putative full-length cDNA from the λPRL2 cDNA library (Stock No. CD4-7), referred to as *GAL1*, the nucleotide sequence of which was determined (accession number AF024623). The sequence of a similar cDNA isolated from the Landsberg *er* ecotype has previously been published [12]. To confirm the predicted function of *GAL1* the cDNA was expressed as a fusion protein from the *lacZ* promoter and translation initiation signals in the Bluescript KS+ vector. When transformed into the *E. coli* galactose kinase mutant strain CA190 (*galK35*, *relA1*, *spoT1*, obtained from the *E. coli* Genetic Stock Center, Yale University, New Haven, CT), this construct, pKS+(*GAL1*), unlike the pKS+ control, was able to complement the inability of CA190 to metabolise galactose. Furthermore, the activity of galactose kinase in crude whole-cell extracts from the CA190 pKS+(*GAL1*) strain and the control was assayed *in vitro* [22, 25]. For the control, essentially no galactose kinase activity was detected ($<0.1\pm$ 0.1 pmol D-galactose-1-phosphate per minute per mg total protein), while in extracts from the CA190 pKS+(*GAL1*) strain galactose kinase activity of 6.2 ± 1.7 pmol min^{-1} mg^{-1} D-galactose-1-phosphate was observed. The map position of *GAL1* was determined by using the recombinant inbred lines described by Lister and Dean [15]. These data indicated *GAL1* mapped on chromosome 3, near the m583 marker between positions 8.4 and 9.6.

Genomic sequences related to ARA1 *and* GAL1

The *GAL1* and *ARA1* cDNAs were hybridised successively to a Southern blot of genomic DNA digested with the restriction enzymes *Eco*RI or *Eco*RV at high and low stringency (Figure 3). From the restriction map of the *ARA1* region, three and six bands were anticipated for the *ARA1* cDNA hybridised at high stringency to *Eco*RI- and *Eco*RV-digested genomic DNA, respectively. At least one additional band was observed in each digest at high stringency and further additional bands appeared at low stringency. For *GAL1* at high stringency a single band was detected for each enzyme indicating that *GAL1* is a single-copy gene. In contrast, hybridisation at low stringency detected a number of additional bands in both digests indicating the presence of related sequences in the genome. The bands detected by the *GAL1* clone did not correspond to those detected by *ARA1*. In addition, three non-overlapping clones which contain sequences which hybridise to *GAL1* at low, but not high, stringency have been isolated from a genomic library (not shown). Together these data indicate the presence of additional genomic sequences related to either *ARA1* or *GAL1*.

Discussion

The positional cloning of the *ARA1*, arabinose kinase gene has led to the identification of a new member of the galactose kinase family of sugar kinases. Fine-scale recombinational mapping of *ARA1* coupled with the analysis of sequences derived from the *Arabidopsis* genome sequencing initiative led to the identification of a candidate for *ARA1*, *ISA1*. The experiments described here, showing that both the *ara1-1* and *ara1-1* sup1 mutants have mutations within *ISA1* and that the *ISA1* cDNA is able to complement the mutant *ara1-1* and *ara-1-1* sup1 phenotypes, demonstrate clearly that *ISA1* is the *ARA1* gene. Transformation of the *ara1-1* mutant with the *ISA1* cDNA complemented both the arabinose-sensitive phenotype and the ability of transformants to metabolise exogenous L-arabinose. In addition, we have isolated a cDNA encoding a galactose kinase and have extended the observations of Kaplan *et al.* [12] by demonstrating the *in vitro* galactose kinase activity of the *GAL1* cDNA expressed in a heterologous organism.

Table 2. Complementation of the *ara1-1* sup1 mutant with the *ISA1* cDNA.

Strain		Kanamycin				^{3}H-L-arabinose incorporation		
		R	S	χ^2 (P)	number of loci	H	L	total
Wild type		0	82			10	0	10
ara1-1 sup1		0	116			0	10	10
ara1-1 sup1 pBI121	-A	106	2		>2	0	20	20
ara1-1 sup1 pBI121	-B	99	32	0.02 (>0.8)	1	0	20	20
ara1-1 sup1 pBI121	-C	105	7	0.0 (>0.9)	2	0	20	20
ara1-1 sup1 pBI-*ISA1*	-A	153	10	0.01 (>0.9)	2	16	4	20
ara1-1 sup1 pBI-*ISA1*	-B	106	30	0.63 (>0.3)	1	9	3	12
ara1-1 sup1 pBI-*ISA1*	-C	102	38	0.34 (>0.5)	1	0	20	20
ara1-1 sup1 pBI-*ISA1*	-D	108	32	0.34 (>0.5)	1	6	14	20
ara1-1 sup1 pBI-*ISA1*	-E	129	7	0.28 (>0.5)	2	10	10	20
ara1-1 sup1pBI-*ISA1*	-F	143	8	0.23 (>0.5)	2	12	8	20

Progeny of kanamycin-resistant transformants were germinated on medium containing kanamycin (50 μg/ml). The ratio of resistant (R) to sensitive (S) individuals was used to estimate the number of segregating transgenic loci. Where the segregation of the sensitive and resistant phenotypes approximated a 3:1 or 15:1 ratio a χ^2 value based on the expected ratios was calculated as shown. In the presence of L-arabinose all seedlings tested were arabinose-resistant (not shown). Unselected individual plants were assayed for ^{3}H-L-arabinose incorporation [5] and compared with wild type (high, H; > 10% incorporation of label) and the *ara1-1* sup1 mutant (low, L; < 1% incorporation of label).

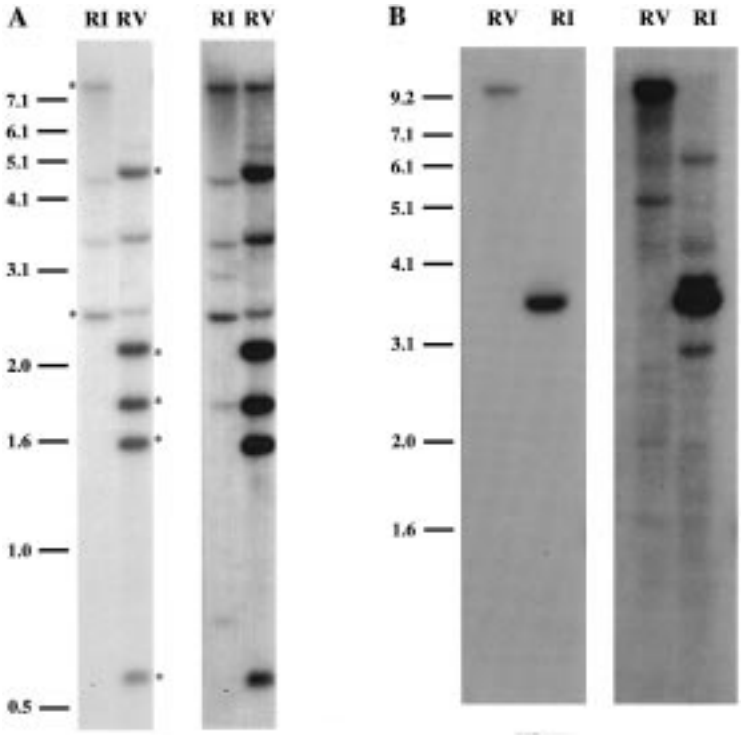

Figure 3. Southern analysis of *A. thaliana* genomic DNA. A Southern blot of genomic DNA digested with restriction enzymes *Eco*RI (RI) and *Eco*RV (RV) was hybridised with the labelled *ARA1* (A) and *GAL1* (B) cDNA clones, respectively, under low-stringency (right) and high-stringency (left) conditions. The positions of size markers (kb) are indicated. In A, bands corresponding to fragment sizes predicted from the genomic DNA sequence are indicated (*). In the *Eco*RV digest an additional fragment of 0.3 kb is expected (not shown) and in the *Eco*RI digest a fragment of 5.6 kb is expected but not observed. This difference may arise from RFLP differences between strains.

While the RFLP mapping of the *ARA1* locus was clearly able to define a small region containing the gene, two noteworthy anomalies were observed. The first of these was the apparent clustering of the recombination events between *ARA1* and *FCA*. The left-hand cluster of 4 recombination events occurred within a region of about 10 kb. This can be inferred from the distance between the 0.8 kb *Bam*HI fragment of cAt1.1 and the *Eco*RI fragments of cAt1.1 which represent yUP3F7(R), although the precise positions of the polymorphic restriction sites detected by these probes have not been determined. The right-hand cluster of 3 recombination events occurred within an interval of ca. 25 kb and between these two clusters was about 80 kb of DNA containing a single recombination event. While these data are derived from a limited set of recombinants it does appear to illustrate the non-uniformity of distribution of recombination events across a relatively small region of the genome. This may reflect localised differences between the genomes of the Columbia and Landsberg ecotypes or may simply reflect sequence-dependent recombination hot-spots.

A considerable anomaly was observed in mapping *ARA1* with respect to *COP9*. The recombination distance observed between these two loci was 1.2 cM. This is surprising because the physical location of *COP9* is between markers mi128 and mi279 both of which are to the left of m326 [24]. Marker mi279 is 1.1 cM from m326 and 6.2 cM from m226 [15]. The comparison of these genetic distances and the observation that all the *COP9-ARA1* recombination events lay to the right of m326 would suggest *COP9* might also lie to the right of m326. Both the *ara1* and *cop9* mutant phenotypes segregated in a Mendelian ratio (data not shown) discounting the possibility of a bias in the recombinant classes selected. The *cop9* allele used in this experiment was a T-DNA insertion [26]. A likely explanation for the mapping data obtained is that the presence of the T-DNA insertion suppressed local recombination events.

Previous comparisons of sugar kinase genes from various organisms have indicated the existence of three evolutionarily independent gene families: the hexose kinase family (with specificities for glucose, fructose, ribulose, xylulose and fucose); the ribose kinase family (with specificities for ribose and fructose); and the galactose kinase family (with specificity for galactose, N-acetylgalactosamine) [2, 20]. Amino acid alignments indicate ARA1 is a member of the galactose kinase family of sugar kinases. That *ARA1* encodes an arabinose kinase activity can be inferred from the fact that the *ara1-1* mutant is deficient both in the ability to metabolise exogenous L-arabinose and in L-arabinose kinase activity [7]. Whether ARA1 is specific for L-arabinose or may also utilise other substrates has not been determined. To confirm the activity of the *ARA1* gene product, we have attempted, without success, to express the *ARA1* cDNA in both *E. coli* and yeast to detect arabinose kinase activity. Early studies of the arabinose kinase activity of mung bean indicated it was a membrane-associated activity, in contrast to galactose kinase which appeared to be a soluble protein [3]. We have speculated previously that the *ARA1* gene product may interact with an arabinose transporter [5]. If this were true it may explain why expression of arabinose kinase activity in a heterologous organism was not successful, in contrast to our experiments and the experiments of others [12] with the galactose kinase. This remains a focus for further experimentation. The most apparent difference between the GAL1 and ARA1 proteins in *Arabidopsis* is that the latter contains a large N-terminal region of about 450 amino acids which is not present in the former. Only the C-terminal region of ARA1 can be aligned with the galactose kinase. The N-terminal region is not similar to any protein in the databases of a known or predicted function.

Arabinose is almost exclusively found in plants and genes encoding kinases with specificity for arabinose have not been previously identified. The isolation and characterisation of the arabinose kinase gene in this study adds a new specificity to the galactose kinase family of sugar kinases. A closer comparison of the ARA1 amino acid sequence with those of the various galactose kinases gives some indication of amino acids which may confer the substrate specificity of the kinase activity. A comparison of the seven conserved blocks defined by BlockSearch in 11 galactose kinase sequences available in the GenPept database shows 38 amino acids are absolutely conserved. Of these, 37 are conserved in GAL1 while only 29 are conserved in ARA1 (Figure 3). It is likely that the amino acids which differ between the two play a role in substrate specificity. Five of the nine differences occur in Block A suggesting that this region may have a particular role in substrate recognition. Further studies on the specificities of both the arabinose and galactose kinase enzymes from *Arabidopsis* may allow a precise definition of the determinants of substrate binding in the two enzymes.

Acknowledgements

We acknowledge the help of F. Creusot, C. Macadré and N. Schwelbel in mapping *GAL1*. We are grateful to the Arabidopsis Biological Resource Center (Ohio State University) and the Nottingham Arabidopsis Stock Centre (Nottingham, UK). S.S. was supported by an Australian Postgraduate Research Award and C.C. was funded by a grant from the Australian Research Council.

References

1. Bechtold N, Ellis J, Pelletier G: *In planta Agrobacterium* mediated gene transfer by infiltration of adult *Arabidopsis* plants. C R Acad Sci Paris, Sciences de la vie/Life Sciences 316: 1194–1199 (1993).
2. Bork P, Sander C, Valencia A: Convergent evolution of similar enzymatic function on different protein folds: the hexokinase, ribokinase and galactokinase families of sugar kinases. Protein Sci 2: 31–40 (1993).
3. Chan PH, Hassid WZ: One step purification of D-galactose and L-arabinose kinases from *Phaseolus aureus* seedlings by ATP-sepharose affinity chromatography. Anal Biochem 64: 372–379 (1975).
4. Chang C, Bowman JL, DeJohn AW, Lander ES, Meyerowitz EM: Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. Proc Natl Acad Sci USA 85: 6856–6859 (1988).
5. Cobbett CS, Medd JM, Dolezal O: Suppressors of an arabinose-sensitive mutant of *Arabidopsis thaliana*. Aust J Plant Physiol 19: 367–375 (1992).
6. Dellaporta SL, Wood J, Hicks JB: A plant DNA minipreparation: version 2. Plant Mol Biol Rep 1: 19–22 (1983).
7. Dolezal O, Cobbett CS: Arabinose kinase-deficient mutant of *Arabidopsis thaliana*. Plant Physiol 96: 1255–1260 (1991).
8. Feingold DS: Aldo (and keto) hexoses and uronic acids. In: Encyclopedia of Plant Physiology (New Series), pp. 3–76. Springer-Verlag, Berlin (1982).
9. Gy I, Aubourg S, Sherson S, Cobbett CS, Cheron A, Kreis M, Lecharny A: Analysis of a 14 kb fragment containing a putative cell wall gene and a candidate for the *ARA1*, arabinose kinase, gene from chromosome IV of *Arabidopsis thaliana*. Gene 209: 201–210 (1998).
10. Henikoff JG, Pietrokovski S, Henikoff S: Recent enhancements to the Blocks Database servers. Nucl Acids Res 25: 222–225 (1997).
11. Jang J-C, Leon P, Zhou L, Sheen J: Hexokinase as a sugar sensor in higher plants. Plant Cell 9: 5–19 (1997).
12. Kaplan CP, Tugal HB, Baker A: Isolation of a cDNA encoding an *Arabidopsis* galactokinase by functional expression in yeast. Plant Mol Biol 34: 497–506 (1997).
13. Lee N, Gielow W, Martin R, Hamilton E, Fowler A: The organization of the *araBAD* operon of *Escherichia coli*. Gene 47: 231–244 (1986).
14. Lee RT, Peterson CL, Calman AF, Herskowitz I, O'Donnell JJ: Cloning of a human galactokinase gene (*GK2*) on chromosome 15 by complementation in yeast. Proc Natl Acad Sci USA 89: 10887–10891 (1992).
15. Lister C, Dean C: Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. Plant J 4: 745–750 (1993).
16. Macknight R, Bancroft I, Page T, Lister C, Schmidt R, Love K, Westphal L, Murphy G, Sherson S, Cobbett C, Dean C: *FCA*, a gene controlling flowering time in *Arabidopsis*, encodes a protein containing RNA-binding motifs and a WW protein interaction domain. Cell 89: 737–745 (1997).
17. Neufeld EF, Feingold DS, Hassid WZ: Phosphorylation of D-galactose and L-arabinose by extracts from *Phaseolus aureus* seedlings. J Biol Chem 235: 906–909 (1960).
18. Neufeld EF, Feingold DS, Ilves SM, Kessler G, Hassid WZ: Phosphorylation of D-galacturonic acid by extracts from germinating seeds of *Phaseolus aureus*. J Biol Chem 236: 3102–3105 (1961).
19. Olszewski N, Martin F, Ausubel F: Specialized binary vector for plant transformation: expression of the *Arabidopsis thaliana AHAS* gene in *Nicotiana tabacum*. Nucl Acids Res 16: 10765–10782 (1988).
20. Pastuszak I, O'Donnell J, Elbein A: Identification of the GalNAc kinase amino acid sequence. J Biol Chem 271: 23653–23656 (1996).
21. Reiter W-D, Chapple CS, Somerville CR: Altered growth and cell walls in a fucose-deficient mutant of *Arabidopsis*. Science 261: 1032–1035 (1993).
22. Robrish SA, Thompson J: Regulation of fructose metabolism and polymer synthesis by *Fusobacterium nucleatum* ATCC 10953. J Bact 172: 5714–5723 (1990).
23. Sambrook J, Fritsch E, Maniatis T: Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).
24. Schmidt R, West J, Love K, Lenehan Z, Lister C, Thompson H, Bouchez D, Dean C: Physical map and organization of *Arabidopsis thaliana* chromosome 4. Science 270: 480–483 (1995).
25. Sherman JR, Alder J: Galactokinase from *Escherichia coli*. J Biol Chem 238: 873–878 (1963).
26. Wei N, Chamovitz DA, Deng W-W: *Arabidopsis COP9* is a component of a novel signaling complex mediating light control of development. Cell 78: 117–124 (1994).
27. Zablackis E, Huang J, Muller B, Darvill AG, Albersheim P: Characterization of the cell-wall polysaccharides of *Arabidopsis thaliana* leaves. Plant Physiol 107: 1129–1138 (1995).