論文 / 著書情報
Article / Book Information

題目(和文)	シロイヌナズナを形成する各組織の特性および光形態形成の制御機構のトランスクリプトーム解析
Title(English)	Transcriptome analysis for characteristics of organs and regulation of photomorphogenesis in Arabidopsis thaliana
著者(和文)	大林武
Author(English)	
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第5633号, 授与年月日:2004年3月26日, 学位の種別:課程博士, 審査員:
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第5633号, Conferred date:2004/3/26, Degree Type:Course doctor, Examiner:
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Contents

Abbreviations

AGRIS	Arabidopsis Gene Regulatory Information Server
ATTED	*Arabidopsis thaliana* Tissue-Specific Expression Database
AraCyc	*Arabidopsis thaliana* Biochemical Pathways
BCL	Background Subtraction, Centering and Lowess
c1 to c9, cx	cluster 1 to cluster 9, cluster x
CEG	Correlation between Expression and Group
EST	Expressed Sequence Tag
GO	Gene Ontology
GRE	Group Relative Expression
KEGG	Kyoto Encyclopedia of Genes and Genomes
LN	Lognormal Fitting
NE	Normalized Expression
RE	Relative Expression
SN	Signal to Noise ratio
SOM	Self-Organizing Map
TAIR	The Arabidopsis Information Resource

Chapter I

General Introduction

I-1.1 Transcriptome

Transcript profiling has been performed on such one by one analysis as Northern hybridization for 25 years (Alwine et al., 1977). Because this method is characterized by well-defined sensitivity, it is suited for the in-depth analysis of a small number of genes. By comparison, current high-throughput transcript profiling technologies have relatively poorly defined sensitivities, so the traditional methods provide both a valuable means of confirming and extending results obtained by the global approaches. Several approaches for the global transcripts profiling are used. Currently, this kind of technologies are promoting day after day. The expression of almost all genes in an intended living organism could be measured. In that sense, this kind of global transcript profiling is called transcriptome, which is now fundamental analysis for overall understanding of organisms.

Fig. I-1 is a conceptual diagram for the difference between transcriptome analysis and traditional in-depth analysis. Traditional analysis deals with restricted target(s) using a lot of methods to clone the gene, analyze the function of protein, investigate the phenotype of mutation and so on. On the other hand, transcriptome analysis deals with an organism as a while only at the transcription level. Transcriptome analysis can not answer such a question, how much the protein accumulates. On the other hand, traditional analysis can not answer such a question; how many genes are up-regulated by the stimulus as a whole. Transcriptome analysis is also valuable with other exhaustive methods like genome sequence. Collaboration of transcriptome and genome enables us to perform promoter analysis. Because genome is common among all cells and unchangeable, the first change linking function of cells is caused at mRNA levels.

The most direct and conventional method is EST (expressed sequence tag, Adams et al., 1991) sequencing. Because EST sequencing does not need genome information and it is also valuable for gene finding, it has been applied for many targets even without determination of the whole genome sequence.

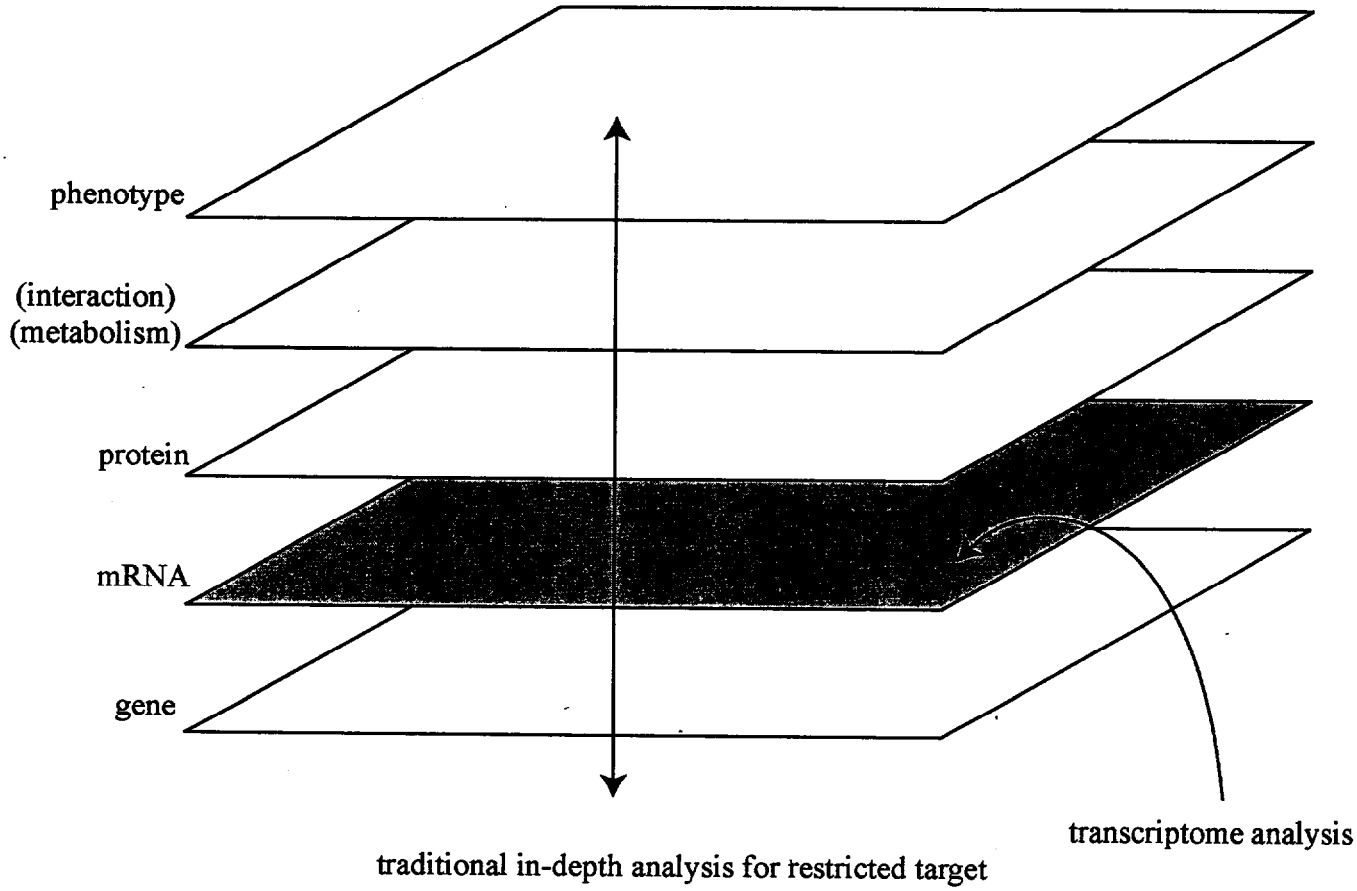


Fig. I-1. Difference between transcriptome analysis and traditional in-depth analysis.

SAGE (serial analysis of gene expression, Velculescu et al., 1995) and MPSS (massively parallel signature sequencing, Brenner et al., 2000) are also sequencing methods, and they are less expensive than EST sequencing. Because genome-wide sequencing needs expensive equipments, it is not suited for small laboratories, while DNA arrays are core methods for global transcript profiling by such laboratories.

The basic idea of spotted nucleic acid arrays for gene expression analysis is not new and has been used in some form for over twenty years (Kafatos et al., 1979). Thus the nylon membrane array is familiar to many biologists. In large scale analysis, such a cDNA macroarray method was reported by Pietu et al. in 1996 using 1091 human skeletal muscle cDNA clone. On the other hand, cDNA microarray using 2-color fluorescence hybridization is more technical method, reported by group Stanford first for 45 Arabidopsis genes (Schena et al., 1995) and then reported for all yeast genes (DeRisi et al., 1997, Chu et al., 1998, Spellman et al., 1998). The commercial oligonucleotide-based arrays are now one of the standards of transcriptome analysis. As far as Arabidopsis is concerned, GeneChip (Affymetrix) with 22000 probes, and arabidopsis2 (agilent technology) with 20500 probes are provided.

Since genomic sequencing for Arabidopsis was finished in 2000 (Arabidopsis Genome Initiative, 2000), the transcriptome analyses are accelerated. So far, a lot of data from cDNA microarray and oligonucleotide array analyses of various physiological events in Arabidopsis have been reported (for review: Donson et al., 2002, Aharoni and Vorst, 2001).

cDNA macroarrays are made using PCR-amplified DNA fragments spotted on nylon membranes (Pietu et al., 1996). They are one of the most convenient tools to handle in a laboratory for global expression analysis. Moreover, their use allows a quantification of expression levels without competition experiments, and the results obtained have been proved to agree well with these obtained by Northern blot analysis (Sasaki et al., 2001).

I-1.2 Strategy for data analysis

Application of computer in biology now opens a break through especially in the field requiring vast calculation. Managements of genome sequence and DNA array data are examples which require

such vast calculation. Construction of gene network using DNA array data of time series or global gene disruptions is one of the most challenging study of bioinformatics. However, it requires almost perfectly accurate data. Moreover in multicellular living organisms, expression data of single tissue or organ are required. Even in that condition, nobody knows how meaningful results those can produce. Several works, however, have been carried out. Abaratani et al. reported network of transcriptional control. They used data of full-genome yeast cDNA microarray about 118 separate gene disruption (Aburatani et al., 2003).

Among several gene monitoring systems, cDNA macroarray system is the cheapest and the easiest tool to handle. However the quality of the data is sometimes worse than such an expensive DNA array system as GeneChip distributed by Affymetrix. Thus, data produced by cDNA macroarray system is not adequate for construction of gene network by gene expression data. Robust analysis allowing errors of data is required. Thus I developed methods for the analysis of the cDNA macroarray data. One major point of the development was application of "gene group", which contains several number of genes. In the global analysis, the characteristics of the gene groups are not varied and thus evident in comparison with characteristics of a single gene.

To analyze the characteristics of organs and relationship between light and cytokinin signals in photomorphogenesis of Arabidopsis, I performed the transcriptome analysis using the analytical methods established here. cDNA macroarray membranes with 13516 cDNA clones (Asamizu et al., 2000), which corresponded to about 9000 loci, were used for the two analyses. The global relationships among the gene expression patterns on the view of functions are discussed.

I-2 REFERENCES

Aburatani S, Tashiro K, Savoie CJ, Nishizawa M, Hayashi K, Ito Y, Muta S, Yamamoto K, Ogawa M, Enomoto A, Masaki M, Watanabe S, Maki Y, Takahashi Y, Eguchi Y, Sakaki Y, Kuhara S (2003) Discovery of novel transcription control relationships with gene regulatory networks generated from multiple-disruption full genome expression libraries. DNA Res **10**: 1-8

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie WR, Venter JC (1991) Complementary DNA Sequencing: Expressed Sequence Tags

and Human Genome Project. Sience **252**: 1651-6

Aharoni A, Vorst O (2001) DNA microarrays for functional plant genomics. Plant Mol Biol **48**: 99-118

Alwine JC, Kemp DJ, Stark GR (1977) Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc Natl Acad Sci U S A **74**: 5350-4.

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

Asamizu E, Nakamura Y, Sato S, Tabata S (2000) A large scale analysis of cDNA in *Arabidopsis thaliana*: Generation of 12,028 non-redundant Expressed Sequence Tags from normalized and size-selected cDNA libraries. DNA Res **7**: 175-80

Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J, Corcoran K (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol **18**: 630-4

Chu S, DeRisi J, Eisen M, Mulholland J, Botstein D, Brown PO, Herskowitz I (1998) The transcriptional program of sporulation in budding yeast. Science **5389**: 699-705

DeRisi JL, Iyer VR, Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. Science **5338**: 680-6

Donson J, Fang Y, Espiritu-Santo G, Xing W, Salazar A, Miyamoto S, Armendarez V, Volkmuth W (2002) Comprehensive gene expression analysis by transcript profiling. Plant Mol Biol **48**: 75-97

Kafatos FC, Jones CW, Efstratiadis A (1979) Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. Nucleic Acids Res **7**: 1541-52

Pietu G, Alibert O, Guichard V, Lamy B, Bois F, Leroy E, Mariage-Sampson R, Houlgatte R, Soularue P, Auffray C (1996) Novel gene transcripts preferentially expressed in human muscles revealed by quantitative hybridization of a high density cDNA array. Genome Res **6**: 492-503

Sasaki Y, Asamizu E, Shibata D, Nakamura Y, Kaneko T, Awai K, Amagai M, Kuwata C, Tsugane T, Masuda T, Shimada H, Takamiya K, Ohta H, Tabata S (2001) Monitoring of Methyl Jasmonate-responsive Genes in Arabidopsis by cDNA Macroarray: Self-activation of Jasmonic Acid Biosynthesis and Crosstalk with Other Phytohormone Signaling Pathways. DNA Res **8**: 153-61

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science **270**: 467-70

Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. Mol Biol Cell **9**: 3273-97

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. Science **5235**: 484-7

Chapter II

Transcriptome Analyses on Characteristics of Organs

II-1 INTORODUCTION

In multicellular living organisms, cells differentiate into many types of tissues, which have specific functions that result from the orchestration of numerous gene products. In higher plants, leaf, stem, flower and root are largely different in morphology and functional roles. For instance, the leaf mainly functions for photosynthesis, the stem provides mechanical strength to the plant, the flower functions for reproduction, and the root functions for uptake of inorganic nutrients. In order to perform these special roles, each organ expresses particular sets of gene groups. Since the genome is common among all cells, the difference of the amount of particular mRNAs is the primary determinant of organ-specific differences, and a major regulatory point for the function of different organs. Although information about organ-specific gene expression has been accumulated for many kinds of genes, the total number of genes investigated is still small. Furthermore, even in Arabidopsis, the overall pattern of organ-specific gene expression has not been determined yet.

Since genomic sequencing for Arabidopsis was finished in 2000 (Arabidopsis Genome Initiative, 2000), a lot of data from cDNA microarray and oligonucleotide array analyses of various physiological events in Arabidopsis have been reported (for review: Donson et al., 2002, Aharoni and Vorst, 2001). However, little has been reported about the global analysis of organ-specific gene expression. MPSS (Massively Parallel Signature Sequencing) data were also reported for libraries from 5 organs (http://mpss.ucdavis.edu/).

I performed transcriptome analyses on characteristics of 7 organs, the young leaf, leaf, stem, bud, flower, silique and root of Arabidopsis, using cDNA macroarray system. The global relationship among the gene expression patterns of various organs and their physiological functions are discussed.

II-2. MATERIALS AND METHODS

II-2.1 Plant materials and growth conditions

For the preparations of leaf (mature rosette), stem, bud, flower and silique, *Arabidopsis thaliana* (accession Colombia) was grown on soil at 22˚C under continuous light for 2-3 months. To perform a set of macroarray experiments for one organ, 40-80 plants were sampled. To obtain young leaf and root, plants were grown on MS medium (Murashige and Skoog, 1962) plates containing 1% sucrose and 0.8% agarose under continuous light. The samples of the young leaf (about 6 leaves per plant except for cotyledons) and root were obtained from 17-day-old plants. For the young leaf and root, 300-600 plants were used for one array experiment.

II-2.2 cDNA macroarray design

The 13516 EST clones (Asamizu et al., 2000) were spotted onto three nylon filters (8 x 12cm) Biodyne A (PALL, U.S.A.) using a Biomek 2000 Laboratory Automation Workstation (Beckman Instruments, Ins, U.S.A.). The spotted cDNAs were fixed on the nylon filter by UV crosslinking. Lambda DNA and human transferrin receptor were also spotted as negative controls.

Assignment of EST clones to Arabidopsis loci was executed using BLAST and BLAT of NCBI. As a result, 13516 ESTs were assigned to 8809 loci. Namely, in many cases, multiple ESTs were assigned to the same locus, although these ESTs have been clustered as non-redundant EST clones (Asamizu et al., 2000).

II-2.3 RNA preparation and radiolabeling

Total RNA was extracted by the phenol/SDS method (Chirgwin et al., 1979). Purification of total RNA was performed using an RNeasy mini kit (Qiagen, U.S.A.) in general or using ultracentrifugation with caesium trifluoroacetate (Amersham, U.S.A.) for silique. Hybridization was performed according to Sasaki et al. (2001).

Hybridized membranes were exposed to an Imaging Plate (Fuji Film, Japan) for 2-3 days under a shield box made of lead in order to reduce the background generated by natural radiation. The use of this shield box significantly improved the raw images of hybridization. Radioactive images were obtained with a high-resolution scanner (Storm, Amersham, U.S.A.), and quantification of the signal

intensity was carried out using Array Vision software (Amersham, U.S.A.).

II-2.4 Evaluation of data quality

I used the signal to noise (SN) ratio to express data quality for the cDNA macroarray. The formula is:

$$SN_{membrane} = \log\left(\frac{\text{median}\{RAW_{ESTclones}\} - vBG_{membrane}}{\text{median}\{RAW_{Negative\ controls}\} - vBG_{membrane}}\right)$$

where $RAW_{ESTclones}$ is the set of raw data of EST clones, $RAW_{negative\ controls}$ is the set of raw data of negative controls. $vBG_{membrane}$ is the virtual background of the membrane (see next paragraph). Typical examples of SN are pictured in Fig. II-1A and II-1B. The SN values in my experiments are shown in the Fig. II-2.

The background of macroarrays using nylon membranes is generally higher than that of microarrays using glass plates. Subtraction of a real value (such as the signal intensity of a non-spotted area) as a membrane background from raw spot intensity leads to non-reliable values with many pseudo-positives. Therefore I used the virtual background (*vBG*). That is

$$vBG_{membrane} = 0.8 \times \min\{RAW_{all\ spots}\}$$

where $RAW_{allspots}$ is the set of raw data of all spots on the membrane. The coefficient 0.8 was optimized for my macroarray system by estimation of the fluctuation scale for weak signals. The SN values were used for normalization with respect to the quality of experiments.

II-2.5 Data normalization

In this study, I applied the LN normalizing method. The main point of this method is the non-parametric fitting of data to the lognormal distribution. The probability density function for the lognormal distribution (in base-10 logarithm) used here was

$$f(x) = \frac{10^x}{\sqrt{2\pi} \cdot 0.75 \cdot (10^x - 20)} \exp\left[\frac{\left(\log_{10}(10^x - 20) - 3\right)^2}{2 \cdot 0.75^2}\right]$$

In order to fit this lognormal distribution, I used Microsoft Excel with the following functions;

RANK = rank (*RAW*, all *RAW*s)

$LN_{fitting}$ = 10^(norminv (((*RANK* – 0.5) / N_{data}), 3, 0.75))+20

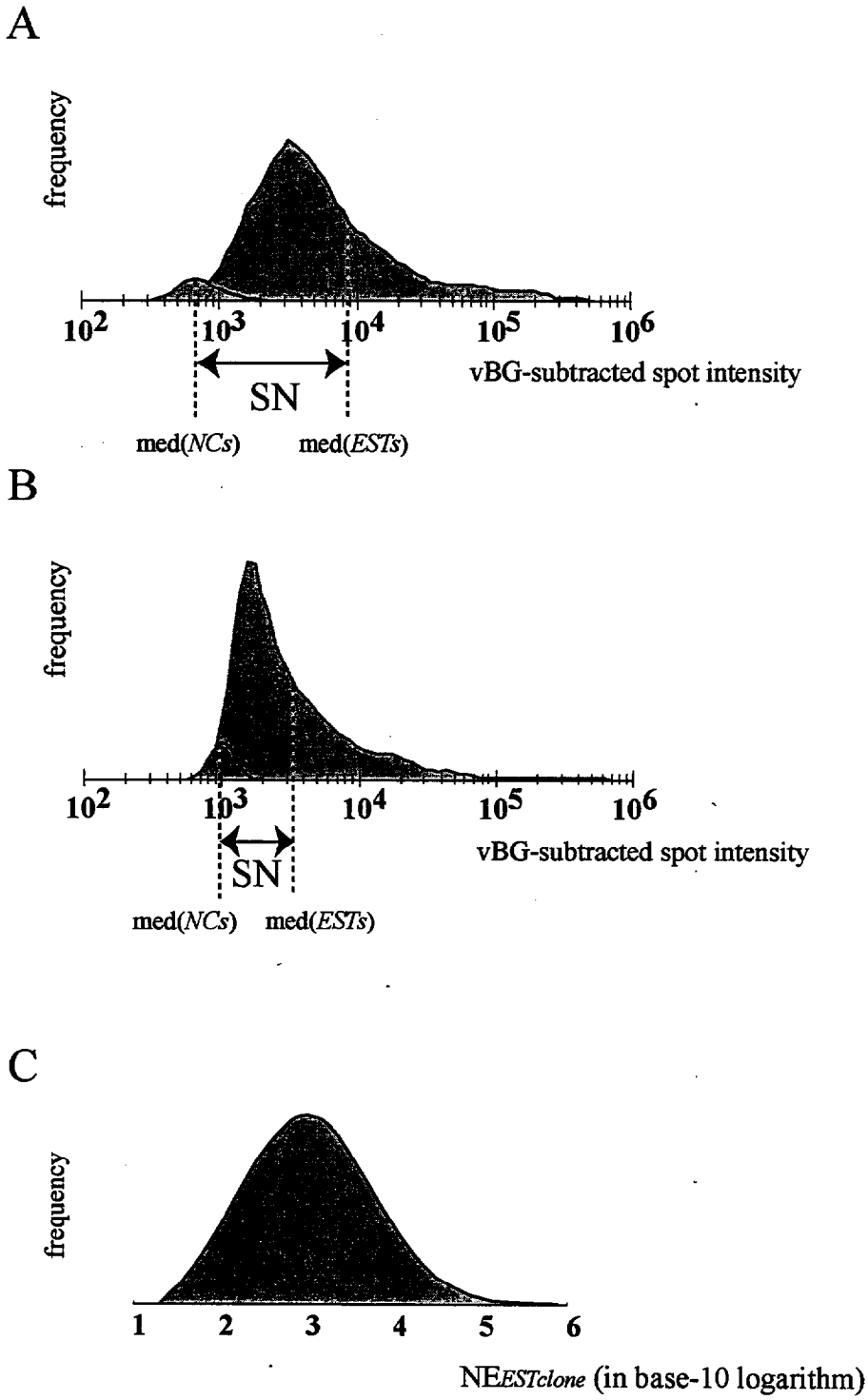


Figure II-1. Distribution of raw and normalized spot intensities. Distributions of raw data with high (A) and low (B) SN values are shown. Both of the raw data sets were normalized by the LN normalization method. The distribution of LN normalized data is presented in (C). SN, signal to noise. med(ESTs), median of signal intensity for EST clones. med(NCs), median of signal intensity for negative controls.

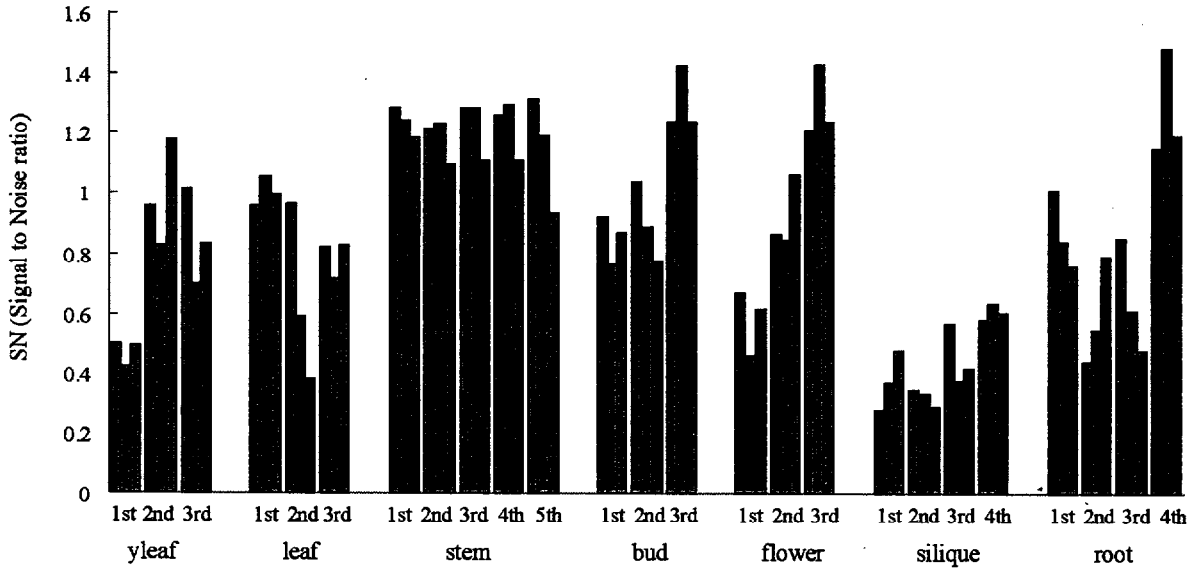


Figure II-2. SN values of macroarray data in this study. Since 13516 ESTs were spotted on 3 membranes, each organ has 3 SN values. Experiments were repeated 3 to 5 times for each organ.

$NE = \log10\ (LN_{fitting})$

Where *RAW* is the raw data of an EST clone, *RANK* is ranked data, N_{data} is the number of EST clones analyzed, $LN_{fitting}$ is the data fitted to the lognormal distribution, and NE is the normalized data. First, ranking from top to bottom was performed, and next, fitting to the lognormal distribution was done where the mean is 3 and the standard deviation is 0.75 (in base-10 logarithm) by emulating the real data distribution on my macroarray, and a small background value (20 in normal) was added to reduce excessive data fluctuation. Finally, transformation to logarithmic value was performed. I call the value obtained the "Normalized Expression (NE)". The distribution of the normalized spot intensities is shown in Fig. II-1C. The NE values reflect the absolute expression levels of genes.

The actual procedure was as follows. First, raw data were fitted to the lognormal distribution.

$$NE_{ESTclone,organ,repetition} = f_{LN}\left(RAW_{ESTclone,organ,repetition}\right)$$

where *RAW* is raw data, and f_{LN} is the function for fitting lognormal distribution mentioned above. The output is Normalized Expression of an EST clone. I usually performed cDNA macroarray analysis 3 to 5 times to obtain a set of data for an organ. The raw data of clone number 36 for leaf in the 2nd experiment, for example, is described as $RAW_{36,leaf,2}$, and the NE value is designated $NE_{36,leaf,2}$. Data with various level of quality were mixed in this repetition. To get the maximal potential information from these data, repeated data were first normalized by the LN normalization method, and then the set of the repeated data for the same tissue were combined by weighted average. The SN value mentioned above was used as the weight. The weighted average for repetition was fitted to the lognormal distribution again. The mathematical expression used was

$$NE_{ESTclone,organ} = f_{LN}\left(\frac{\sum_{repetition}\left(W_{organ,repetition} \times NE_{ESTclone,organ,repetition}\right)}{\sum_{repetition} W_{organ,repetition}}\right)$$

where *W* is weight. The output $NE_{ESTclone,organ}$ is the unified Normalized Expression for an EST clone. Then the data for EST clones were transformed to data for the corresponding locus.

$$NE_{gene,organ} = f_{locus}\left(NE_{ESTclone,organ}\right)$$

where *flocus* is a function for transformation from the EST clone to the corresponding locus. If a locus has only one EST clone, $NE_{gene,organ} = NE_{ESTclone,organ}$. However in my macroarray system, about 2000 loci had multi-EST clones. Thus, the representative value for the locus is calculated as the median of the values for the corresponding clones for each organ. The output is NE for a gene (locus). The NE value reflects the absolute expression level. The NE values are shown in several tables in this report.

For comparison of the expression profiles in 7 organs, the ratio of an expression value in a given organ to its average in 7 organs was calculated for each locus. Using the logarithm, the ratio corresponds to the difference as follows:

$$RE_{gene,organ} = NE_{gene,organ} - \frac{1}{N_{organ}} \sum_{organ} NE_{gene,organ}$$

I call this value the "Relative expression (RE)". Although NE is valuable for evaluating the absolute expression levels, RE is generally more useful for categorizing the expression profiles.

For evaluation of the organ specificity of gene groups, GRE (Group Relative Expression) was calculated as the average of the RE values of genes which belong to a gene group.

$$GRE_{group,organ} = \frac{1}{N_{groupgenes}} \sum_{groupgenes} RE_{gene,organ}$$

II-2.6 Classification of expression patterns

For classification of gene expression patterns, self-organizing map (SOM) was carried out using GeneCluster2 released by the Whitehead Institute (http://www-genome.wi.mit.edu/cancer/software/software.html). To perform SOM clustering, determination of an appropriate number of category was required. As a result of trials for altering the number from 5 to 100, we preliminarily found 8 outstanding expression patterns, in addition to weak organ-specific expression profiles. From the primary observation, we determined the cluster number to 9. Generally in the SOM clustering step, intrinsic groups that include larger numbers of genes tend to be divided into smaller categories. To prevent the appearance of less-significant categories mainly composed of genes with weak organ specificity, genes with strong organ specificity were given an appropriate weight in such a way that genes with weak organ specificity could be clustered into just one

category in addition to 8 major categories (18 was applied for 2755 genes which have REs more than 3). Consequently, 9 expected clusters (c1 - c9) were obtained. In repetition of clustering, some genes were unstably sorted to the 9 categories probably due to their distinct organ specificity from the 9 clusters. These genes were moved to another cluster, cx. Finally, the order of the 10 clusters obtained was partially changed for presentation. The adequacy of SOM categories was also shown by Principal Component Analysis of the 7 organs (Fig. II-3). The existence of each SOM category reflects the distances among organs in the figure. Principal Component Analysis was done using Black-Box made by Prof. Aoki in Gunma University (http://aoki2.si.gunma-u.ac.jp/BlackBox/BlackBox.html).

II-2.7 Public data used

The following files (and their release date) were downloaded from the TAIR ftp site (ftp://ftp.arabidopsis.org): Gene model (31, Jul, 2002), Gene Ontology (14, Apr, 2003), global analysis by TargetP (9, Oct, 2002), probes for GeneChip (23, Dec, 2002). Pathways data of KEGG (19, May, 2003) was downloaded from GenomeNet (ftp://ftp.genome.ad.jp/pub/kegg/). A gene list for Arabidopsis transcription factors was downloaded from AGRIS (http://arabidopsis.med.ohio-state.edu/).

II-3 RESULTS

II-3.1 Data processing and confirmation of cDNA macroarray data

To evaluate the quality of array data for 7 organs in Arabidopsis, I calculated SN (signal to noise ratio) values for each membrane (Fig. II-2). Figure II-1A and II-1B show examples of the distribution of the intensity in repeated experiments using macroarrays. The former is an example of data with high SN, and the latter is one with lower SN. The data from former experiment with the higher SN value are regarded as being more reliable. Each distribution was normalized by the LN (lognormal distribution fitting) normalization method as shown in Figure II-1C. The normalized data of the repeated experiments were then unified by calculating the weighted average using the SN values as the weight. The average Pearson's correlation coefficients for the repeated experiments was also calculated to be 0.85. Although the value is slightly small compared with that for repeated data produced using

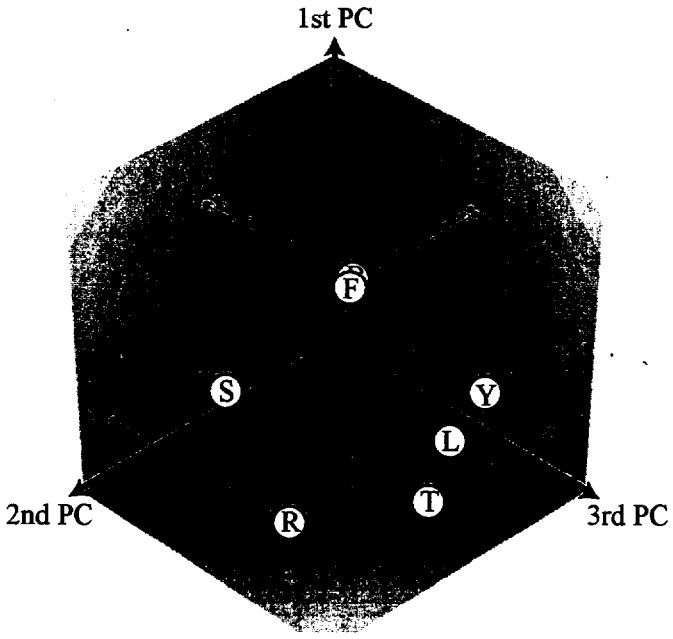


Figure II-3. Principal Component Analysis (PCA) of 7 organs. The 1st, 2nd and 3rd Principal Components (PC) are shown, which covered 29%, 23% and 19% of the total variance, respectively. Y, young leaf. L, leaf. T, stem. B, bud. F, flower. S, silique. R, root.

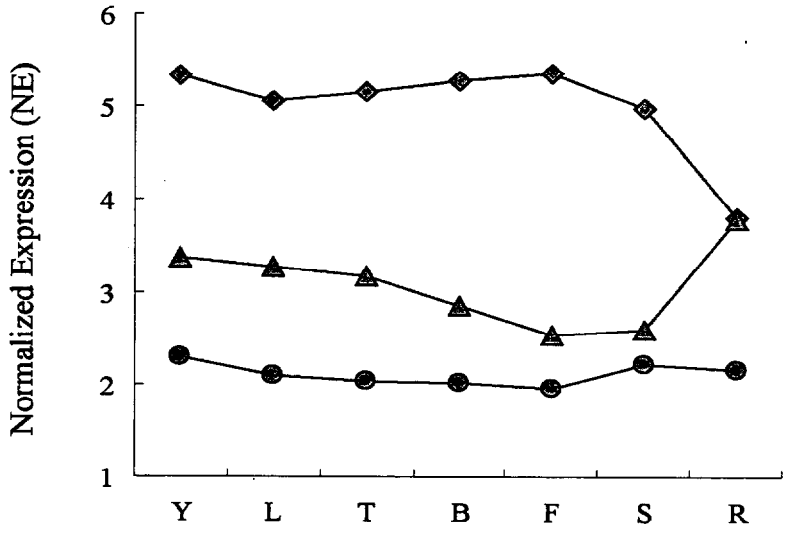
microarrays, the reliability of my data was well guaranteed by the performance of 3-5 experimental repetitions.

After subsequent data processing, "Normalized Expression (NE)" values, which reflect the absolute expression level of a gene as a base-10 logarithmic value, were obtained (see Materials and Methods). Examples of NE values are shown in Fig. II-4A. As shown in the figure, the Rubisco small subunit gene was expressed in the leaf almost 100 fold more strongly than in the root. However, it is also obvious that even in the root, the expression level was higher than that of ubiquitin-specific protease 2. The ratio of the NE value to the average of the NE values in the 7 organs was then calculated for each locus. I called the resultant value the "Relative Expression (RE)", which reflects the specificity of gene expression among organs (Fig. II-4B). To reveal distinctive feature of organs, I mainly used the RE values in following analyses.

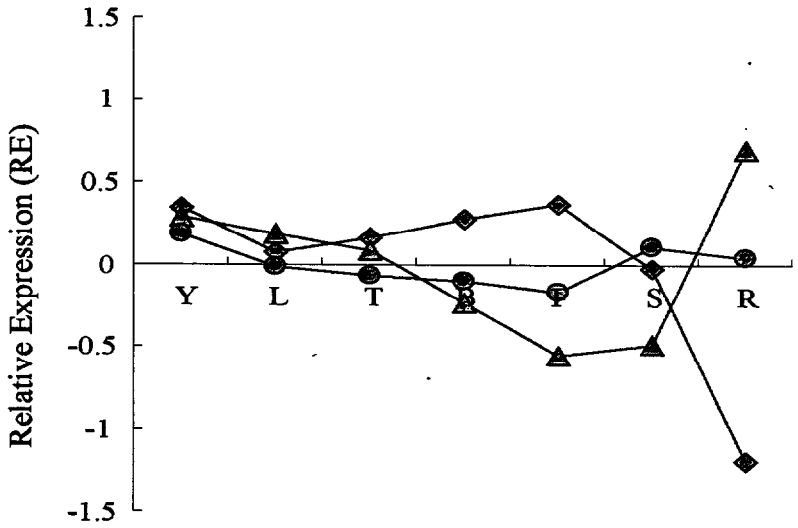
Northern blot analysis was performed using the same probe which was spotted on the macroarray membrane. After data quantification and background subtraction, RE was calculated for the Northern analysis. Figure II-5 shows the RE values from Northern blotting data and the corresponding macroarray data for 8 clones. The two lines for the RE values in each figure showed a very good correlation, indicating that the results of cDNA macroarray analysis agreed well with those of the Northern blot analysis. This correlation also indicates that data normalization could be successfully performed for the macroarray results.

Chen et al. (2002) demonstrated tissue-specific gene expression of 404 transcription factors using GeneChip. Here, I investigated the consistency of the results obtained by macroarray and GeneChip analyses. My macroarray and their GeneChip shared loci for 99 genes. Each set of data was transformed into the RE value. Then correlation between my data (7 organs) and their data (19 tissues) was analyzed for the 107 probes corresponding to 99 loci. Figure II-6 shows that my data for a particular organ had a good correlation to their data for the corresponding tissues. For example, the most correlative group of their tissues to stem in this study was stem, and the second-most was inflorescence. In the same way, my results for young leaf, leaf, bud, flower and root showed good correlation to the corresponding

A

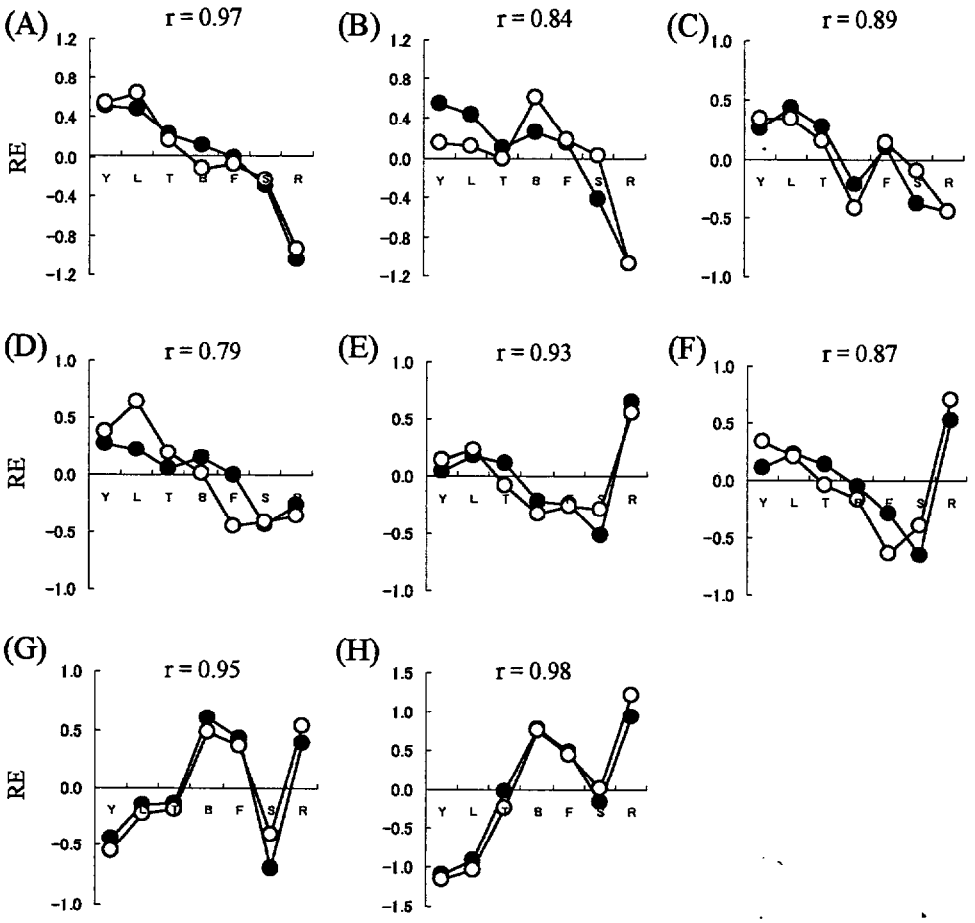


B



locus	description	Normalized Expression (NE) Y	L	T	B	F	S	R
—◆—	At5g38410 Rubisco small subunit	5.3	5.1	5.2	5.3	5.4	5.0	3.8
—▲—	At1g76680 12-oxophytodienoate reductase 1	3.4	3.3	3.2	2.8	2.5	2.6	3.8
—●—	At1g04860 ubiquitin-specific protease 2	2.3	2.1	2.0	2.0	2.0	2.2	2.2

Figure II-4. Examples of "Normalized Expression (NE)" (A) and "Relative Expression (RE)" (B) in 7 organs. NE and RE are base-10 logarithmic values. The key gives the locus descriptions and NE values. Y, young leaf. L, leaf. T, stem. B, bud. F, flower. S, silique. R, root.



	clone name	locus	description	TargetP	NE of macroarray Y	L	T	B	F	S	R	SOM
(A)	APD12a12	At3g14420	glycolate oxidase, putative	_2	4.4	4.5	4.0	3.7	3.7	3.6	2.9	c1
(B)	APD16h01	At1g70700	hypothetical protein	_1	3.5	3.5	3.3	4.0	3.5	3.4	2.3	c1
(C)	APD69a07	At5g14780	formate dehydrogenase	M4	3.6	3.6	3.4	2.8	3.4	3.2	2.8	c1
(D)	APZ27h12	At5g42650	allene oxide synthase	C2	3.4	3.7	3.2	3.0	2.6	2.6	2.7	c2
(E)	APZ27e09	At5g11670	NADP dependent malic enzyme - like	_1	3.8	3.9	3.6	3.4	3.4	3.4	4.2	c3
(F)	APD09b08	At1g76680	12-oxophytodienoate reductase 1	_5	3.4	3.3	3.0	2.9	2.4	2.7	3.7	c4
(G)	APD32f11	At3g16470	putative lectin	_2	2.8	3.1	3.2	3.9	3.7	3.0	3.9	c5
(H)	APZ12b10	At1g28290	prolin-rich protein, putative	S4	2.2	2.3	3.1	4.1	3.8	3.3	4.5	c5

Figure II-5. Agreement of macroarray (white circles) with Northern blot (black circles) analysis of 8 EST clones. Y-axis is RE value. Pearson's correlation coefficients (r) are also indicated. Y, young leaf. L, leaf. T, stem. B, bud. F, flower. S, silique. R, root. Information about the EST clones is shown. TargetP, prediction of subcellular localization; C, chloroplast. M, mitochondria. S, secretory. _, others. Number, prediction reliability. A smaller number means a more reliable prediction. RE, Relative Expression. NE, Normalized Expression. SOM, SOM category to which the gene belongs.

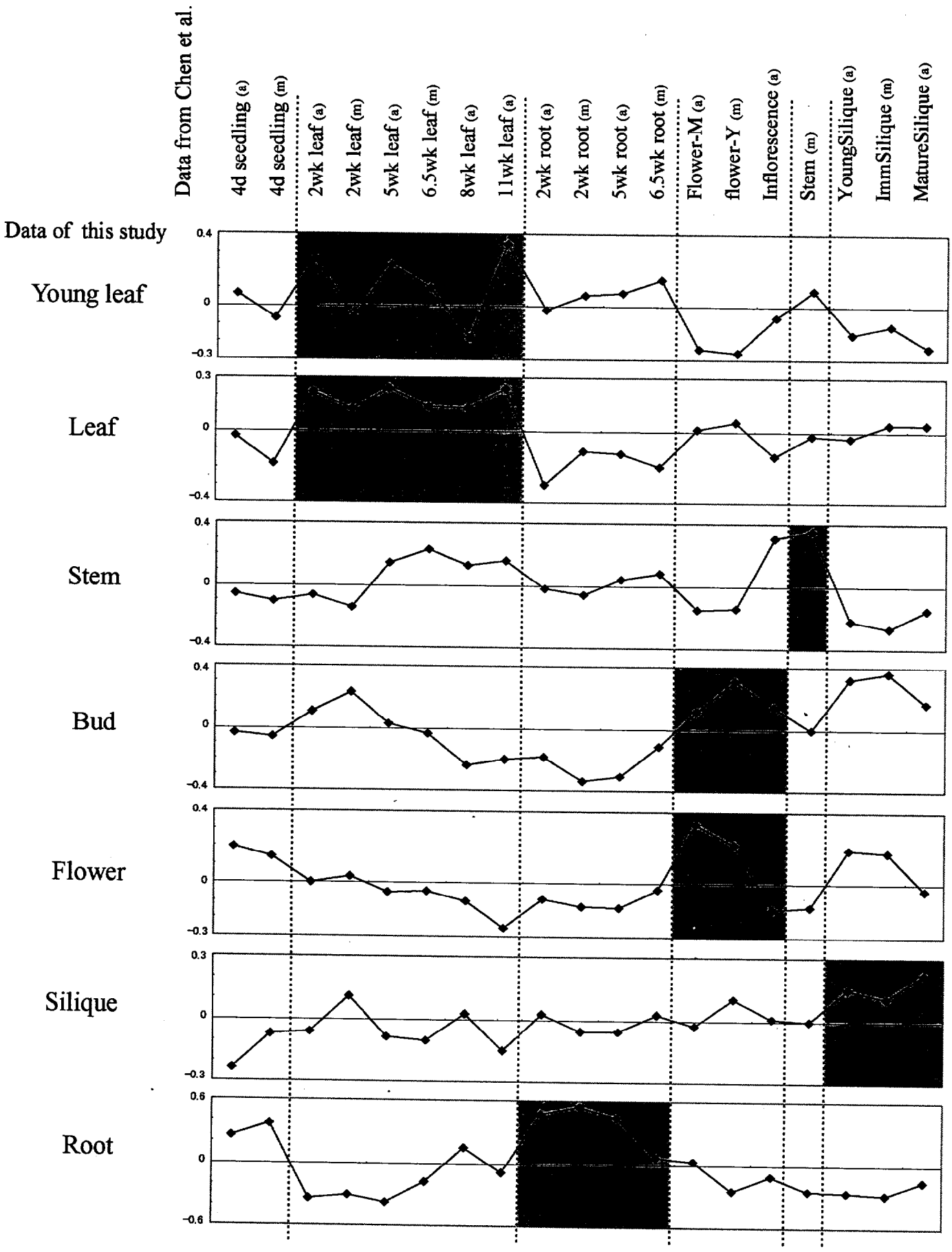


Figure II-6. Correlation of gene expression profiles between macroarray and GeneChip analyses. The Y-axis is Pearson's correlation coefficients. The most correlative combinations are emphasized by shading. In the data of Chen et al. (2002), (m) and (a) correspond to samples collected in the morning and afternoon, respectively.

data from GeneChip.

II-3.2 Organ specificity of genes

To characterize organs at the mRNA level, I first made a gene ranking of organ specificity. Table II-1 shows a list of the top 15 genes with the highest RE values in each organ. Genes for typical physiological functions of each organ were generally found in the lists. For instance, two nitrate reductase genes were found in the young leaf, two cellulose synthase genes in stem, genes for floral homeotic proteins such as AGAMOUS and PISTILLATA in bud and flower, genes for storage proteins (2S albumins and 12S cruciferins) in silique and three lectin genes in root. The occurrence of such characteristic members in each organ also ensured the reliability of my data.

However, two photosynthetic organs, young leaf and leaf, did not contain typical photosynthesis-related genes in the top 15 RE. In fact, photosynthesis-related genes are expressed not only in leaves but also in other green-colored organs such as stem and buds to some extent. Because RE is a value for organ specificity not for expression level, such genes with broad specificity were absent in these lists, even if their NE levels were high in some organs. On the other hand, 5 genes for defense-related proteins (two disease resistance proteins, myrosinase-associated protein, PR2, and ECS1) were found in leaf.

In bud and flower, glycine rich protein GRP17, glycosyl hydrolase family 17, anther development protein ATA20 and vegetative storage protein Vsp1 were found in the top 15 RE. They are known for their specific expression in the floral organs (Oliveira et al., 1993, Hird et al., 1993, Rubinelli et al., 1998, Berger et al., 1995). Several genes related to floral organ-specific metabolism, such as the terpene synthase, pectinesterase and polygalacturonase genes, were also seen in the bud and flower. Moreover, it is evident that genes related to lipid metabolism such as family II extracellular lipase and 4 lipid transfer protein genes were specifically expressed in these organs. This result indicates that particular lipid degradation mediated by the lipase and lipid transfer proteins might occur in flower development.

In silique, it is remarkable that all of the top five genes encode putative storage proteins, and

Table II-1. *List of the genes with the top 15 REs for each organ.*

TargetP, prediction of subcellur localization; C, chloroplast. M, mitochondria. S, secretory. _, others. Number, prediction reliability. A smaller number means a more reliable prediction. RE, Relative Expression. NE, Normalized Expression. SOM, SOM category to which the gene belongs. Y, young leaf. L, leaf. T, stem. B, bud. F, flower. S, silique. R, root.

organ	locus	description	TargetP	RE	NE							SOM
				Y	Y	L	T	B	F	S	R	
young leaf	At3g48360	unknown	_2	1.5	4.3	2.5	2.2	2.1	2.1	2.7	3.3	cx
	At5g63160	unknown	_2	1.2	4.0	2.5	2.7	2.4	2.7	2.2	3.2	c3
	At5g20250	glycosyl hydrolase family 36 (DIN10)	_4	1.1	4.4	2.9	2.7	3.4	3.4	3.3	3.3	c9
	At1g37130	nitrate reductase (NR2)	_5	1.0	4.5	4.2	3.1	2.8	2.8	2.8	4.1	c2
	At1g21120	O-methyltransferase 1, putative	_1	1.0	4.0	3.7	3.6	2.3	2.0	2.1	3.4	c2
	At1g76740	unknown	_3	1.0	4.4	3.9	2.7	2.4	2.7	3.6	4.4	c4
	At1g68520	CONSTANS B-box zinc finger family protein	M5	0.9	3.7	3.6	3.1	2.4	2.6	2.3	1.6	c1
	At1g21110	O-methyltransferase 1, putative	_1	0.9	3.8	3.7	3.5	1.6	1.6	2.3	3.4	c2
	At4g19530	disease resistance protein (TIR-NBS-LRR class)	_5	0.9	2.9	2.2	1.7	1.5	1.6	2.0	1.7	c2
	At1g77760	nitrate reductase 1 (NR1)	_5	0.9	4.0	3.0	2.5	2.3	2.9	2.6	4.0	c4
	At4g36850	unknown	_2	0.9	3.7	2.6	2.5	2.4	2.5	3.3	2.4	c8
	At2g41100	calmodulin-like protein (TCH3)	_1	0.9	4.6	4.2	4.0	3.5	3.5	2.6	3.2	c2
	At5g59250	D-xylose-H+ symporter - like protein	C1	0.9	2.9	2.1	1.9	1.9	1.8	2.0	1.8	cx
	At1g76140	prolyl endopeptidase, putative	_3	0.9	4.2	2.3	3.5	3.3	4.1	2.4	3.3	c3
	At1g21910	TINY-like protein	C1	0.8	3.1	2.0	2.3	2.0	1.8	2.5	2.3	c9
organ	locus	description	TargetP	RE	NE							SOM
				L	Y	L	T	B	F	S	R	
leaf	At5g46480	disease resistance protein (TIR class), putative	_2	1.2	3.3	3.9	2.9	2.2	2.2	2.1	2.2	c2
	At1g31580	ORF1, putative (ECS1)	S1	1.2	4.4	5.0	4.2	3.9	3.4	3.5	2.6	c1
	At3g57260	glycosyl hydrolase family 17 (BGL2, PR2)	S1	1.1	3.0	3.9	3.2	2.1	2.1	2.6	2.2	c2
	At4g02520	glutathione transferase	_4	1.1	4.2	4.5	4.1	2.3	2.2	2.7	3.9	c2
	At5g05600	leucoanthocyanidin dioxygenase-like protein	_3	1.1	3.7	4.3	3.2	3.1	3.0	2.9	2.6	c2
	At2g25510	unknown	M2	1.1	4.1	4.8	3.7	3.6	3.3	3.4	3.2	c2
	At4g33210	F-box protein family, AtFBL15	_2	1.1	2.6	3.4	2.2	1.9	2.1	2.2	2.0	c2
	At3g22231	unknown	C5	1.0	4.0	4.5	3.5	2.8	3.1	3.1	3.4	c2
	At3g50480	unknown	S5	1.0	3.4	3.8	3.1	2.6	2.1	2.9	2.0	c2
	At3g14210	myrosinase-associated protein, putative	S1	1.0	3.9	4.0	2.6	2.9	3.0	2.6	2.4	c1
	At2g32680	disease resistance protein family	S1	0.9	2.4	3.3	2.5	2.3	2.2	2.0	2.1	c2
	At5g42530	unknown	S1	0.9	4.2	4.7	3.9	3.5	3.3	3.5	3.4	c2
	At1g13110	cytochrome p450 family	S5	0.9	3.0	3.6	2.8	2.0	2.2	2.3	3.0	c2
	At5g23010	2-isopropylmalate synthase (IMS3)	C1	0.9	3.0	3.7	3.4	1.9	1.8	2.9	2.8	c2
	At1g49630	hydrogenase protein, putative	C3	0.9	3.1	3.7	3.1	2.4	2.5	1.9	2.7	c2
organ	locus	description	TargetP	RE	NE							SOM
				T	Y	L	T	B	F	S	R	
stem	At3g16920	glycosyl hydrolase family 19 (chitinase)	S1	1.5	2.3	1.8	4.3	2.4	2.2	3.6	2.7	c7
	At5g59320	lipid-transfer protein 3 (LTP3)	S1	1.4	2.5	2.5	5.2	4.3	4.9	4.0	3.1	c6
	At3g42180	exostosin family	_1	1.3	1.6	1.7	3.3	2.0	1.6	2.2	1.7	c7
	At5g17420	cellulose synthase catalytic subunit (IRX3)	_2	1.3	1.8	1.6	3.3	1.7	1.5	2.5	1.9	c7
	At5g60490	fasciclin-like arabinogalactan-protein (FLA12)	S2	1.3	2.1	2.4	4.0	2.5	2.6	3.4	2.3	c7
	At4g18780	cellulose synthase - like protein	_2	1.1	2.4	2.0	3.8	2.3	2.3	3.2	2.5	c7
	At2g38080	laccase (diphenol oxidase), putative	S1	1.1	1.9	2.0	3.6	2.7	2.0	3.1	2.4	c7
	At1g20850	cysteine proteinase XCP2	S1	1.0	1.8	1.5	3.3	2.1	2.3	2.1	3.2	c5
	At1g61810	glycosyl hydrolase family 1	S1	1.0	2.3	2.1	3.4	2.3	2.1	2.3	2.6	c7
	At5g48560	unknown	_2	1.0	2.1	1.8	3.3	1.7	1.9	2.4	3.2	c4
	At2g39270	similar to putative adenylate kinase	M5	0.9	1.9	2.3	3.3	2.4	2.0	2.6	2.2	c7
	At1g52150	HD-Zip transcription factor	_2	0.9	2.7	1.6	3.4	2.5	2.4	2.2	2.7	c3
	At5g63580	flavonol synthase	_3	0.9	1.6	1.9	3.3	2.4	2.6	2.7	2.4	c7
	At2g29130	laccase (diphenol oxidase), putative	S1	0.9	2.0	2.4	3.5	2.4	1.9	3.4	2.4	c7
	At5g25610	dehydration-induced protein RD22	S1	0.9	2.7	3.0	4.0	3.2	3.4	3.1	2.6	c7

Table II-1. (continued)

organ	locus	description	TargetP	RE	NE							SOM
				B	Y	L	T	B	F	S	R	
bud	At5g07530	glycine-rich protein GRP17	C5	1.6	2.6	2.4	2.1	4.3	2.4	2.4	2.4	c6
	At3g12145	pseudogene, polygalacturonase inhibitor		1.5	2.5	2.3	2.4	4.7	4.3	3.2	2.7	c6
	At5g22430	unknown	S1	1.5	2.5	2.4	2.2	4.3	3.0	2.3	2.7	c6
	At2g20870	unknown	S1	1.4	2.0	2.2	1.7	3.7	2.4	2.4	2.0	c6
	At1g75910	family II extracellular lipase 4 (EXL4)	S1	1.4	2.3	1.9	1.9	3.9	3.5	2.2	2.1	c6
	At5g33370	unknown	S1	1.3	2.5	2.3	2.5	4.1	3.1	2.6	2.1	c6
	At4g14080	glycosyl hydrolase family 17 (anther-specific protein, A6)	S4	1.3	2.1	1.9	1.9	3.6	2.0	2.4	1.9	c6
	At5g24780	vegetative storage protein Vsp1	S1	1.3	3.3	3.3	3.6	5.4	5.4	4.0	3.7	c6
	At1g66850	lipid transfer protein, putative	S1	1.3	2.2	2.9	2.1	4.0	2.5	2.9	2.4	c6
	At3g15400	anther development protein, ATA20	S1	1.2	2.4	2.3	2.2	3.7	2.4	2.4	2.1	c6
	At4g16590	cellulose synthase like protein	_5	1.2	2.0	2.2	2.0	3.7	2.9	2.7	2.0	c6
	At1g52030	myrosinase binding protein (MBP1.2)	_2	1.2	2.6	2.7	2.8	4.3	4.2	2.8	2.5	c6
	At3g14040	exopolygalacturonase	S1	1.1	2.0	2.2	2.0	3.7	3.5	2.5	1.9	c6
	At5g20240	MADS-box protein (PISTILLATA)	_1	1.1	1.9	2.1	2.0	3.7	3.5	2.4	2.3	c6
	At4g18960	floral homeotic protein agamous (AGAMOUS)	S5	1.1	2.0	1.8	1.7	3.3	3.0	2.2	1.5	c6
organ	locus	description	TargetP	RE	NE							SOM
				F	Y	L	T	B	F	S	R	
flower	At2g36190	glycosyl hydrolase family 32	S1	1.4	2.0	2.1	1.9	2.6	3.7	2.1	1.8	c6
	At5g24780	vegetative storage protein Vsp1	S1	1.3	3.3	3.3	3.6	5.4	5.4	4.0	3.7	c6
	At4g12870	unknown	S1	1.2	2.3	2.5	2.0	3.5	4.2	3.7	2.2	c6
	At3g12145	pseudogene, polygalacturonase inhibitor		1.1	2.5	2.3	2.4	4.7	4.3	3.2	2.7	c6
	At5g59320	lipid-transfer protein (LTP3)	S1	1.1	2.5	2.5	5.2	4.3	4.9	4.0	3.1	c6
	At2g38530	lipid-transfer protein (LTP2)	S1	1.1	3.4	2.8	3.5	4.4	4.8	3.7	3.1	c6
	At2g47040	pectinesterase family	S1	1.1	2.1	2.1	2.2	3.3	3.7	2.4	2.0	c6
	At5g44630	terpene synthase/cyclase family	_2	1.1	2.1	2.1	1.7	3.4	3.6	2.7	2.0	c6
	At5g20720	chloroplast Cpn21 protein	C2	1.1	2.4	2.3	2.5	3.6	3.8	2.8	1.8	c6
	At1g52030	myrosinase binding protein (MBP1.2)	_2	1.1	2.6	2.7	2.8	4.3	4.2	2.8	2.5	c6
	At3g08770	lipid-transfer protein (LTP6)	S1	1.1	2.8	2.5	2.5	4.2	4.3	3.8	2.8	c6
	At4g19430	unknown	C3	1.1	1.8	2.5	2.1	3.3	3.7	2.9	2.1	c6
	At1g02205	unknown	_3	1.0	2.4	2.0	3.7	4.0	4.1	3.3	1.9	c6
	At5g48140	polygalacturonase, putative	S1	1.0	2.3	2.3	2.0	2.8	3.4	1.8	2.1	c6
	At5g09730	glycosyl hydrolase family 3	S1	1.0	2.7	2.8	2.4	3.5	3.9	2.8	2.3	c6
organ	locus	description	TargetP	RE	NE							SOM
				S	Y	L	T	B	F	S	R	
silique	At4g27150	NWMU2 - 2S albumin 2 precursor	S1	2.7	1.9	2.2	2.0	1.8	2.1	5.1	1.8	c8
	At4g27140	NWMU1 - 2S albumin 1 precursor	S1	2.6	2.5	1.9	2.2	2.2	2.1	5.2	1.9	c8
	At5g44120	legumin-like protein (CRA1)	S1	2.1	1.9	1.9	2.0	1.8	1.9	4.3	1.6	c8
	At4g28520	12S cruciferin seed storage protein	S1	2.1	2.4	2.4	2.3	2.7	2.4	4.9	2.2	c8
	At1g03880	putative cruciferin 12S seed storage protein	S1	1.8	2.2	2.4	2.1	2.2	2.2	4.3	2.0	c8
	At5g40420	oleosin	_3	1.7	2.4	1.9	1.9	2.0	2.1	4.0	1.8	c8
	At4g25140	oleosin	_2	1.6	2.1	2.6	2.4	2.4	2.6	4.3	2.4	c8
	At1g13450	DNA binding protein GT-1, putative	_2	1.6	2.1	2.4	2.2	1.8	1.9	4.0	2.2	c8
	At3g22640	unknown	S1	1.4	2.3	1.8	1.6	2.0	1.7	3.5	1.8	c8
	At4g36700	globulin-like protein	S1	1.4	1.9	2.2	1.9	1.9	2.0	3.6	2.3	c8
	At5g59170	cell wall protein	S1	1.4	2.1	2.0	2.4	1.8	1.9	3.7	2.8	c7
	At1g71691	unknown	S2	1.3	1.7	2.0	1.8	1.9	2.4	3.4	1.6	c8
	At3g09740	syntaxin (SYP71)	_1	1.3	2.2	2.6	2.1	2.4	2.9	3.9	2.2	c8
	At5g26910	unknown	M4	1.3	1.8	1.9	2.4	2.0	2.0	3.5	2.0	c7
	At3g62730	unknown	S1	1.2	1.9	1.6	2.1	1.8	1.8	3.3	1.7	cx
organ	locus	description	TargetP	RE	NE							SOM
				R	Y	L	T	B	F	S	R	
root	At1g66270	glycosyl hydrolase family 1, beta-glucosidase	S1	2.2	2.4	2.4	2.0	3.1	3.3	2.6	5.2	c4
	At3g09260	glycosyl hydrolase family 1	S1	2.2	3.0	3.0	2.6	3.3	3.4	3.2	5.6	c4
	At3g16450	putative lectin	_2	2.1	1.9	1.9	2.0	1.8	1.9	2.0	4.4	c4
	At3g16460	putative lectin	_3	1.9	2.7	2.1	1.8	1.9	1.7	2.5	4.4	c4
	At1g13940	unknown	_2	1.9	2.3	1.9	2.9	2.2	2.1	1.9	4.4	c4
	At4g37410	cytochrome p450, putative	S4	1.9	1.9	2.3	2.1	2.2	2.1	2.0	4.3	c4
	At3g16430	putative lectin	_1	1.8	2.8	2.4	2.4	1.9	2.1	2.8	4.5	c4
	At5g14020	unknown	_4	1.8	2.0	2.0	2.3	2.3	1.8	2.0	4.1	c4
	At4g24100	unknown	_4	1.8	3.2	3.3	3.5	3.5	3.4	3.9	5.6	c4
	At3g28550	unknown	S3	1.7	3.0	2.8	2.9	2.6	2.8	2.8	4.8	c4
	At5g18490	unknown	_4	1.7	2.8	2.8	3.3	2.0	2.5	2.3	4.6	c4
	At4g28480	heat-shock protein	_2	1.6	3.0	2.7	2.8	2.4	2.4	2.1	4.5	c4
	At1g66280	glycosyl hydrolase family 1	S1	1.6	2.9	3.2	3.2	3.0	3.0	2.8	4.9	c4
	At4g11320	cysteine proteinase	S1	1.6	2.2	2.2	2.5	3.6	3.4	2.8	4.7	c5
	At1g70320	ubiquitin-protein ligase 2 (UPL2)	M5	1.6	2.3	1.9	2.0	2.4	2.2	2.5	4.1	c4

the next two genes correspond to oleosin, a major component of the lipid body. In particular, two 2S albumin precursor genes showed very high expression, whose NE values were larger than 5.

Root highly expressed the glycosyl hydrolase 1 families. One glycosyl hydrolase, At3g09260, is a β-glucosidase recently identified as a major protein of the ER body, a compartment newly found by Matsushima et al. (2003). Table II-1 indicates that the mRNA is extraordinary accumulated in root. In fact, it was reported that the ER body was frequently observed in root tissues (Matsushima et al., 2003). It is possible that the other two members of the glycosyl hydrolase 1 family expressed in root are also ER body proteins. These three genes were also induced by exogenous application of methyl jasmonate (Sasaki-Sekimoto, 2003), which is in agreement with the report showing that the ER body is induced by methyl jasmonate treatment (Matsushima et al., 2003). Although the physiological role of the ER body is still unknown, such high expression in the root probably reflects the importance of the gene family in the root.

II-3.3 Organ specificity of gene groups

To identify major physiological events which were transcriptionally and systemically regulated in each organ, I tried to find which groups of genes are specifically expressed in an organ. Gene groups composed of more than 10 of the 8809 genes were selected from several databases. The groups obtained were as follows: 194 gene groups from GO (Gene Ontology, Gene Ontology Consortium, 2000) defined by TAIR (The Arabidopsis Information Resource, Rhee et al., 2003), 30 gene groups from AraCyc (*Arabidopsis thaliana* Biochemical Pathways) also defined by TAIR, and 67 gene groups from KEGG (Kyoto Encyclopedia of Genes and Genomes, Kanehisa et al., 2002). To compare the organ specificity of these gene groups, the GRE (Group Relative Expression) value was defined as the average of the RE values for all genes which belong to a gene group. Then the GRE values of various gene groups were compared in each organ. The groups particularly up-regulated in the 7 organs are listed in Table II-2. In this comparison, photosynthesis-related gene groups, such as "chlorophyll binding activity" and "reductive pentose-phosphate cycle" in GO, were ranked higher in position in the young leaf and leaf. "Nitrogen metabolism" in KEGG was also found in the young leaf. Gene groups related to the cell wall,

Table II-2. *Gene groups (GO, AraCyc, KEGG) specifically expressed in Arabidopsis organs.*
Gene groups whose GRE value were more than 0.20 (about 1.6-fold in normal value) are shown. Almost identical gene groups sharing with more than 80% of genes in the smallest category were represented by one category.

organ	GRE	gene group ID	description	# of gene in macroarray	# of gene in Arabidopsis
young leaf	0.40	GO:0016168	chlorophyll binding activity	14	19
	0.23	ath00910	nitrogen metabolism	26	29
	0.23	GO:0009570	chloroplast stroma	13	15
	0.22	GO:0019253	reductive pentose-phosphate cycle	22	30
leaf	0.39	GO:0016168	chlorophyll binding activity	14	19
stem	0.29	GO:0007155	cell adhesion	15	27
	0.22	AraCyc	lignin biosynthesis	12	19
bud	0.40	GO:0008289	lipid binding activity	13	20
	0.26	GO:0005718	nucleosome	28	60
	0.25	GO:0005874	microtubule	17	32
	0.24	GO:0000074	regulation of cell cycle	11	35
	0.20	GO:0009570	chloroplast stroma	13	15
	0.20	GO:0006633	fatty acid biosynthesis	15	36
flower	0.25	GO:0009341	beta-galactosidase complex	11	21
	0.21	GO:0005874	microtubule	17	32
	0.21	GO:0005718	nucleosome	28	60
silique	0.23	GO:0005529	sugar binding activity	15	62
	0.22	GO:0005576	extracellular	16	64
	0.20	GO:0004194	pepsin A activity	26	63
root	0.51	ath00360	phenylalanine metabolism	51	91
	0.32	ath00272	cysteine metabolism	18	27
	0.31	AraCyc	lignin biosynthesis	12	19
	0.26	ath00271	methionine metabolism	14	20
	0.25	ath00480	glutathione metabolism	21	40
	0.24	AraCyc	TCA cycle, aerobic respiration	18	22
	0.24	GO:0015250	water channel activity	21	35
	0.23	GO:0005829	cytosol	12	20
	0.21	ath00410	β-alanine metabolism	11	20
	0.21	GO:0006520	amino acid metabolism	21	40
	0.21	ath00920	sulfur metabolism	14	21
	0.20	ath00350	tyrosine metabolism	27	37

which gives mechanical strength to the plant ("cell adhesion" in GO and "lignin biosynthesis" in AraCyc), were generally up-regulated in the stem. Interestingly, gene groups associated with sulfur metabolism, including "cysteine metabolism", "methionine metabolism" and "glutathione metabolism", were all ranked in a higher position in the root. The ranking of the groups for amino acid metabolism was also high in the root. This is probably due to the fact that root has important roles in the uptake of inorganic nutrients and their assimilation into organic compounds.

II-3.4 Clustering of organ-specific gene expression profiles

In order to investigate the relationships among the 7 organs, I did clustering for organ-specific expression patterns for all 8809 genes using Self Organizing Map (SOM, Tamayo et al., 1999). The 8809 loci were clustered into 10 categories from cluster 1 (c1) to cluster x (cx). In Fig. II-7, each plot in the 10 graphs shows the GRE values for all genes categorized in each cluster. The small dots represent the SDs. The first line in the table just below the 10 graphs shows the distribution of the 8809 loci in the 10 different clusters. The biggest cluster was c9 (3451 loci), which showed constitutive expression in all 7 organs. Almost 40% of the total genes in Arabidopsis belonged to this category. On the other hand, c4 (266 loci) and c6 (265 loci) were rather small categories showing specific expression in root (c4) or in bud and flower (c6). The lower lines in the table in Fig. II-7 indicate the distribution of the top 100 genes with the highest RE values in each organ. It appears that the top 100 genes in young leaf were mainly categorized into c1, those for leaf into c2, stem into c2 and c7, bud: c6, flower: c6, silique: c8 and root: c4. Most of the top 100 genes in leaf, bud, flower, silique and root belonged to one SOM category. On the other hand, two organs, young leaf and stem, had various expression patterns. There were two major expression patterns for highly expressed genes in stem. A major group categorized to c2 showed relatively high expression in young leaf and leaf, in addition to stem. This cluster is mainly composed of genes related to defense responses or secondary metabolism (Table II-3 c2). Another category, c7, was particularly expressed in both stem and silique. Interestingly, the members of latter cluster were completely different from the members of the former. Genes involved in cell wall and lignin biosynthesis, such as cellulose synthases, fasciclin-like arabinogalactan proteins and cinnamyl alcohol dehydrogenase,

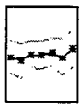
	c1	c2	c3	c4	c5	c6	c7	c8	c9	cx
										
all 8809 loci	581	456	1074	266	888	265	697	1005	3451	126
top 100 RE in Young leaf	[redacted]	25	10	7	0	0	0	2	17	3
top 100 RE in Leaf	25	[redacted]	0	2	0	0	0	3	1	5
top 100 RE in Stem	2	[redacted]	20	3	4	5	26	0	0	9
top 100 RE in Bud	5	0	0	0	7	[redacted]	0	1	3	1
top 100 RE in Flower	4	0	4	0	7	[redacted]	0	1	0	3
top 100 RE in Silique	0	0	0	0	0	2	30	[redacted]	0	3
top 100 RE in Root	0	1	0	[redacted]	4	0	0	0	0	0

Figure II-7. Gene groups classified by pattern of RE using 1D-SOM. Dark-shaded boxes with bold letters show the largest-sized clusters. The upper figures show the average of RE for each gene category. The two gray lines in each figure show the SD values for the expression of each gene group.

Table II-3. *Genes of SOM categories c2 and c7 in top 100 REs of stem.*
TargetP, prediction of subcellur localization; C, chloroplast. M, mitochondria. S, secretory. _, others. Number, prediction reliability. A smaller number means a more reliable prediction.

SOM	locus	TargetP	description
c2	At5g23010	C1	2-isopropylmalate synthase-like (IMS3)
	At5g13200	_2	ABA-responsive protein - like
	At5g19140	C3	aluminium-induced protein - like
	At1g62810	S1	amine oxidase, putative
	At5g09220	_4	amino acid transporter AAP2
	At5g48930	_3	anthranilate N-benzoyltransferase
	At3g19710	_3	branched-chain amino acid aminotransferase, putative
	At5g39670	_5	calcium-binding protein - like
	At4g35310	_2	calmodulin-domain protein kinase CDPK isoform 5 (CPK5)
	At1g20620	_3	catalase 3
	At4g31500	S5	cytochrome p450 (CYP83B1)
	At1g16410	_2	cytochrome p450, putative
	At5g41740	C4	disease resistance protein (TIR-NBS-LRR class)
	At4g02520	_4	glutathione transferase
	At2g21260	_3	mannose 6-phosphate reductase (NADPH-dependent)
	At1g18570	_2	myb family transcription factor (MYB51)
	At3g11280	_4	myb family transcription factor, putative
	At1g21110	_1	O-methyltransferase 1, putative
	At1g21120	_1	O-methyltransferase 1, putative
	At1g10070	M4	tat-binding protein, putative
	At2g01060	_4	transfactor-like protein
	At1g13640	_4	ubiquitin, putative
	At1g21250	S1	wall-associated kinase 1 (WAK1)
	-	-	8 unknown genes
c7	At2g39270	M5	adenylate kinase, putative
	At3g28270	_3	integrin, putative (At14a)
	At4g18780	_2	cellulose synthase, putative
	At5g17420	_2	cellulose synthase catalytic subunit (IRX3)
	At4g34230	_3	cinnamyl alcohol dehydrogenase, putative
	At3g42180	_1	exostosin family
	At5g60490	S2	fasciclin-like arabinogalactan-protein (FLA12)
	At1g61810	S1	glycosyl hydrolase family 1
	At3g16920	S1	glycosyl hydrolase family 19 (chitinase)
	At2g29130	S1	laccase (diphenol oxidase), putative
	At2g38080	S1	laccase (diphenol oxidase), putative
	At5g21930	C3	metal-transporting P-type ATPase, putative
	At5g60660	_2	aquaporin
	At5g38000	_3	oxidoreductase-like protein
	At2g43050	S5	pectinesterase family
	At5g15630	_3	phytochelatin synthetase, putative
	At1g11080	S2	serine carboxypeptidase isolog
	-	-	9 unknown genes

were commonly found in the latter cluster (Table II-3 c7). From this point of view, it is possible that two Myb-family transcription factors, At1g18570 and At3g11280, are candidates for factors which regulate some metabolism operated by genes found in this c2 category of stem. The results clearly demonstrated that, although the relative intensity of expression in each organ (RE value) is important information by itself, clustering by SOM could provide much more information for characterizing their gene functions.

As described above, the SOM categories were suggested to have some relations to particular functions. I further analyzed gene groups whose members were mainly categorized in one SOM cluster with a high population. The top 5 gene groups that had the largest population in each cluster are listed in Fig. II-8. Generally, gene groups related to photosynthesis were present in c1 but not in c2, showing that c1 is a major cluster for photosynthesis-related genes. In contrast, genes related to defense responses were mainly clustered in c2, which may be genes for the basal response to some stresses such as light, although the plants were grown under medium light conditions. Many gene groups related to amino acid synthesis were found in c3, probably because I used mature siliques in these experiments, and therefore primary metabolism such as amino acid synthesis was relatively repressed in this organ. Lipase and lipid binding activity, including lipid transfer proteins, were in c6. Cellulose synthase and lignin synthesis were in c7. Unexpectedly, significant organ specificity in the gene groups of "ribosome" and "protein biosynthesis" was found (c5) which were specifically expressed in bud, flower and root; nevertheless, these genes are generally believed to be genes constitutively expressed in all organs.

II-3.5 Organ specificities of transcription factors and plastid-targeting proteins

Analysis of SOM clustering for various transcription factors revealed that they were mainly classified into c9, revealing constitutive expression of this gene family (Table II-4, left column). However, several transcription factors were found in other clusters, including MADS box genes, which showed stronger expression in floral organs. Table II-4 lists the transcription factor genes clustered into c4 and c6. As shown in the table, transcription factors involved in flower development, such as AGAMOUS, AP3, AGL2, AGL5, PISTILLATA (Steven et al., 1995, Honma and Goto, 2001, Theißen and Saedler, 2001), were all classified into category c6. Some transcription factors whose function was

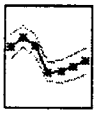
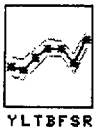
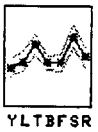
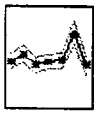
c1 (7% of total genes)  YLTBFSR	groupID	description	% of c1	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	number of genes in cluster	number of genes in Arabidopsis
	GO:0016168	chlorophyll binding activity	100%	14	0	0	0	0	0	0	0	0	0	14	19
	GO:0009570	chloroplast stroma	46%	6	0	1	0	1	0	1	1	3	0	13	15
	GO:0019253	reductive pentose-phosphate cycle	45%	10	2	1	0	2	0	0	1	6	0	22	30
	AraCyc	biosynthesis of chlorophyll	44%	8	0	1	0	0	0	0	2	7	0	18	26
	ath00195	photosynthesis	37%	16	0	5	0	4	0	1	0	17	0	43	89
c2 (5% of total genes)  YLTBFSR	groupID	description	% of c2	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	number of genes in cluster	number of genes in Arabidopsis
	GO:0006952	defense response	29%	4	10	3	0	1	2	2	5	6	2	35	151
	AraCyc	tryptophan biosynthesis	21%	0	3	4	0	0	0	1	0	6	0	14	25
	ath00480	glutathione metabolism	19%	1	4	7	1	3	0	1	0	4	0	21	40
	GO:0006725	aromatic compound metabolism	19%	3	3	1	1	1	1	1	2	3	0	16	34
	GO:0015144	carbohydrate transporter activity	18%	1	4	2	0	1	2	2	2	7	1	22	54
c3 (12% of total genes)  YLTBFSR	groupID	description	% of c3	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	number of genes in cluster	number of genes in Arabidopsis
	ath00410	beta-alanine metabolism	55%	0	0	6	0	2	0	1	1	1	0	11	20
	ath00450	selenoamino acid metabolism	47%	0	2	9	0	2	0	0	1	4	1	19	26
	ath00272	cysteine metabolism	44%	0	0	8	1	4	0	0	0	4	1	18	27
	ath00640	propanoate metabolism	44%	0	0	7	0	3	0	1	1	4	0	16	30
	ath00271	methionine metabolism	43%	0	0	6	0	4	1	0	0	3	0	14	20
c4 (3% of total genes)  YLTBFSR	groupID	description	% of c4	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	number of genes in cluster	number of genes in Arabidopsis
	ath00360	phenylalanine metabolism	39%	2	3	3	20	6	2	5	4	4	2	51	91
	ath00350	tyrosine metabolism	15%	2	3	4	4	3	1	1	1	6	2	27	37
	GO:0005507	copper ion binding activity	14%	0	3	1	3	2	2	3	2	6	0	22	55
	GO:0003779	actin binding activity	13%	0	0	4	3	4	1	3	2	5	1	23	49
	ath00910	nitrogen metabolism	12%	3	1	4	3	1	0	3	0	8	3	26	29
c5 (10% of total genes)  YLTBFSR	groupID	description	% of c5	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	number of genes in cluster	number of genes in Arabidopsis
	GO:0005874	microtubule	76%	0	0	2	0	13	0	0	0	2	0	17	32
	GO:0005840	ribosome	56%	23	1	13	0	96	2	0	7	27	3	172	296
	GO:0000074	regulation of cell cycle	55%	0	0	2	0	6	1	0	1	1	0	11	35
	GO:0006414	translational elongation	44%	3	0	6	0	15	0	1	1	8	0	34	53
	GO:0008248	pre-mRNA splicing factor activity	42%	0	0	3	0	5	0	0	0	4	0	12	30
c6 (3% of total genes)  YLTBFSR	groupID	description	% of c6	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	number of genes in cluster	number of genes in Arabidopsis
	GO:0016298	lipase activity	33%	2	1	1	2	0	6	2	3	1	0	18	45
	GO:0008289	lipid binding activity	31%	0	0	2	0	1	4	0	3	3	0	13	20
	GO:0004185	serine carboxypeptidase activity	20%	1	1	2	0	4	4	1	3	4	0	20	53
	GO:0008415	acyltransferase activity	18%	0	0	3	0	4	4	1	1	9	0	22	36
	GO:0005975	carbohydrate metabolism	15%	11	5	7	6	14	23	12	27	48	2	155	370
c7 (8% of total genes)  YLTBFSR	groupID	description	% of c7	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	number of genes in cluster	number of genes in Arabidopsis
	AraCyc	lignin biosynthesis	33%	0	0	0	2	2	1	4	0	2	1	12	19
	GO:0016829	lyase activity	29%	0	0	2	0	2	1	6	2	6	2	21	53
	GO:0007155	cell adhesion	27%	0	0	2	0	3	1	4	1	3	1	15	27
	GO:0016759	cellulose synthase activity	26%	0	1	1	0	1	2	5	0	9	0	19	39
	AraCyc	trehalose anabolism	25%	1	0	1	0	2	1	3	2	2	0	12	24
c8 (11% of total genes)  YLTBFSR	groupID	description	% of c8	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	number of genes in cluster	number of genes in Arabidopsis
	GO:0006857	oligopeptide transport	30%	1	1	1	2	1	2	3	7	4	1	23	49
	GO:0006629	lipid metabolism	27%	1	3	1	1	0	0	2	6	8	0	22	65
	GO:0006457	protein folding	27%	1	0	1	0	0	0	0	3	6	0	11	28
	GO:0005529	sugar binding activity	27%	0	1	2	0	0	0	3	4	4	1	15	62
	GO:0005990	lactose catabolism	26%	2	1	1	0	4	2	0	5	4	0	19	24

Figure II-8. List of 5 gene groups which had the largest population in each SOM cluster. Almost identical gene groups sharing with more than 80% of genes in the smallest category were represented by one category. Gene groups which had less than 3 genes in an intended SOM category were omitted.

Table II-4. *Genes categorized into c4 (root) or c6 (bud and flower).*
(below) in each SOM category. NE, Normalized Expression. Y, young leaf. L, leaf. T, stem. B, bud. F, flower. S, silique. R, root.

gene group	SOM	locus	description	NE						
				Y	L	T	B	F	S	R
Transcription factors (total 407 loci) c1:27, c2: 24, c3: 55, c4: 4, c5: 27, c6: 21, c7: 37, c8: 50, c9: 163, cx: 4	c4	At1g62300	WRKY family transcription factor (WRKY6)	2.6	2.6	2.0	1.9	2.1	2.3	2.9
		At3g24520	heat shock transcription factor HSF1, putative	3.9	3.9	4.0	3.6	3.8	4.0	4.5
		At5g48560	putative protein	2.1	1.8	3.3	1.7	1.9	2.4	3.2
		At5g65790	myb DNA-binding protein (MYB68)	2.0	1.9	2.1	2.1	1.9	2.2	2.8
	c6	At1g24260	SEPALLATA 3, MADS-box protein (SEP3)	2.0	2.7	1.9	3.7	3.5	3.4	2.2
		At1g27360	putative squamosa-promoter binding protein 2	2.4	2.6	2.5	3.0	3.2	2.7	2.2
		At1g27370	putative squamosa-promoter binding protein 2	2.3	2.2	2.4	2.7	2.8	2.4	2.3
		At1g69690	expressed protein	2.5	2.0	2.7	2.9	2.9	2.3	1.9
		At1g70790	zinc finger and C2 domain protein, putative	2.1	2.1	2.3	2.6	2.8	2.6	2.2
		At2g42830	floral homeodomain transcription factor (AGL5)	1.8	2.4	2.3	3.1	3.4	2.8	2.0
		At3g05400	zinc finger protein, putative	2.0	2.0	2.1	2.5	2.5	2.5	2.0
		At3g15500	putative jasmonic acid regulatory NAC family protein	2.8	2.4	2.3	2.8	3.2	2.9	2.2
		At3g21175	GATA zinc finger protein	2.6	2.5	2.7	3.6	3.4	2.7	2.8
		At3g54340	floral homeotic protein APETALA3 (AP3)	2.0	2.5	2.0	3.3	3.2	2.9	2.0
		At3g61950	bHLH family protein	2.1	2.3	2.3	2.4	2.9	2.3	1.9
		At4g09960	MADS-box protein (AGL11)	2.6	2.4	2.2	3.0	3.4	3.1	2.4
		At4g18960	floral homeotic protein agamous (AGAMOUS)	2.0	1.8	1.7	3.3	3.0	2.2	1.5
		At4g29030	trihelix family protein	2.5	2.4	2.4	3.4	3.2	2.6	2.5
		At4g31060	AP2-domain transcription factor	3.0	2.9	3.0	3.3	3.4	2.8	2.7
		At4g34410	AP2-domain transcription factor	2.5	3.0	2.9	3.3	3.1	3.3	2.6
		At5g15310	myb family transcription factor	2.4	2.5	2.1	3.0	2.7	2.4	2.1
		At5g15800	MADS-box protein AGL2	2.2	2.3	2.1	3.6	3.6	3.0	2.0
		At5g20240	PISTILLATA, MADS-box protein	1.9	2.1	2.0	3.7	3.5	2.4	2.3
		At5g37020	auxin response transcription factor (ARF8)	2.6	2.7	2.5	3.1	3.3	3.3	2.8
		At5g60910	MADS-box protein (AGL8)	2.5	2.9	3.3	3.4	3.5	3.0	2.5
Putative plastid targeting proteins (total 746 loci) c1:166, c2: 32, c3: 50, c4: 5, c5: 46, c6: 10, c7: 52, c8: 47, c9: 326, cx: 12	c4	At2g31920	hypothetical protein	1.6	2.0	2.1	1.7	1.8	2.5	3.4
		At2g36300	expressed protein	1.6	2.2	1.6	1.7	1.5	1.9	3.1
		At3g04340	unknown protein	2.5	2.3	2.7	2.4	2.4	2.6	3.4
		At3g23940	dihydroxyacid dehydratase, putative	2.8	1.8	1.8	2.2	2.0	1.7	2.9
		At4g22240	fibrillin precursor-like protein	1.9	2.4	2.3	2.2	2.3	2.3	2.6
	c6	At1g66430	fructokinase, putative	2.9	2.4	2.7	3.3	3.2	3.1	2.7
		At1g69690	expressed protein	2.5	2.0	2.7	2.9	2.9	2.3	1.9
		At2g12940	putative VSF-1-like b-ZIP transcription factor	2.1	2.3	2.4	2.5	2.7	2.4	1.8
		At2g24490	putative replication protein A1	2.3	2.2	2.4	2.6	2.9	2.5	2.2
		At3g10690	putative DNA gyrase subunit A	2.1	2.4	2.0	2.8	3.0	2.9	2.5
		At3g56090	putative protein	3.1	2.6	3.0	3.4	3.5	3.2	3.0
		At5g04710	aspartyl aminopeptidase -like protein	2.7	2.8	2.8	3.2	3.5	3.3	3.0
		At5g13840	fizzy-related (FZR), putative	2.6	1.7	2.0	2.8	2.5	1.8	2.0
		At5g18660	putative protein	2.8	2.3	2.5	3.0	3.2	2.7	2.4
		At5g20720	chloroplast Cpn21 protein	2.4	2.3	2.5	3.6	3.8	2.8	1.8

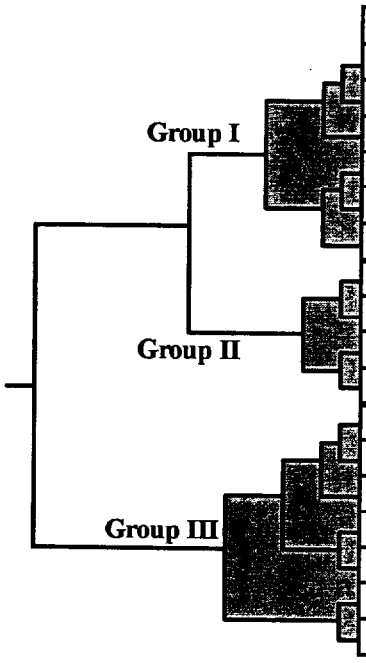
not identified were also found in this category.

Then, various types of transcription factors in AGRIS (Arabidopsis Gene Regulatory Information Server, Davuluri et al., 2003) were further analyzed according to these SOM categories. In addition to MADS box proteins, which tended to be in c6, a gene family of C2C2-CO-like transcription factors had a clear tendency to be in c1 (data not shown). This gene group is constituted by CONSTANS and CONSTANS-like proteins. CONSTANS has an important role in the regulation of flowering by photoperiod. It was reported that CONSTANS-like 1 and 2 genes were strongly expressed in leaf and rarely expressed in root (Ledger et al., 2001). Figure II-9 shows that CONSTANS and 16 CONSTANS-like genes consist of three subgroups, as previously reported (Griffiths et al., 2003). Interestingly, all genes in groups I and II were expressed with a c1 type expression pattern, and no genes in group III were expressed with a c1 type profile. Namely, gene groups I and II evolved with strong organ specificity (c1), suggesting their conserved function in photosynthetic tissues.

I also analyzed the organ specificity of putative plastid-targeted proteins. A gene list of putative nuclear-encoded plastid proteins was obtained from the TAIR website, in which each destination of the location site has been predicted by TargetP (Emanuelsson et al., 2000). Seven hundred forty-six genes with high reliability (Classes 1 and 2 of putative chloroplast proteins in TargetP) were selected, and their SOM categories were analyzed. Basically, most of the genes for nuclear-encoded plastid proteins were classified into c1 (photosynthesis-related gene category) or c9 (constitutively expressed gene category), indicating that their gene products function mainly or constitutively in plastids (Table II-4). However, a few genes were specifically expressed in other organs (root c4, or bud and flower, c6). These genes probably play specific roles in particular types of plastids like leucoplasts, chlomoplasts or amyloplasts.

II-3.6 Data representation by Tissue-Specific Expression Plot (TEP)

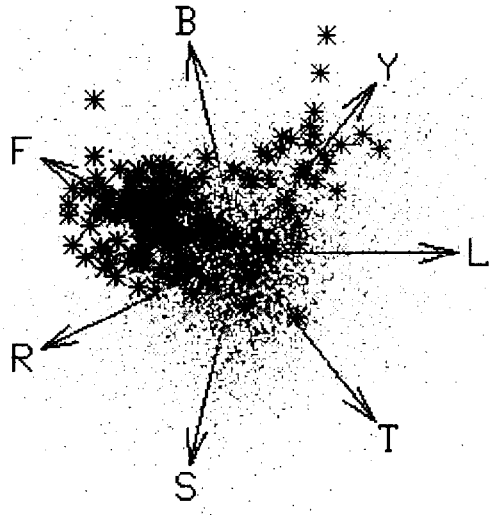
To visualize particular tendencies of the organ-specific gene expression of some gene groups, the specificity of each gene expression was represented using linear mapping of the 7 NE values for the organs to a 2-dimensional plane. Figure II-10A shows the distribution of the expression plots for 174 genes which belong to the gene groups of "ribosome (GO:0005840)". This plot, the Tissue-Specific



locus	gene name	SOM category
At5g15840	CO	-
At5g15850	COL1	-
At3g02380	COL2	c1
At2g24790	COL3	c1
At5g24930	COL4	c1
At5g57660	COL5	c1
At1g25440	COL16	c1
At1g68520	COL6	c1
At1g49130	COL8	-
At1g73870	COL7	-
At1g28050	COL15	c9
At2g33500	COL14	c9
At2g47890	COL13	c9
At3g21880	COL12	-
At4g15250	COL11	-
At3g07650	COL9	c2
At5g48250	COL10	c8

Figure II-9. Relationship between phylogenetic tree and SOM expression categories for CONSTANS (CO) and CONSTANS-like (COL) genes. -, no EST clone in macroarray.

A



B

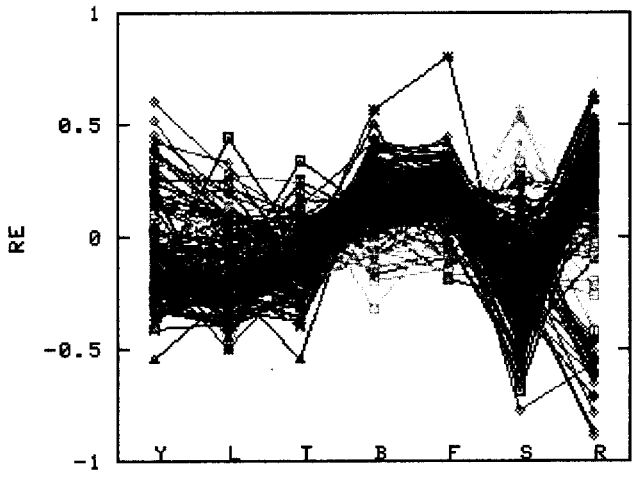


Figure II-10. "Tissue-Specific Expression Plot (TEP)" (A) and line plot of RE (B) for 172 genes in the gene family "ribosome (GO:0005840)". Genes in different SOM categories are represented by different colors; c1, lime. c2, green. c3, maroon. c4, red. c5, blue. c6, magenta. c7, navy. c8, aqua. c9 and cx, black. Gray dots in TEP represent all 8809 genes.

Expression Plot (TEP) was composed of 7 vectors derived from NE in the 7 organs. In Fig. II-10A, the sum of the 7 vectors is plotted. In this representation, genes constitutively expressed in all organs are indistinguishably plotted in the center with unexpressed genes. Basically, more closely related organs, such as bud and flower, were located in the same neighborhood to reduce interference, since the position for each plot is affected by the combination of coordinates with an opposite pair.

As shown in Fig. II-10B, a large number of ribosome genes showed relatively higher expression in bud, flower and root than in other organs, which corresponds to SOM category c5 (red color). Moreover, it is also evident that another major gene group of the ribosome is present, which corresponds to SOM category c1 (lime color). When I analyzed their targeting signals by TargetP, most of the gene products (22/23) were predicted to be targeted into plastids, revealing that these are plastid-targeted ribosomal proteins. The results indicated that visualization of organ specificity of genes by TEP is helpful for finding some tendency of expression in a gene group systematically, even when a simple line plot (Fig II-10B) has some limitations in this respect because the number of the genes is too large.

II-4 DISCUSSION

II-4.1 The method for normalization used in this study

I developed an LN (lognormal distribution fitting) method for data normalization. The LN method reconstructs all data using non-parametric fitting to a lognormal distribution based on the hypothesis that populations of mRNA contents in organs ideally match a lognormal distribution. Actually, matching to the lognormal distribution is commonly observed for various microarray data (Konishi, 2001). On the other hand, to screen genes responsive to some phytohormones or environmental changes using macroarrays, I usually use another conventional normalization method, which is designated as BC (Background subtraction and Centering). The BC method is constituted by two steps, subtracting an appropriate value as background (BG) and centering by the median of all signal intensities. Major problems in normalization are occurrences of fluctuation and bias in normalized data. Both fluctuation

and bias cause pseudo-positives. Although fluctuation of BC-normalized data depends on the definition of background, BC-normalized data generally have small fluctuations. However, they sometimes show a large bias in distributions of normalized signal intensities due to the different range of absolute signal intensities between normalized data groups. In contrast, with the fitting to lognormal distribution, no bias occurs in the normalized data (see Fig. II-1C), although it occasionally produces large fluctuations.

There were several reasons to use the LN method in this study. (1) Pseudo-positives caused by data fluctuation are less serious in the analysis of the global tendencies of expression than in that of gene screening. For instance, I analyzed the occurrence of organ specificity of expression in some gene groups using the GRE value. The averaging step of RE values in a gene group to obtain GRE nullifies the fluctuations of data that occur in normalization, while data bias remains if it happens in the normalization. (2) The LN method does not need any reference of expression. Therefore, it is suitable for the determination of organ specificity obtained using one-dye arrays because there is no common reference for the analysis of organ specificity. (3) Repetition of data collection generally reduces fluctuation, but in many cases, it has no ability to reduce a large bias. The use of macroarrays makes it relatively easy to get repeated data because the membrane is cheaper and easier to handle with than in other microarray systems. (4) The estimation of fold-enhancement is accurate after LN normalization. As shown in Fig.II-5, data of macroarray and Northern blot analysis are quantitatively consistent.

II-4.2 Clustering of expression profiles by 1D-SOM

Several algorithms are commonly used for clustering array data. Since I use a line graph to represent each cluster for easy understanding, SOM (Tamayo et al., 1999) was the best among the typical clustering algorithms. I did not employ two-dimensional SOM, which is commonly used for clustering of microarray data, but employed one-dimensional SOM (1D-SOM), because the latter procedure is simpler to understand and adequate for categorizing the organ specificity. In clustering, a smaller number of categories is more useful for understanding the physiological characteristics in each cluster. In my case, the expression profiles were principally categorized into 9 clusters. When the number of categories is so small, multi-dimensional visualization of the relationship among different categories (self-organizing

view) is not so important. More importantly, 1D-SOM consumes only one dimension, and thus enables us to represent additional information using another dimension, as shown in Fig. II-7 and II-8. Therefore, my results demonstrated that 1D-SOM is a useful tool for categorizing organ-specific expression patterns.

II-4.3 Advantages of macroarrays for studies of organ-specific gene expression

When I evaluate the contributions of some genes to a particular function, information about the absolute expression level of each gene is important for determining the exact contribution. For instance, when an organ has multiple gene copies for the same function, the contribution of each gene mainly depends on their absolute expression levels in the organ, not on the relative expression levels compared to the other organs. Since in competition assays using two fluorescence dyes (Cy3 and Cy5) which are generally applied in typical microarray systems (Schena et al., 1995), the relative expression level compared to that in another organ is obtained from the data, it is not sufficient for this purpose. Since radioactive compounds are used to quantify the data in cDNA macroarray experiments, quantification can be obtained without competition experiments (Sasaki et al., 2001). As described above, from macroarray experiments, both absolute and relative expression levels in the organs (NE and RE) are obtained, and thus the exact importance of genes in each organ can be more precisely evaluated. For instance, all NE values of the top 15 genes in 7 organs were shown in Table II-1. Although the top 15 genes were selected based on the RE values in an organ where the genes are strongly expressed, the NE values in all organs could be determined. As shown for silique in Table II-1, the NE values of two 2S albumin genes in silique (NE=5.1, 5.2) are 100-fold higher than the expression levels of major genes (NE=3, approximately equal to average expression level), indicating the very high expression level of these two genes in this organ. It was also noted that for these genes, the expression level in other organs are about 10-fold lower than the expression levels of major genes.

The overall data of organ-specific gene expression obtained here will also be useful for identifying consensus promoter sequences for organ-specific expression. For this purpose, more appropriate and refined clustering of the expression profiles will be required.

II-4.4 A web site for *Arabidopsis thaliana* Tissue-Specific Expression Database (ATTED)

These organ-specific gene expression data are available in the web site for *Arabidopsis thaliana* Tissue-Specific Expression Database (ATTED) (http://www.atted.bio.titech.ac.jp/). See Appendix about ATTED.

II-5 REFERENCES

Aharoni A, Vorst O (2001) DNA microarrays for functional plant genomics. Plant Mol Biol **48**: 99-118

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

Asamizu E, Nakamura Y, Sato S, Tabata S (2000) A large scale analysis of cDNA in *Arabidopsis thaliana*: Generation of 12,028 non-redundant Expressed Sequence Tags from normalized and size-selected cDNA libraries. DNA Res **7**: 175-80

Berger S, Bell E, Sadka A, Mullet JE (1995) *Arabidopsis thaliana* Atvsp is homologous to soybean *VspA* and *VspB*, genes encoding vegetative storage protein acid phosphatases, and is regulated similarly by methyl jasmonate, wounding, sugars, light and phosphate. Plant Mol Biol **27**: 933-42

Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J, Corcoran K (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol **18**: 630-4

Chen W, Provart NJ, Glazebrook J, Katagiri F, Chang H-S, Eulgem T, Mauch F, Luan S, Zou G, Whitham SA, Budworth PR, Tao Y, Xie Z, Chen X, Lam S, Kreps JA, Harper JF, Si-Ammour A, Mauch-Mani B, Heinlein M, Kobayashi K, Hohn T, Dangl JL, Wang X, Zhu T (2002) Expression profile matrix of Arabidopsis transcription factor genes suggests their putative functions in response to environmental stresses. Plant Cell **14**: 559-74

Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**: 5294-9

Davuluri RV, Sun H, Palaniswamy SK, Matthews N, Molina C, Kurtz M, Grotewold E (2003) AGRIS: Arabidopsis Gene Regulatory Information Server, an information resource of Arabidopsis cis-regulatory elements and transcription factors. BMC Bioinformatics **4**: 25

Donson J, Fang Y, Espiritu-Santo G, Xing W, Salazar A, Miyamoto S, Armendarez V, Volkmuth W (2002) Comprehensive gene expression analysis by transcript profiling. Plant Mol Biol **48**: 75-97

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol **300**: 1005-16

Gene Ontology Consortium (2000) Gene Ontology: tool for the unification of biology. Nature genetics **25**: 25-9

Griffiths S, Dunford RP, Coupland G, Laurie DA (2003) The Evolution of *CONSTANS*-Like Gene Families in Barley, Rice, and Arabidopsis. Plant Physiol **131**: 1855-67

Hird DL, Worrall D, Hodge R, Smartt S, Paul W, Scott R (1993) The anther-specific protein encoded by the *Brassica napus* and *Arabidopsis thaliana* A6 gene displays similarity to beta-1,3-glucanases. Plant J **4**: 1023-33

Honma T, Goto K (2001) Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. Nature **409**: 525-9

Kanehisa M, Goto S, Kawashima S, Nakaya A (2002) The KEGG databases at GenomeNet. Nucleic Acids Res **30**: 42-6

Konishi T (2002) Parametric Treatment of cDNA Microarray Data. Genome Informatics **13**: 280-1

Ledger S, Strayer C, Ashton F, Kay SA, Putterill J (2001) Analysis of the function of two circadian-regulated *CONSTANS-LIKE* genes. Plant J **26**: 15-22

Matsushima R, Hayashi Y, Yamada K, Shimada T, Nishimura M, Hara-Nishimura I (2003) The ER Body, a Novel Endoplasmic Reticulum-Derived Structure in Arabidopsis. Plant Cell Physiol **44**: 661-6

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol Plant **15**: 473- 93

Oliveira DE, Franco LO, Simoens C, Seurinck J, Coppieters J, Botterman J, Van Montagu M (1993) Inflorescence-specific genes from *Arabidopsis thaliana* encoding glycine-rich proteins. Plant J **3**: 495-507

Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M, Miller N, Mueller LA, Mundodi S, Reiser L, Tacklind J, Weems DC, Wu Y, Xu I, Yoo D, Yoon J, Zhang P (2003) The *Arabidopsis* Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. Nucleic Acids Res. **31**: 224-8

Rubinelli P, Hu Y, Ma H (1998) Identification, sequence analysis and expression studies of novel anther-specific genes of *Arabidopsis thaliana*. Plant Mol Biol **37**: 607-19

Sasaki Y, Asamizu E, Shibata D, Nakamura Y, Kaneko T, Awai K, Amagai M, Kuwata C, Tsugane T, Masuda T, Shimada H, Takamiya K, Ohta H, Tabata S (2001) Monitoring of Methyl Jasmonate-responsive Genes in Arabidopsis by cDNA Macroarray: Self-activation of Jasmonic Acid Biosynthesis and Crosstalk with Other Phytohormone Signaling Pathways. DNA Res **8**: 153-61

Sasaki-Sekimoto Y (2004) Studies on jasmonates responsive gene expressions and metabolic flow in *Arabidopsis thaliana*. Doctoral Thesis, Tokyo Institute of Technology.

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science **270**: 467-70

Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR (1999) Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc Natl Acad Sci USA **96**: 2907-12

Theißen G, Saedler H (2001) Floral quartets. Nature **409**: 469-71

Chapter III

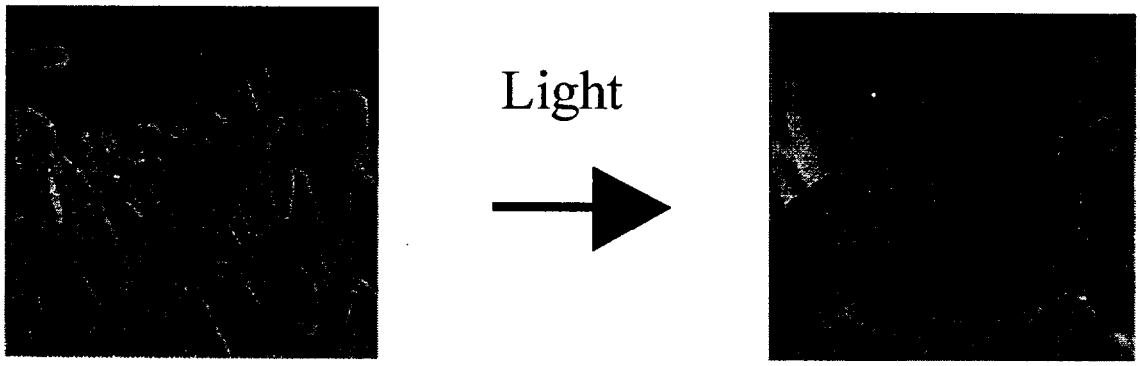
Transcriptome Analysis on Photomorphogenesis

III-1 INTRODUCTION

Light has profound effects on the development of plants. The most striking effects of light are observed at the first exposure of the light, namely, seed germination from the soil. This morphogenesis induced by light were called photomorphogenesis (Fig. III-1). Normally after the seedling root emerges first from the seed, and the root is established, the shoot appears. Later, with growth of the shoot in the light there is increased secondary root formation and branching. To a large degree, these coordinated differential growth responses are hormone-mediated events such as cytokinin and auxin. In the absence of light, plants develop an etiolated growth pattern. Etiolation of the seedling adapts it to emerging from the soil. As reaching the surface of the soil, de-etiolation of etiolated seedlings is induced by light. Typically, plants are responsive to wavelengths of light in the blue, red and far-red regions of the spectrum through the action of several different photosensory systems. Although all the three types of the light induce different aspects of de-etiolation, red and far-red lights have main function for the de-etiolation. The photoreceptors for red and far-red wavelengths are known as phytochromes. Phytochromes are proteins with a light absorbing pigment attached (chromophore). The chromophore is a linear tetrapyrrole called phytochromobilin. Mutations of phytochromes or phytochromobilin biosynthesis defeat de-etiolation process.

Cytokinins promote plant cell division in conjunction with auxin. They also influence differentiation of plant cells in culture: A high cytokinin/auxin ratio promotes shoot production; auxin alone initiates root growth; and approximately equimolar amounts of cytokinin and auxin cause largely undifferentiated callus cells. Cytokinins also are known to induce opening of stomata, suppress auxin-induced apical dominance, and inhibit senescence of plant organs.

Although cytokinin is noted for a variety of effects on plant growth and development, it is frequently associated with the action of light. The interaction between cytokinin and light responses was



Etiolated characteristics	De-etiolated characteristics
Distinct apical hook	Apical hook opened
No leaf growth	Leaf growth promoted
No chlorophyll	Chlorophyll produced
Rapid stem elongation	Stem elongation suppressed
Limited radial expansion of stem	Radial expansion of stem
Limited root elongation	Root elongation promoted
Limited production of lateral roots	Lateral root development accelerated

Figure III-1. Photomorphogenesis. Pictures of cucumber are represented as an example.

first reported by Miller (1956), who found that cytokinin in the dark could elicit several responses that were induced by red light, such as leaf disc expansion in bean and seed germination in lettuce. Many observations were reported about the interaction between light and cytokinin responses in the context of processes such as hypocotyl elongation (Su and Howell, 1995), structural chloroplast development (Chory et al., 1994), and chloroplast membrane synthesis (Yamaryo et al., 2003). For example, light illumination changes etioplast in etiolated seeding to chloroplast, while cytokinin treatment also changes etioplast to chloroplast (Chory et al., 1994).

Although such many observations for cooperative effects of light and cytokinin were reported, I thought global analysis was needed for the overall understanding of light and cytokinin on the effects of photomorphogenesis. Moreover, understanding in relation with organ functions examined in Chapter II is also important. Although cytokinin is synthesized in each tissue needed, cytokinin biosynthesis in root is the most active (Chen et al., 1985). I used cDNA macroarray system for the global analysis of light and cytokinin effects on the major event of photomorphogenesis, de-etiolation process.

Cytokinins can be defined structurally as adenine derivatives with an isopentenyl-based side chain attached to the N6 amino group. Although plant cytokinins with a benzyl or hydroxybenzyl group at N6 are rare, the synthetic cytokinin, N6-benzyladenine (BA) has been used frequently in biochemical and physiological studies. In this study, BA was used as the cytokinin.

III-2 MATERIALS AND METHODS

III-2.1 Plant materials

Sterilized seeds of Arabidopsis (accession Columbia) were germinated in MS medium (Murashige and Skoog, 1962). After cold treatment for over night, the seedlings were incubated on an orbital shaker in the dark at 22°C. After 5 days, the plants were treated with 10μM benzyl-aminopurine (BA) for cytokinin experiments or treated with $2W/m^2$ illumination.

III-2.2 RNA isolation

Total RNAs were extracted using an RNeasy mini kit (Qiagen, U.S.A.) for Northern analysis or

guanidine thiocyanate and phenol-chloroform extraction (Chomczynski and Sacchi, 1987) and then purified using RNeasy mini kit (Qiagen, U.S.A.) for the macroarray experiment.

III-2.3 Northern blot analysis

Total RNAs were separated by electrophoresis, blotted onto nylon membrane and probed with [α-^{32}P]dCTP-labeled DNA. Hybridization was carried out using Church phosphate buffer (Church and Gilbert, 1984). A probe for ARR5 (At3g48100) was provided from Kazusa DNA Research Institute as EST clone (Genbank AV536923, Asamizu et al., 2000), and a probe for Ubiquitin10 (At4g05320) was provided from Dr. Awai (Awai, 2001).

III-2.4 cDNA macroarray system

The cDNA macroarray system used in Chapter II was also used for this light and cytokinin transcriptome analysis (II-2.2, II-2.3) except for reassignment of all ESTs to the genome. The 13516 cDNA clones were assigned to 8372 loci.

III-2.5 BCL normalization

Qualified macroarray data were normalized by BCL normalizing method. BCL normalizing method and the related methods are one of standard normalizing method. The BCL method is constituted by three steps, subtracting an appropriate value as background (BG), centering by the median of all signal intensities, and fitting average distribution of a time series of data using lowess (locally weighted scatter plot smooth). The actual procedure was as follows. First, raw data were subtracted virtual back ground.

$$NE_{ESTclone,time,repetition}(B) = \log(RAW_{ESTclone,time,repetition} - vBG_{membrane})$$

where vBG is defined 0.8 times of minimum value of all spot including negative control as described in II-2.4. Then, subtract median of the all EST clones in the membrane.

$$NE_{ESTclone,time,repetition}(BC) = NE_{ESTclone,time,repetition}(B) - median(NEs_{ESTclone})$$

Finally, fitting to average distribution using lowess.

$$NE_{ESTclone,time,repetition}(BCL) = f_{lowess}(NE_{ESTclone,time,repetition}(BC))$$

where *f*lowess is function to fit average distribution. Average distribution is distribution of averaged data of a time series.

$$Averaged\_data = \frac{1}{N_{time}} \sum_{time} \left(NE_{ESTclone,time,repetition}(BC) \right)$$

Where N_{time} is number of time points for the experiments. Lowess was achieved using a statistical programming language R. Unification of repeated data and follows are the same procedure to LN method described in Chapter II.

$$NE_{ESTclone,time} = \frac{\sum_{repetition} \left(W_{time,repetition} \times NE_{ESTclone,time,repetition} \right)}{\sum_{repetition} W_{time,repetition}}$$

$$NE_{gene,time} = f_{locus}\left(NE_{ESTclone,time} \right)$$

$$RE_{gene,time} = NE_{gene,time} - \frac{1}{N_{time}} \sum_{time} NE_{gene,time}$$

$$GRE_{group,time} = \frac{1}{N_{genes.in.group}} \sum_{genes.in.group} RE_{gene,time}$$

where W is weight. LN method described in Chapter II (II-2.5) and this BCL method have almost the same effect.

III-2.6 Gene groups

"Gene groups" were defined the groups of genes. In this chapter, for detail information, gene group was defined using the group of genes provided by public societies like GO and the intercellur localization predicted by TargetP. All genes have predictions of 4 types of subcellur localization, chloroplast (C), mitochondria (M), secretory (S) and others (O). Each prediction also has reliability defined from 1 (high) to 5 (low). The lowest reliability, namely the value is 5, was omitted for the purpose. Thus one gene group could be divided into 4 subgroups like Fig. III-2. The subgroups were used in this chapter.

III-3 RESULTS

III-3.1 Effect of sucrose on cytokinin response

Genes in the group "ABC transporter activity (GO:0004009)"

TargetP reliability	Chloroplast (C)	Mitochondria (M)	Secretory (S)	Others (O)
1	At4g04770	At2g41700	At5g39390	At1g27940 At1g28010 At1g30400 At1g30410 At1g30420 At2g13610 At2g34660 At2g36910 At2g39350 At4g27420 At5g32621
2	At1g28530 At1g54350 At2g28070	At2g07680 At3g47730 At4g28620 At5g58270	At1g04120 At3g13090	At1g02520 At1g02530 At1g10680 At1g31770 At1g51460 At1g53270 At1g71330 At2g31970 At2g37280 At2g47000 At3g13220 At3g28345 At3g28860 At3g55090 At3g55100 At3g55110 At3g60160 At3g62150 At4g15230 At4g25960
3	At1g15210 At1g15520 At1g59870 At1g70610 At3g52310 At5g03910 At5g19410	At4g28630	At5g60740	At1g17840 At2g26910 At2g29940 At2g37360 At3g13080 At3g13100 At3g47780 At3g47790 At4g18050 At5g06530 At5g07660At1g17840 At2g26910 At2g29940 At2g37360 At3g13080 At3g13100 At3g47780 At3g47790 At4g18050 At5g06530 At5g07660
4	At1g66950 At2g46630 At3g25620	At2g27170 At2g36380 At5g38640 At5g61730	At3g60970	At1g51500 At2g01320 At2g47800 At3g47740 At3g53480 At3g53510 At3g59140 At4g01830 At4g25750 At5g13580 At5g39040 At5g46540
5	At3g16340 At3g21250 At3g24600 At5g52860 At5g61460	At3g28360 At3g47750 At5g61690 At5g61700	At1g53390	At1g71960 At2g39480 At3g21090 At3g28390 At3g47760 At3g47770 At3g55130 At3g55320 At3g62700 At4g01660 At4g01820 At4g25450 At5g61740

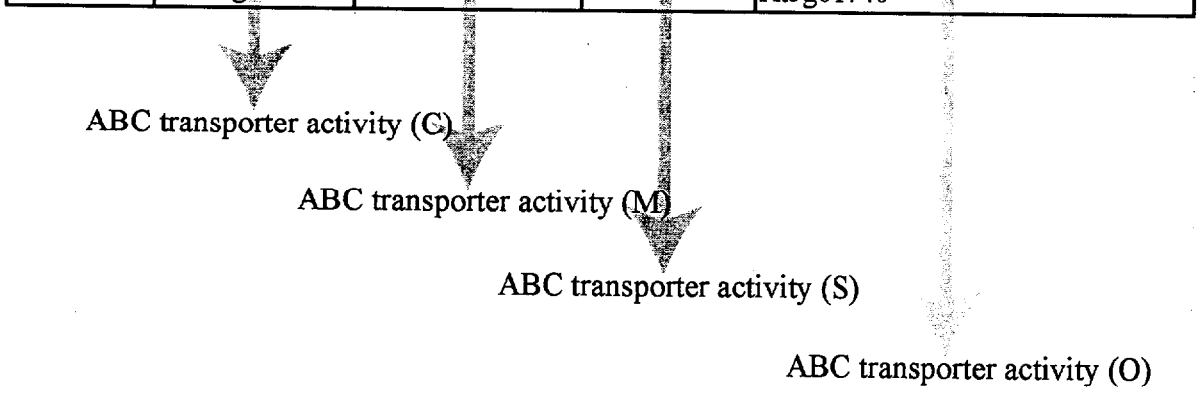


Figure III-2. A gene group divided by subcellular localization.

Before the analysis of cytokinin-responsive genes, I established an optimal condition of cytokinin treatment of etiolated seedling of Arabidopsis. Sucrose is known to affect plant responses to phytohormones especially during greening period (Sheen et al., 1999). Cytokinin responses were partially repressed by external sucrose. Because Arabidopsis is generally grown in a culture medium containing 1% sucrose, the effects of various concentrations of sucrose on cytokinin response were examined. Northern blot analysis of a known cytokinin responsive gene, ARR5 (Arabidopsis Response Regulator 5), showed that the cytokinin response was observed even under no sucrose condition (Fig. III-3). Although plant yields were lower under lower sucrose condition (data not shown), it was enough to carry out the transcriptome analysis. Thus in this study, plants were grown in MS solution without addition of sucrose.

III-3.2 Obtaining data by macroarray method

Macroarray data for cytokinin treatment and light illumination were obtained. The SN values of the data were shown in Fig. III-4. After BCL-normalization and succeeding data treatments, the 8372 gene expression data were obtained. The result was represented as plot of NE at 0h vs. fold (Fig. III-5). In the figure III-5, representative genes were highlighted. Expression of all the representative genes which are known as the responses to light and cytokinins was induced, indicating the expected response in the experiments. For example, *FAD7* encodes an enzyme for fatty acid desaturation in chloroplast, which is known to be up-regulated responding to the light (Nishiuchi et al., 1995). The enzyme catalyzes ω-3 desaturation of fatty acids such as linoleic acid. ω-3 desaturation of fatty acids is a typical phenomenon during chloroplast development. In global point of view, two characters were observed in this figure. The responses in 3h were bigger than those in 1h as expected. The responses to light were bigger than those to cytokinin. The reason of smaller response of cytokinin is not clear.

III-3.3 Correlation between light and cytokinin responses

As described in Introduction, cytokinin treatment mimics light illumination. Figure III-6 shows relationship between light and cytokinin responses in transcriptome level. Although 1 h treatment showed no correlation ($r = -0.07$), 3 h treatment had significant correlation ($r = 0.23$) with stronger

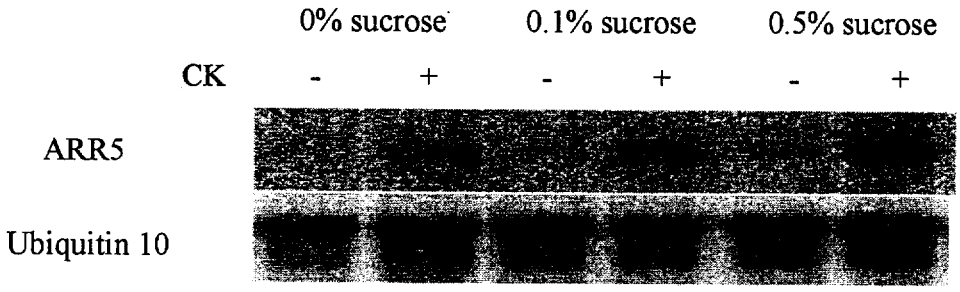


Figure III-3. Effect of sucrose on cytokinin response at transcription level. 10μM BA was used as cytokinin. ARR5 (Arabidopsis Response Regulator 5) and Ubiqutin 10 were positive and negative control of cytokinin response, respectively.

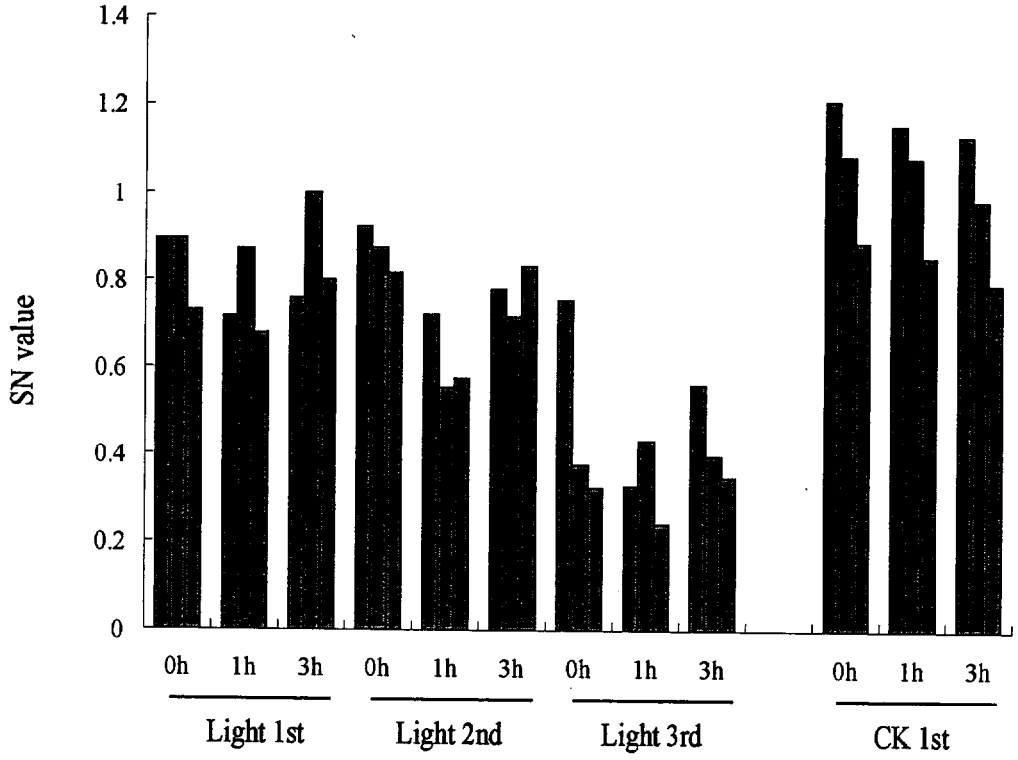


Figure III-4. SN (Signal to Noise ratio) value of 3 light experiments and 1 cytokinin experiment. Each experiment has 3 membranes representing 3 bars in each time point.

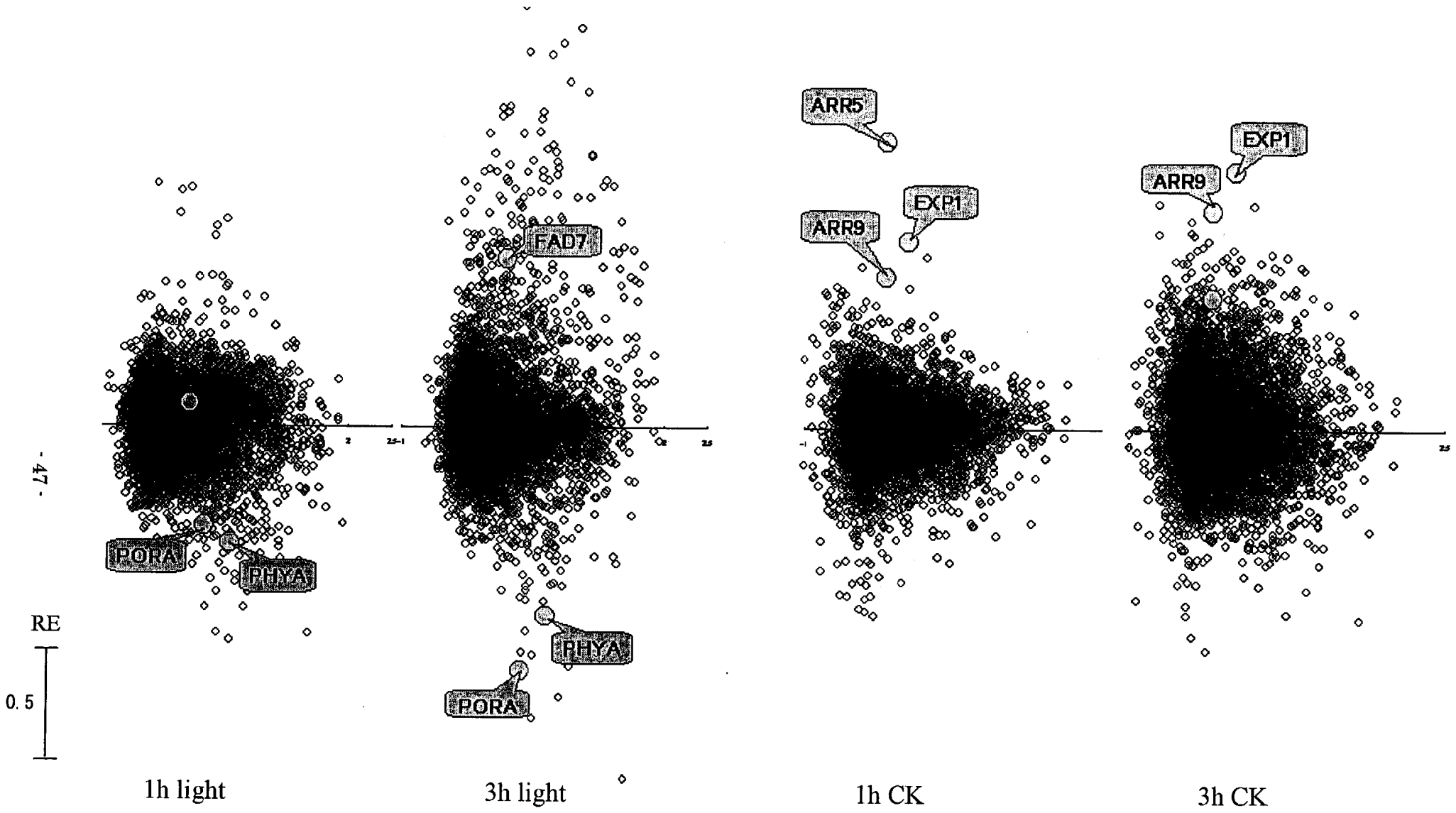


Figure III-5. 0h NE (Normalized Expression) vs fold plot about 8372 loci. Representative genes were highlighted. x-axis, NE at 0h. y-axis, RE (Relative Expression) using same scale in the 4 graph.

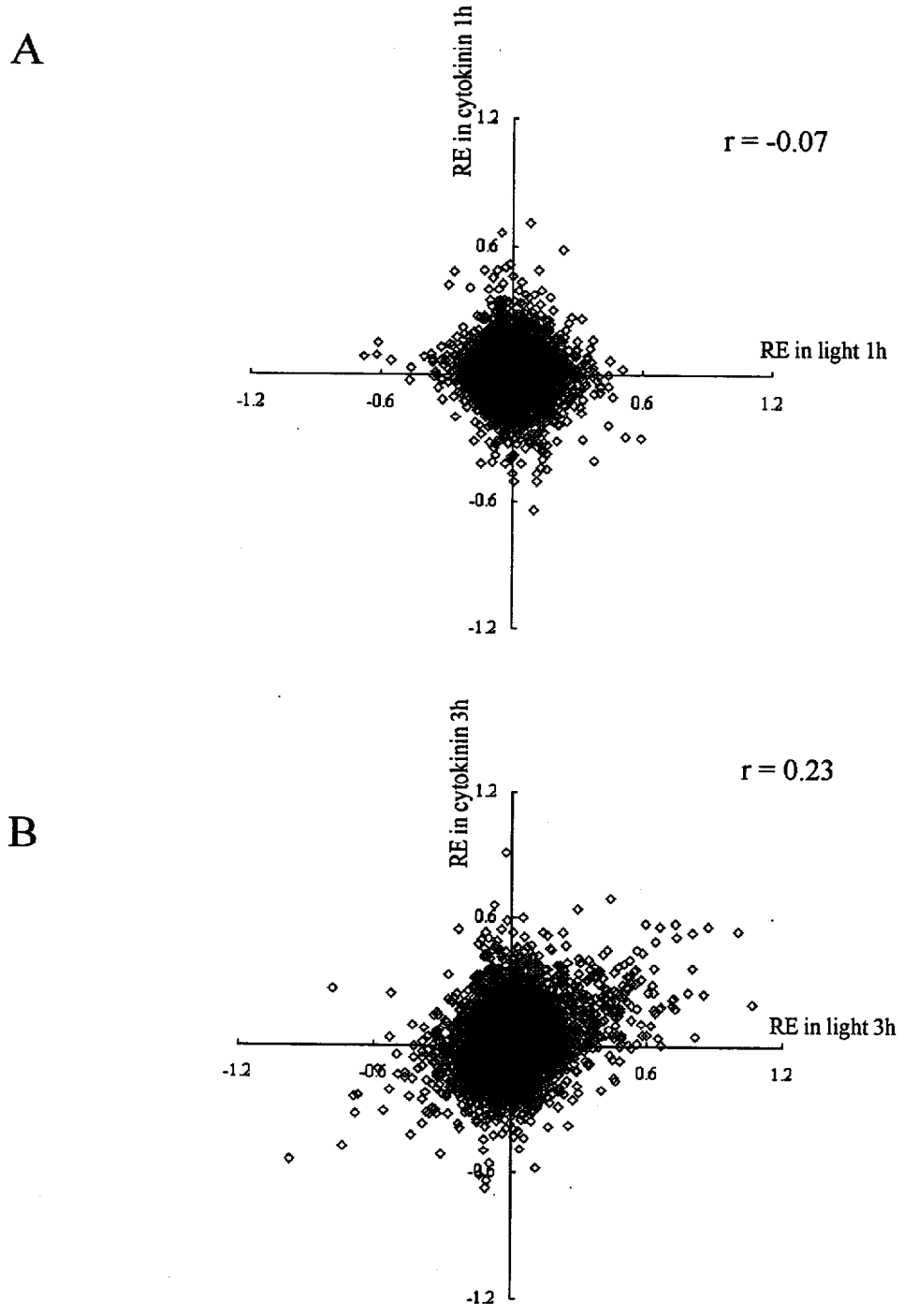


Figure III-6. Correlation between light and cytokinin responses at 1h (A) and 3h (B) about 8372 loci. Pearson's correlation coefficients are shown as r value. X-axis is RE of light. Y-axis is RE of cytokinin.

response to the light. While cytokinin-specific response was observed on y-axis, light-specific response was not apparent in the figure, because it was hidden in the genes responding to both light and cytokinin.

III-3.4 Gene list

Table III-1 is a list of the genes with the top 15 REs for light or cytokinin treatments. In 1 h of light response, a gene for a putative chloroplast ferredoxin was up-regulated. Many proteins of chloroplast were regulated by redox change via ferredoxin (Balmer et al., 2003). The ferredoxin was thought to be a key protein in the redox regulation. Three genes related to stress response, SOUL-related protein, glutathione reductase and glutathione transferase, were up-regulated strongly at 1h light illumination. In 3h of light illumination, many genes which encode proteins directly involved in photosynthesis (particularly components for electron transfer in photosynthesis such as light-harvesting chlorophyll a/b binding proteins) were up-regulated.

Both *ARR5* and *ARR9* encode a response regulator involved in cytokinin signal transduction and known as cytokinin-inducible genes (Kiba et al., 2002). In the case of *ARR5*, its expression was transient, as reported previously. *EXP1* is a gene for cell wall synthesis, which is also known to respond to cytokinin (Downes and Crowell, 1998, Rashotte et al., 2003).

On the other hand, Table III-2 is a list of down-regulated genes. In the light response, genes for phytochrome A (*PHYA*, Canton and Quail, 1999) and NADPH:protochlorophyllide oxidoreductase (*PORA* and *PORB*, Oosawa et al., 2000) were down regulated, as reported. PHYA is a light receptor and PORA and PORB are the enzymes of light-dependent reaction in chlorophyll biosynthesis. Transcripts of the latter two genes were accumulated in the dark and rapidly degraded in response to light illumination. In cytokinin response, cell wall-related genes (*EXP13* and 2 *XTR*s) were down-regulated.

III-3.5 GRE analysis

To know physiological events which are globally regulated by cytokinin and light, GRE analysis was carried out according to the same procedure described in Chapter II. Table III-3 shows gene groups selected by GRE analysis. It was clustered using hierarchical clustering of GRE value of light, cytokinin and 7 organs. Type of responses was classified into 5 groups, indicating clear relationship

Table III-1. *List of the genes with the top 15 REs for light illumination or cytokinin treatment.*

RE values more than Log(2) are bolded. SOM, SOM category of 7 organs defined in Chapter II. targetP, predicted subcelluar localization: C, chloroplast. M, mitochondria. S, secretory. _, Others. number, reliability (1: highest).

		RE						
	locus	Light1h	Light3h	CK1h	CK3h	description	SOM	targetP
Light 1h	At1g17100	**0.6**	**0.5**	0.1	0.0	SOUL-related protein	c8	S1
	At1g06040	**0.6**	**0.6**	0.2	0.2	salt-tolerance protein (STO)	c9	_5
	At4g26850	**0.6**	**0.5**	0.1	0.2	expressed protein (VTC2)	c1	_3
	At2g30510	**0.5**	**0.7**	0.1	0.0	RPT2	cx	_2
	At2g05100	**0.5**	**1.1**	**0.4**	**0.6**	light-harvesting chlorophyll a/b binding protein (LHCB2)	c1	C1
	At1g10960	**0.5**	**0.7**	0.1	0.3	ferredoxin, chloroplast, putative	c9	C2
	At2g28670	**0.5**	0.3	0.0	0.1	fibroin-related	c3	S1
	At3g47470	**0.4**	**1.0**	0.2	**0.4**	light-harvesting chlorophyll a/b binding protein	c1	C2
	At5g44190	**0.4**	0.2	0.0	0.0	myb family transcription factor (GLK2, GPRI2)	c2	_2
	At3g13100	**0.4**	**1.1**	**0.4**	**0.4**	ABC transporter family protein (MRP7)	c1	_3
	At3g24170	**0.4**	0.2	0.0	0.0	glutathione reductase, putative	c9	_3
	At2g47730	**0.4**	0.2	0.0	0.1	glutathione transferase (GST6)	c9	C1
	At3g17800	**0.4**	0.0	-0.1	-0.2	expressed protein	c1	C2
	At4g16980	**0.4**	**0.8**	0.1	0.0	arabinogalactan-protein family		
	At1g59750	**0.3**	**1.0**	0.2	**0.4**	auxin response transcription factor (ARF1)	c1	_4
Light 3h	At3g13100	**0.4**	**1.1**	**0.4**	**0.4**	ABC transporter family protein (MRP7)	c1	_3
	At2g05100	**0.5**	**1.1**	**0.4**	**0.6**	light-harvesting chlorophyll a/b binding protein (LHCB2)	c1	C1
	At3g54890	**0.3**	**1.1**	0.1	0.2	light-harvesting chlorophyll a/b binding protein	c1	C4
	At1g29910	**0.3**	**1.0**	0.2	**0.5**	photosystem II type I chlorophyll a /b binding protein (CAB3)		C3
	At1g59750	**0.3**	**1.0**	0.2	**0.4**	auxin response transcription factor (ARF1)	c1	_4
	At3g47470	**0.4**	**1.0**	0.2	**0.4**	light-harvesting chlorophyll a/b binding protein	c1	C2
	At5g54270	0.3	**1.0**	0.1	0.2	light-harvesting chlorophyll a/b binding protein (LHCB3)	c1	C3
	At2g34430	0.3	**0.9**	0.2	**0.4**	photosystem II type I chlorophyll a /b binding protein	c1	C3
	At1g29920	0.2	**0.9**	0.2	**0.4**	photosystem II type I chlorophyll a /b binding protein (CAB2)	c1	C3
	At4g38970	0.2	**0.8**	0.1	**0.4**	fructose-bisphosphate aldolase, putative	c1	C2
	At5g13630	0.3	**0.8**	0.1	0.2	Mg chelatase H sub unit (CHLH)	c1	C2
	At1g44575	0.2	**0.8**	0.0	0.2	photosystem II 22kDa protein -related (NPQ4, PSBS)	c1	C5
	At4g16980	**0.4**	**0.8**	0.1	0.0	arabinogalactan-protein family		
	At2g38980	0.2	**0.8**	0.0	**0.5**	expressed protein	c1	M4
	At5g14530	0.3	**0.8**	0.1	0.2	transducin / WD-40 repeat protein family	c1	_2
CK1h	At3g48100	-0.1	-0.2	**0.7**	**0.3**	Arabidopsis response regulator 5 (ARR5)	c5	_4
	At1g69530	0.1	**0.5**	**0.5**	**0.7**	expansin (EXP1)	c9	S4
	At2g05100	**0.5**	**1.1**	**0.4**	**0.6**	light-harvesting chlorophyll a/b binding protein (LHCB2)	c1	C1
	At5g40390	0.1	0.1	**0.4**	0.3	glycosyl hydrolase family 36	c4	_2
	At3g57040	0.0	0.1	**0.4**	**0.6**	Arabidopsis responce regulator 9 (ARR9)	cx	_3
	At5g45170	-0.1	0.1	**0.4**	-0.1	expressed protein		_5
	At3g13100	**0.4**	**1.1**	**0.4**	**0.4**	ABC transporter family protein (MRP7)	c1	_3
	At3g04605	-0.1	0.0	**0.4**	0.0	Mutator-related transposase	c9	M4
	At5g06530	0.1	0.1	**0.4**	-0.3	ABC transporter family protein	c2	_3
	At2g29130	0.1	0.1	**0.3**	**0.4**	laccase (diphenol oxidase), putative	c7	S1
	At1g76240	-0.1	-0.2	**0.3**	0.1	expressed protein	c9	C3
	At1g20780	-0.1	0.1	**0.3**	0.0	expressed protein	c9	_1
	At5g48850	0.0	0.0	**0.3**	0.2	male sterility MS5 family	c6	_2
	At1g79080	0.0	0.0	**0.3**	**0.3**	pentatricopeptide (PPR) repeat-containing protein	c9	C4
	At3g06480	0.0	0.0	**0.3**	**0.5**	DEAD box RNA helicase, putative	c9	_2
CK3h	At1g69530	0.1	**0.5**	**0.5**	**0.7**	expansin (EXP1)	c9	S4
	At5g35210	0.0	0.0	0.0	**0.6**	PHD finger transcription factor, putative	c9	_2
	At3g28630	0.1	0.0	0.2	**0.6**	expressed protein	c9	_1
	At2g05100	**0.5**	**1.1**	**0.4**	**0.6**	light-harvesting chlorophyll a/b binding protein (LHCB2)	c1	C1
	At3g57040	0.0	0.1	**0.4**	**0.6**	Arabidopsis responce regulator 9 (ARR9)	cx	_3
	At1g80270	0.0	0.0	0.0	**0.5**	DNA binding protein, putative	c3	M2
	At1g53120	0.0	-0.1	-0.1	**0.5**	expressed protein	c9	M4
	At3g06480	0.0	0.0	**0.3**	**0.5**	DEAD box RNA helicase, putative	c9	_2
	At5g11480	0.0	0.0	0.0	**0.5**	GTP-binding protein -related	c9	C2
	At5g41460	0.2	**0.8**	0.1	**0.5**	fringe-related protein	c1	_1
	At3g05030	0.0	-0.2	0.0	**0.5**	sodium proton exchanger (NHX2)	c6	S2
	At2g38980	0.2	**0.8**	0.0	**0.5**	expressed protein	c1	M4
	At3g07630	0.0	0.0	0.0	**0.5**	prephenate dehydratase family	c9	C4
	At1g29910	**0.3**	**1.0**	0.2	**0.5**	photosystem II type I chlorophyll a /b binding protein (CAB3)		C3
	At1g10880	0.0	0.0	0.0	**0.4**	expressed protein	c8	_1

Table III-2. *List of the genes with the bottom 15 REs for light illumination or cytokinin treatment.*

RE values more than Log(2) are bolded. SOM, SOM category of 7 organs defined in Chapter II. targetP, predicted subcelluar localization: C, chloroplast. M, mitochondria. S, secretory. _, others. number, reliability (1: highest).

		RE						
	locus	Light1h	Light3h	CK1h	CK3h	description	SOM	targetP
Light 1h	At4g27450	**-0.6**	-0.3	0.0	-0.1	expressed protein	c2	_3
	At4g14130	**-0.5**	**-0.9**	-0.2	**-0.5**	xyloglucan endotransglycosylase (XTR7)	c9	S1
	At3g63510	**-0.5**	**-0.8**	-0.2	**-0.4**	nitrogen regulation protein family (NIFR3)	c9	C5
	At5g02020	**-0.5**	**-0.7**	-0.1	-0.3	expressed protein	c9	C1
	At5g07010	**-0.5**	**-0.6**	0.0	-0.3	sulfotransferase family	c9	_4
	At1g74670	**-0.4**	0.5	0.0	0.2	GAST1-related protein	c1	_2
	At3g23730	**-0.4**	-0.3	-0.1	**-0.4**	xyloglucan endotransglycosylase, putative	c6	S2
	At2g20670	**-0.4**	-0.3	0.0	-0.1	expressed protein	c9	_2
	At4g16515	**-0.4**	**-0.3**	-0.1	0.0	expressed protein	c6	S1
	At3g48360	**-0.4**	-0.3	0.0	-0.2	expressed protein	cx	_2
	At4g08950	**-0.4**	0.1	**-0.3**	-0.2	phi-1 phosphate-induced protein -related	c3	S2
	At5g07000	**-0.4**	-0.3	0.0	-0.1	sulfotransferase family	c9	_1
	At1g80440	**-0.4**	-0.1	0.0	-0.1	Kelch repeat containing F-box protein family	c2	_4
	At3g15450	**-0.4**	**-0.4**	0.0	0.0	expressed protein	cx	_4
	AtCg00480	**-0.3**	-0.2	0.0	-0.1	ATP synthase CF1 beta chain (ATPB)		
Light 3h	At4g14130	**-0.5**	**-0.9**	-0.2	**-0.5**	xyloglucan endotransglycosylase (XTR7)	c9	S1
	At3g63510	**-0.5**	**-0.8**	-0.2	**-0.4**	nitrogen regulation protein family (NIFR3)	c9	C5
	At5g02020	**-0.5**	**-0.7**	-0.1	-0.3	expressed protein	c9	C1
	At5g54190	-0.3	**-0.6**	0.0	0.3	protochlorophyllide reductase A (PORA)	c1	C1
	At3g47340	-0.3	**-0.6**	0.0	**-0.3**	glutamine-dependent asparagine synthetase (ASN1)	c9	_5
	At1g10070	-0.3	**-0.6**	0.0	-0.2	tat-binding protein -related	c2	M4
	At5g07010	**-0.5**	**-0.6**	0.0	-0.3	sulfotransferase family	c9	_4
	At1g21400	-0.3	**-0.5**	-0.2	-0.3	branched-chain alpha keto-acid dehydrogenase -related	c1	M2
	At1g09570	**-0.3**	**-0.5**	0.1	0.0	phytochrome A (PHYA)	c3	M5
	At3g17100	**-0.3**	**-0.5**	0.0	-0.1	bHLH protein	c9	C2
	At3g60530	-0.2	**-0.5**	-0.3	-0.2	GATA zinc finger protein	c9	_2
	At5g63160	**-0.3**	**-0.4**	0.0	**-0.3**	expressed protein	c3	_2
	At4g36780	-0.2	**-0.4**	0.0	-0.1	expressed protein	c2	_5
	At1g03090	-0.3	**-0.4**	0.0	-0.1	3-methylcrotonyl-CoA carboxylase -related (MCCA)	c9	M2
	At4g27440	-0.3	**-0.4**	0.0	0.3	protochlorophyllide reductase B (PORB)	c9	C1
CK 1h	At3g45650	0.0	0.0	**-0.5**	0.0	hypothetical protein	c4	_1
	At3g03220	-0.1	0.0	**-0.5**	0.0	expansin (EXP13)	c5	C4
	At5g59020	0.0	0.0	**-0.5**	0.0	expressed protein	c9	_1
	At2g21850	0.1	0.0	**-0.4**	0.1	CHP-rich zinc finger protein, putative	c9	_3
	At3g18890	0.0	0.3	**-0.4**	0.1	expressed protein	c1	C3
	At3g25640	-0.1	0.0	**-0.4**	-0.2	expressed protein	c4	_2
	At3g01270	0.0	0.1	**-0.4**	0.0	polysaccharide lyase family 1 (pectate lyase)	c6	S1
	At5g04270	-0.1	-0.1	**-0.4**	-0.1	DHHC-type zinc finger domain-containing protein	c3	S2
	At4g14220	0.0	0.1	**-0.4**	0.0	C3HC4-type zinc finger protein family	c9	C4
	At5g01210	-0.2	-0.2	**-0.4**	-0.3	transferase family	c9	_5
	At1g03310	0.1	0.0	**-0.4**	-0.2	glycoside hydrolase family 13	c9	C2
	At5g49830	0.0	0.1	**-0.4**	0.0	expressed protein		_2
	At1g68300	-0.2	-0.1	**-0.4**	0.0	expressed protein	c3	_2
	At2g27060	0.0	0.0	**-0.4**	0.0	leucine-rich repeat transmembrane protein kinase, putative	c9	_2
	At3g08960	-0.2	-0.1	**-0.4**	0.0	importin beta family	c9	C5
CK 3h	At1g69570	0.0	-0.2	-0.3	**-0.6**	Dof zinc finger protein	c2	_2
	At2g18500	0.0	0.0	0.2	**-0.5**	expressed protein	c7	C3
	At4g14130	**-0.5**	**-0.9**	-0.2	**-0.5**	xyloglucan endotransglycosylase (XTR7)	c9	S1
	At5g27960	-0.1	0.0	-0.1	**-0.5**	MADS-box protein		_4
	At5g15950	0.1	-0.1	-0.1	**-0.5**	S-adenosylmethionine decarboxylase (adoMetDC2)	c9	_2
	At3g53980	-0.1	-0.1	0.2	**-0.5**	protease inhibitor/seed storage/lipid transfer protein (LTP) family	c4	S1
	At1g27910	0.1	-0.1	0.0	**-0.4**	armadillo repeat containing protein	c7	_1
	At3g23730	**-0.4**	-0.3	-0.1	**-0.4**	xyloglucan endotransglycosylase, putative	c6	S2
	At3g63510	**-0.5**	**-0.8**	-0.2	**-0.4**	nitrogen regulation protein family (NIFR3)	c9	C5
	At1g29140	0.1	0.1	0.0	**-0.4**	expressed protein	c9	S1
	At3g54790	0.0	-0.1	0.0	**-0.4**	armadillo repeat containing protein	c9	_5
	At1g16950	0.0	0.0	-0.2	**-0.4**	expressed protein	c7	S1
	At1g78680	0.0	0.0	-0.1	**-0.4**	γ-glutamyl hydrolase	c9	C4
	At5g08440	-0.1	0.0	0.0	**-0.4**	expressed protein	c9	_2
	At1g10970	-0.1	0.0	-0.1	**-0.3**	metal transporter (ZIP4)	c7	_1

Table III-3. *GRE analysis about light and cytokinin responses with organ GRE.*

Gene groups were clustered by hierarchical clustering of 11 GRE values. GRE value larger than 0.10 (about 1.26 fold in normal value, bolded and shaded) and which have genes in the macroarray more than 5 are shown. L1, 1h light. L3, 3h light. C1, 1h cytokinin. C3, 3h cytokinin. Y, young leaf. L, leaf. T, stem. B, bud. F, flower. S, silique. R, root. Target, subgroups devided using results of targetP prediction as described III-2.6: C, chloroplast. M, mitochondria. S, secretory. O, others.

Response Type	GRE value											description	Target	groupID	gene #
	L1	L3	C1	C3	Y	L	T	B	F	S	R				
A	0						-0	-0	-0	-0	-1	chlorophyll binding activity	C	GO:0016168	12
	0						-0	-0	0.1	-0	-1	photosynthesis	C	GO:0015979	7
	0		0.1				-0	0	0.1	-0	-1	photosynthesis	C	ath00195	19
B	0		0	0.1		0.1	-0	0	-0	-0	-0	glycine, serine and threonine metabolism	M	ath00260	7
	0.1		-0	0			-0	-0	-0	-0	-0	electron transfer flavoprotein	O	GO:0008246	6
	0		0	0			-0	-0	-0	-0	-0	glyoxylate and dicarboxylate metabolism	O	ath00630	9
	0		-0	-0	0		-0	0	-0	-0	0.1	glyceraldehyde 3-phosphate catabolism	O	AraCyc	6
	0		-0	-0	0.1		-0	-0	-0	-0	0	gluconeogenesis	O	AraCyc	10
	0		-0	-0			-0	-0	-0	-0	-0	calvin cycle	O	AraCyc	6
	0		0	-0		0	0	-0	-0	-0	0	response to abscisic acid stimulus	O	GO:0009737	9
	0		0	0.1		0.1	-0	-0	-0	-0	-0	electron transport	C	GO:0006118	26
	-0		0	0			-0	-0	-0	-0	-0	pentose phosphate pathway	C	ath00030	14
	-0		-0	0	0	0.1	-0	-0	-0	-0	0.1	glycolysis	O	GO:0006096	6
	-0		0	0	0	0	-0	-0	-0	-0	0.1	acetate fermentation	O	GO:0019654	8
	0		0			0.1	-0	0	0	-0	-0	glycolysis	C	AraCyc	12
	0		0			0.1	-0	0.1	0	-0	-0	glycolysis / gluconeogenesis	C	ath00010	18
	0		0	0.1	0	0	-0	0	0	-0	-0	biosynthesis	C	GO:0009058	10
	-0		0	0.1			-0	0.1	0.1	-0	-0	carbon fixation	C	ath00710	22
	-0		0	0.1		0	-0	0.1	0.1	-0	-0	aerobic glycerol catabolism	C	AraCyc	7
	0		0				-0	0	0	-0	-0	gluconeogenesis	C	AraCyc	10
	0		0				-0	0	0	-0	-0	calvin cycle	C	AraCyc	13
	0		0.1				-0	0	-0	0	-0	fructose and mannose metabolism	C	ath00051	7
	0.1		0				-0	-0	-0	0.1	-0	reductive pentose-phosphate cycle	C	GO:0019253	6
	0		0			0.1	-0	-0	-0	0.1	-0	anaerobic glycolysis	C	GO:0019642	6
	0		0			0.1	-0	-0	-0	0.1	-0	butanediol fermentation	C	GO:0019650	6
	0		0			0.1	-0	-0	-0	-0	-0	glycolysis	C	GO:0006096	8
	0		0			0	-0	0	-0	0.1	-0	acetate fermentation	C	GO:0019654	8
	0.1		-0	0.1		0	-0	-0	-0	0.1	-0	translational elongation	C	GO:0006414	6
	0		0	0	-0	-0	0	-0	-0	0		flavonoids, stilbene and lignin biosynthesis	O	ath00940	12
	0	0	0		-0	-0	-0	0	-0	0.1	0	RNA processing	O	GO:0006396	9
C	-0	0			-0	-0		0	-0	0	0.1	two-component response regulator activity	O	GO:0000156	6
	-0	-0		0.1	-0	-0	0.1	-0	-0	-0	0.1	two-component signal transduction system	O	GO:0000160	11
	-0	0	0.1		-0	-0		-0	-0	-0		methionine metabolism	O	ath00271	11
	-0	0	0.1		-0	-0	-0	-0	-0	-0		lectin	O	GO:0005530	7
D	-0	0	0		-0	-0	-0			-0		ribosome	M	ath03010	27
	-0	0	0		-0	-0	-0			-0		intracellular	M	GO:0005622	27
	-0	0	0		-0	-0	-0			-0		protein biosynthesis	M	GO:0006412	29
	-0	0	0		-0	-0	-0			-0		ribosome	M	GO:0005840	27
	-0	0	0		-0	-0	-0			-0		structural constituent of ribosome	M	GO:0003735	29
	-0	0	0		-0	-0	-0			-0		structural constituent of ribosome	O	GO:0003735	55
	-0	0	0		-0	-0	-0			-0		protein biosynthesis	O	GO:0006412	62
	-0	0	0		-0	-0	-0			-0		ribosome	O	GO:0005840	63
	-0	0	0		-0	-0	-0			-0		ribosome	O	ath03010	86
	-0	0	-0		-0	-0	-0	0.1	0	-0		translational elongation	O	GO:0006414	15
E	-0	0.1	0			0	-0	0	0	-0	-0	RNA binding activity	C	GO:0003723	16
	-0	0.1	0			0	-0	0	0.1	-0	-0	biosynthesis of proto- and siroheme	C	AraCyc	9
	-0	0	-0			0.1	-0	0	0.1	-0	-0	biosynthesis of chlorophyll	C	AraCyc	16
	-0	0	-0			0.1	-0	0	0	-0	-0	porphyrin and chlorophyll metabolism	C	ath00860	17
	-0	0	-0			0	-0			-0	-0	ribosome	C	GO:0005840	22
	-0	0	0		0.1	-0	-0			-0	-0	ribosome	C	ath03010	18
	-0	0	0			0	-0			-0	-0	structural constituent of ribosome	C	GO:0003735	29
	-0	0	0			0	-0	0.1	0.1	0	-0	protein biosynthesis	C	GO:0006412	31
	-0	0	0.1		-0	-0	-0	0.1	0.1	-0	0	N-methyltransferase activity	O	GO:0008170	6
	-0	-0	0.1		-0	-0	-0	0.1	0	0	-0	histone deacetylase activity	O	GO:0004407	7

between light, cytokinin and organs. Response type A, which was induced by both light and cytokinin very strongly. Because chlorophyll biosynthesis rapidly progresses after illumination using pre-accumulated key enzyme POR and its substrates protochlorophyllide and NADPH, chlorophyll binding protein may rapidly be required. Type B which was induced by light was composed of gene groups related to carbon fixation mainly in chloroplast. Gene group of mitochondria, "Glycine, serine and threonine metabolism" was also found in this category. This group was composed of genes involved in photorespiration in mitochondria (Fig.III-7, Srinivasan et al., 1995). Gene groups in type A and B had a clear tendency to leaves in the 7 organs, indicating chloroplast and photosynthesis are main organelle and function in leaves. While type C, which responded to 1h cytokinin, has no significant tendency to particular organs. Type D and E, which were induced by 3h cytokinin, were related to protein biosynthesis like ribosome proteins. Type D was composed of ribosomes in mitochondria and other intercellular location probably cytoplasm, which had tendency to proliferative organs; bud, flower and root. On the other hand, type E, which was also induced by 3h cytokinin, was composed of chloroplast genes. Most of gene groups in type E had tendency to young leaf, indicating proliferation of chloroplast was the strongest in young leaf.

This GRE analysis clearly showed typical characteristics of two different signals of light as energy captured by chloroplast and cytokinin as phytohormone for reproduction.

III-3.6 Correlation between cytokinin response and organ specificity

Type D in Table III-3 shows clear relationship between cytokinin response and organ specificity about ribosome genes. In simple idea, the ribosome genes may be induced in bud, flower and root using cytokinin signal transduction. To assess the possibility, I examined the correlation of REs of cytokinin and organ specificity. Figure III-8 shows that only root has correlation with cytokinin response, indicating cytokinin-dependent induction of ribosomes in root. On the other hand, bad and flower did not have correlation with cytokinin response, indicating cytokinin-independent induction of ribosomes in bud and flower. Auxin and brassinosteroids are also known as phytohormones to induce reproduction of cells apart from cytokinin. Thus in bud and flower, ribosomes may be induced via those phytohormones.

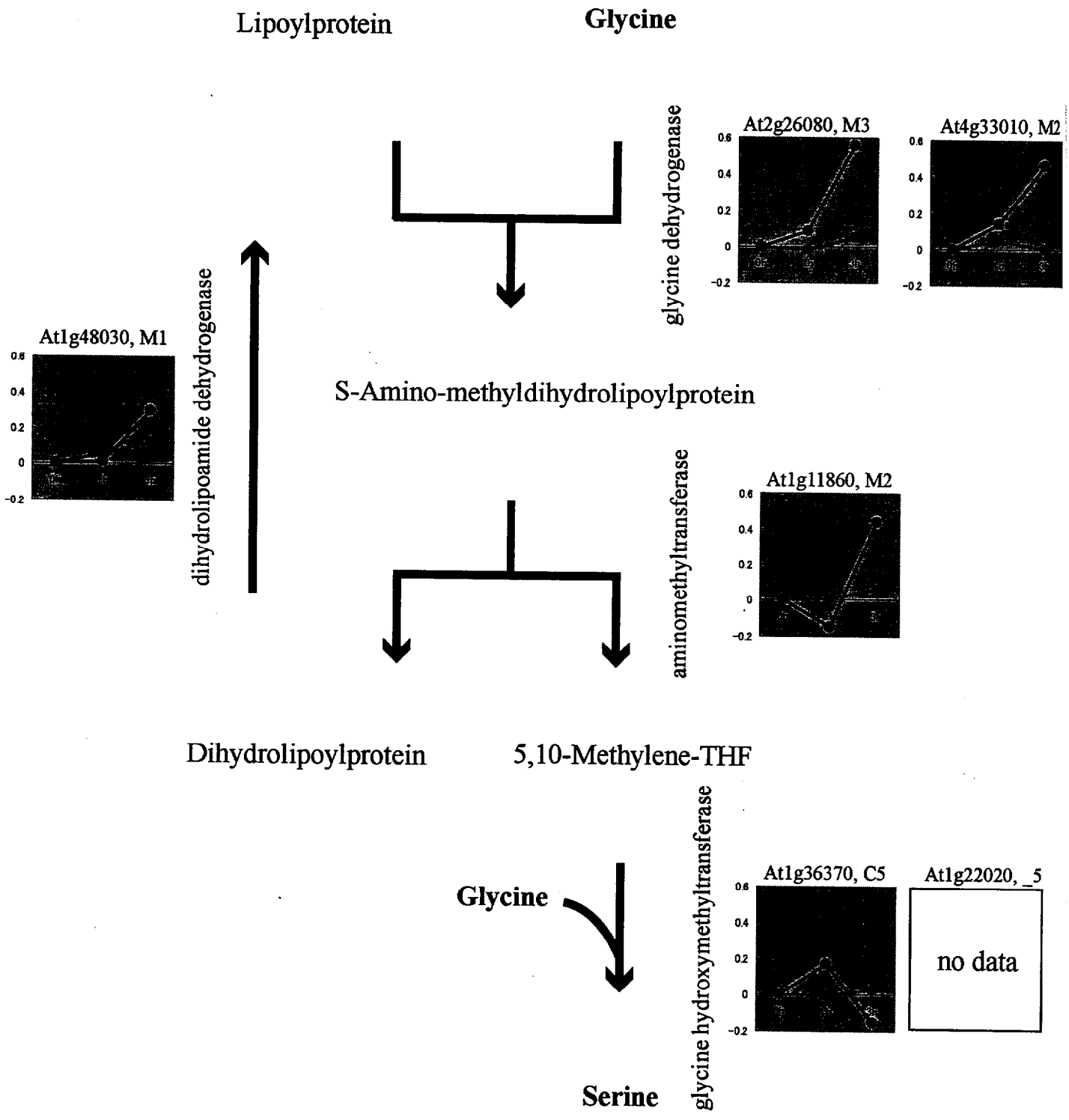
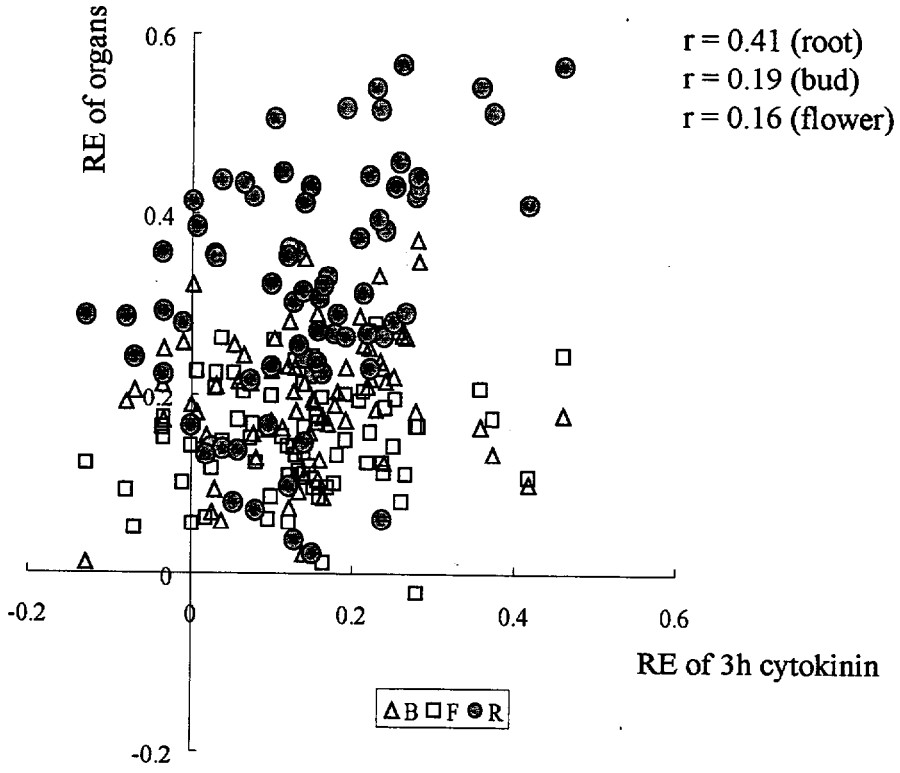


Fig. III-7. Photorespiration induced by light in mitochondria. Corresponding locus with the data for TargetP and RE (black, light. gray, cytokinin). TargetP, predicted subcellular localization: C, chloroplast. M, mitochondria. _, others. number, reliability (1: highest).



	# of genes
Ribosome (ath03010)	161
with B-F-R specificity	116

Fig. III-8. Correlation between cytokinin response and organ specificity of ribosome genes which have organ specificity to bud-flower-root. All 116 ribosome genes with specificity to bud-flower-root indicating only root have strong correlation to cytokinin response. Triangle, bud. Square, flower. Circle, root.

Auxin is rich in apex of stem, while brassinosteroids distribute all tissues, especially exist in pollen and seeds.

III-4 DISCUSSION

The most important point of the analysis on the light and cytokinin gene expression is a high correlation between light and cytokinin responses on the level of both individual genes and functional gene groups. Cytokinin treatment well mimics light illumination at least first step of photomorphogenesis with regard to cell expansion, opening of cotyledons, development of thylakoid membrane in chloroplast. But photomorphogenesis can not be completed by cytokinin, because of the presence of a strictly light-dependent step of the enzyme reaction in chlorophyll biosynthesis.

Although main photomorphogenesis processes were promoted by both light and cytokinin, the functional specificities of the two stimuli were also observed. Light illumination is real energy for photosynthesis, thus a lot of enzymes for carbon metabolism were induced in chloroplast and probably cytoplasm, while fundamental function of cytokinin was thought to promote proliferation which included chloroplast proliferation.

III-4.1 Conjunction point of light and cytokinin signal transductions

A question arose about the high correlation of light and cytokinin responses. How did the two responses so resemble each other? Three possibilities for this question could be answered. (A) Light illumination induces *de novo* synthesis of cytokinin. (B) Light and cytokinin use a common signal transduction pathway to regulate common genes. (C) Light and cytokinin use independent signal transduction pathways to regulate common genes (Fig. III-9).

About the A-type possibility, the last step of cytokinin biosynthesis is catalyzed by an enzyme, isopentenyl transferase (IPT), whose homologs are coded by 9 genes in Arabidopsis. The expressions of *IPT2* and *IPT9* were observed in the macroarray system. The REs are represented in Table III-4. Both of the IPTs, which are reported as enzymes for tRNA synthesis (Takei et al., 2001), did not respond to light or cytokinin in this study. Su and Howell (1995) examined effects of cytokinin and light on hypocotyl

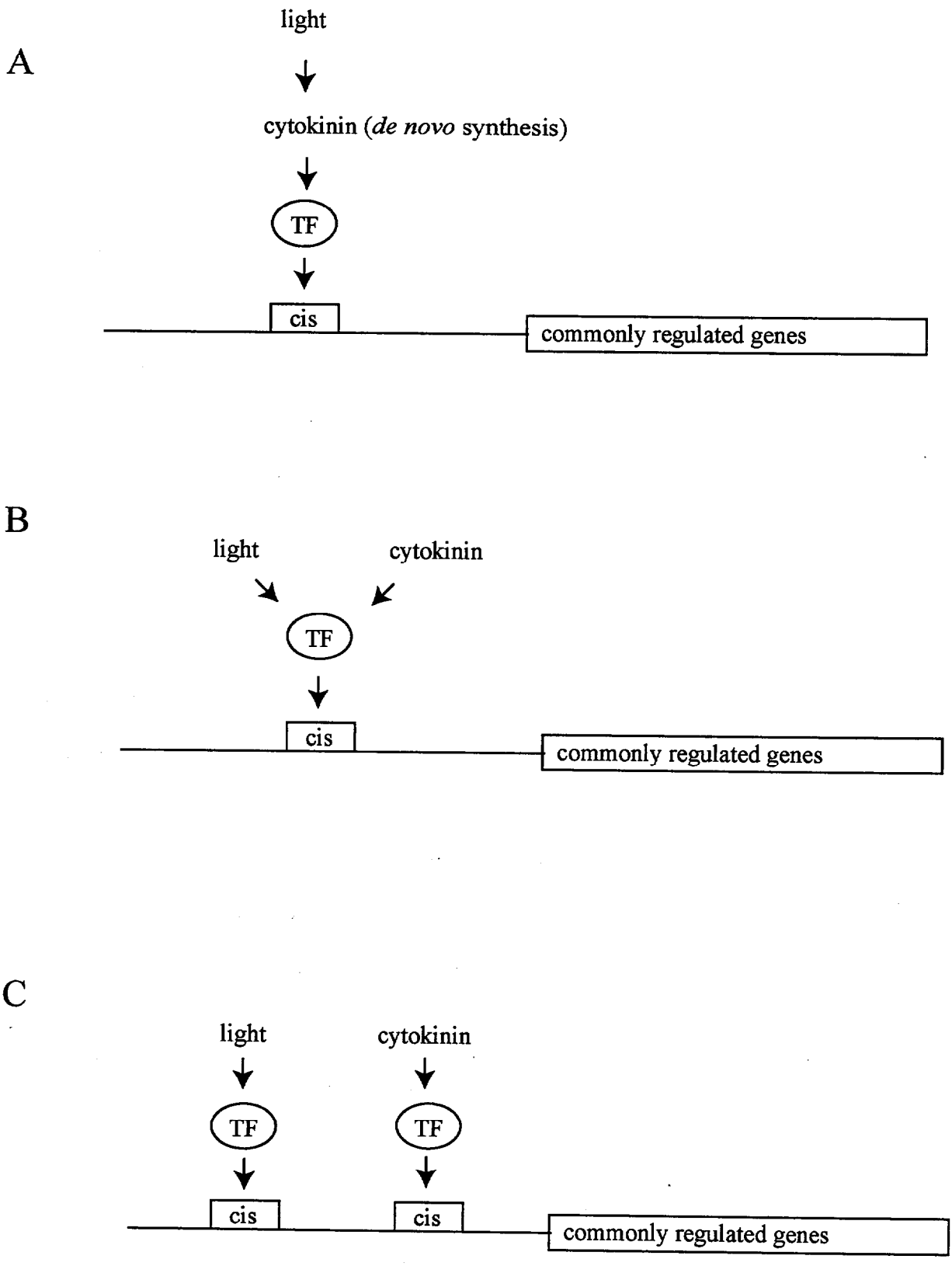


Fig. III-9. Three possibilities for commonly regulated genes by light and cytokinin. TF, trans factor. cis, cis regulatory element.

Table III-4. *RE of light- and cytokinin-related genes.*

Absolute RE value larger than 0.25 (about 1.8 fold in normal value) are bolded and shaded. targetP, predicted subcelluar localization: C, chloroplast. M, mitochondria. S, secretory. _, others. number, reliability (1: highest).

			RE				
Locus	alias	targetP	L1	L3	CK1	CK3	remarks
Cytokinin synthesis							
At1g68460	AtIPT1	C1					activity confirmed by recombinant
At2g27760	AtIPT2	_1	0.14	0.14	-0.01	-0.07	iPA synthesis for tRNA (Takei et al., 2001)
At3g63110	AtIPT3	_4					
At4g24650	AtIPT4	_4					activity confirmed by recombinant (Kakimoto, 2001)
At5g19040	AtIPT5	M4					
At1g25410	AtIPT6	C4					
At3g23630	AtIPT7	S5					
At3g19160	AtIPT8	C2					*de novo* synthesis (Sun et al., 2003)
At5g20040	AtIPT9	M2	-0.04	-0.02	-0.07	-0.05	iPA synthesis for tRNA (Takei et al., 2001)
Cytokinin oxidase, CKX							
At2g19500	CKX2	S5					
At5g56970	CKX3	M5					
At4g29740	CKX4	S2					
At1g75450	CKX6	S1					
cytokinin receptor							
At2g01830	CRE1	_2	-0.03	-0.09	0.08	0.05	
At5g35750	AHK2	_3	-0.07	0.06	-0.03	-0.02	
At1g27320	AHK3	S1					
histidine-containing phosphotransmitter (AHP)							
At3g21510	AHP1	_3	0.01	0.00	0.04	0.09	
At3g29350	AHP2	_3					
At5g39340	AHP3	_3					
Arabidopsis response regulator (ARR)							
At3g16857	ARR1	_2					type B
At4g16110	ARR2	_3					type B
At1g59940	ARR3	_2					type A
At1g10470	ARR4	_4	-0.05	**-0.36**	0.20	**0.27**	type A
At3g48100	ARR5	_4	-0.08	-0.16	**-0.51**	**-0.51**	type A
At5g62920	ARR6	_3					type A
At1g19050	ARR7	_2					type A
At2g41310	ARR8	_3					type A
At3g57040	ARR9	_3	0.04	0.06	**0.48**	**0.60**	type A
At4g31920	ARR10	_2					type B
At1g67710	ARR11	_2					type B
At2g25180	ARR12	_2					type B
At2g27070	ARR13	S5					type B
At2g01760	ARR14	_2					type B
At1g74890	ARR15	_4					type A
At2g40670	ARR16	_3					type A
At3g56380	ARR17	_2					type A
photoreceptor							
At1g09570	PHYA	M5	**-0.55**	**-0.53**	0.09	0.04	from cytosol into nucleus by light
At2g18790	PHYB	C5					from cytosol into nucleus by light
At5g35840	PHYC	M5	0.07	-0.01	-0.02	-0.07	
At4g16250	PHYD	C3					
At4g18130	PHYE	_4	-0.01	0.08	-0.05	0.01	
At4g08920	CRY1	_5	0.06	-0.02	0.00	-0.24	
At1g04400	CRY2	_3					
At3g45780	PHOT1	C3	-0.10	-0.02	0.10	0.13	
At5g58140	PHOT2	C5	0.19	0.08	0.06	-0.02	

elongation in Arabidopsis seedling. Both light and cytokinin are known to cause inhibition of hypocotyl elongation. Even if using mutants of cytokinin signal transduction (ckr1/ein2), light could cause inhibition of hypocotyl elongation, indicating light-mediated inhibition of hypocotyl elongation did not use cytokinin signal transduction. And also Chory et al. (1994) reported that cytokinins accumulated to similar levels grown in either the light or the dark in Arabidopsis. Although their data were not sampled at the moment of light illumination but 7d-old plant grown in the light or in the dark, these results suggested that biosynthesis of cytokinin in developmental seedlings is independent of light-illumination. For these reasons, this possibility is less feasible.

For the B-type possibility, Arabidopsis Response Regulators (ARRs) are key transducer for light and cytokinin responses. ARRs are classified into 2 types, type A and type B. Generally, type B ARRs, which are transcription factors themselves, induce type A ARRs in response to cytokinin (Haberer and Kieber, 2002). ARR4 is the one of such type A ARRs, and ARR4 showed cytokinin response with other type A ARR genes (Table III-4). In addition, the ARR4 gene was also induced in response to light, although neither ARR4 nor ARR5 in etiolated seedling has been reported as light-induced ARR (Brandstatter and Kieber, 1998). Moreover, ARR4 induction was earlier to cytokinin than to light (Fig. III-10), suggesting *de novo* protein synthesis of transcription factor for light response, because cytokinin response did not need *de novo* protein synthesis (Haberer and Kieber, 2002). ARR4 overexpressor was reported to show response of cytokinin without external cytokinin treatment, indicating ARR4 plays a role in cytokinin signal transduction (Osakabe et al, 2002). Moreover, interaction between ARR4 and PHYB, a red light photoreceptor, was reported (Sweere et al, 2001). ARR4 is considered to stabilize Pfr, an active form of PHYB. Therefore, ARR4 is now regarded as one of the junction points of light and cytokinin, and it commonly induced the genes responding both light and cytokinin. It should be noted that no obvious changes of other cytokinin signal transduction related genes were observed (Table III-4).

Under the B-type possibility, several transcription factors responding both light and cytokinin were expected, although some of early responses did not seem to need *de novo* transcription. Table III-5 shows the list of transcription factors induced or repressed by light or cytokinin. Expression of 35

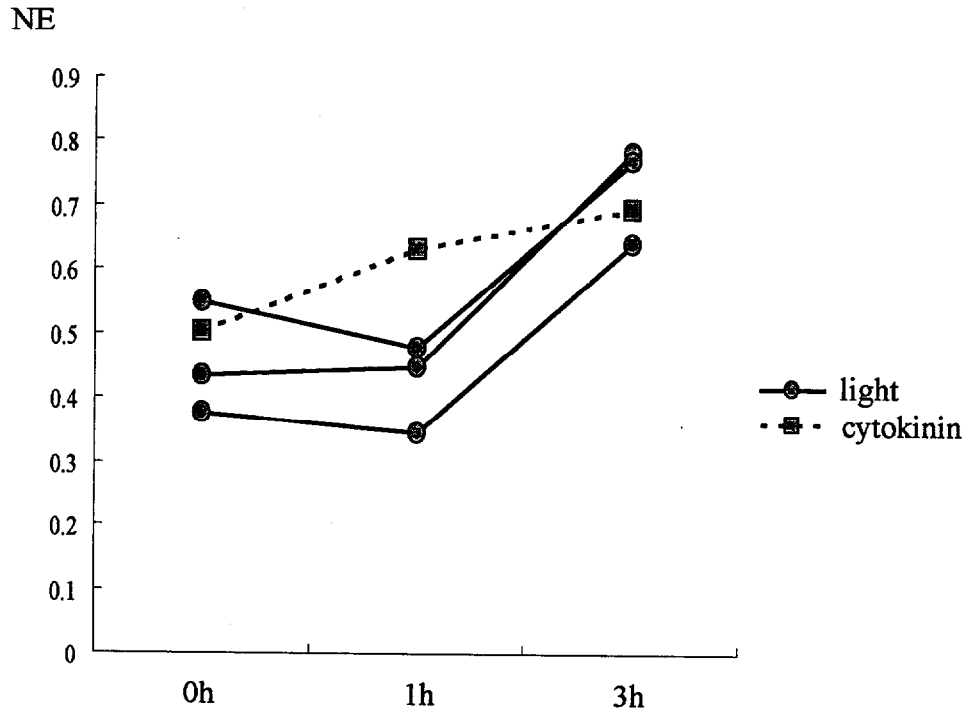


Fig. III-10. mRNA levels of ARR4 (At1g10470) were increased by light as well as cytokinin. Macroarray data of 3 experiments for light and one experiment for cytokinin are shown.

Table III-5. *Transcription factors induced or repressed by light or cytokinin.*
RE value more than 0.25 or less than -0.25 are shown (bolded and shaded).

induction	locus	description	L1	L3	C1	C3
light and cytokinin (3h)	At1g59750	auxin response transcription factor (ARF1)	0.2	**1.6**	**0.5**	**0.5**
	At5g11510	myb family transcription factor (MYB3R-4)	0.0	**1.5**	0.2	**0.5**
light (1h)	At5g44190	myb family transcription factor (GLK2, GPRI2)	**1.5**	0.1	0.1	0.0
light (3h)	At5g60890	receptor-like protein kinase (ATR1)	0.1	**1.6**	0.1	0.2
	At1g69180	transcription factor CRC	0.1	**1.5**	0.1	0.2
	At4g14410	bHLH protein	0.1	**1.5**	0.1	0.2
	At2g28350	auxin response transcription factor	0.2	**1.5**	0.0	0.1
	At3g25710	bHLH protein family	0.1	**1.5**	-0.1	0.0
cytokinin (1h)	At3g48100	Arabidopsis response regulator 5 (ARR5)	-0.1	-0.2	**0.7**	**0.3**
	At3g25990	DNA-binding protein, GT-1	0.0	0.0	**0.5**	**0.5**
	At3g02150	TCP family transcription factor TFPD	0.0	0.1	**0.5**	0.2
	At1g54160	CCAAT-binding factor B subunit	0.0	-0.1	**0.5**	0.1
	At4g17460	homeobox-leucine zipper protein HAT1	-0.2	-0.2	**0.5**	0.2
cytokinin (3h)	At5g61430	NAM, no apical meristem, - like protein	0.0	0.0	0.1	**0.6**
	At1g21970	CCAAT-box binding factor HAP3 homolog	0.0	0.1	0.0	**0.5**
	At5g63780	C3HC4-type zinc finger protein family	0.0	0.1	0.2	**0.5**
	At4g13100	C3HC4-type zinc finger protein family	0.0	0.1	0.2	**0.5**
	At1g14580	zinc finger protein	0.0	0.1	0.1	**0.5**
	At1g30650	WRKY family transcription factor (WRKY14)	0.0	0.0	0.1	**0.5**
	At1g13450	DNA binding protein GT-1	-0.2	-0.2	0.1	**0.5**
	At3g16090	C3HC4-type zinc finger protein family	-0.1	0.0	0.0	**0.5**

repression	locus	description	L1	L3	C1	C3
light (1h)	At4g39070	CONSTANS B-box zinc finger family protein	**-1.5**	-0.2	0.2	0.1
light (3h)	At3g60530	GATA zinc finger protein	**-1.5**	**-0.5**	-0.2	-0.2
	At2g45050	GATA zinc finger protein	-0.1	**-0.5**	0.0	-0.1
	At1g13260	AP2 domain transcription factor (RAV1)	-0.2	**-0.5**	0.0	0.0
	At4g32890	GATA zinc finger protein	-0.1	**-0.5**	0.0	0.0
	At2g40970	myb family transcription factor	-0.1	**-0.5**	0.0	-0.2
	At2g01570	gibberellin response modulator (RGA1)	-0.1	**-0.5**	0.0	-0.1
cytokinin (1h)	At3g07940	zinc finger and C2 domain protein	-0.1	0.0	**-0.5**	-0.2
	At5g63790	No apical meristem (NAM) protein family	0.0	0.0	**-0.5**	-0.2
	At1g73410	myb family transcription factor (MYB54)	-0.1	0.0	**-0.5**	-0.1
cytokinin (3h)	At5g27960	MADS-box protein	-0.1	0.0	0.0	**-0.6**
	At5g65790	myb DNA-binding protein (MYB68)	0.0	-0.1	0.0	**-0.5**
	At4g17880	bHLH protein family	-0.1	0.0	0.0	**-0.5**
	At3g62610	myb family transcription factor	0.1	0.0	-0.1	**-0.5**

transcription factors changed in 391 transcription factors on the macroarray. Existence of 2 light- and cytokinin-commonly-regulated transcription factors suggested B-type possibility, while existence of 33 light- or cytokinin-specifically regulated transcription factors suggested C-type possibility. However the list includes 2 CCAAT-box binding factors. The cis element CAAT-box, which is the other name of CCAAT-box more familiar for plant researcher, was fundamental cis element found not only in plants but also in animals. In plant, CAAT-box on promoter of an ATP synthase gene (*AtpC*), is essential region for both light and cytokinin induction of the gene via CAAT-box binding factor (CBF) and repressor protein ATPC-2 (Kusnetsov et al., 1999, Bezhani et al., 2001). Thus regulation of CAAT-box binding factor causes regulation of commonly responding gene, namely, regulation via CAAT-box could be one of regulatory point of B-type possibility.

Although assessment of the B- and C-type possibility is difficult from expression profile, analysis of cis elements provides other aspects, namely, commonly responding cis elements supported B-type possibility and light- or cytokinin-specific cis elements are supported C-type possibility. Thus promoter analysis are conducted in Chapter IV for light and cytokinin regulation at photomorphogenesis.

III-5 REFERENCES

Aharoni A, Vorst O (2001) DNA microarrays for functional plant genomics. Plant Mol Biol **48**: 99-118

Asamizu E, Nakamura Y, Sato S, Tabata S (2000) A large scale analysis of cDNA in *Arabidopsis thaliana*: Generation of 12,028 non-redundant Expressed Sequence Tags from normalized and size-selected cDNA libraries. DNA Res 7: 175-80

Awai K (2001) Molecular biological studies on multigene family of MGDG synthase in *Arabidopsis thaliana*. Doctoral Thesis, Tokyo Institute of Technology.

Balmer Y, Koller A, del Val G, Manieri W, Schurmann P, Buchanan BB (2003) Proteomics gives insight into the regulatory function of chloroplast thioredoxins. Proc Natl Acad Sci U S A 7: 370-5

Bezhani S, Sherameti I, Pfannschmidt T, Oelmuller R (2001) A repressor with similarities to prokaryotic and eukaryotic DNA helicases controls the assembly of the CAAT box binding complex at a photosynthesis gene promoter. J Biol Chem **276**: 23785-9

Brandstatter I, Kieber JJ (1998) Two genes with similarity to bacterial response regulators are rapidly and specifically induced by cytokinin in Arabidopsis. Plant Cell **10**: 1009-19

Canton FR, Quail PH (1999) Both phyA and phyB mediate light-imposed repression of PHYA gene expression in Arabidopsis. Plant Physiol **121**: 1207-16

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem **162**: 156-9

Chen CM, Ertl JR, Leisner SM, Chang CC (1985) Localization of cytokinin biosynthetic site in pea plants and carrot roots. Plant Physiol **78**: 510-3

Chory J, Reinecke D, Sim S, Washburn T, Brenner M (1994) A Role for Cytokinins in De-Etiolation in Arabidopsis. Plant Physiol **104**: 339-47

Church GM, Gilbert W (1984) Genomic sequencing. Proc Natl Acad Sci U S A **81**: 1991-5

Downes BP, Crowell DN (1998) Cytokinin regulates the expression of a soybean beta-expansin gene by a post-transcriptional mechanism. Plant Mol Biol **37**: 437-44

Haberer G, Kieber JJ (2002) Cytokinins. New insights into a classic phytohormone. Plant Physiol **128**: 354-62

Hoth S, Ikeda Y, Morgante M, Wang X, Zuo J, Hanafey MK, Gaasterland T, Tingey SV, Chua NH (2003) Monitoring genome-wide changes in gene expression in response to endogenous cytokinin reveals targets in Arabidopsis thaliana. FEBS Lett **554**: 373-80

Kiba T, Yamada H, Mizuno T (2002) Characterization of the ARR15 and ARR16 response regulators with special reference to the cytokinin signaling pathway mediated by the AHK4 histidine kinase in roots of *Arabidopsis thaliana*. Plant Cell Physiol **43**: 1059-66

Kusnetsov V, Landsberger M, Meurer J, Oelmuller R (1999) The assembly of the CAAT-box binding complex at a photosynthesis gene promoter is regulated by light, cytokinin, and the stage of the plastids. J Biol Chem **274**: 36009-14

Oosawa N, Masuda T, Awai K, Fusada N, Shimada H, Ohta H, Takamiya K (2000) Identification and light-induced expression of a novel gene of NADPH-protochlorophyllide oxidoreductase isoform in Arabidopsis thaliana. FEBS Lett **474**: 133-6

Osakabe Y, Miyata S, Urao T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2002) Overexpression of Arabidopsis response regulators, ARR4/ATRR1/IBC7 and ARR8/ATRR3, alters cytokinin responses differentially in the shoot and in callus formation. Biochem Biophys Res Commun **293**: 806-15

Miller CO (1956) Similarity of kinetin and red light effects. Plant Physiol **31**: 318-9

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol Plant **15**: 473- 93

Nishiuchi T, Nakamura T, Abe T, Kodama H, Nishimura M, Iba K (1995) Tissue-specific and light-responsive regulation of the promoter region of the Arabidopsis thaliana chloroplast omega-3 fatty acid desaturase gene (FAD7). Plant Mol Biol **29**: 599-609

Rashotte AM, Carson SD, To JP, Kieber JJ (2003) Expression profiling of cytokinin action in Arabidopsis. Plant Physiol **132**: 1998-2011

Sheen J, Zhou L, Jang JC (1999) Sugars as signaling molecules. Curr Opin Plant Biol **2**: 410-8

Srinivasan R, Oliver DJ (1995) Light-dependent and tissue-specific expression of the H-protein of the glycine decarboxylase complex. Plant Physiol **109**: 161-8

Su W, Howell SH (1995) The Effects of Cytokinin and Light on Hypocotyl Elongation in Arabidopsis Seedlings Are Independent and Additive. Plant Physiol **108**: 1423-30

Sweere U, Eichenberg K, Lohrmann J, Mira-Rodado V, Baurle I, Kudla J, Nagy F, Schafer E, Harter K (2001) Interaction

of the response regulator ARR4 with phytochrome B in modulating red light signaling. Science **294**: 1108-11

Takei K, Sakakibara H, Sugiyama T (2001) Identification of genes encoding adenylate isopentenyltransferase, a cytokinin biosynthesis enzyme, in *Arabidopsis thaliana*. J Biol Chem **276**: 26405-10

Yamaryo Y, Kanai D, Awai K, Shimojima M, Masuda T, Shimada H, Takamiya K, Ohta H (2003) Light and cytokinin play a co-operative role in MGDG synthesis in greening cucumber cotyledons. Plant Cell Physiol **44**: 844-55

Chapter IV

Comprehensive Analysis for cis-regulatory elements for Light- and Cytokinin-Responsive Genes

IV-1 INTORODUCTION

mRNA levels in a cell mainly depend on the balance between its synthesis and degradation. In general, the transcription of genes is activated or repressed by transcription factors which bind to a regulatory region of a gene, especially to a upstream region. Many efforts to discover cis regulatory elements using experimental methods such as reporter assay of deletion construct, gel sift assay, foot print and so on. In addition to such experimental approaches, recent advances in whole genome sequencing and transcriptome analysis provided new strategy for the promoter analysis. DNA sequence without any function is basically randomized in evolutional process. Conversely, conserved sequences may function for some biological activities. Thus, searching conserved sequences in genomic sequence(s) is one method to discover important regions in the genome function. DNA array data are also useful for such kind of purpose. Conserved sequences of upstream region of commonly regulated gene groups by some stimuli are probably good candidates for cis-regulatory elements. Several programs for searching conserved sequences have been reported. Some of them are available in their web sites. MEME (Muptiple Em for Mutif Elicitation, Bailey and Elkan, 1994), CONSENSUS, GIBBS sampler and CORESEARCH are such examples (Hertz and Stormo, 1999, Lawrence et al, 1993, Wolfersetter et al, 1996). A web site named "Melina" for running such programs at once is also provided (Poluliakh et al., 2003). Thus we easily get the conserved sequences among sequences commonly regulated by some stimuli. However, does the conserved sequence really work for the regulation? True cis elements for the regulation are expected to exist only in the promoter region of the regulated genes, and not exist in other genes. Thus large bias of the presence of cis elements is expected among regulated and non-regulated genes. I developed an easy method for analyzing such regions, and employed the method to discover some motifs in DNA sequences about light and cytokinin responses in Chapter III.

IV-2 MATERIALS AND METHODS

IV-2.1 Definition of gene groups

The same "gene group"s related to function defined in Chapter III were used. Namely, groups defined by public societies were divided into 4 intercellular localization; chloroplast (C), mitochondria (M), secretory (S) and others (O).

IV-2.2 CEG analysis

CEG (Correlation between Gene Expression and Gene Group) analysis was developed to search cis element with which genes are up-regulated by light or cytokinin. Figure IV-1 shows RE of 3h light on x-axis with value of presence (1) or absence (0) of a particular sequence. In this CEG analysis, the Pearson's correlation coefficient between the x-axis and the y-axis is used as a criterion. Since I used the correlation coefficient in unusual way, and the possible range of the value is much lower than the usual, CEG value was defined as 100 fold of the correlation coefficient. Figure IV-1 shows 3 examples. "AAAAAAA" was generally found in 500 base up-stream region from predicted transcriptional initiation site with a small CEG value, 1, because the presence of "AAAAAAA" has no relation to light response. In contrast, "G.CACGTG" (". " means any of "A", "C", "G" or "T") was often found upstream region of up-regulated genes by 3h light, showing relatively a large CEG value, 6. On the other hand, "CGTGGGTG" were rarely found in genes with tendency of down regulation, showing oppositely a large CEG value, -6. The CEG values enable us to rank the possibility.

IV-2.3 Correlation between cis elements and function of regulated genes

General statistical test for independence was applied to examine correlation between cis elements and function of regulated genes. Note that this test is compatible with χ^2 test. A value, Z, is criterion for degree of independence about the altered gene number between the total gene set and a gene group. The mathematical formula of Z is;

$$Z = \frac{p1 - p2}{\sqrt{p(1-p)(1/n1 + 1/n2)}}$$

where, n1 is the number of genes in the gene group, n2 is the number of genes not belonged to the gene

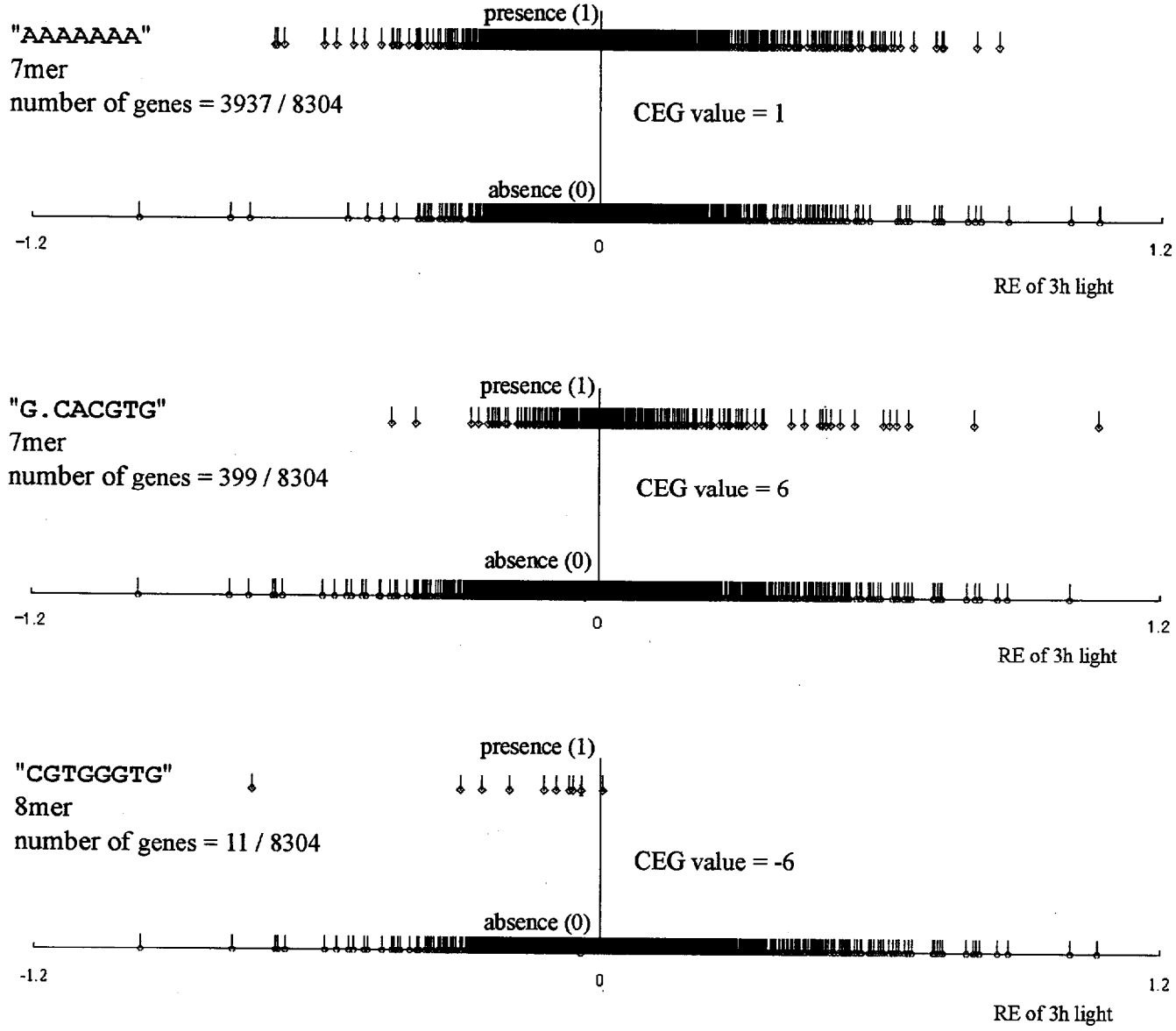


Fig. IV-1. Three examples of CEG (Correlation between Expression and Gene Group) analysis for searching conserved sequence responding to 3h light. "AAAAAAA" is an example for non-related sequence, inspite of its highly frequent appearance. "G.CACGTG" and "CGTGGGTG" are examples for the sequence which has correlation with light induction and repression, respectively. x-axis, RE value of 3h light. y-axis, presense or absense of a sequence in 500 upstream region of 8304 loci on macroarray. ".", any single nucleotide.

group, p1 is the population of altered genes in the gene group, p2 is the population of altered genes not belonged to the gene group, and p is the population of total altered genes in all genes. The calculated Z values follow standard normal distribution. In this study, for presentation purpose, $Z > 7$ was considered as significant correlation corresponding 99.9999999999% with sufficient significance. This Z value was adopted when I estimated the correlation between cis elements and function of regulated genes (see IV-3.3).

IV-2.4 Analysis for position specificity

Analysis for position specificity was done as follows. First, 3000b upstream regions from putative transcription initiation site were divided into 60 segments, each of which has 50bp in length, and total number of the cis elements was counted for all segments (Fig. IV-2A). Figure IV-2C1 shows appearance of "CTATAATAC" was strictly limited around 100 bp upstream region. To assess this bias of position, skewness was used, which is a measurement of symmetry. The value is positive when distribution is asymmetric with a tendency to have larger values than average (Fig. IV-2C1), oppositely negative when the distribution tends to have smaller values than average (Fig. IV-2C2), and almost zero under symmetric distribution (Fig IV-2C3). Skewness is defined as follows:

$$skewness = \frac{n}{(n-1)(n-2)} \sum_i \frac{(x_i - \bar{x})^3}{s^3}$$

where n is number of data (60 in this example), x with top bar is average of data, and s is standard deviation of data. Because skewness has bias to size of data in this analysis (the smaller in size, the higher in skewness) (Fig. IV-2B1), it is normalized as follows; each spot was subtracted by median of 41 points around the spot in the graph shown in Fig. IV-2B1 (Fig. IV-2B2). For 20 spots in the top or bottom, the medians for top or bottom 41 spots were respectively applied for the normalization.

IV-2.5 Public data used

The following files (and their release date) were downloaded from the TAIR ftp site (ftp://ftp.arabidopsis.org): Upstream sequence (17, Apr, 2003), Gene Ontology (4, Sep, 2003), global analysis by TargetP (9, Oct, 2002). Pathway data of KEGG (11, Sep, 2003) was downloaded from GenomeNet (ftp://ftp.genome.ad.jp/pub/kegg/). cis element sequences (30, Sep, 2003) were downloaded

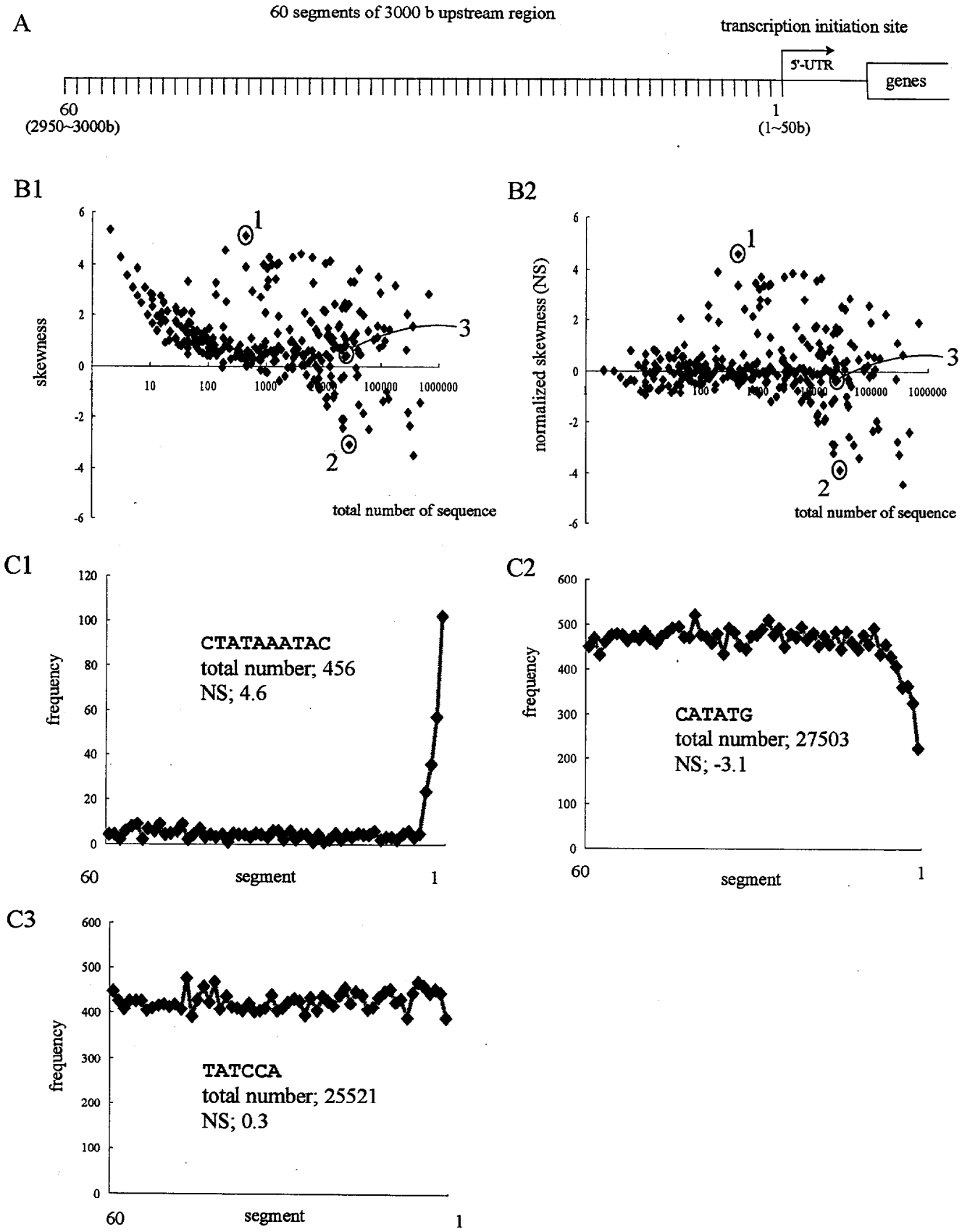


Fig. IV-2. Analysis for position specificity. Three handred ninety cis elements of PLACE were shown as example. A, schematic representation of 60 segments for counting appearance. B, distribution of 390 cis elements with total number of appearance of a cis element in 27819 Arabidopsis genes (x-axis) and skewness of frequency in 60 segments of the cis element (y-axis). B1, before size normalization. B2, after size normalization. C, 3 examples of frequency polygon of positions. NS, normalized skewness.

from PLACE (ftp://ftp.dna.affrc.go.jp/pub/dna_place/).

IV-3 RESULTS

IV-3.1 CEG analysis to light and cytokinin specific cis-elements

For strict regulation of gene expression, the presence of specific cis-elements in the upstream regions could be expected. To analyze the specific cis-elements for light and cytokinin responses, CEG analysis was achieved. Namely, 87,040 of all possible sequence combinations of 5-8mer composed of "A", "C", "G" and "T", and 122,880 combinations of 6-8mer composed of "A", "C", "G" and "T" which contain single insertion of "." (any) were combined, and correlation between the presence of these sequences in promoter regions and their responses by light or cytokinin were calculated by their CEG values. All sequences showing the CEG values more than 5 were as follows; 6 for 1h light and 18 for 3h light, 5 sequences for 1h cytokinin, 17 for 3h cytokinin. The result was reasonable because the responses at 1h were generally weaker than those at 3h, and thus difficult to find conserved sequences. Sequences for 3h light and 3h cytokinin were summarized in Table IV-1. From the sequences obtained, four major sequence groups could be categorized. Since these 4 groups were extracted in high frequency with slight modifications, the groups could be defined as more reliable groups than others. Two sequences "GCCACGTG" and "ACCACA" for 3h light contained the common core sequence "CCAC", however sequences for the light and those for the cytokinin had no similarity.

These sequences were compared with previously reported cis elements stored in PLACE (Plant Cis-acting Regulatory DNA Elements, Higo et al., 1999). Upper two boxes in Table IV-1 have been reported as cis-element; one sequence, "GCCACGTG", which was present in up-regulated genes by 3h light, is known as G-box, "CACGTG" (McKendree et al., 1990), the other, "AAACCCTA", which was frequently observed in up-regulated genes by cytokinin, was reported as "telo-box" (Regad et al., 1994, Tremousaygue et al., 1999, Manevski et al., 2000). These findings supported this CEG analysis. For other two sequences, "ACCACA", and "ATGGGCCT", similar motifs have not been reported so far, indicating that these are motifs firstly found by this approach.

Table IV-1. *CEG promoter analysis of 500bp upstream regions from putative transcription initiation site.*

Twenty five elements with CEG values more than 5 are mainly categorized in 4 groups. CEG values are represented; L1, 1h light. L3, 3h light. CK1, 1h cytokinin. CK3, 3h cytokinin.

for 3h light response						for 3h cytokinin response					
GCCACGTG	L1	L3	CK1	CK3	# of genes	AAACCCTA	L1	L3	CK1	CK3	# of genes
G.CACGTG	3	6	1	0	399	AAACCCTA	1	1	0	6	482
GCCACGTG	3	6	1	0	154	AACCCTA	1	2	1	6	672
GCCACG.G	3	5	1	0	188	A.ACCCTA	1	1	0	6	562
GC.ACGTG	2	5	1	0	200	AA.CCCTA	0	1	0	6	634
GCCAC	3	5	1	1	1461	A.CCCTA	1	2	2	6	970
AAGCC.CA	0	5	0	2	149	AAACCC.A	1	2	0	6	975
						AAACCCT	1	1	-1	5	788
ACCACA	L1	L3	CK1	CK3	# of genes	AACCC.A	0	2	1	5	1571
ACCACA	5	6	0	1	1043						
ACCACA.A	5	6	0	-1	451	ATGGGCT	L1	L3	CK1	CK3	# of genes
AACCACA	4	6	0	1	422	TGGGC.T	0	0	2	6	1171
ACCACAG	3	6	0	3	121	ATGGGC	0	0	3	6	964
CCACA.A	5	6	0	1	1213	TGGGC	0	0	2	5	1993
CCACA.AA	4	5	0	0	532	ATGGGC.T	-1	-1	2	5	557
GAACCACA	3	5	-1	1	62	TG.GCC	1	2	2	5	1697
						ATG.GC	0	-1	2	5	1816
and 5 sequences						and 3 sequences					

IV-3.2 Comparison of gene population which has particular cis element

To assure the elements searched by this new approach, I further analyzed correlation between intensity of RE values and the presence of these motifs. Top-500 or 50 genes in RE values for the light response were selected, and frequency of the appearance of these motifs was counted, and then, compared to that value for total genes (Fig. IV-3A). The figure indicates that top 50 genes have higher population of two cis elements, "GCCACGTG" and "ACCACA". This indicates that presence of these sequences have strong correlation with light inducibility of the genes.

In similar way, such frequency of 390 known cis elements stored in PLACE was counted in the sequences of light-induced genes. Calculating difference of RE values between total genes and of top 50 or 500 genes, top 20 cis elements with the largest difference were presented in Fig.IV-3B. In the 20 cis elements, well-known light-responsive elements, CAAT-box, GATA-box, I-box and circadian-related cis element were found, in addition to G-box that I also found by the present study. However, except for the G-box those showed smaller increase in frequency even when the target sequences were limited for top 50 genes with high RE values. This result indicates that CEG analysis was effective tool to find such cis elements which produce large expression bias in accord with their presence in promoter sequence. The results also showed that the cis elements newly found by this approach had equivalent frequency with other known motifs as light-responsive elements.

By the same way, cis elements in cytokinin-responsive genes were pictured in Fig. IV-4. The sequences also tended to be associated with cytokinin response, although the correlation to the intensity of RE values was smaller than light-responsive elements. It probably reflected lower induction of genes by cytokinin than by light. Top 20 cis elements obtained from PLACE contained I-box and GATA-box, which is related to light response, in addition to CAAT-box, a more general cis-elements for gene transcription in eukaryotes. As described in Chapter III, CAAT-box has been reported as an essential sequence for light and cytokinin responses. In fact, CAAT-box was also found in Fig. IV-3B as an element which is frequently observed in regulatory regions of light responsive genes.

IV-3.3 Correlation between cis elements and function of regulated genes

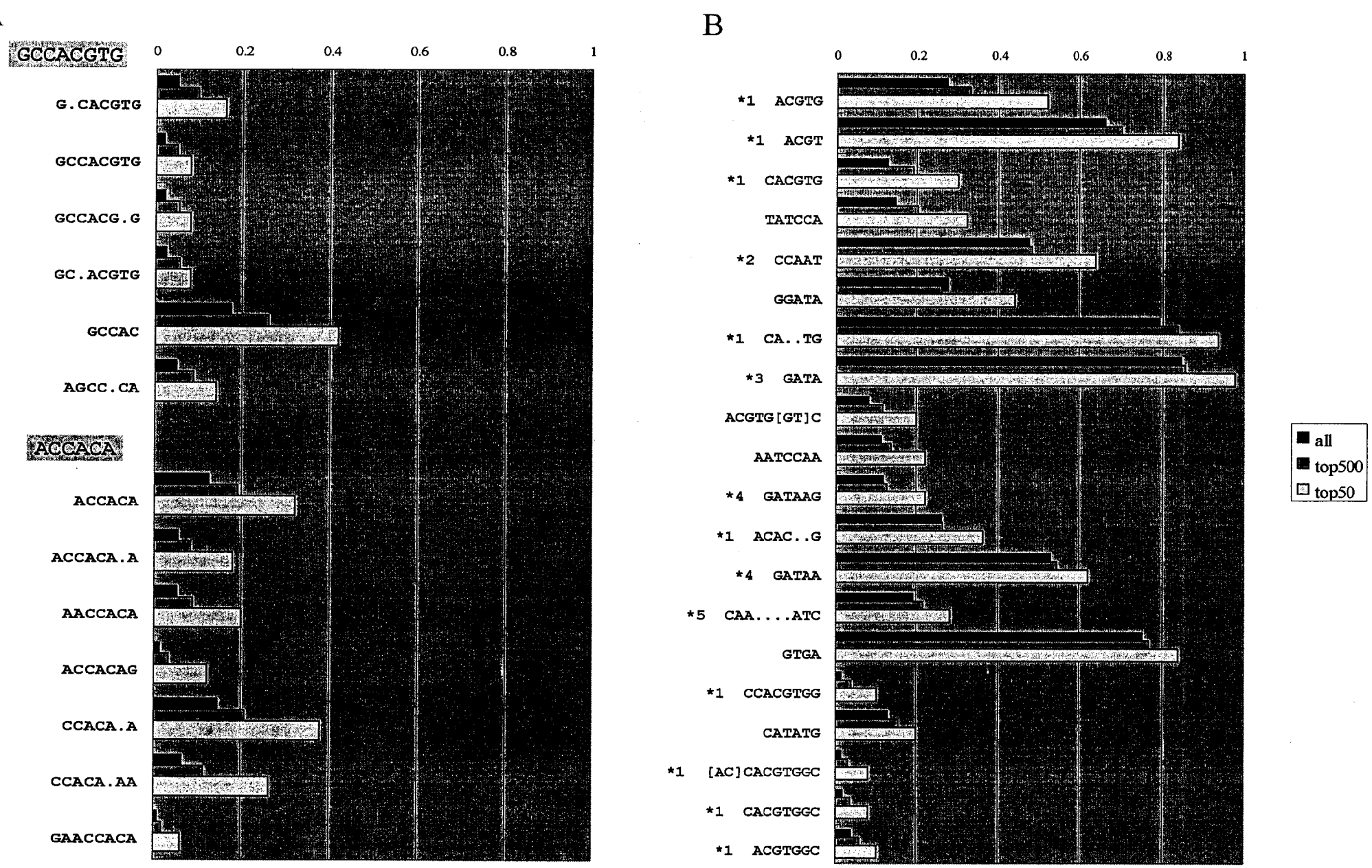


Fig. IV-3. Population of existence of cis elements in total loci and genes with top 50 and 500 REs of response to 3h light. A, selected sequences in this study. B, cis elements reported in PLACE. *1, G-box. *2, CAAT-box. *3, GATA-box. *4, I-box, *5, circadian related.

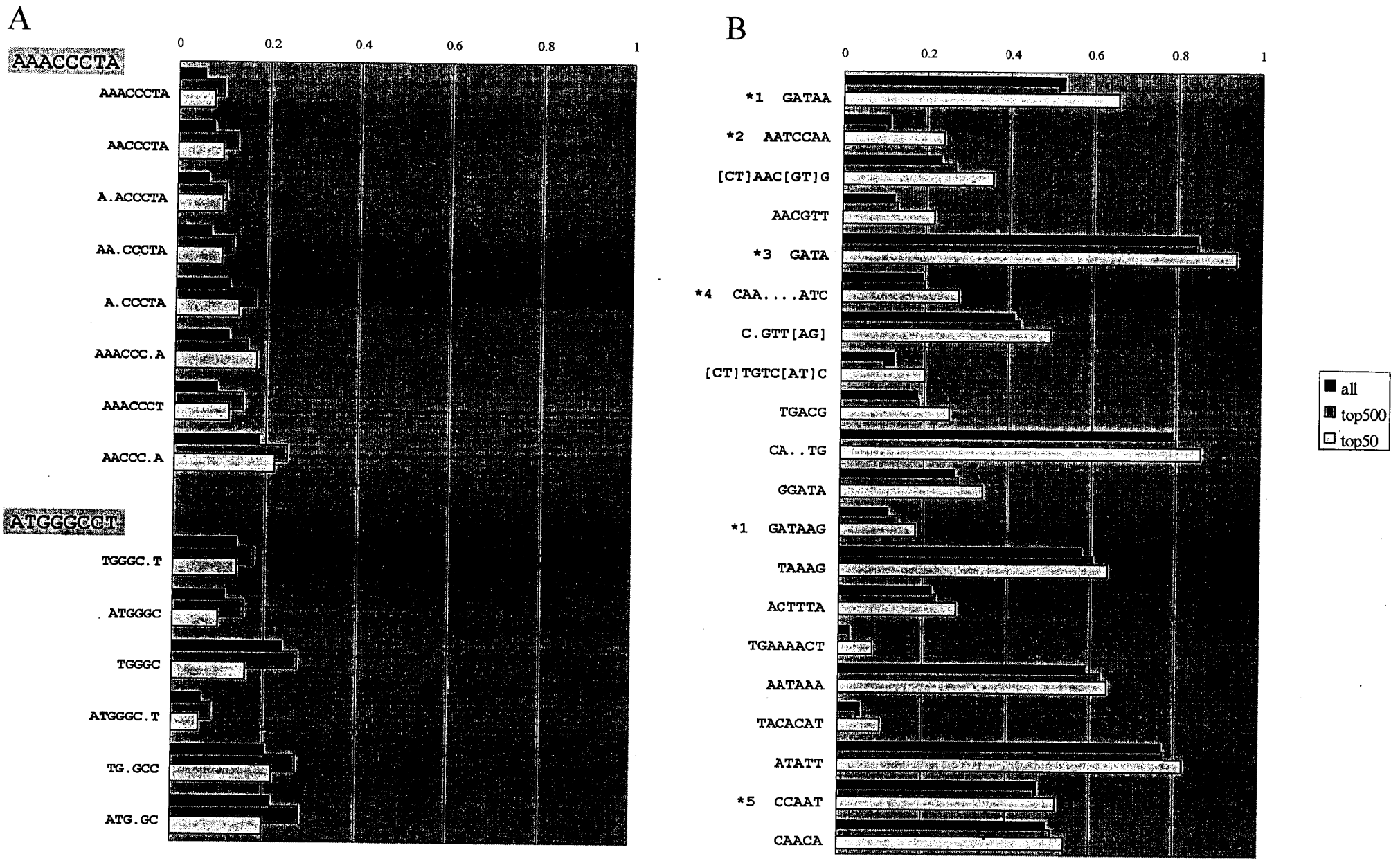


Fig. IV-4. Population of existence of cis elements in total loci and genes with top 50 and 500 REs of response to 3h cytokinin. A, selected sequences in this study. B, cis elements reported in PLACE. *1, I-box. *2, Rubisco related. *3, GATA-box. *4, circadian related. *5, CAAT-box.

I picked up 4 sequences strongly associated with light and cytokinin responses. As described in Chapter III, several function groups were up-regulated by light and cytokinin. What are the functions of the genes with the cis elements? To answer this question, relationship between cis elements and function of regulated genes were analyzed. Statistical test for independence were used about all genes in Arabidopsis (see IV-2.3). First, relationship of cis elements in PLACE with functional gene groups (see IV-2.1) was examined. The gene groups with the Z value more than 7 were presented in Fig. IV-5. In these criteria, 7 cis elements in 390 elements of PLACE showed strong correlation with 8 functional gene groups. For example, 3 cis elements corresponding to G-boxes were associated with a gene group "photosynthesis". Telo-box, "AAACCCTAA", showed strong correlation with 4 gene groups related to ribosome, in agreement with previous reports that telo-box was found in the 5' region of numerous genes encoding components of translation machinery (Tremousaygue et al., 1999). Since these results confirmed validity of this statistical test, cis elements listed in Table IV-1 were further examined. In the 27 conserved sequences which were shown in Table IV-1, 14 sequences were associated with 12 functional gene groups in the same criteria above (Fig. IV-6). In the light-responsive elements, "G.CACGTG", namely G-box itself, was related to the gene group "photosynthesis", as expected. The other 13 sequences were clearly associated with translation especially ribosomes. Interestingly, "AAACCCTA", telo-box has strong specificity to cytosolic ribosomes, whereas "TGGGC" showed to both cytosolic and mitochondorial ribosomes. These results indicated that the two boxes were involved in the cytokinin responses of ribosomal protein genes. Note that since this statistical test tends to give higher Z value to larger gene groups, "TGGGC" could not be determined whether ribosome-specific or global element for cytokinin response.

IV-3.4 Search for sequences showing synergic effects on "G.CACGTG"

Although "G.CACGTG" were the most distinct motif in the association with the light responsive genes, all genes possessing this sequence are not responsive to light. To search the 2nd cis element that helps the function of "G.CACGTG" in regulation of the light response of the genes, CEG analysis was performed again. Namely, PLACE cis elements whose co-existence with "G.CACGTG"

group ID	gene #	description	target	[CT]ACGTGGC *1	CACGTGGC *1	[AC]CACGTGGC *1	AAACCCTAA *2	TAGTGGAT	CTATAAATAC	CC[AT]ACC
ath00195	23	photosynthesis	C	8	10	11	0	0	0	0
ath03010	143	ribosome	O	-1	-1	-1	10	0	0	-2
GO:0003735	107	structural constituent of ribosome	O	-1	-1	-1	7	0	0	-1
GO:0005840	114	ribosome	O	-1	-1	-1	9	0	0	-2
GO:0006412	115	protein biosynthesis	O	0	0	0	8	0	0	-1
GO:0003676	210	nucleic acid binding activity	O	0	-1	0	1	7	4	0
GO:0006310	104	DNA recombination	O	-1	-1	-1	0	10	7	-1
AraCyc	14	lignin biosynthesis	O	1	1	2	0	0	0	7

Fig. IV-5. Cis elements in PLACE related to functional gene groups with Z value more than 7 (which corresponds 99.9999999999% level of significance, and which were highlighted by bold and shade). 27819 nucleic loci were analyzed. Z values rounded down were represented.*1, G-box related sequence. *2, telo-box. Target, subgroups devided using results of targetP prediction as described III-2.6: C, chloroplast. M,

				L3	C3 (AAACCCTA; telo-box)								C3 (TGGGC)				
group ID	gene #	description	target	G.CACGTG	AAACCCTA	AACCCTA	A.ACCCTA	AA.CCCTA	A.CCCTA	AAACCC.A	AAACCCT	AACCC.A	TGGGC.T	ATGGGC	TGGGC	ATGGGC.T	TG.GCC
ath00195	23	photosynthesis	C	7	-1	-1	-1	-1	-1	-1	0	-1	0	2	0	1	1
ath03010	46	ribosome	M	-1	6	6	6	5	5	4	5	4	7	7	8	8	8
GO:0003735	63	structural constituent of ribosome	M	-1	6	6	6	6	6	5	5	4	7	7	9	7	7
GO:0006412	63	protein biosynthesis	M	-1	5	5	5	5	5	4	5	3	6	6	9	7	7
GO:0005840	58	ribosome	M	-1	5	5	4	5	5	5	5	4	7	6	8	7	6
GO:0005622	61	intracellular	M	0	4	5	4	4	5	4	4	4	6	5	7	6	6
ath03010	143	ribosome	O	0	16	13	14	13	11	10	11	8	11	9	11	7	6
GO:0003735	107	structural constituent of ribosome	O	1	10	9	9	9	8	6	7	5	10	6	9	8	4
GO:0005840	114	ribosome	O	0	11	11	11	10	9	7	9	6	9	5	8	5	4
GO:0006412	115	protein biosynthesis	O	0	10	10	10	9	8	6	8	6	9	5	8	5	5
GO:0005622	277	intracellular	O	0	9	9	8	7	7	4	7	4	6	2	5	4	4
GO:0006414	28	translational elongation	O	0	7	5	6	5	4	5	5	3	1	0	1	2	2

Fig. IV-6. Cis elements (found in this study) related to functional gene groups with Z value more than 7 (which corresponds 99.9999999999% level of significance, and which were highlighted by bold and shade). 27819 nucleic loci were analyzed. Z values rounded down were represented. L3, 3h light. C3, 3h cytokinin. Target, subgroups devided using results of targetP prediction as described III-2.6: C, chloroplast. M, mitochondria. S, secretory. O, others.

gave higher CEG value than "G.CACGTG" alone were searched. As a result, 3 cis elements which gave higher CEG values were selected from 390 elements of PLACE (Table IV-2). The top 2 cis elements corresponded to I-boxes. Actually, the coexistence of I-box with G-box has been reported to be enough to show light response (Puente et al., 1996, Chattopadhyay et al., 1998). In addition to searching PLACE elements, random 8mer were also searched (upto two "." were allowed only in this analysis). The sequence which gave a top CEG value was "A.GAT.AG". It was basically same as the I-box motif, although newly found sequence allowed slight replacements of the nucleotides. The modified sequence probably corresponds to a true consensus motif for Arabidopsis I-box binding factors.

IV-3.5 Position specificity of two motifs "G.CACGTG" and "A.GAT.AG" in promoter regions

In light-dependent gene expression in Arabidopsis, the two sequences "G.CACGTG" and "A.GAT.AG" were suggested to have important roles. In Fig. IV-7, the position of "G.CACGTG" and "A.GAT.AG" was illustrated. Fifty five genes which contain the two elements were sorted by RE values of 3h light. The element "G.CACGTG" clearly preferred to about 100b upstream of transcription start site irrespective of the RE values. Moreover, the other element "A.GAT.AG" tends to be placed around the position of "G.CACGTG", namely about 50 b or 170b upstream, suggesting direct interaction of the transcription factors which may bind to the two cis elements. The results also indicated that the presence of G-box was strictly determined around 100 bp upstream independent of light response. In fact, G-box is also generally found in ABA-responsive genes, which is not regulated by light (Shen and Ho, 1995). A list for all genes presented in Fig. IV-7 is shown in Table IV-3. Most genes that have high RE values were photosynthesis-related, and thus, expected to be regulated through the two cis elements. In addition, there are possible candidates for ABA-responsive genes, which are regulated through ABA-signal transduction via the G-box motif.

IV-3.6 Position specificity of all known cis elements (PLACE) in promoter regions

Interestingly, as mentioned above, "G.CACGTG" box located around 100b upstresm region of transcription start site. However, do cis elements generally have any position preferences? I analyzed position specificity of 390 cis elements stored in PLACE for 3000b upstream region of all nuclear genes

Table IV-2. *Synergic effect between "G.CACGTG" and 2nd cis elements on 3h light response.*

I-box-related cis elements were found by 2nd CEG analysis. Gene #, number of genes which have cis elements in 500b upstream regions in 8304 genes for macroarray.

CEG value	cis element	gene #	description of 2nd cis element
5.57	G.CACGTG	399	this study
5.93	G.CACGTG & GATAAG	57	"I-box"; Conserved sequence upstream of light-regulated genes; Sequence found in the promoter region of rbcS of tomato and Arabidopsis; I box (Giuliano et al. 1988); Binding site of LeMYB1, that is a member of a novel class of myb-like proteins; LeMYBI act as a transcriptional activator.
5.66	G.CACGTG & GATAA	217	"I-box"; Conserved sequence upstream of light-regulated genes; Conserved sequence upstream of light-regulated genes of both monocots and dicots.
5.62	G.CACGTG & AGAAA	324	POLLEN1LELAT52; One of two co-dependent regulatory elements responsible for pollen specific activation of tomato lat52 gene; Found at -72 to -68 region;AGAAA and TCCACCATA are required for pollen specific expression.
7.86	G.CACGTG & A.GAT.AG	55	this study

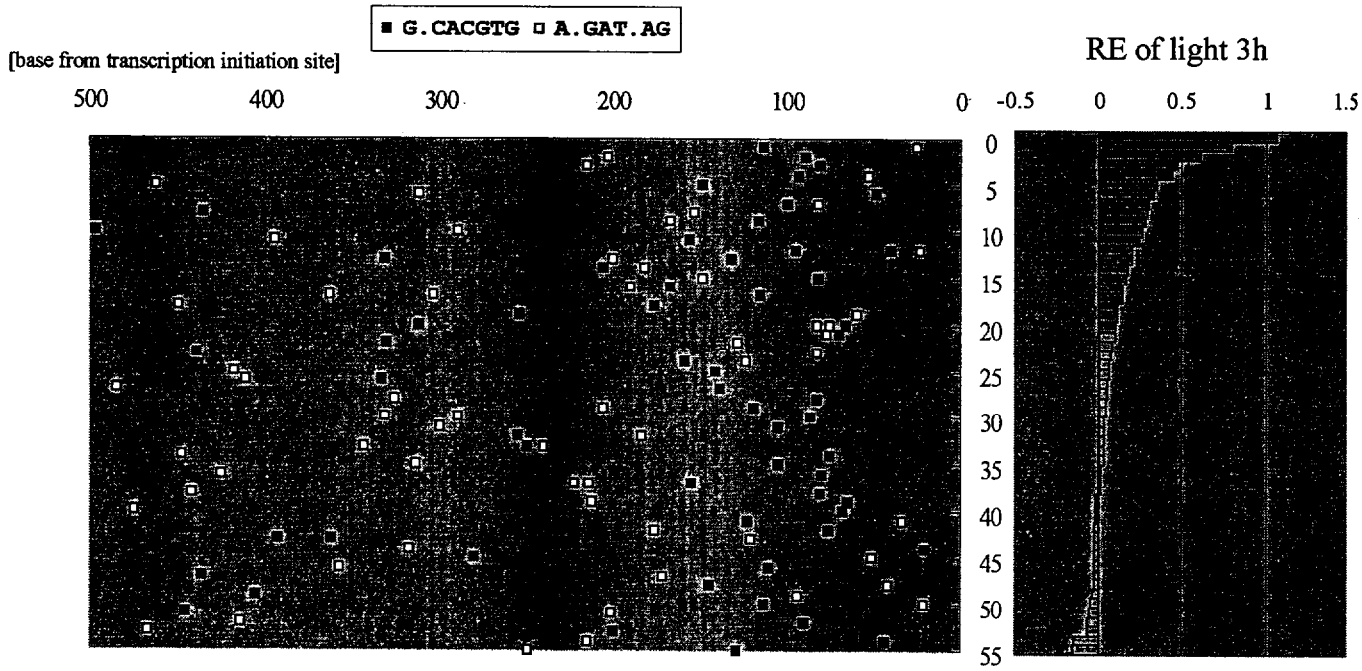


Fig. IV-7 Position of the two cis elements (left) with RE of light 3h (right). y-axis, 55 genes which have two cis elements "G.CACGTG" and "A.GAT.AG" in 500b upstream region included in 8809 genes for macroarray. Genes were sorted by RE of light 3h.

Table IV-3. *List of genes which have two cis elements "G.CACGTG" and "A.GAT.AG".*

Many photosynthesis-related and ABA-responsive genes were found. Gene were sorted by RE of 3h light.

locus	description	targetP	RE (L3)
At3g54890	light-harvesting chlorophyll a/b binding protein	C4	1.1
At3g47470	light-harvesting chlorophyll a/b binding protein	C2	0.8
At5g23120	photosystem II stability/assembly factor HCF136	C2	0.6
At3g56940	leucine zipper-containing protein AT103	C1	0.5
At4g05180	oxygen-evolving complex protein 16, chloroplast precursor (OEC16)	C3	0.4
At3g16140	photosystem I subunit VI precursor	C3	0.4
At3g54050	fructose-bisphosphatase precursor	C1	0.3
At2g45290	transketolase precursor -related	_2	0.3
At5g11510	myb family transcription factor (MYB3R4)	_5	0.3
At3g63160	expressed protein	_2	0.3
At2g36630	expressed protein	S1	0.3
At1g52220	expressed protein	C1	0.2
At4g16370	isp4 like protein (ATOPT3)	_3	0.2
At1g78510	geranyl diphosphate synthase (GPPS)(dimethylallyltransferase)	M3	0.2
At5g64940	ABC transporter-related (ATATH13)	C1	0.2
At5g38410	Rubisco small subunit 3b	C4	0.2
At4g28440	expressed protein	C5	0.1
At4g31060	AP2 domain transcription factor, putative	C5	0.1
At5g22730	F-box protein family	_4	0.1
At1g03600	photosystem II protein family	C2	0.1
At1g61110	No apical meristem (NAM) protein family	_1	0.1
At5g61930	expressed protein	_3	0.1
At1g18900	pentatricopeptide (PPR) repeat-containing protein	M5	0.1
At5g28640	glycine-rich protein	_2	0.1
At5g26600	expressed protein		0.1
At5g02820	meiosis specific - like protein (RHL2)	_1	0.1
At2g44610	GTP-binding protein (RAB6)	_5	0.1
At1g58340	MATE efflux protein (ZF14)	_5	0.1
At1g43670	fructose 1,6-bisphosphatase -related	_1	0.0
At1g69960	serine/threonine protein phosphatase type 2A (PP2A)	_1	0.0
At1g53770	hypothetical protein	_3	0.0
At5g03070	importin alpha subunit -related	C4	0.0
At4g00700	C2 domain-containing protein	_3	0.0
At5g57900	SKP1 interacting partner 1 (SKIP1)	_2	0.0
At1g32560	late-embryogenesis abundant protein -related	_2	0.0
At5g25100	endomembrane protein 70, putative	S3	0.0
At2g42750	DnaJ protein family	C1	0.0
At5g18670	glycosyl hydrolase family 14 (beta-amylase) (BMY3)	_5	0.0
At5g17190	expressed protein	S3	0.0
At5g48655	C3HC4-type zinc finger protein family	_4	0.0
At4g15110	cytochrome p450 (CYP97B3)	M4	0.0
At5g67250	SKP1 interacting partner 2 (SKIP2)	_3	0.0
At3g12320	expressed protein	_2	0.0
At3g03160	expressed protein	S3	0.0
At4g28330	expressed protein	M2	0.0
At2g39930	isoamylase, putative	C3	0.0
At1g80300	adenine nucleotide translocase	C5	0.0
At5g49100	expressed protein	_2	0.0
At4g33220	pectinesterase family	_4	0.0
At3g08640	glycine-rich protein	C2	0.0
At1g50450	expressed protein	M2	-0.1
At5g24970	expressed protein	_5	-0.1
At3g04240	glycosyltransferase - related	C5	-0.1
At4g17190	farnesyl pyrophosphate synthetase 2 (FPS2)	_3	-0.1
At3g05520	alpha subunit of F-actin capping protein	_1	-0.2

in Arabidopsis (27819 genes). To examine the position preference, 3000b upstream regions were divided into 60 segments, each of which has 50bp in length, and total number of the cis elements was counted for all segments. For example, Figure IV-8C1 shows appearance of "CACGTG" (G-box) was strictly limited around 100 bp upstream of transcription start sites. About 30% of reported cis elements have position preference with normalized skewness more than 0.5 or less than -0.5. Moreover, the preferential position was restricted from 1b to 300b region as shown in the figures.

IV-3.7 Position specificity of random 5-8mer in cytokinin-responsive ribosomal protein genes

As described in Chapter III, cytokinin-responsive ribosomal protein genes showed particular organ specificity in Bud-Flower-Root (which belongs c5 of SOM cluster). To find other cytokinin-responsive element(s) which cooperates with "AAACCCTA" or "TGGGC" boxes, I searched random 5-8mers which showed strong position specificity in the cytokinin-regulated ribosomal protein genes. Fig. IV-9 showed that in 3 major motifs, high position specificity was observed. One of them corresponded to telo-box. This box is specifically located in around 50bp upstream from transcription start site. Interestingly, their complementary sequences were also obtained by this analysis, suggesting that complementary sequence is also used as cis-element. In addition, a novel motif, "AGGCCCAA" was found by the specific location in around 100bp upstream just beside with the telo-box. Surprisingly, this novel motif is a complementary sequence of "TGGGC" which was identified in Table IV-1. Actually, the "TGGGC" box was also found in this procedure. As expected, TATAA-box was found just beside transcription start site. These data strongly suggested that the telo-box and "AGGCCCAA" box cooperatively work in cytokinin-dependent expression of these ribosomal protein genes. These boxes are also probably one of determinants for organ specificity of the ribosome genes and thus, the genes were mainly expressed in bud, flower and roots.

IV-4 DISCUSSION

In this chapter, promoter analysis was performed using the data of transcriptome analysis for light and cytokinin responses. Promoter analysis succeeding DNA array experiments is one of the major

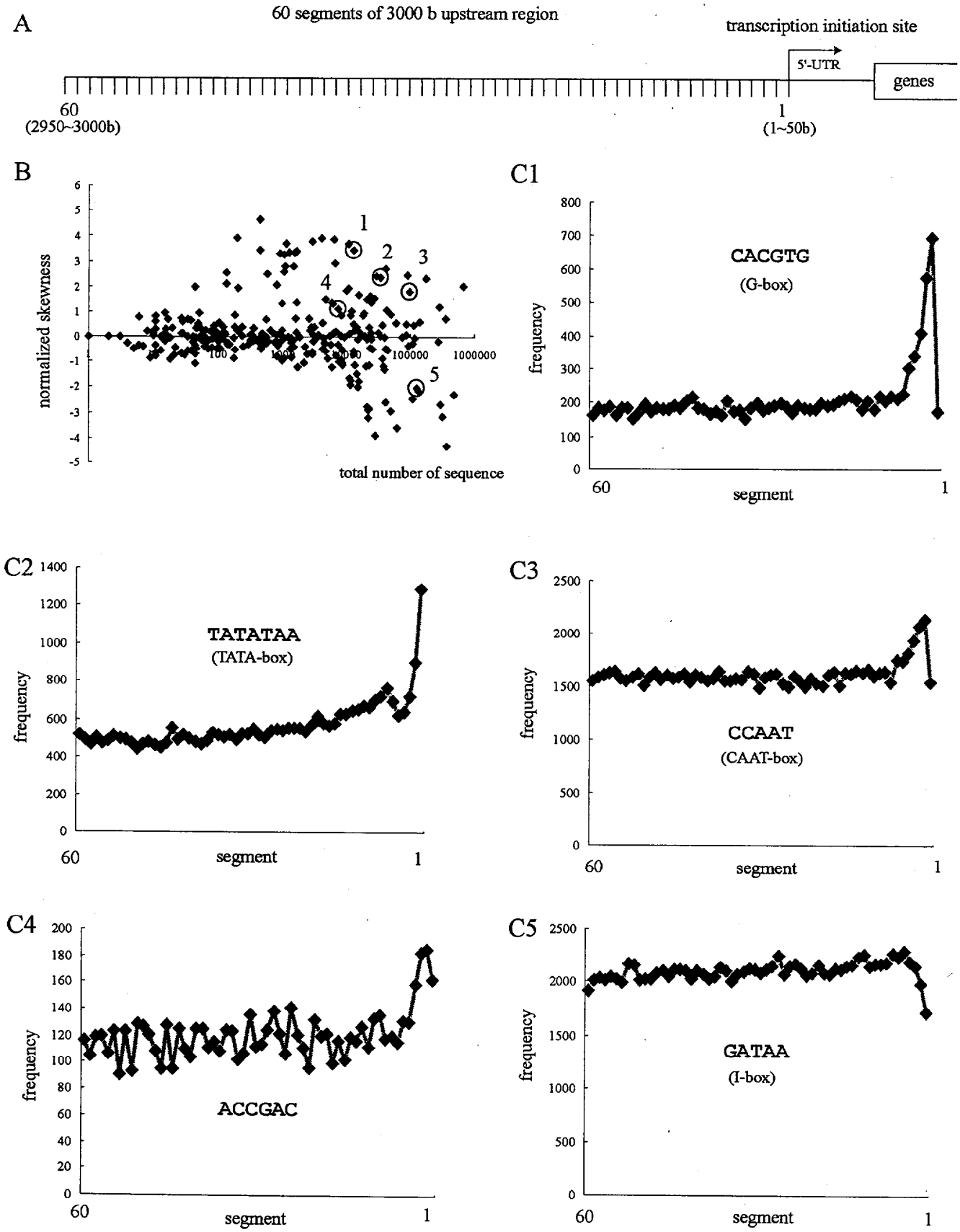


Fig. IV-8. Analysis for position specificity of cis elements (PLACE). A, schematic representation of 60 segments for counting appearance. B, distribution of total number of appearance of a cis element in 27819 Arabidopsis genes (x-axis) and skewness of frequency of 60 segments of the cis element (y-axis) for 390 elements in PLACE. C, five examples of frequency polygon of position of cis elements in upstream regions.

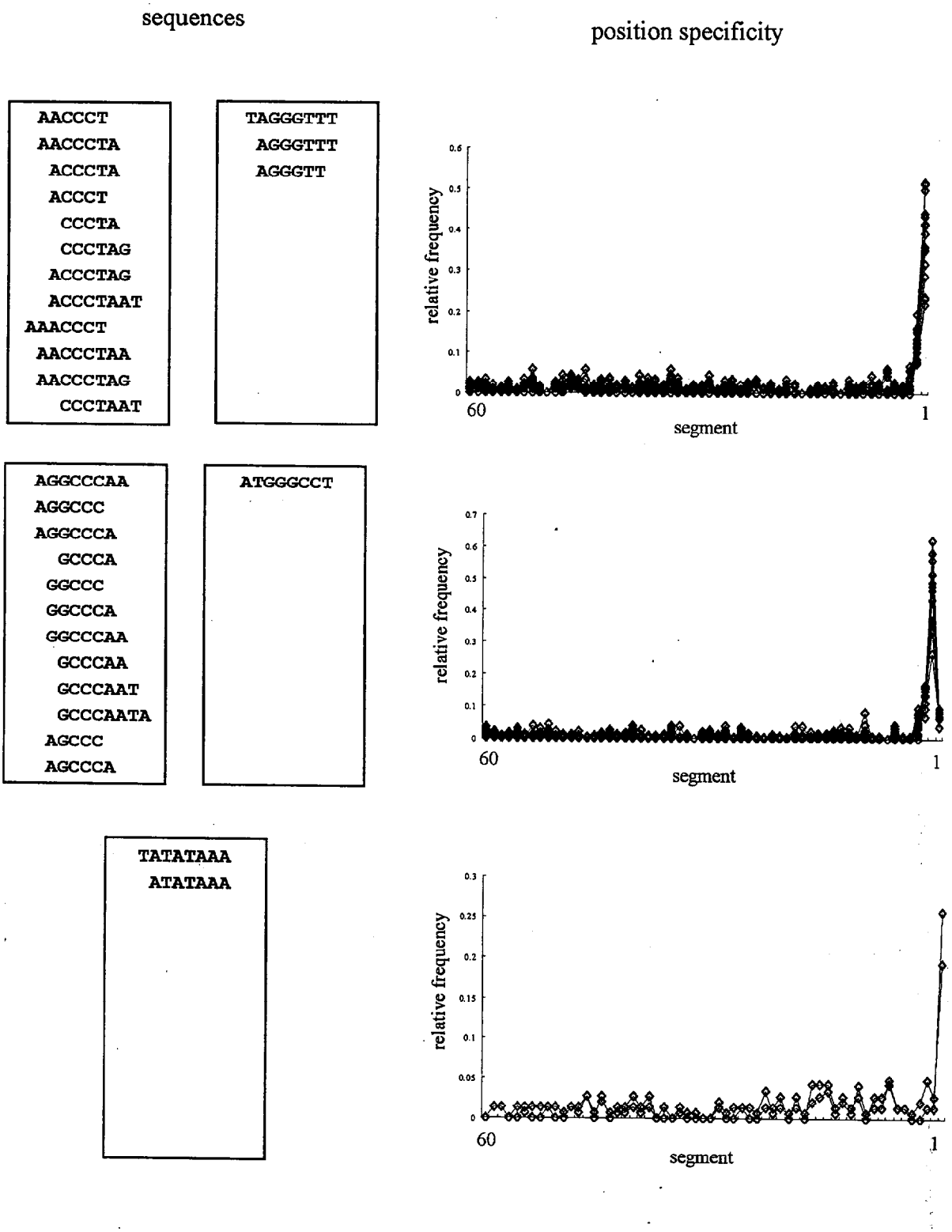


Fig IV-9. Search of cis elements which have position specificity in 116 ribosome genes with bud-flower-root specificity. The result revealed 3 types of cis elements (see also Fig III-7). Top 30 cis sequences with highest skewness and those position specificity are shown. Sequences in each box were sorted by skewness (top: highest). Frequency polygon of positions of each type of all cis elements are overwritten in right figure. The same 3 types of cis elements were observed even if top 100 cis sequences were analyzed. segment, see Fig. IV-7A.

trends in this high-throughput age. Global promoter analysis needs many techniques of informatics, which is generally difficult for biologist. However, information of promoter is very important for biological study. I challenged the promoter analysis using macroarray data and got important information to understand mechanism for light- and cytokinin-responsive gene expression.

As reported previously, G-box and I-box were important in the light-dependent gene expression (Puente et al., 1996, Chattopadhyay et al., 1998). The present results indicated that the conserved sequences of the two elements were slightly replaced in Arabidopsis. Consensus sequence for G-box was "G.CACGTG" and I-box was "A.GAT.AG". Strict positional specificity of known and novel motifs I found here was also revealed in this study. Most interestingly, location of G-box is critically determined, and I-box and CAAT-box probably cooperate with G-box and its binding trans-acting factors such as Hy5 and PIF3. G-box was also known as ABA-responsive cis elements. Like I-box in light regulation, some other factor may be required to determine specificity for ABA signal transduction.

In the case of cytokinin response, I found two major motifs which frequently observed in upstream regions of cytokinin-responsive genes. One of them, "AAACCCTA" has been already reported as telo-box which is specifically observed in upstream regions of nuclear-encoded ribosome genes (Tremousaygue et al., 1999). Interestingly, this telo-box was particularly associated with cytosolic ribosomal protein genes, not with chloroplast and mitochondorial ribosomes. In contrast, "AGGCCCAA" box (which probably equivalent to "TGGGC" box) is more common for both cytosolic and mitochondorial ribosomes, and cooperates with the telo-box in cytokinin response and organ-specific expression. Based on the present results, I propose working hypothesis for light- and cytokinin-dependent gene expression in Chapter V.

IV-5 REFERENCES

Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol 2:28-36

Chattopadhyay S, Puente P, Deng XW, Wei N (1998) Combinatorial interaction of light-responsive elements plays a critical role in determining the response characteristics of light-regulated promoters in Arabidopsis. Plant J **15**: 69-77

Hertz GZ, Stormo GD (1999) Identifying DNA and protein patterns with statistically significant alignments of multiple

sequences. Bioinformatics **15** :563-77

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

Manevski A, Bertoni G, Bardet C, Tremousaygue D, Lescure B (2000) In synergy with various cis-acting elements, plant insterstitial telomere motifs regulate gene expression in Arabidopsis root meristems. FEBS Lett **483**: 43-6

McKendree WL, Paul AL, DeLisle AJ, Ferl RJ (1990) In vivo and in vitro characterization of protein interactions with the dyad G-box of the Arabidopsis Adh gene. Plant Cell **2**: 207-14

Lawrence CE, Altschul SF, Boguski MS, Liu JS, Neuwald AF, Wootton JC (1993) Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. Science **262**: 208-14

Poluliakh N, Takagi T, Nakai K (2003) Melina: motif extraction from promoter regions of potentially co-regulated genes. Bioinformatics **19**: 423-4

Puente P, Wei N, Deng XW (1996) Combinatorial interplay of promoter elements constitutes the minimal determinants for light and developmental control of gene expression in Arabidopsis. EMBO J **15**: 3732-43

Regad F, Lebas M, Lescure B (1994) Interstitial telomeric repeats within the Arabidopsis thaliana genome. J Mol Biol **239**: 163-9

Shen Q, Ho TH (1995) Functional dissection of an abscisic acid (ABA)-inducible gene reveals two independent ABA-responsive complexes each containing a G-box and a novel cis-acting element. Plant Cell **7**: 295-307

Tremousaygue D, Manevski A, Bardet C, Lescure N, Lescure B (1999) Plant interstitial telomere motifs participate in the control of gene expression in root meristems. Plant J **20**: 553-61

Wolfertstetter F, Frech K, Herrmann G, Werner T (1996) Identification of functional elements in unaligned nucleic acid sequences by a novel tuple search algorithm. Comput Appl Biosci **12**: 71-80

Chapter V

General Discussion

In my study, two transcriptome analyses for characteristics of organs and regulation of photomorphogenesis were performed.

The first transcriptome analysis for characteristics of organs revealed distinctive features of organs. Typical features of leaves for photosynthesis, stem for mechanical strength, bud and flower for reproduction, silique for energy storage and root for metabolism were observed at transcription level using GRE analysis. Moreover using clustering method, particular gene groups which showed multi-organ-specific expression profiles, such as bud-flower-specific, stem-silique-specific or bud-flower-root-specific profiles, could be effectively identified. The organ-specific gene expression data are released at a web site, ATTED (see Appendix).

The second transcriptome analysis for regulation of photomorphogenesis using light and cytokinin treatments to etiolated seedling revealed common and different features between the two stimuli to photomorphogenesis. Cytokinin basically induced genes for photomorphogenesis similarly to light. Besides, light induced function for carbon fixation, and cytokinin induced protein synthesis for reproduction. Expression profiles responding to light and cytokinin suggested common and independent signal transductions while evidence for *de novo* cytokinin synthesis was not observed.

For examination of the regulatory mechanism by light and cytokinin, promoter analysis was achieved. Using three approaches, CEG, functional correlation and position specificity, important cis elements and their relationships were revealed.

V-1.1 Cytokinin-mediated photomorphogenesis

Yamaryo et al. (2003) reported regulation of synthesis of chloroplast membrane responding to light or cytokinin. As described, chloroplast development was promoted by light or cytokinin. A key enzyme of chloroplast membrane biosynthesis, MGDG synthase, was also induced by light and cytokinin. They showed importance of root for light-induced biosynthesis of chloroplast membrane and also

showed deficiency of root was complemented in addition of cytokinin, suggesting *de novo* biosynthesis of cytokinin in root responding to light stimulus. In my study, cytokinin biosynthesis and signal transduction in root was expected. Only root in the bud, flower and root, where cytosolic-type ribosomes were highly expressed, had correlation to cytokinin response, indicating cytokinin-dependent induction of ribosomes in root. Unfortunately, to study this relationship between cytokinin and root, our cDNA macroarray system has small number of genes encoding cytokinin signal transduction related proteins. However, some important evidences were obtained as described below.

V-1.2 A model for regulation of photomorphogenesis by light and cytokinin

Remember the discussion in Chapter III. In Fig. III-9, I presented three possibilities for common regulation of light and cytokinin. A-type possibility was not supported by my data and previous reports, although nobody has definitive evidence. About C-type possibility, independent regulation of gene expression responding to light and cytokinin, only 2 transcription factors, At1g59750 and At5g11510, were observed responding to both light and cytokinin treatments in 391 transcription factors on the macroarray, although 13 and 20 transcription factors were specifically responded to light and cytokinin, respectively. Promoter analysis revealed several cis elements and their relationships, however common element(s) for light and cytokinin responses was not picked up. These results supported the C-type possibility in the 3 hypotheses in Fig III-9. In contrast, ARR4 is commonly induced by light and cytokinin (Fig III-10). Although ARR4 is not transcription factor in itself, it has a role for cytokinin signal transduction via binding PHYB (Sweere et al., 2001). Besides, as described in Chapter IV, although cis elements commonly responding to light and cytokinin were not searched by CEG method, CAAT-box was commonly observed by a simple method (Fig IV-2B, 3B). Because at least 2 CAAT-box binding factors were induced by cytokinin (Table III-5), it was possible that the CAAT-box and its binding factors modulated commonly regulated genes to light and cytokinin. G-box is very important for light and other signal transduction. ARR4 and CAAT system supported the B-type possibility. From these evidences, B- and C-type possibilities seem to co-exist in plant. Although the usage of the two-type regulations was unknown so far, for example, it may be possible that some key functions are regulated by

B-type pathway and other fine functions are regulated by C-type signal transduction system.

Fig. V-1 shows a prospective model of light and cytokinin regulation of photomorphogenesis combined with results of promoter analysis. As described above, ARR4 and CAAT system are two key junction points of signal conjugation. Light signal was mainly accepted via G-box with help of I-box or CAAT-box, which were positioned on around 100b upstream region.

V-1.3 Transcriptome analysis

In this thesis, I achieved transcriptome analysis using cDNA macroarray. Transcriptome analysis deals with all genes at once at mRNA level, which was only one reflection of state of an organism. One of the advantages of transcriptome analysis is to look whole picture of an organism. Up-regulated functions were easily revealed in the studies. On the other hand, it is difficult to elucidate the mechanism of regulation just by examining mRNA level. However, using promoter analysis, transcriptome analysis can approach gene regulation. Since mRNA levels are determined on the balance between its synthesis and degradation, sequence analysis for regulation of degradation, namely, analyses for 5'- and 3'-untranslated region is also needed. Events concerning subcellular localization are also pictured using prediction of subcellular localization. In this thesis, only 4 subcellular lacalizations were used. More fine subcellular prediction will lead to draw more fine picture.

V-2 REFERENCES

Sweere U, Eichenberg K, Lohrmann J, Mira-Rodado V, Baurle I, Kudla J, Nagy F, Schafer E, Harter K (2001) Interaction of the response regulator ARR4 with phytochrome B in modulating red light signaling. Science **294**: 1108-11

Yamaryo Y, Kanai D, Awai K, Shimojima M, Masuda T, Shimada H, Takamiya K, Ohta H (2003) Light and cytokinin play a co-operative role in MGDG synthesis in greening cucumber cotyledons. Plant Cell Physiol **44**: 844-55

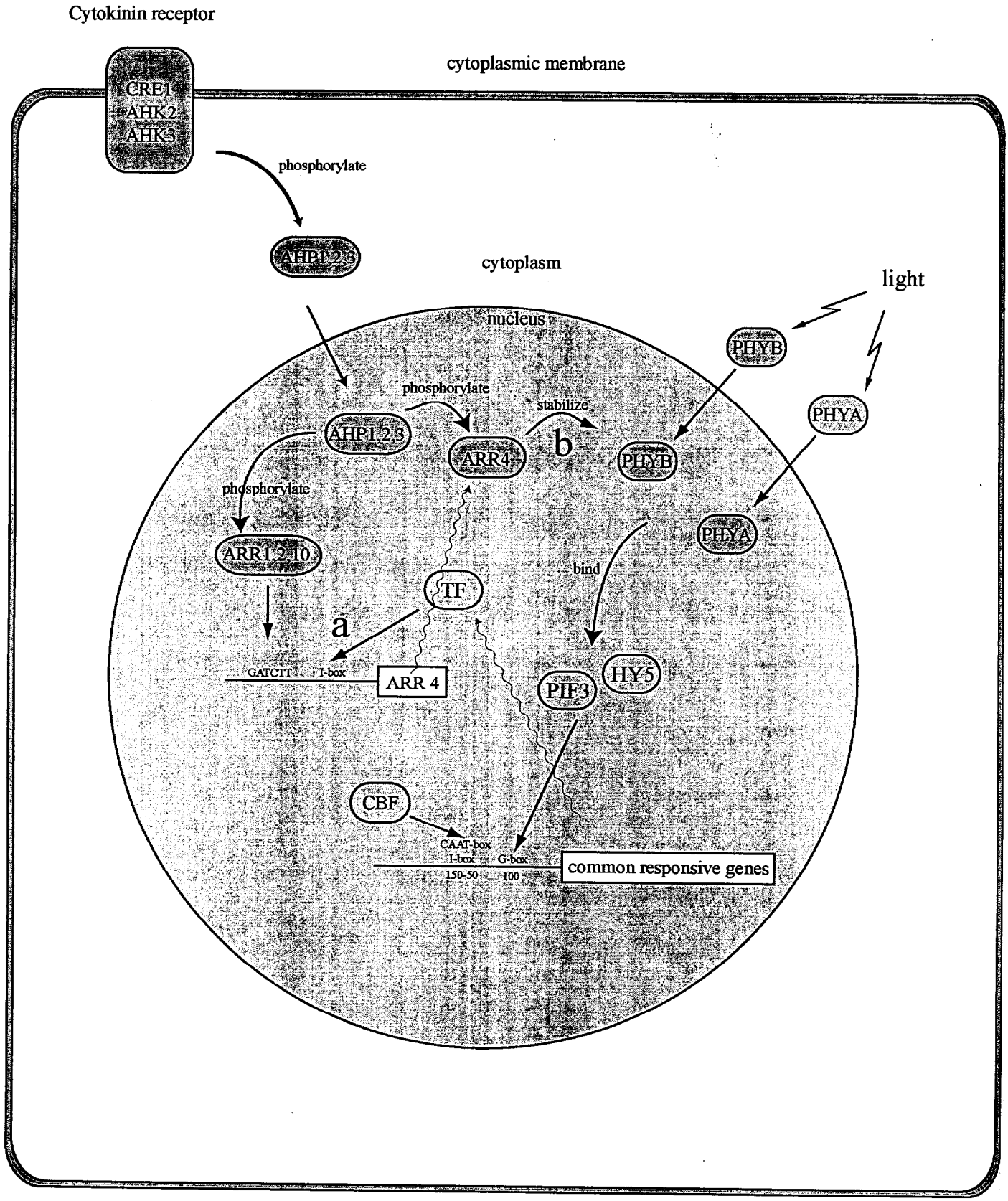


Fig. V-1. Prospected model of interaction between light and cytokinin. a, prediction according to this study and previous reports. b, prediction according to previous reports. HY5 and PIF3 are G-box binding transcription factors previously reported. Straight arrow, moving of protein. Curved arrow, protein-protein interaction. Wavy arrow, translation.

Appendix

Data Opening Site for Organ-Specific Gene Expression

A-1 INTRODUCTION

Genome sequence and the related data such as exon-intron organizations, ESTs, mRNA expressions, gene disruptants, gene functions, related papers and so on, are fundamental information in the modern world of biological sciences. Although main Arabidopsis data are managed by TAIR, more specifical data are released by each researcher. For example, PLACE (plant cis-acting regulatory DNA elements, Higo et al., 1999, http://www.dna.affrc.go.jp/htdocs/PLACE/) used in Chapter IV is for plant researchers. READ (RIKEN expression Array Database, Bono et al., 2002, http://read.gsc.riken.go.jp/) is a database for mouse cDNA microarray. Organ-specific gene expression data are one of the most fundamental information for plant science. For scientific value of the data, I made data opening site named ATTED (*Arabidopsis thaliana* Tissue-Specific Expression Database, http://www.atted.bio.titech.ac.jp/, Fig. A-1).

A-2 MATERIALS AND METHODS

A-2.1 Data storage

Data used in this database were shown in Table A-I.

A-2.2 Systems

Because the ATTED system is implemented on the Web, it is easily accessible via typical Web browsers, which enables the system to be accessed without any special software and any OS restriction. All data are stored in a flat-file format (tab-delimitated text). Perl script language (http://www.cpan.org/) is used for formatting data, and the scripts are combined in UNIX shell scripts. An Apache Web server (http://www.apache.org/) makes it easy to query the stored data via the Web. BLAST released by NCBI (http://www3.ncbi.nlm.nih.gov/) is used to search homologs when users desired. The system is built on a PC-UNIX server running Vine Linux 2.6 (http://www.vinelovux.org/).

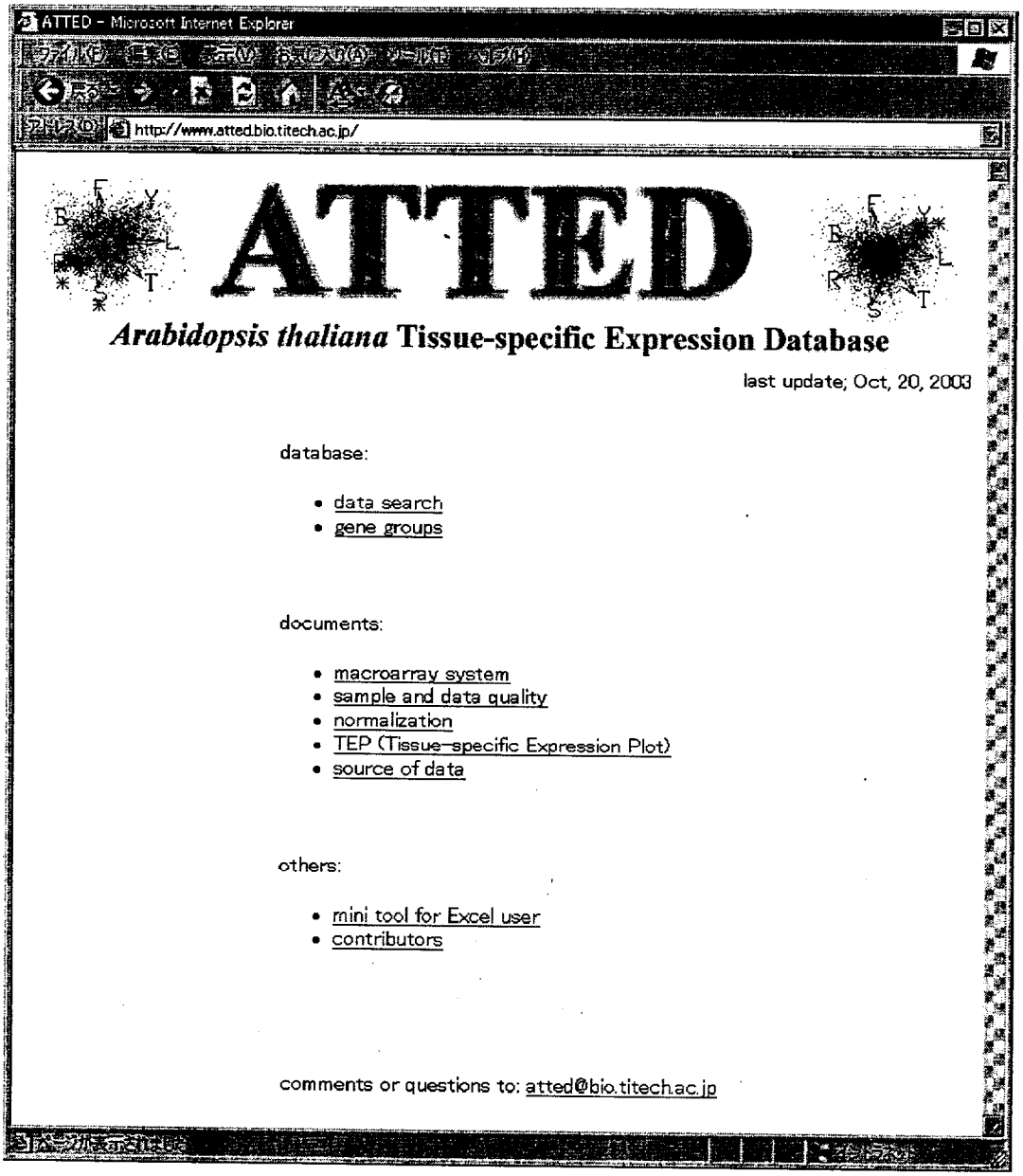


Figure A-1. Top page of ATTED (*Arabidopsis thaliana* Tissue-specific Expression Database).

Table A-1. *Data used in ATTED.*

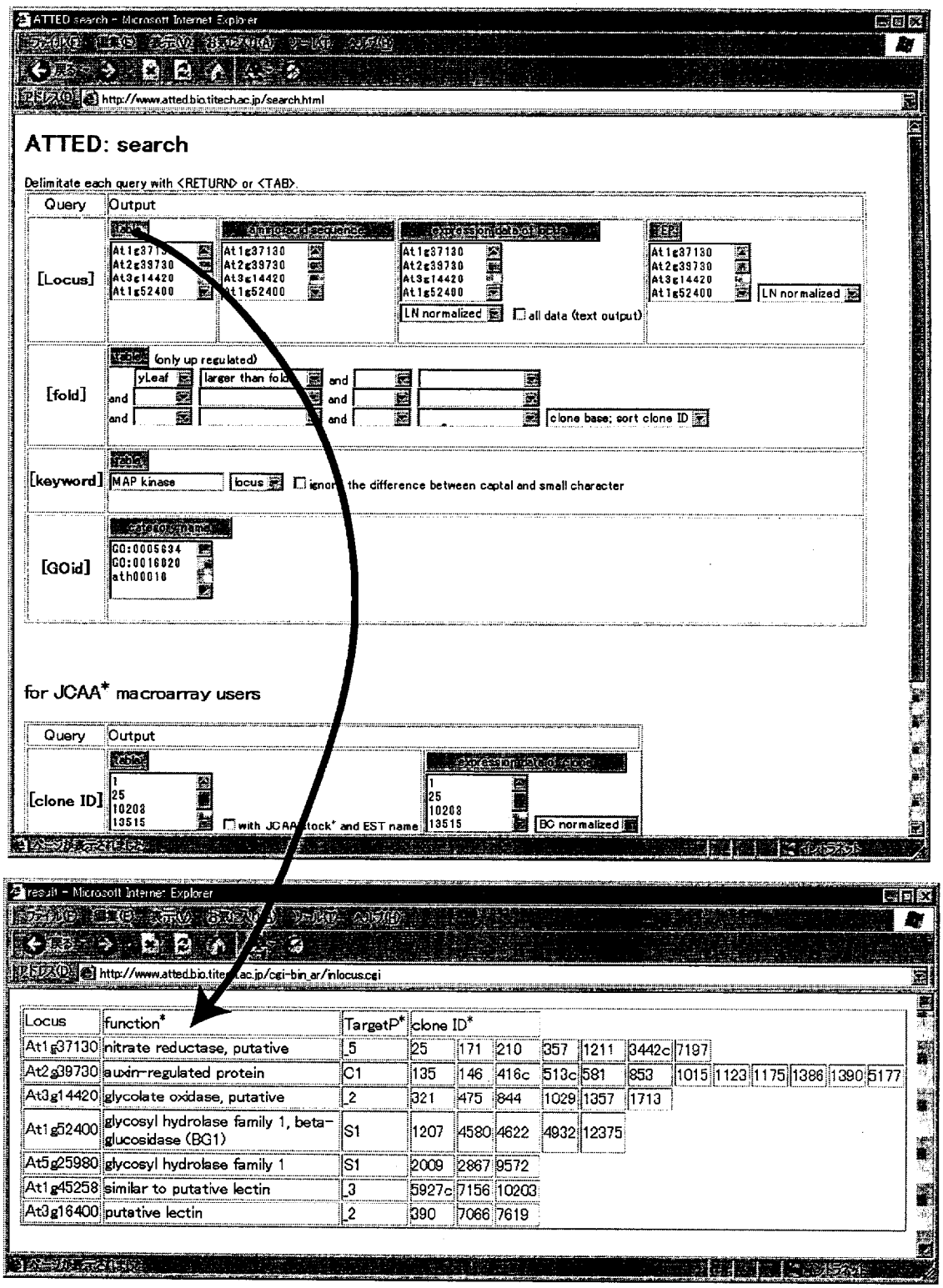
data	source	last update
Organ expression data	original data	18-Jan-03
EST clone	Kazusa DNA research Inst.	
Gene Annotation	TIGR	31-Jul-02
TargetP	TIAR	9-Oct-02
Gene Ontology	TIAR	2-Dec-02
KEGG pathway	GenomeNet	6-Jan-03

A-3 RESULTS

A-3.1 Appearance and use of ATTED

Because origin of this database was in-laboratory database for us to use cDNA macroarray data, this database is especially recommended for researchers of global analysis like DNA array. First labor-consuming point after data normalization of DNA array is making gene list induced or repressed with appropriate annotations. The appropriate information is various, which is related to not only gene functions but also array system itself, for example, spot conditions and EST accuracies. Our cDNA macroarray system is compatible with cDNA macroarray of JCAA (The Japanese Consortium for *Arabidopsis thaliana* DNA Array, Fujiwara et al., 2002), thus users of JCAA cDNA array are another users of ATTED.

Each researcher has his own data usually in general purpose software like Microsoft Excel. Up- and down-regulated gene list is thought to be in Excel. In ATTED, it is easily achieved that users get information about the loci with coping the list from Excel to the textbox (Fig. A-2, top). For different purposes, different query boxes were prepared. And also different types of query were accepted, namely locus ID, fold of gene expression, keyword search of gene function, Gene Ontology ID and for JCAA users clone ID and coordinates of clone stored in 96-well plates. For making the gene list, when user clicks "table" button, the gene list will appear (Fig A-2, bottom). In the list, each locus is clickable to get detail locus information. Locus information includes functions, clone information, expression data of line plot and table. And also query box for Arabidopsis homolog search and links to several databases are available (Fig. A-3). As described in Chapter II, some loci have multi-corresponding clones (In the Fig. A-3, 3 clones correspond to the locus), each expression of which is visualized in line plot. Each EST clone respectively has information, which is accessible to click "clone ID". Then EST clone information of accession number, EST sequence and expression data in all repeated experiment data (Fig. A-4). Because EST sequence was conducted by one pass sequencing, some of EST sequences were not good quality, resulting in miss-corresponding to locus. In such case, users can confirm the correspondence of



Locus	function*	TargetP*	clone ID*											
At1g37130	nitrate reductase, putative	_5	25	171	210	357	1211	3442c	7197					
At2g39730	auxin-regulated protein	C1	135	146	416c	513c	581	853	1015	1123	1175	1386	1390	5177
At3g14420	glycolate oxidase, putative	_2	321	475	844	1029	1357	1713						
At1g52400	glycosyl hydrolase family 1, beta-glucosidase (BG1)	S1	1207	4580	4622	4932	12375							
At5g25980	glycosyl hydrolase family 1	S1	2009	2867	9572									
At1g45258	similar to putative lectin	_3	5927c	7156	10203									
At3g16400	putative lectin	_2	390	7066	7619									

Figure A-2. Making gene list.

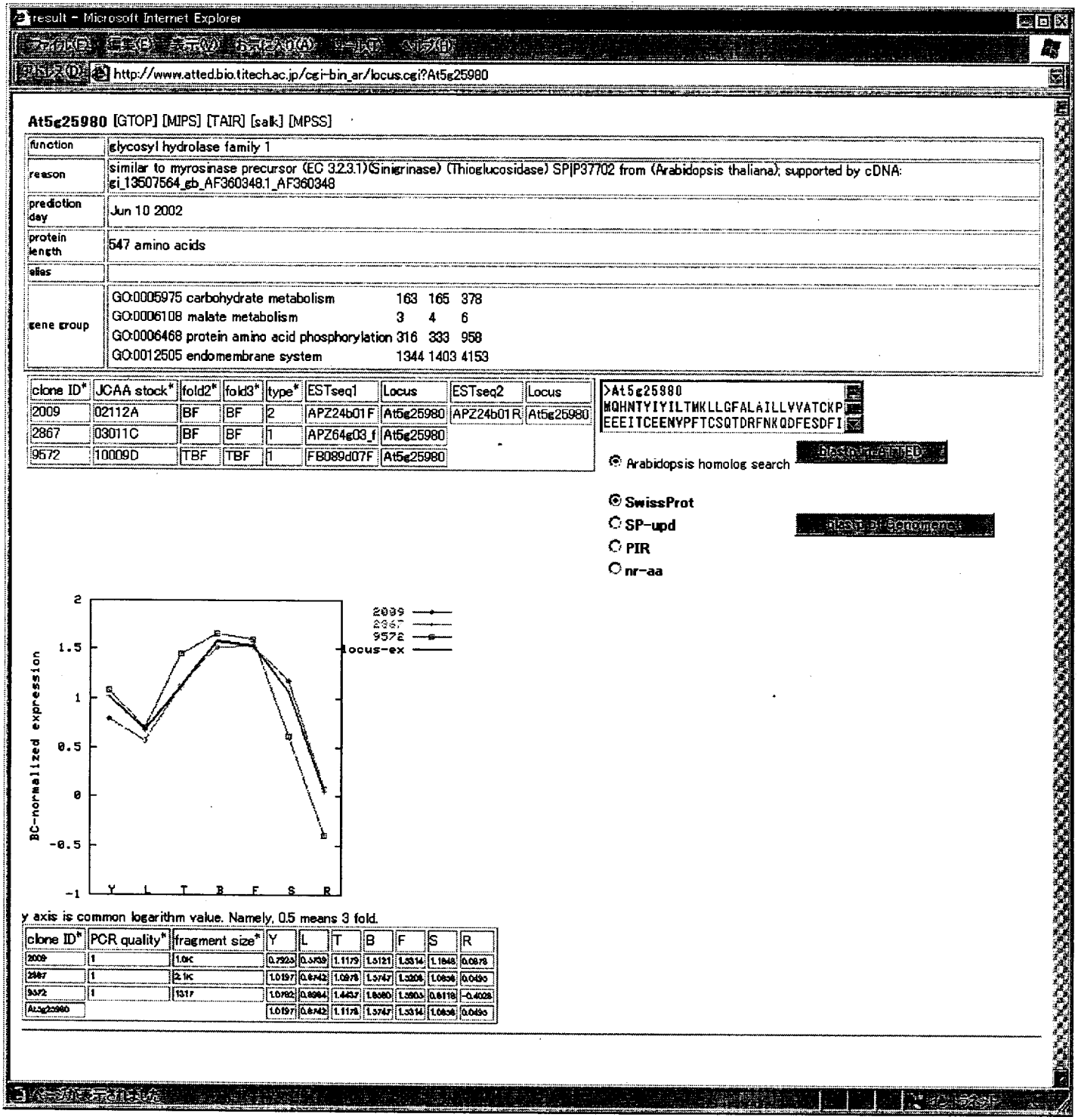


Fig. A-3 Information for a locus.

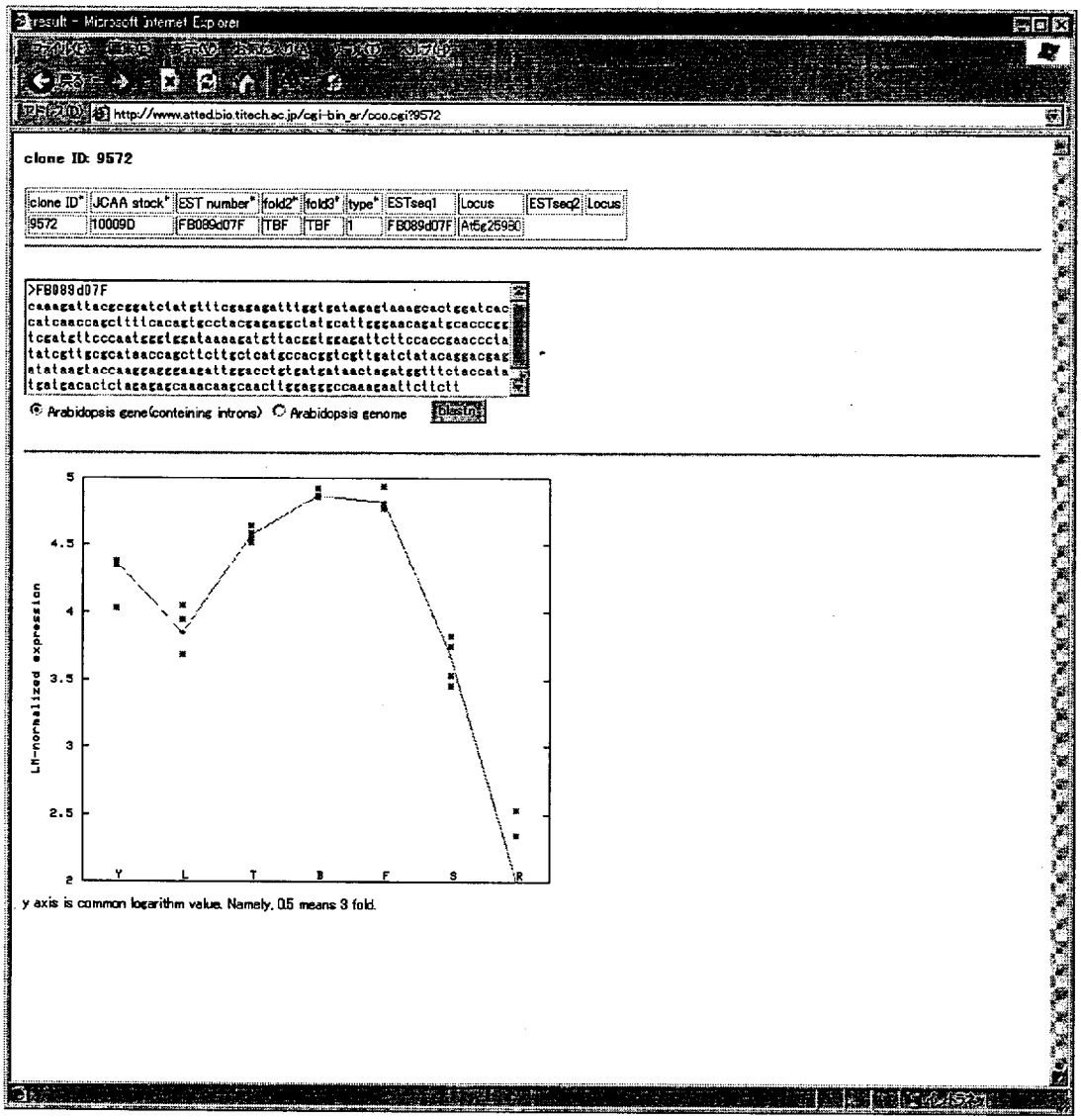


Figure A-4. Information for an EST clone.

EST and locus using BLAST (Fig. A-4).

Gene expressions of gene groups are important as described. Another view for expression is discussed as TEP with SOM categories in Chapter II. TEP for any genes and gene groups which users want to get are also available (Fig A-5).

A-3.2 Utilization

About 10,000 files per month are provided according to access log for about 6 months.

A-4 REFERENCES

Bono H, Kasukawa T, Hayashizaki Y, Okazaki Y (2002) READ: RIKEN Expression Array Database. Nucleic Acids Res **30**: 211-3

Fujiwara T, Ohta H, and Kanamaru K (2002) Activity of The Japanese Consortium for *Arabidopsis thaliana* DNA array (JCAA). Tanpakushitu Kakusan Koso **27**: 91-3

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

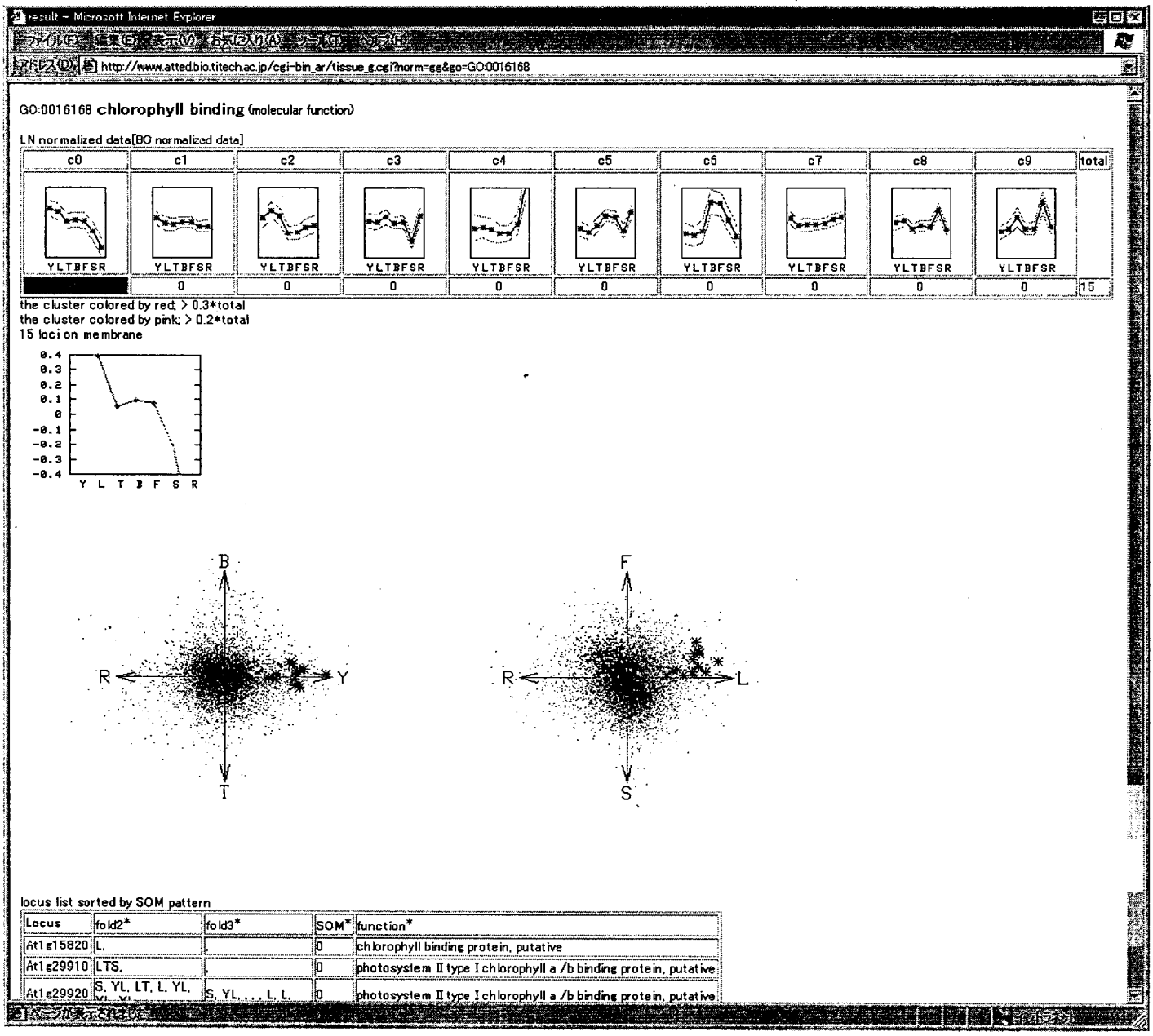


Figure A-5. TEP (Tissue-Specific Expression Plot).

Acknowledgments

I greatly appreciate Professor Ken-ichiro Takamiya, Tokyo Institute of Technology, for his generous suggestion and cheerful encouragement during my researches. I also wish to express my sincere gratitude to Associate professor Hiroyuki Ohta, Tokyo Institute of Technology, for his continuous and indispensable discussion.

I am grateful to Dr. Tatsuru Masuda, Tokyo institute of Technology, for his helpful and elegant advise. I also wish to express his gratitude to Dr. Hiroshi Shimada, Tokyo Institute of Technology, for his kind counsel and generous suggestion.

I am also grateful to the colleagues of Takamiya-Ohta laboratory of Tokyo Institute of Technology, especially to Takashi Okegawa and following persons; Makiko Horikoshi, Mariko Mochizuki, Nozomi Taki, Shu-ichiro Takahashi, Shun-ichi Miyachi and Yuko Sasaki-Sekimoto, for valuable discussions, with which the methods for data analysis in my study were developed. Thanks are also due to my families and friends for their kind supports during my school days.