

Analysis of randomly isolated cDNAs from developing endosperm of rice (*Oryza sativa* L.): evaluation of expressed sequence tags, and expression levels of mRNAs

Jinyuan Liu^{1,2}, Chikage Hara¹, Masaaki Umeda¹, Yuan Zhao³, Thomas W. Okita³ and Hirofumi Uchimiya^{1,*}

¹*Institute of Molecular and Cellular Biosciences, University of Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113, Japan (*author for correspondence);* ²*Institute of Life Sciences and Biotechnology, Tsinghua University, Beijing 100084, China;* ³*Institute of Biological Chemistry, Washington State University, Pullman, WA 99164, USA*

Received 12 April 1995; accepted in revised form 25 July 1995

Key words: rice, endosperm, cDNA, ESTs

Abstract

Using a cDNA library prepared from poly(A)⁺ RNA from 10-day-old rice endosperm, partial nucleotide sequences of randomly isolated clones were analyzed. A total of 153 (30.6%) out of 500 cDNA clones showed high amino acid identity to previously identified genes. There was significant redundancy in cDNAs encoding prolamine and glutelin. About 21.0% of the cDNA clones were found to code for seed storage protein genes. Consequently, 37 independent genes were identified. Using cDNA clones encoding glutelin, prolamine, seed allergen, α -1,4-glucan branching enzyme, glycine-rich RNA binding protein, metallothionein, non-specific lipid-transfer protein and ubiquitin conjugating enzyme the accumulation of mRNA during rice seed development was compared. Genes associated with seed storage protein and starch biosynthesis were expressed according to expected developmental stages. Glycine-rich RNA binding protein genes as well as metallothionein-like protein genes were highly expressed in developing seeds, but low in leaves of whole plants.

Introduction

Numerous genes have recently been identified by random nucleotide sequencing of cDNA clones in higher plants [1, 2, 3, 4, 5, 6, 7]. Such methods have proven valuable in terms of a rapid screening of putative genes from given cDNA li-

braries. Genes identified as expressed sequence tags (ESTs) can be useful for the construction of RFLP maps and for the isolation of full-length cDNA clones. Previously we reported the isolation of a variety of putative genes from cDNA libraries produced from cultured cells of rice [1, 7]. In this study, we determined the partial

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers listed in Table 1 (D29993–D30029).

nucleotide sequences of 500 cDNA clones of 10-day-old rice endosperm tissues, in which 153 clones showed homology with known genes on the basis of nucleotide or protein database searches. As expected, there was a tremendous redundancy of genes encoding seed proteins such as prolamin and glutelin. Nevertheless, other genes identified were distributed to a broad spectrum of plant metabolic pathways. Comparative analysis of gene expression pattern during seed development using several cDNA probes will also be discussed.

Materials and methods

cDNA library

Endosperms of rice (*Oryza sativa* L. cv. M20) seed isolated 10 days after anthesis were used for the extraction of mRNA. Poly(A)⁺ RNA was obtained according to Logemann *et al.* [8]. A cDNA library was constructed in Uni-ZAP XR vector (Stratagene). *In vivo* excision of pBlue-script plasmids was performed in the *Escherichia coli* K-12 strain XL1-Blue.

Nucleotide sequence determination

Plasmid DNA for nucleotide sequencing was prepared using a DNA purification column (Wizard Minipreps, Promega). Partial nucleotide sequences of cDNA inserts were determined by the fluorescence detection method using dye-labeled T3 primer. Cycle sequencing was performed with a CATALYST 800 Molecular Biology LabStation (Perkin Elmer). A Model 373A sequencer (Perkin Elmer) was used for automated sequence analysis. The homology of each partial cDNA sequence was searched as described previously [1].

Northern blot analysis

Total RNA was isolated by the phenol-SDS method from developing seeds as described ear-

lier [9, 10]. After electrophoresis of 10 µg of total RNA per lane in a 1.2% formaldehyde agarose gel, RNA was transferred to a Hybond-N⁺ membrane (Amersham) according to the manufacturer's specifications. A ³²P-labeled DNA probe was added to the solution, which consisted of 1% (w/v) SDS, 1 M NaCl, 10% (w/v) dextran sulfate, and 0.1 mg/ml heat-denatured salmon sperm DNA. After hybridization at 65 °C, the filter was washed in 2 × SSC for 10 min at room temperature with vigorous shaking and then twice in 2 × SSC and 1% SDS for 30 min at 65 °C. The filter was dried and exposed to X-ray film.

Results and discussion

This research was carried out to analyze ESTs of cDNA libraries from developing endosperms of rice plants. Such libraries were abundant in storage protein cDNAs [11]. Table 1 shows a list of putative genes identified in this study. As expected we found high degree of redundancy of specific genes (Table 2). cDNA clones corresponding to glutelin and prolamin constituted 12.6% and 8.4% of the 500 cDNAs analyzed, respectively. In view of the distribution of other genes engaged in various metabolic pathways, genes related to translational processes are most abundant (e.g. elongation factors and ribosomal proteins). This substantiates reports that show a significant amount of protein synthesis is taking place during protein synthesis [12].

Northern blot analysis of glutelin and prolamin mRNA accumulation indicated that both genes were expressed in a similar manner 10–30 days after anthesis (Fig. 1). In contrast, genes encoding α-1,4-glucan branching enzyme [13] was expressed rather early in seed development. Adachi *et al.* [14] reported that transcripts of a rice seed allergenic protein gene accumulated maximally between 15 and 20 days after flowering, which is also consistent with our results. The ubiquitin conjugating enzyme gene was not expressed at a substantial level in developing seeds, although significant expression was observed in young and mature leaves of rice plants. Interestingly, genes

Table 1. Summary of EST's isolated from the cDNA libraries of developing rice endosperm.

Accession No.	Gene	Length compared (aa)	% matched (aa)	Source of comparison
D29995	Acyl carrier protein III precursor	85	75	ACP3-HORVU (SWISS)
D30010	ADP,ATP carrier protein precursor	32	96	ADT-ORYSA (SWISS)
D30002	ADP-ribosylation factor	40	85	S28875 (PIR)
D29993	Alanine transaminase	91	91	S28429 (PIR)
D30024	Alcohol dehydrogenase	97	97	ADH1-ORYSA (SWISS)
D29996	α -1,4-glucan branching enzyme	71	97	GLGB-ORYSA (SWISS)
D30007	Asparagine synthase	48	85	AJPMN2 (PIR)
D30018	Elongation factor 1- α	74	66	EF1A-WHEAT (SWISS)
D30000	Elongation factor TU	46	65	EFTU-BACSU (SWISS)
D29999	Glutelin	83	97	JR0002 (PIP)
D30025	Glycine-rich RNA-binding protein	48	75	S30147 (PIR)
D30021	Histidine decarboxylase	66	54	LEHDC (GB)
D30022	Histone H4	56	66	S25638 (PIR)
D30006	IM30 protein	105	80	S27765 (PIR)
D30029	Initiation factor eIF-2	50	74	HUMIEF2G (GB)
D30020	Malate dehydrogenase	78	57	MDH-THEFL (SWISS)
D29998	Metallothionein-like protein	48	62	S16534 (PIR)
D30023	NADPH-ferrihemoprotein reductase	37	70	HTL2NFR (GB)
D29994	Nonspecific lipid-transfer protein	26	88	NLTP-ORYSA (SWISS)
D30017	Phenylalanine ammonia-lyase	24	95	S28185 (PIR)
D30003	Phosphoprotein phosphatase	125	88	S31088 (PIR)
D30004	Procyclin	33	69	TOBPROCYC (GB)
D30008	Prolamin	83	92	PRO2-ORYSA (SWISS)
D30016	Ribosomal protein L7	43	79	1909359B (PRF)
D30026	Ribosomal protein L17	32	84	S32579 (PIR)
D30019	Ribosomal protein S4	48	45	R3RT4 (PIR)
D30013	Ribosomal protein S15A	60	90	RS15-BRANA (SWISS)
D30028	40S ribosomal protein	42	97	RIC40SSRPD (GB)
D30011	50S ribosomal protein L14	28	96	RL14-ORYSA (SWISS)
D30012	60S ribosomal protein L25	58	67	TOBRPL25A (GB)
D30001	Seed allergen	38	81	S31078 (PIR)
D29997	Stearoyl acyl carrier protein desaturase	27	70	STAD-BRANA (SWISS)
D30005	Sucrose synthase	112	91	JT0281 (PIR)
D30009	α -tubulin	31	80	TBA-NAEGR (SWISS)
D30027	Ubiquitin conjugating enzyme	99	85	S31971 (PIR)
D30015	Wound-inducible basic protein	25	80	JS0731 (PIR)

Quality of match is given as percent identity in amino acid residues. Sources of comparison were based on the recorded sequences in the SWISS-PROT (SWISS), PIR, PRF and GenBank (GB) databases which are indicated in parenthesis.

like glycine-rich RNA-binding protein and metallothionein like protein were highly expressed in developing seeds when compared with leaf tissues. Since the discovery of the ABA-inducible glycine-rich protein gene in the maize embryo [15], a number of similar glycine-rich RNA binding protein genes have been found in other plants. However, expression of this gene in developing

seeds has not been reported. Metallothioneins (MTs) are a class of cystein-rich, metal-binding proteins important for heavy metal tolerance in animals and fungi [16]. Genes encoding putative MTs were reported in a variety of plant species. Among them, a zinc-binding protein (Ec) was known as a class II MT [17] which accumulated during wheat embryogenesis [18]. Although the

Table 2. A list of identified cDNA clones showing redundancy.

Putative genes	Total numbers of cDNA clones	(%) ^a
Glutelin	63	12.6
Avenin precursor	1	0.2
Globulin precursor	1	0.2
Glutelin	8	1.6
Glutelin precursor	14	2.8
Glutelin 22	1	0.2
Glutelin I	12	2.4
Glutelin II	4	0.8
Glutelin B-1	15	3.0
Glutelin B-2	7	1.4
Prolamin	42	8.4
Prolamin	9	1.8
Prolamin precursor	7	1.4
13 kDa prolamin	14	2.8
14 kDa prolamin	2	0.4
Ribosomal protein	7	1.4
Elongation factor	4	0.8
Metallothionein	3	0.6
Sucrose synthase	3	0.6
Ubiquitin conjugating enzyme	3	0.6
ADP-ribosylation factor	2	0.4
α -1,4-glucan branching enzyme	2	0.4
Malate dehydrogenase	2	0.4
Other putative genes appearing once	22	4.4

^a Calculation was made on the basis of 500 cDNA clones analyzed.

biological significance of seed related expression of both genes is not clearly understood, they may have significant roles during seed development.

In conclusion, random nucleotide sequencing of endosperm cDNA clones provides us with an opportunity to isolate numerous genes including the highly expressed seed protein genes and other metabolically important genes. It would be worthwhile to map such cDNAs to their respective chromosomes of the rice genome.

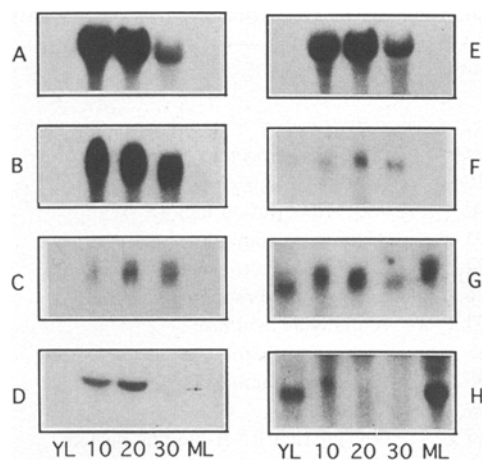


Fig. 1. Autoradiograms showing northern blot analysis of mRNA accumulation during rice seed development at 10, 20 and 30 days after anthesis, in leaves of 10-day-old seedlings (YL), and in leaves of 3-month-old rice plants (ML). Probes used were: glutelin (A), prolamin (B), seed allergen (C), α -1,4-glucan branching enzyme (D), glycine-rich RNA-binding protein (E), metallothionein-like protein (F), non-specific lipid transfer protein (G) and ubiquitin conjugating enzyme (H).

Acknowledgements

This research was supported by Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan, and by a grant from the Rockefeller Foundation. J.L. is a recipient of the fellowship from the Sasakawa Peace Foundation, and from the Association of International Education, Japan.

References

1. Uchimiya H, Kidou S, Shimazaki T, Aotsuka S, Takamatsu S, Nishi R, Hashimoto H, Matsubayashi Y, Kidou N, Umeda M, Kato A: Random sequencing of cDNA libraries reveals a variety of expressed genes in cultured cells of rice (*Oryza sativa* L.). *Plant J* 2: 1005–1009 (1992).
2. Keith CS, Hoang DO, Barret BM, Feigelman B, Nelson MC, Thai H, Baysdorfer C: Partial sequence analysis of 130 randomly selected maize cDNA clones. *Plant Physiol* 101: 329–332 (1993).
3. Park YS, Kwak JM, Kwon O-Y, Kim YS, Lee DS, Cho MJ, Lee HH, Nam HG: Generation of expressed sequence tags of random root cDNA clones of *Brassica napus* by single-run partial sequencing. *Plant Physiol* 103: 359–370 (1993).

4. Höfte H, Desprez T, Amselem J, Chiapello H, Caboche M, Moisan A, Jourjon M-F, Charpentau J-L, Berthomieu P, Guerrier D, Giraudat J, Quigley F, Thomas F, Yu D-Y, Mache R, Raynal M, Cooke R, Grellet F, Delseny M, Parmentier Y, de Marcillac G, Gigot C, Fleck J, Philipps G, Axelos M, Bardet C, Tremousaygue D, Lescure B: An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. *Plant J* 4: 1051–1061 (1993).
5. Sasaki T, Song J, Koga-Ban Y, Matsui E, Fang F, Higo H, Nagasaki H, Hori M, Miya M, Murayama-Kayano E, Takiguchi T, Takasuga A, Niki T, Ishimaru K, Ikeda H, Yamamoto Y, Mukai Y, Ohta I, Miyadera N, Havakkala I, Minobe Y: Toward cataloguing all rice genes: large-scale sequencing of randomly chosen rice cDNAs from a callus cDNA library. *Plant J* 6: 615–624 (1994).
6. Shen B, Carneiro N, Torres-Jarez I, Stevenson B, McCreery T, Helentjaris T, Baysdorfer C, Almira E, Ferl RJ, Habben JE, Larkins B: Partial sequencing and mapping of clones from two maize cDNA libraries. *Plant Mol Biol* 26: 1085–1101 (1994).
7. Umeda M, Hara C, Matsubayashi Y, Li H-H, Liu Q, Tadokoro F, Aotsuka S, Uchimiya H: Expressed sequence tags from cultured cells of rice (*Oryza sativa* L.) under stressed conditions: analysis of transcripts of genes engaged in ATP-generating pathways. *Plant Mol Biol* 25: 469–478 (1994).
8. Logemann J, Schell J, Willmtzer L: Improved method for the isolation of RNA from plant tissues. *Anal Biochem* 163: 16–20 (1987).
9. Palmiter RD: Magnesium precipitation of ribonucleoprotein complexes: expedient techniques for the isolation of undergraded polysomes and messenger ribonucleic acid. *Biochemistry* 13: 3606–3675 (1974).
10. Umeda M, Uchimiya H: Differential transcript levels of genes associated with glycolysis and alcohol fermentation in rice plants (*Oryza sativa* L.) under submergence stress. *Plant Physiol* 106: 1015–1022 (1994).
11. Kim WT, Li X, Okita TW: Expression of storage protein multigene families in developing rice endosperm. *Plant Cell Physiol* 34: 595–603 (1993).
12. Li X, Okita TW: Accumulation of prolamines and glutelins during rice seed development: A quantitative evaluation. *Plant Cell Physiol* 34: 385–390 (1993).
13. Nakamura Y, Yamanouchi H: Nucleotide sequence of a cDNA encoding starch-branching enzyme, or Q-enzyme I, from rice endosperm. *Plant Physiol* 99: 1265–1266 (1992).
14. Adachi T, Izumi H, Yamada T, Tanaka K, Takeuchi S, Nakamura R, Matsuda T: Gene structure and expression of rice seed allergenic proteins belonging to the α -amylase/trypsin inhibitor family. *Plant Mol Biol* 21: 239–248 (1993).
15. Gómez J, Sánchez-Martínez D, Stiefel V, Rigau J, Puigdomènech P, Pagès M: A gene induced by the plant hormone abscisic acid in response to water stress encodes a glycine-rich protein. *Nature* 334: 262–264 (1988).
16. Hamer DH: Metallothionein. *Annu Rev Biochem* 55: 913–951 (1986).
17. Kagi JHR, Schaffer A: Biochemistry of metallothionein. *Biochemistry* 27: 8509–8515 (1988).
18. Kawashima I, Kennedy TD, Chino M, Kimura M, Lane BG: Wheat E_c metallothionein genes: Like mammalian Zn^{2+} metallothionein genes, wheat Zn^{2+} metallothionein genes are conspicuously expressed during embryogenesis. *Eur J Biochem* 209: 971–976 (1992).