A new family of lipolytic plant enzymes with members in rice, arabidopsis and maize

David J. Brick, Michael J. Brumlik, J. Thomas Buckley, Jing-Xin Cao, Paula C. Davies, Santosh Misra, Timothy J. Tranbarger, Chris Upton*

Department of Biochemistry and Microbiology, 150 Petch Building, University of Victoria, Victoria, B.C. V8W 3P6, Canada

Received 23 October 1995; revised version received 20 November 1995

Abstract We have noted a striking similarity between the sequences of proteins in a novel family of lipases we recently reported [Upton, C. and Buckley, J.T. (1995) Trends Biol. Sci. 20, 178-9] and more than 120 sequences from the database of Expressed Sequence Tags (dbEST) which correspond to at least 30 unique genes from arabidopsis, rice and maize. A cDNA (Arab-1) corresponding to one of these sequences was isolated, sequenced and translated. There was significant similarity to sequences in the new lipase family over the entire open reading frame of Arab-1 and when expressed in *E. coli*, the gene product was lipolytic. Arab-1 and genes for some of the other plant proteins appear to be differentially expressed. They may play a role in the regulation of lipid metabolism during plant development.

Key words: Lipase; dbEST; Lipolytic; Motif; Esterase

1. Introduction

The synthesis and hydrolysis of plant lipids are tightly regulated processes and are of immense importance in world agriculture [1]. Vegetative plant cells produce fatty acids primarily to supply the needs of membrane biogenesis, while developing seeds deposit large amounts of fatty acids as triacylglycerols in oil bodies [2]. In seeds, specific lipases control triacylglycerol hydrolysis, and may regulate germination. These lipases are absent from ungerminated seeds, but appear post germination when their activity increases [3]. The fatty acids released by lipolysis are further metabolized in glyoxysomes to provide energy for embryonic growth and development. Investigation of the importance of lipolytic enzymes in plant development has been hampered by a lack of information at a molecular level. Despite intense studies of true lipases in germinating oil seeds, and of a variety of lipolytic enzymes in other tissues [4,5], few have been purified or cloned [6].

We have previously reported a distinctive set of motifs in a small group of proteins from plants and bacteria, and we speculated that the proteins might be members of a new family of lipolytic enzymes [7]. Within the highly conserved blocks of amino acids in each protein (Fig. 1) are the same invariant residues (serine, aspartic acid and histidine) that make up the catalytic triad of classical lipases [8,9]. Several of the bacterial proteins in this group are known to possess lipolytic activity. These include a lipase/acyltransferase from *Aeromonas hydrophila* [10], a hemolysin from *Vibrio parahaemolyticus* [11] and a lipase from *Xenorhabdus luminescens* [12]. In addition, we have already used our motif scheme to identify the catalytic triad of the *Aeromonas hydrophila* enzyme [13]. Nevertheless, there is little sequence similarity between this new family of lipolytic proteins and the classical lipases. The key GxSxG motif that contains the active site serine residue in the classical lipases is usually replaced by GxSxSxxG (Block I; Fig. 2), and although the order of the motifs that contain the amino acids of the predicted catalytic triad is the same in the new family, their spacial arrangement is quite different. The two plant proteins we placed in this group, one from *Arabidopsis thaliana* and one from *Brassica napus*, both contain a proline rich amino-terminal domain and their expression is anther-specific [14]. However, no enzymatic activity has been previously associated with either of these proteins.

Here, we show the presence of these motifs in a large group of plant dbEST sequences and demonstrate that one of these plant genes does indeed encode a lipolytic activity.

2. Materials and methods

2.1. Isolation of clone Arab-1

Nested oligonucleotide primers A-1 and A-1-1 were designed using the DNA sequence of dbEST [15] clone 47138 (Fig. 1). Two rounds of anchored PCR (standard conditions) were performed using nested primers and a vector primer with *A. thaliana*-Versailles (VA, VB and VC) cDNA libraries (generously provided by Dr. T. Desprez; [16]). In the first round of PCR, target DNA was supplied by a sample from an overnight culture of the cDNA plasmid library or from pools of colonies from plates. The complexity of the positive clone pools was reduced until a single clone was isolated. Clone Arab-1, isolated from the VB library, was sequenced completely on both strands by dideoxynucleotide chemistry using double stranded DNA template, Sequenase (Amersham), [^{35}S]dATP (Dupont) and specific primers (DNAgency). Database searches were performed remotely using BLAST [17] and sequences were retrieved from the database using the World Wide Web form interface of BLAST. Protein sequences were manually aligned using SeqVu (Garvan Institute for Medical Research, Australia).

2.2. Expression of Arab-1 protein

A 1.2 kb *Nco*I–*Xho*I fragment from pARAB-1 was ligated into compatible sites in plasmid pET19b (Novagen) to create pARAB2.1 and expression of Arab-1 gene product was achieved in *E. coli* strain K38(pGP1-2) following heat induction of the T7 RNA polymerase [18]. Lipase production was detected using Tween-80 plates [19] streaked with *E. coli* K38(pGP1-2) containing pARAB2.1 or pET19b and incubated at either 30°C or 42°C for 48 h followed by a 24 h incubation at 4°C. Esterase activity was assayed in clarified cell extracts following sonication of induced and uninduced cultures. 1 mM *p*-nitrophenyl butyrate was used as the substrate in 3% acetonitrile, 20 mM HEPES

*Corresponding author. Fax: (1) (604) 721-8855.
E-mail: cupton@uvic.ca

Authors are listed alphabetically.

Genbank accession number: U38916.

SSDI 0014-5793(95)01405-5

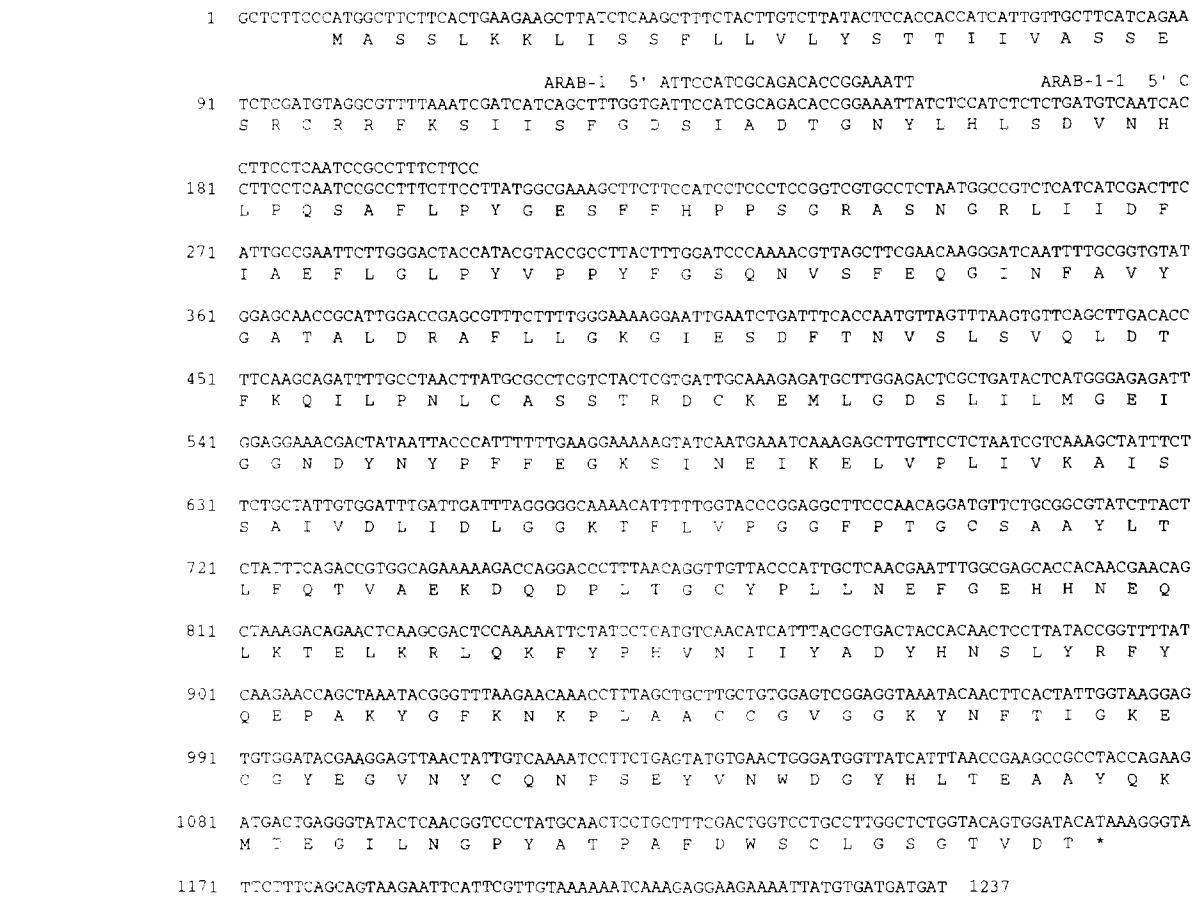

Fig. 1. Complete DNA sequence and translation of Arab-1 cDNA insert. Oligonucleotides used in PCR reactions for clone identification are indicated.

(pH 7.4), 100 mM NaCl and 1% sodium deoxycholate. The increase in absorbance at 400 nm was monitored spectrophotometrically.

2.3. Transcription analysis

Plants were grown from seed (strain RLD) in continuous light at 24°C unless otherwise indicated. Tissue samples were frozen in liquid nitrogen immediately after collection and total RNA was isolated [20]. Total RNA (20 μg/lane) was electrophoretically separated in agarose gels containing formaldehyde [21] and transferred to a Zeta-Probe GT membrane (Bio-Rad). An Arab-1 specific probe was made by random priming with [^{32}P]dCTP (Bio-Rad). Membranes were prehybridized at 65°C for 10 min and hybridized for 16 h in 0.25 M Na_2HPO_4, pH 7.2, containing 7% SDS. Membranes were washed at 65°C, twice in 20 mM Na_2HPO_4, pH 7.2, containing 5% SDS and twice in 20 mM Na_2HPO_4, pH 7.2, containing 1% SDS for 30 min each. The 0.24–9.5 kb RNA marker ladder was purchased from BRL (Burlington).

3. Results

During analysis of the new lipase family, each protein was compared to the dbEST by using TBLASTN [17] to translate the DNA database in all six possible reading frames. We were surprised to find a total of more than 120 significant matches with these translated cDNA sequences, all of which derived from rice, arabidopsis or maize. Most of the matches were against Block I, which contains the predicted active site serine (Fig. 1). Although the limited size of the dbEST sequences prohibits extensive sequence alignment, it was apparent that the great majority of these translated dbEST sequences also contained a PYG motif shortly after Block I. In subsequent database searches we discovered that the simple, single motif [FT]G[ND]SxxDxG[NG]x(10,20)PY, a modified Block I motif, is absolutely specific for the translated cDNAs from the dbEST and the two plant proteins previously identified. This indicates that the database search results are significant and that these sequences are indeed related.

Multiple rounds of database searching with these 120 DNA dbEST sequences revealed that many are identical except for obvious DNA sequencing errors which are characteristic of data in the dbEST. We have therefore organized the dbEST clones into groups (Table 1; [22]) representing 40 distinct sequences from rice [22], arabidopsis [1] and maize [1]. Upon translation, 30 groups of the dbEST sequences gave different peptide sequences each containing the Block I motif. We were therefore able to unequivocally identify these genes as unique, although the rice [15], arabidopsis [14] and maize [1] genes may be homologues. Translation of the remaining 10 groups produced protein sequences that were clearly from the middle of genes (presumably originating from incomplete cDNAs) or was carboxy-terminal peptide sequence from cDNAs cloned and sequenced in the opposite orientation.

In order to determine if the presence of the motif by itself could be used to ascribe dbEST sequences to the new lipase family, the cDNA clone of one arabidopsis dbEST entry (NCBI ID47138) was isolated from a cDNA library and its complete DNA sequence was determined (Fig. 1). The predicted transla-

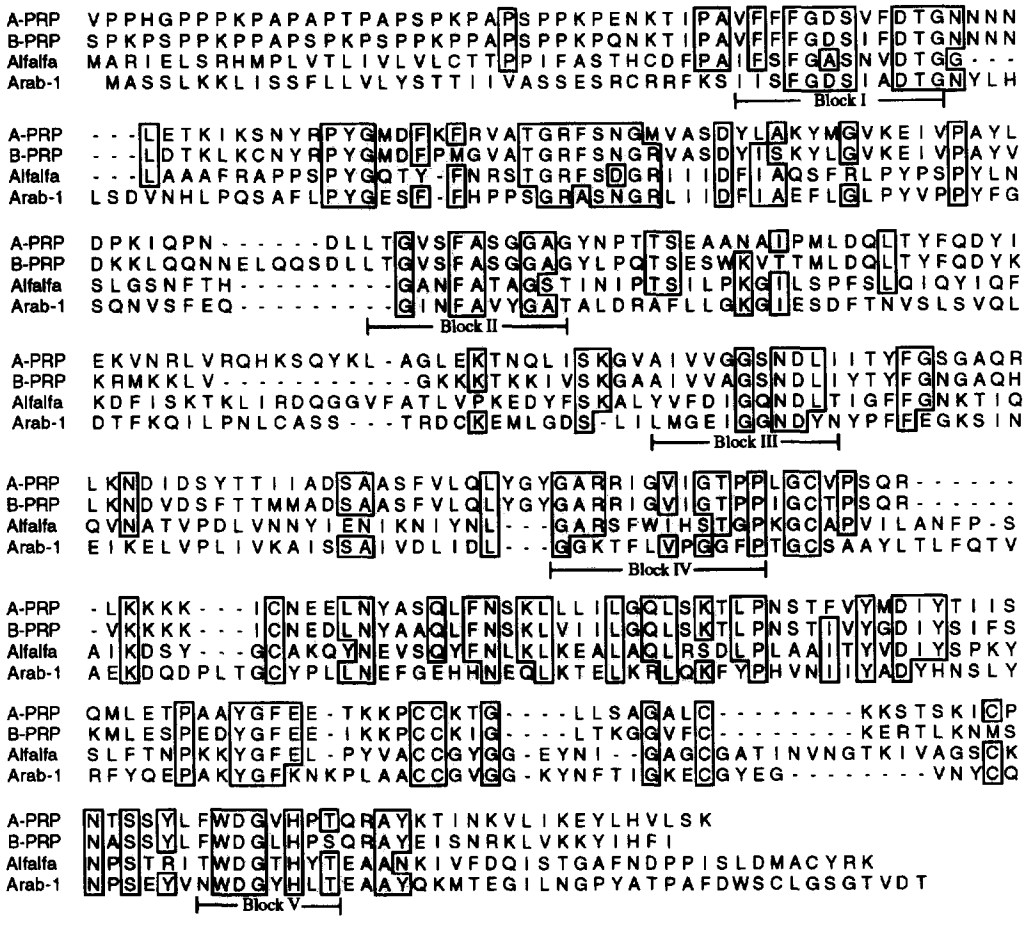

Fig. 2. Alignment of Arab-1 protein sequence with anther-specific proteins from *Arabidopsis thaliana* and *Brassica napus* (A-PRP and B-PRP, respectively) and an early nodule-specific alfalfa protein [23]. The N-terminal, proline-rich domains of A-PRP and B-PRP have been omitted. Amino acids conserved among 3 of the 4 proteins are boxed.

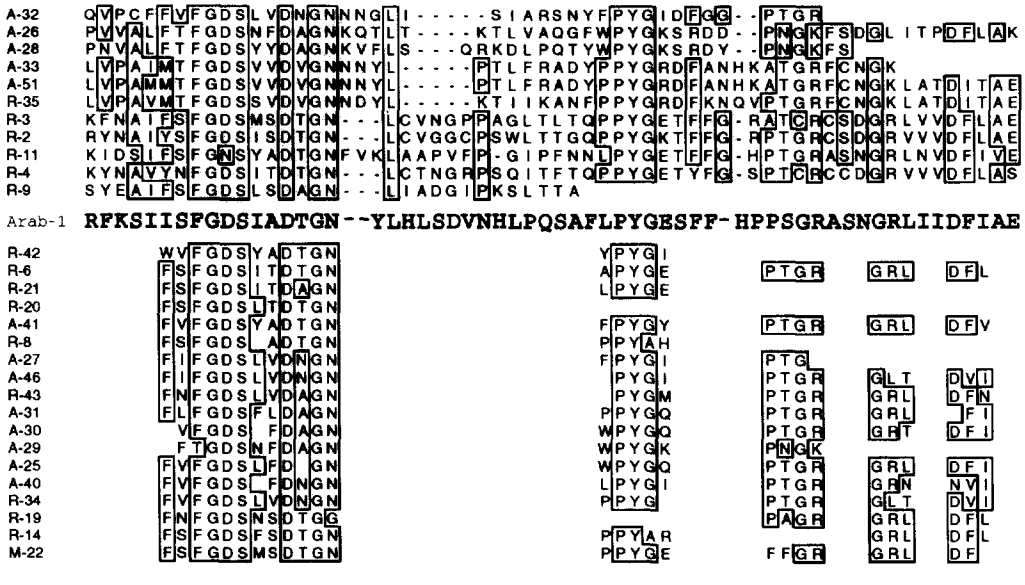

Fig. 3. Alignment of Arab-1 related protein sequences translated from dbEST DNA sequences. The predicted amino acid sequence of clone Arab-1 is shown (bold) in the center. Each sequence above Arab-1 has been translated from a consensus of three or more dbEST sequences (gaps are indicated, –). Each sequence below Arab-1 has been translated from one or two dbEST sequences.

Table 1
Distribution and source of dbEST clones related to Arab-1 (A-1)

Gene	Total	R-s	R-r	R-c	M-e	A-m	A-s	A-e	A-f
A-1	2						1	1	
R-2	12	12							
R-3	13	13							
R-4	20	19	1						
R-6	2		2						
R-8	2	2							
R-9	4	2		2					
R-10	2			2					
R-11	3		3						
R-13	4	4							
R-14	2	2							
R-16	5	3	2						
R-17	3		3						
R-19	1			1					
R-20	1	1							
R-21	1		1						
M-22	1				1				
R-23	1	1							
A-25	2					2			
A-26	3					3			
A-27	2					2			
A-28	3					3			
A-29	2					2			
A-30	1					1			
A-31	1					1			
A-32	6					4	1		1
A-33	3					3			
R-34	1	1							
R-35	3	3							
A-36	2					1		1	
A-40	1					1			
A-41	1					1			
R-42	1	1							
R-43	1	1							
R-44	1	1							
A-45	2							1	1
A-46	2						1	1	
A-48	2					2			
A-49	2					2			
R-50	1	1							
A-51	3					3			
Total	125	67	12	5	1	31	3	4	2

R-s, rice shoot; R-r, rice root; R-c, rice callus; M-e, maize etiolated shoots; A-m, arabidopsis mixed tissues; A-s, arabidopsis siliques; A-e, arabidopsis etiolated shoots; A-f, arabidopsis flowers.

tion product from this clone (Arab-1) is aligned with homologous regions of the two proline-rich-proteins together with a related protein from alfalfa in Fig. 2. The length of the Arab-1 protein is similar to that of the alfalfa protein which is expressed in empty root nodules but is of unknown function [23]. Neither protein possesses the amino-terminal proline-rich domain, a feature of the other two proteins. Each of the five motif blocks previously described as characteristic of the new family of lipases [7] are present in Arab-1 (Fig. 2) and it is clear that there are several other highly conserved regions. Thus, we are confident that the detection of a [FT]G[ND]SxxDxG[NG]x-(10,20)PY motif is sufficient to include genes in the new family of lipases.

The conservation of several blocks of protein sequence, including Block I, at the amino-termini of these proteins is more dramatically demonstrated by an alignment of the predicted protein sequences from the 30 unique genes (Fig. 3). These have been divided into two groups based on the number of dbEST

Fig. 4. Lipolytic activity of the Arab-1 gene product. Tween-80 plates were streaked with *E. coli* K38(pGP1–2) containing pARAB2.1 (panel A) or vector pET19b (panel B) and were incubated at 42°C for 48 h followed by a 24 h incubation at 4°C. The zone of precipitation indicating lipolytic activity is shown by a bar in panel A.

sequences available for analysis. Eleven are derived from three or more dbEST sequences and each of the predicted amino acid sequences is likely to be accurate since it was translated from a consensus DNA sequence. The other translations are from one or two dbEST sequences and while the conserved blocks (presented in Fig. 3) are probably correct, errors in the dbEST sequences make accurate translations of the intermediate sequences impossible and these regions have therefore been omitted from the alignment.

The significant similarity to several bacterial lipases suggested that the Arab-1 gene product would possess lipolytic activity. To test this hypothesis, the Arab-1 gene product was expressed in *E. coli* using an inducible T7 RNA polymerase vector system. A new 38 kDa protein was detected by SDS-PAGE (not shown) and lipolytic activity was correlated with the synthesis of the Arab-1 gene product by screening on Tween-80 plates (Fig. 4). This detects the hydrolysis of oleic acid esters of a polyoxyalkene derivative of sorbitan by precipitation of calcium salts of the liberated fatty acids in the agar. In addition, extracts from induced and uninduced cultures were assayed for esterase activity using *p*-nitrophenyl butyrate as a substrate. Significant activity was associated with synthesis of the Arab-1 gene product upon thermal induction (Table 2).

The dbEST sequences that make up this gene family are from several cDNA libraries, some of which were constructed using mRNA isolated from different plant tissues, allowing us to analyze the source of the individual dbEST sequences (Table 1). It is apparent that some of the genes are differentially expressed. For example, 13 of 13 R-3 sequences are from rice shoot; 3 of 3 R-11 sequences are from rice root; 2 of 2 R-10 sequences are from rice callus and 3 of 5 and 2 of 5 R-16 sequences are from rice shoot and rice root respectively. A Northern blot of mRNA from eight arabidopsis tissues detected Arab-1 mRNA only in the sample isolated from etiolated

Table 2
Esterase activity assayed using *p*-nitrophenyl butyrate substrate

	nmol/min/mg protein
Control	0.380 ± 0.017
Induced	0.730 ± 0.012

Assays were performed in triplicate. Substrate conversion was linear over the time course of the experiment (105 min).

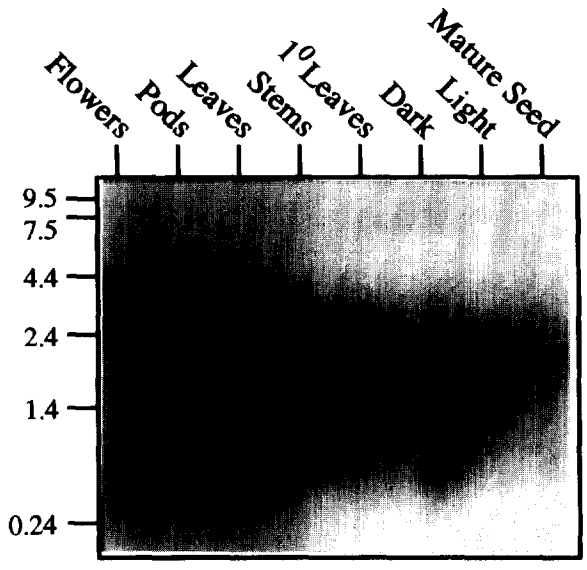

Fig. 5. Northern blot of RNA from different arabidopsis tissues or whole seedlings grown in the light or dark probed with Arab-1. Molecular weight markers indicate kilobases.

shoots. This tissue was use in the construction of the cDNA library from which Arab-1 was isolated (Fig. 5). Together with the data in Table 1, this supports the hypothesis that many of the sequences in the large group of cDNAs described here may represent differentially expressed genes. The single band in Fig. 5 has a mobility of approximately 1.3 kb and suggests that clone Arab-1 (1237 bp) is, or is very close to, a complete cDNA. Thus, these data support our previous conclusion that Arab-1 is more similar to the alfalfa protein, shown in Fig. 2, than to the proline-rich proteins from *Arabidopsis* and *Brassica sp.*

4. Discussion

We believe that we have uncovered a large group of plant lipases based on the similarity of their sequences to those of a new family of lipases that we recently identified [7]. Expression of the Arab-1 protein was associated with lipolytic activity. These results, together with the significant primary sequence conservation with previously well characterized lipases [7] suggests that these plant genes encode lipolytic proteins. With the exception of our present study of the Arabidopsis enzyme, no activity has been assigned to any of the other plant proteins in this family. This work may suggest future directions for their study.

The plant proteins we have described possess, and were in fact identified by, significant similarity to a group of bacterial lipases. However, alignments of the protein primary sequences suggest that the bacterial and plant enzymes may actually represent related groups in a larger family which itself is quite distinct from the classical lipases. Testing this will require the isolation of other complete gene sequences, expression of their gene products, and structural studies using methodologies that have been employed successfully with the mammalian and microbial lipases.

An interesting feature of this group of plant proteins is the fact that its members appear to be differentially expressed. It suggests that lipases play an important role in the regulation of plant development and morphogenesis. We also noticed that alignment of some dbEST sequences from genes R-16 and A-33 required the introduction of gaps of unit codon length in otherwise identical DNA sequences (not shown), suggesting that the diversity of this gene family may also be enhanced by differential splicing of primary mRNA transcripts.

Currently nothing is known about the natural substrate(s) for these plant enzymes, although it is likely that novel substrate specificities will be found and that these may have agricultural or industrial significance. Future studies on this new family of lipases will certainly advance our understanding of the important roles that lipids play in plant metabolism and development.

Acknowledgements: This work was supported by Operating Grants from The Natural Sciences and Engineering Research Council of Canada (NSERC) to C.U. and S.M and from The Heart and Stroke Foundation of British Columbia and the Yukon to J.T.B.. P.C.D. was supported by an NSERC Summer Studentship.

References

[1] Somerville, C. and Browse, J. (1991) Science 252, 80–87.
[2] Huang, A.H.C. (1992) Annu. Rev. Plant Physiol. Plant Mol. Biol. 43, 177–200.
[3] Huang, A.H.C. (1993) in: Lipid Metabolism in Plants, T.S. Moore Ed. (Academic Press Inc., Florida, USA), pp. 473–503.
[4] Sparace, S.A. and Kleppinger-Sparace, K.F. (1993) in: Lipid Metabolism in Plants, T.S. Moore Ed. (Academic Press Inc., Florida, USA), pp. 569–589.
[5] Wang, X. (1993) in: Lipid Metabolism in Plants, T.S. Moore Ed. (Academic Press Inc., Florida, USA), pp. 505–525.
[6] Moulin, A., Teissere, M., Bernard, C. and Pieroni, G. Proc. Natl. Acad. Sci. USA (1994) 91, 11328–11332.
[7] Upton, C. and Buckley, J.T. (1995) Trends Biol. Sci. 20, 178–9.
[8] Cygler, M., Schrag, J.D., Sussman, J.L., Harel, M., Silman, I., Gentry, M.K. and Doctor, B.P. (1993) Protein Sci. 2, 366–382.
[9] Derewenda, Z.S. and Derewenda, U. (1991) Biochem. Cell Biol. 69, 842–851.
[10] Thornton, J., Howard, S.P. and Buckley, J.T. (1988) Biochim. Biophys. Acta 959, 153–150.
[11] Shinoda, S., Matsuoka, H., Tsuchie, T., Miyoshi, S., Yamamoto, S., Taniguchi, H. and Mizuguchi, Y. (1991) J. Gen. Microbiol. 137, 2705–2711.
[12] Wang, H., and Dowds, B.C.A. (1993) J. Bacteriol. 175, 1665–1673.
[13] Brumlik, M.J. and Buckley, J.T. (submitted).
[14] Roberts, M.R. et al., (1993) Plant J. 3, 111–120.
[15] Boguski, M.S., Lowe, T.M. and Tolstoshev, C.M. (1993) Nat. Genet. 4, 332–333.
[16] Hofte, H., Desprez, T., Amselem, J., Chiapello, H., Caboche, M., Moisan, A., Jourjon, M.F., Charpenteau, J.L., Berthomieu, P., Guerrier, D. (1993) Plant J. 4, 1051–1061.
[17] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.
[18] Tabor, S. and Richardson, C.C. (1985) Proc. Natl. Acad. Sci. USA 82, 1074–1078.
[19] Sierra, G.J. (1957) Microbiol. Serol. 23, 15–22.
[20] Verwoerd, T.C., Dekker, M.M. and Hoekema, A. (1989) Nucleic Acids Res. 17, 2362.
[21] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) in Molecular Cloning: A Laboratory Manual. (Cold Spring Harbor Laboratory. Cold Spring Harbor, N.Y., edn. 2.).
[22] NCBI IDs of dbEST sequences. Group names used in Table 1 are shown in bold. **A1** = 47138, 52088; **R2** = 72573, 72201, 72336, 75854, 75966, 145908, 145801, 75354, 71463, 71400, 75968, 75189; **R3** = 146127, 75831, 145197, 72477, 145962, 144786, 145322, 146345, 71401, 72477, 146345, 145322, 71401; **R4** = 75878, 146710, 75254, 72471, 146415, 75890, 146277, 71694, 144179, 75602, 75151,

71129, 75382, 71215, 71351, 71261, 35519, 70829, 146105, 71228; **R6** = 37748, 37989; **R8** = 71812, 75939; **R9** = 35343, 35456, 71679, 75927; **R10** = 36804, 36824; **R11** = 37977, 38279, 38295; **R13** = 71781, 71848, 71152, 75836; **R14** = 71332, 71794; **R16** = 71799, 72018, 70777, 75634, 70700; **R17** = 70752, 70676, 37511; **R19** = 36073; **R20** = 72326; **R21** = 37696; **M22** = 40259; **R23** = 146461; **A25** = 200269, 35046; **A26** = 21630, 97255, 95715; **A27** = 48963, 95787; **A28** = 200404, 93247, 200615; **A29** = 96977, 96392; **A30** = 199980; **A31** = 97345; **A32** = 97166, 21171, 94373, 31710, 97098, 62584; **A33** = 97704, 93363, 94191, 98088, 96754, 21028, 48300; **R34** = 75258; **R35** = 146356, 75978, 75922; **A36** = 34958, 52056; **A40** = 949575; **A41** = 209675; **R42** = 717641; **R43** = 717231; **R44** = 1451481; **A45** = 142488, 317119; **A46** = 143138, 625847; **A48** = 349865, 2001835; **A49** = 2004945, 1382285; **R50** = 725321.

[23] Dickstein, R., Prusty, R., Peng, T., Ngo, W. and Smith, M.E. (1993) Molecular Plant-Microbe Interactions 6, 715–721.