Review Plant Tissue Culture Letters, **12**(1), 1-7(1995)

Expressed Sequence Tags of cDNA Libraries in Higher Plants

Shin-ichiro KIDOU*, Masaaki UMEDA** and Hirofumi UCHIMIYA**

(Accepted January 12, 1995)

Introduction

Since the analyses of cDNA libraries play an important role in genome project, many groups around the world have focused their attention in this area. Partial sequencing of randomly selected cDNA clones generates many expressed sequence tags (ESTs) which can be used for genome mapping. Sequence-tagged sites (STSs) are also becoming standard markers in genome mapping. ESTs can provide the opportunity to analyze expression levels of functional genes in different tissues and under various conditions[1]. The first detailed analysis of ESTs was reported in human using brain cDNA library[2]. Since then, similar analyses have been reported in several organisms, including *Caenorhabditis elegans*[3], mammals[4-8] and plants[9-12].

We have initiated a project to identify genes from randomly selected cDNA clones by large-scale sequencing[13-14]. As an experimental material, we have chosen graminaceous monocot, rice (*Oryza sativa* L.), since, along with wheat and corn, it is one of the most important crops. In addition, classical genetics has already succeeded in mapping numerous morphological and isozyme markers on their respective chromosomes[15]. The genome size (4×10^8 kb) of rice is relatively small, about 2.4 times larger than that of *Arabidopsis thaliana*, indicating that molecular techniques can be applied. Recent advances have presented RFLP maps[16,17], and also developed gene transfer technique into rice plants which will help molecular manipulation[18,19].

In the first section of this review, we will report the ESTs identified by a cDNA project which focused on gene expression under different culture conditions[13] and various abiotic stresses[14]. In the second section, we will summarize cDNA projects in other plant species and the ESTs that have already been identified.

Expressed sequence tags (ESTs) from rice suspension cultured cells

1. *Summary of the experimental method*

Fig. 1 shows a flow chart of the basic method used in our rice cDNA project. Total RNA was prepared from suspension cultured cells (4-day-old) according to the method of Palmiter[20], with slight modifications. Poly(A)$^+$RNA was purified by an oligo(dT)-binding latex particle, and cDNAs were synthesized with either cDNA Synthesis System Plus (Amersham) or a ZAP-cDNA Synthesis (Stratagene) kit. cDNA clones were obtained by direct ligation of cDNAs to bacterial plasmids or by *in vivo* excision from the cDNA library constructed with λZAP II vector. The

木藤新一郎*・梅田正明**・内宮博文**
高等植物における cDNA ライブラリーの EST 化研究の現状
* *Institute for Cell Biology and Genetics, Faculty of Agriculture, Iwate University,* 3-18-8 Morioka 020, Japan
** *Institute of Molecular and Cellular Biosciences, University of Tokyo,* Bunkyo-ku, Tokyo 113, Japan
* 岩手大学農学部附属細胞育種実験施設(〒020 岩手県盛岡市上田 3-18-8)
** 東京大学分子細胞生物学研究所 細胞機能研究室(〒113 東京都文京区弥生 1-1-1)

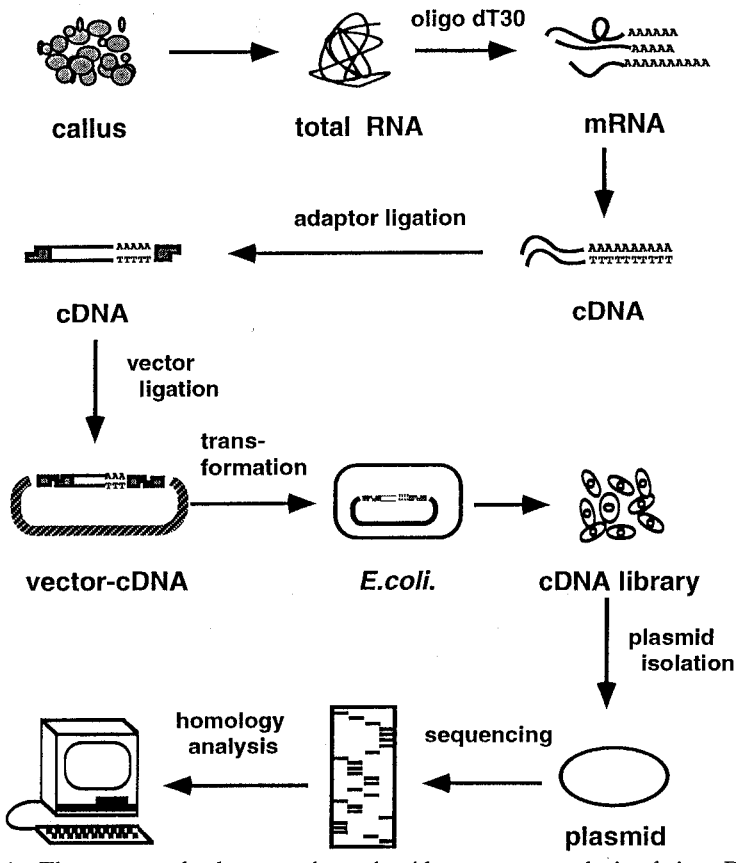

Fig. 1 The strategy for large-scale nucleotide sequence analysis of rice cDNA.

plasmid DNAs of randomly selected cDNA clones were prepared by alkali lysis[21], and the partial nucleotide sequence of each cDNA insert was determined by the chain termination method[22] or by the fluorescence detection method[23]. For automated DNA sequence analysis, a Model 370A sequencer (Applied Biosystems) was used. The GenBank database was searched for nucleic acid sequence homology using FASTA software. Sequences with homology scores over 160 were then evaluated for amino acid homology.

2. *ESTs from rice suspension cultured cells under different sucrose concentrations*[13]

We constructed three cDNA libraries prepared from rice cells cultured in different sucrose concentrations. Although rice cells are usually cultured in AA liquid medium with 2% sucrose, in this experiment, culture media with two additional concentrations of sucrose (6% and 20%) were also used to know the effect of osmotic pressure.

The partial nucleotide sequences (average length 300 bp) of 830 randomly selected cDNA clone were determined from these libraries and subsequently compared to the DNA database. As a result, 68 cDNA clones showed significant homology to known genes, most of which have not yet been isolated from rice plants. As summarized in **Table 1**, the cDNA clones identified in this project were distributed across several categories. Overall, genes related to energy metabolism (*e.g.* enzymes involved in carbohydrate metabolism) and protein biosyntheses (*e.g.* initiation factors and ribosomal proteins) were most abundant. These housekeeping genes were frequently identified in all three cDNA libraries. However, with regard to genes related to signal transduction (*e.g.* small GTP-binding protein), cell cycle (*e.g. cdc2* kinase) and stress response (*e.g.* heat-shock proteins), distinct differences among the three libraries were found. Under the 2% and 6% sucrose conditions, only few genes related to these functions were identified, whereas at 20% sucrose (high

Table 1. Categories of ESTs identified in cDNA libraries of different sucrose concentrations.

Category	Sucrose (%)			Total
	2	6	20	
Carbohydrate metabolism	6	2	3	11
Electron transport system	0	0	2	2
Fatty acid biosynthesis	1	1	1	3
Amino acid biosynthesis	0	0	2	2
Protein biosynthesis	4	3	10	17
Signal transduction and Cell cycle regulation	2	0	8	10
Stress response	1	0	5	6
Others	6	3	8	17
Unknown	290	63	409	762
Total	310	72	448	830

stress condition), more genes were isolated from the library. This may indicate that these genes were expressed at a high level under stressful conditions, resulting in efficient isolation of different types of genes. Using cloned cDNAs, we have reported the nucleotide sequence of the full length cDNA clones such as adenylate kinase[24], 14-3-3 protein[25], ribosomal proteins[26-29], ubiquitin protein[30], ATP/ADP translocator[31], initiation factor 4A[32], *ras*-related small GTP-binding protein[33], nucleoside diphosphate kinase[34], *cyc07*[35], *cdc2* kinase[36] and NADP-dependent malic enzyme[37].

3. *ESTs from rice cultured cells: The influence of external stress*[14]

cDNA libraries were also prepared from rice cultured cells that were subjected to salt stress (2% NaCl) or nitrogen-starvation stress by eliminating nitrogen sources, *L*-glutamine and *L*-aspartic acid from AA medium. Four-day-old cultured cells were subjected to either of these stresses for 24 h before total RNA extraction.

We determined the partial nucleotide sequences of 780 randomly selected cDNA clones, 472 clones from saline condition and 308 clones from nitrogen-starvation conditions. A total of 117 cDNA clones were identified as putative genes (**Table 2**). Many genes involved in carbohydrate metabolism and translation were identified from both libraries. In the library of salt stress, several signal transduction-related genes, such as MAP kinase, receptor protein kinase, ADP-ribosylation factor, and others, were isolated. Such genes have not been identified in other libraries. This result

Table 2. Categories of ESTs identified in cDNA libraries of salt stress and nitrogen-starvation stress.

Category	Stress		Total
	Salt	Nitrogen-Stravation	
Carbohydrate metabolism	8	17	25
Electron transport system	4	0	4
Fatty acid biosynthesis	0	1	1
Amino acid biosynthesis	3	3	6
Protein biosynthesis	14	8	22
Signal transduction and Cell cycle regulation	11	1	12
Stress response	9	15	24
Others	16	7	23
Unknown	407	256	663
Total	472	308	780

Table 3. Examples of expressed sequence tags (ESTs) from plants.

Plant	Tissue	Vector	No. of ESTs	No. of putative genes	% redundancy
Oryza sativa[13]	suspension cultured cells	pIBI-31			
	2% sucrose		310	20	
	6% sucrose		72	9	
	20% sucrose		448	39	
	(total		830	68	17)
Oryza sativa[14]	suspension cultured cells	λZAP II			
	2% sodium chloride		472	65	
	eliminating the nitrogen source		308	52	
	(total		780	117	19)
Oryza sativa[12]	callus	pBluescript II	2778	690	—
Arabidopsis thaliana[9]	flower buds	λZAP II	234	60	34
	developing siliques	λZAP II	318	155	29
	suspension cultured cells	λZAP II	353	94	9
	cultured leaf strips	λZAP II	60	15	0
	etiolated seedlings	pHD-1	169	51	2
	(total		1134	375	31)
Brassica napus[10]	root	pUC 19	191	21	—
Zea mays[11]	leaf	Uni-ZAP	130	25	—

suggests that several signal transduction-related genes that are induced by salt stress might be identified. From both libraries, we succeeded in isolating many stress-related genes, such as ascorbate peroxidase, β-glucanase, heat-shock proteins and superoxide dismutase. To examine the usefulness of such tagged cDNAs for the study of gene expression in a specific metabolic pathway, we studied mRNA levels of genes involved in ATP-generating pathways in rice cultured cells under different stress conditions, such as 20% sucrose, salt stress, cold stress and nitrogen-starvation stress. Northern blot analysis using ESTs as probe revealed the coordinated induction of several genes in key steps under stress conditions. This suggests that activation of the entire energy-producing pathway may require coordinated expression of key enzymes. Furthermore, this result indicates that ESTs can be used to generate a transcript map of rice gene.

Other cDNA projects in plants

Several other cDNA projects in plants are underway[9-12], and many ESTs have been identified (**Table 3**). The cDNA project of *Arabidopsis thaliana* is being conducted as a joint venture of France and the United States. In this project, the partial sequence of 1152 randomly selected cDNA clones from different tissues (flower buds, developing siliques, suspension cultured cells, cultured leaf strips and etiolated seedlings) have been determined, and a total of 375 ESTs (32.6%) has been identified as putative genes that are expressed during floral development, embryogenesis, seed maturation, development of etiolated plants and cell cycle[9]. The number of identified genes in this project is surprisingly high compared with our rice project. The difference might be explained by the type of method used for sequence data analysis. For example, the data analysis of this project was performed by BLASTN software with FASTA software. The cut-off homology score (150) used in their project is low compared to 160 which we used in this project. For instance, in *Brassica napus*, 192 randomly selected cDNA clones from root tissue have been sequenced and 21 ESTs (11.0%) have been identified using FASTDB software[10]. In other words, this program is more sensitive than FASTA software because it can detect local homology regions. In maize, 130 randomly selected cDNA clones from leaf tissue have been sequenced and 25 ESTs (19.2%) have been identified[11]. In Japan, another cDNA project of rice has made great advances recently; 2778 randomly selected cDNA clones from callus tissue have been sequenced and 690 ESTs (24.8%) have been identified[12]. The accumulation of these plants ESTs will present much useful information in knowing the function of plants, and will make possible the comparisons of expressed genes among plants.

Future prospects and conclusions

Recent developments in plant molecular biology, such as chromosome walking and gene tagging, have made it possible to isolate and characterize specific genes from plants. These methods are useful when one is interested in a particular phenotype or gene. However, isolation of numerous genes from one plant species is essential for preparation of genome mapping and systematic analysis of gene expressions. We have initiated a rice cDNA project, and thus far identified many genes engaged in energy metabolism, protein biosynthesis, signal transduction, and the cell cycle. These results suggest that this approach can be used to isolate many functional genes rapidly. Moreover, these cDNA clones can serve as expressed sequence tags (ESTs), which will facilitate the identification of candidate genes on RFLP maps for breeding. However, several problems remain unsolved. For example, in most cases, the percentage of identified genes is less than 20% of the total cDNA clones sequenced, and the extent of repeated sequencing of the same cDNA clones

increases with the progress of this project. Therefore, improvement of softwares is necessary for efficient gene identification by nucleotide or amino acid homology search, and the normalization of cDNA libraries is important to reduce the effort of repetitive sequencing and to have access to the rare mRNA sequences.

Acknowledgments

We thank Dr. A. Kato and other colleagues who have been participating to this research. This research was supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture and Science, Japan, and by a grant from the Rockefeller Foundation.

References

1) Umeda, M., H. Uchimiya, 1994. Plant Physiol., **106**: 1015-1022.
2) Adams, M. D., J. M. Kelley, J. D. Gocayne, M. Dubnick, M. H. Polymeropoulos, H. Xiao, C. R. Merril, A. Wu, B. Olde, R. F. Moreno, A. R. Kerlavage, W. R. McCombie, J. C. Venter, 1991. Science, **252**: 1651-1656.
3) McCombie, W. R., M. D. Adams, J. M. Kelley, M. G. Fitzgerald, T. R. Utterback, M. Khan, M. Dubnick, A. R. Kerlavage, J. C. Venter, C. Fields, 1992. Nature Genet., **1**: 124-131.
4) Hoog, C., 1991. Nucleic. Acids. Res., **19**: 6123-6127.
5) Adams, M. D., M. Dubnik, A. R. Kerlavage, R. Moreno, J. M. Kelley, T. R. Utterback, J. W. Nagle, C. Fields, V. Venter, 1992. Nature, **355**: 632-634.
6) Okubo, K., N. Hori, R. Matoba, T. Niiyama, A. Fukushima, Y. Kojima, K. Matsubara, 1992. Nature Genet., **2**: 173-179.
7) Waterston, R., C. Martin, M. Craxton, C. Huynh, A. Coulson, L. Hillier, R. Durbin, P. Green, R. Shownkeen, N. Halloran, M. Metzstein, T. Hawkins, R. Wilson, M. Berks, Z. Du, K. Thomas, J. Thierry-Mieg, J. Sulston, 1992. Nature Genet., **1**: 114-123.
8) Adams, M. D., A. R. Kerlavage, C. Fields, J. C. Venter, 1993., Nature Genet., **4**: 256-267.
9) Hofte, H., T. Desprez, J. Amselem, H. Chiapello, M. Caboche, A. Moisan, M. F. Jourjon, J. L. Charpenteau, P. Berthomieu, D. Guerrier, J. Giraudat, F. Quigley, F. Thomas, D. Y. Yu, R. Mache, M. Raynal, R. Cooke, F. Grellet, M. Delseny, Y. Parmentier, G. Marcillac, C. Gigot, J. Fleck, G. Philipps, M. Axelos, C. Bardet, D. Tremousaygue, B. Lescure, 1993. Plant J., **4**: 1051-1061.
10) Keith, C. S., D. O. Hoang, B. M. Barrett, B. Feigelman, M. C. Nelson, H. Thai, C. Baysdorfer, 1993. Plant Physiol., **101**: 329-332.
11) Park, Y. S., J. M. Kwak, O. Y. Kim, D. S. Lee, M. J. Cho, H. H. Lee, H. G. Nam, 1993. Plant Physiol., **103**: 359-370.
12) Sasaki, T., J. Song, Y. K. Ban, E. Matsui, F. Fang, H. Higo, H. Nagasaki, M. Hori, M. Miya, E. M. Kayano, T. Takiguchi, A. Takasuga, T. Niki, K. Ishimaru, H. Ikeda, Y. Yamamoto, Y. Mukai, I. Ohta, N. Miyadera, I. Havukkala, Y. Minobe, 1994. Plant J., **6**: 615-624.
13) Uchimiya, H., S. Kidou, T. Shimazaki, S. Aotsuka, S. Takamatsu, R. Nishi, H. Hashimoto, Y. Matsubayashi, N. Kidou, M. Umeda, A. Kato, 1992. Plant J., **2**: 1005-1009.
14) Umeda, M., C. Hara, Y. Matsubayashi, H. H. Li, Q. Liu, F. Tadokoro, S. Aotsuka, H. Uchimiya, 1994. Plant Mol. Biol., **25**: 469-478.
15) Kinoshita, T., 1990. Rice Genetics Newsletter, **7**: 22-23.
16) McCouch, S. R., G. Kochert, Z. H. Yu, Z. Y. Wang, G. S. Khush, W. R. Coffman, S. D. Tanksley, 1988. Theor. Appl. Genet., **76**: 815-829.
17) Saito, A., M. Yano, N. Kishimoto, M. Nakagahra, A. Yoshimura, K. Saito, S. Kuhara, Y. Ukai, M. Kawase, T. Nagamine, S. Yoshimura, O. Ideta, R. Ohsawa, Y. Hayano, N. Iwata, M. Sugiura, 1991. Jpn. J. Breed., **41**: 665-670.
18) Toriyama, K., Y. Arimoto, H. Uchimiya, K. Hinata, 1988. Biotechnology, **6**: 1072-1074.

19) Uchimiya, H., T. Handa, D. S. Brar, 1989. J. Biotechnol., **12**: 1-20.
20) Palmiter, R. D., 1974. Biochemistry, **13**: 3606.
21) Birnboim, H. C., J. Doly, 1979. Nucleic Acids Res., **7**: 1513-1523.
22) Sanger, F., 1981. Science, **214**: 1205-1210.
23) Smith, L. M., J. Z. Sanders, R. J. Kaiser, P. Hughes, C. Dodd, C. R. Connell, C. Heiner, S. B. H. Kent, L. E. Hood, 1986. Nature, **321**: 674-679.
24) Kawai, M., S. Kidou, A. Kato, H. Uchimiya, 1992. Plant J., **2**: 845-854.
25) Kidou, S., M. Umeda, A. Kato, H. Uchimiya, 1993. Plant Mol. Biol., **21**: 191-194.
26) Kidou, S., M. Umeda, A. Kato, H. Uchimiya, 1993. Nuc. Acid Res., **21**: 2013.
27) Nishi, R., S. Kidou, H. Uchimiya, A. Kato, 1993. Biochimi. Biophysi. Acta, **1216**: 110-112.
28) Nishi, R., H. Hashimoto, H. Uchimiya, A. Kato, 1993. Biochimi. Biophysi. Acta, **1216**: 113-114.
29) Hihara, Y., M. Umeda, C. Hara, K. Toriyama, H. Uchimiya, 1994. Plant Physiol., **105**: 753-754.
30) Nishi, R., H. Hashimoto, S. Kidou, H. Uchimiya, A. Kato, 1993. Plant Mol. Biol., **22**: 159-161.
31) Hashimoto, H., R. Nishi, M. Umeda, H. Uchimiya, A. Kato, 1993. Plant Mol. Biol., **22**: 163-164.
32) Nishi, R., S. Kidou, H. Uchimiya, A. Kato, 1993. Biochimi. Biophysi. Acta, **1174**: 293-294.
33) Kidou, S., T. Anai, M. Umeda, S. Aotsuka, T. Tsuge, A. Kato, H. Uchimiya, 1993. FEBS Letter, **332**: 282-286.
34) Yano, A., T. Shimazaki, A. Kato, M. Umeda, 1993. Plant Mol. Biol., **23**: 1087-1090.
35) Kidou, S., M. Umeda, T. Tsuge, A. Kato, H. Uchimiya, 1994. Plant Mol. Biol., **24**: 545-547.
36) Kidou, S., M. Umeda, H. Uchimiya, 1994. DNA Sequence (in press).
37) Fushimi, T., M. Umeda, T. Shimazaki, A. Kato, K. Toriyama, H. Uchimiya, 1994. Plant Mol. Biol., **24**: 965-967.