ORIGINAL PAPER

Lukas Bürkle · Svenja Meyer · Hakan Dortay ·
Hans Lehrach · Alexander Heyl

In vitro recombination cloning of entire cDNA libraries in *Arabidopsis thaliana* and its application to the yeast two-hybrid system

Abstract In the postgenomic era many experiments rely on the availability of transcript sequence for cloning. As these clones usually originate from cDNA libraries, the quality of these libraries is crucial. If a good library is generated it is desirable to use a versatile cloning system suitable for many different kinds of applications. The cloning systems based on in vitro recombination proves fitting for this task. However, the use of this method for shuttling entire cDNA libraries between different vectors has not yet been studied in great detail. Here we describe the construction of four cDNA libraries from different tissues of *Arabidopsis thaliana*, the shuttling of the libraries into expression vectors, and evaluation of this method as well as its suitability for downstream applications. Libraries were constructed from seedlings, hormone treated seedlings, flowers, developing seeds and primary leaves in the "entry vector" of the Gateway™ cloning system. After initial characterization of the libraries, they were shuttled into an expression vector (a yeast two-hybrid prey vector). To monitor for a size bias generally assumed to be inherent to in vitro recombination methods, the libraries were char-acterized before and after the transfer into the expression vector. However no significant difference could be detected. The functionality of the in vitro recombination system for the shuttling of entire libraries was then further tested by protein-protein interaction screens. The results of the library characterization and of the yeast two-hybrid screens and their implications for large-scale proteomic approaches are discussed.

Keywords Yeast two-hybrid · In vitro recombination cloning · *Arabidopsis* · cDNA libraries

L. Bürkle · H. Lehrach · A. Heyl
Max Planck Institute for Molecular Genetics,
Ihnestrasse 73,
14195 Berlin, Germany

S. Meyer
RZPD, German Resource Centre for Genome Research,
Heubnerweg 6,
14059 Berlin, Germany

L. Bürkle
Institute for Veterinary Biochemistry and Molecular Biology,
University of Zurich,
Winterthurerstrasse 190,
8057 Zurich, Switzerland

H. Dortay · A. Heyl (✉)
Institute for Biology, Applied Genetics, Free University Berlin,
Albrecht-Thaer-Weg 6,
14195 Berlin, Germany
e-mail: heyl@zedat.fu-berlin.de
Tel.: +49-30-83856550
Fax: +49-30-83854345

Introduction

The availability of complete genome sequences of a rapidly increasing number of species allows biological questions to be asked in greater complexity than ever before. Although it is exciting to do research in system biology, it has to be grounded in the molecular-level understanding of the processes in question (for review see Kitano 2002). Unicellular organisms and animal model systems, like *Escherichia coli*, *Saccharomyces cerevisiae*, *Caenorhab-ditis elegans*, or *Drosophila melanogaster* have been the focus of attention in the first of these large-scale experiments. However the sequencing of the model plants *Arabidopsis* and rice have enabled large-scale approaches to also be conducted in plants (Arabidopsis Genome 2000; Bürkle et al. 2003; Fang et al. 2002; Goff et al. 2002; Kersten et al. 2002; Roberts 2002; Yu et al. 2002). The need for large-scale investigation of biological processes becomes especially evident when considering the high number of genes of unknown function and often inaccurate bioinformatic predictions. For example, in *A. thaliana* the function of only 2,500 out of the roughly 28,000 genes have been investigated experimentally (Hilson et al. 2003) and whole genome transcription analysis reveals that up to 30% of the genome is wrongly annotated (Yamada et al. 2003). As functional studies on the genomic level are dependent on the availability of full-length cDNA clones, the quality of cDNA libraries as a source for those ORF

clones is critical. Thus the value of a cDNA library increases if its quality in terms of complexity and redundancy has been characterized. While cDNA libraries can be obtained from commercial and noncommercial sources, most libraries available from stock centres are not characterized at all.

Another important factor in large-scale approaches is the cloning strategy to capture cDNA into vectors designed for various purposes—a potentially time and resource consuming effort. Various methods have been developed to circumvent this bottleneck, such as homologous recombination in yeast or the application of in vitro recombination systems (e.g. Creator™ or Gateway™). The great advantage of these cloning systems is the possibility to transfer the gene sequences of an entire library quickly and easily from the original vector to different expression vector systems by in vitro recombination in a single tube reaction. This is particularly important as a good cDNA library can then be used in many different applications.

To address both points, the aims of this study were twofold: (1) to create a valuable resource for plant molecular biologists by generating high-quality and characterized *Arabidopsis* cDNA libraries in the pENTR1A vector from several tissues using the Gateway™ system, and (2) to test the effectiveness of in vitro recombination systems for the shuttling of cDNA libraries used in large-scale approaches. Towards those aims, we constructed oligo-(dT) primed, Gateway-compatible *Arabidopsis* cDNA libraries representing RNA from seedlings, inflorescences, hormone-treated seedlings, primary leaves and developing seeds. To evaluate the suitability of the Gateway™ system for shuttling entire libraries, the cDNA libraries were characterized before and after the in vitro recombination step by restriction analysis and sequence examination of randomly picked and sequenced clones. To assess the value of the newly generated libraries for proteomic experiments, we chose four proteins of the *Arabidopsis* cytokinin signal transduction pathway and two other genes of interest as bait proteins, which were subsequently screened for interaction partners using the cDNA libraries. The quality of the generated cDNA libraries, the suitability of the in vitro recombination system for entire cDNA libraries and the results of the yeast two-hybrid screens are discussed.

Materials and methods

Plant material

For the generation of the cDNA libraries, plants and tissues from *A. thaliana*, Col-0, were used. For seedlings, seeds were sterilized (in 50% ethanol, 25% H_2O_2, 25% H_2O) and placed on Whatman paper, incubated for 2 days at 4°C in the dark, then grown until the cotyledons fully expanded (5–7 days) at 26°C in 16 h light/8 h dark cycle on MS liquid medium. The seedlings were harvested and frozen directly in liquid nitrogen. For the hormone treatment, seedlings were grown on Whatman paper soaked in MS. After 5–7 days, when the cotyledons were fully expanded, the MS was replaced by MS supplemented with appropriate plant hormones (1 μM abscisic acid, 1-aminocyclopropane-1-carboxylic acid, benzyladenine, gibberellic acid, or α-naphthalene acetic acid; Sigma). After 2 h 50% of the plants were harvested and frozen directly in liquid nitrogen, with the other 50% being collected after 24 h in the same manner. This should assure the presence of mRNA of primary response genes as well as of genes upregulated after several hours in response to the hormone treatment. For seed RNA isolation, siliques at all stages of development were harvested, frozen in liquid nitrogen and subsequently freeze-dried. The seeds were collected out of the siliques by vacuum (Gibon et al. 2002). To isolate RNA from primary leaves, seeds were incubated for 2 days at 4°C and then sown on soil. After 7–10 days of growth under standard conditions (see above) in the greenhouse, the primary leaves of the seedlings were cut off and collected in liquid nitrogen. For the flower mRNA, inflorescences ranging from young flower buds to young siliques were harvested and immediately frozen in liquid nitrogen.

RNA extraction

The total RNA was extracted with TRIzol (Sigma) following the manufacturer's manual. In short, the tissue was ground with mortar and pestle cooled by liquid nitrogen. Of the ground tissue, 200 mg was incubated with 1 ml TRIzol for 5 min at room temperature (RT). Cell debris was removed by centrifugation (12,000×g at 4°C for 10 min) and 0.4 ml chloroform was added. After vortexing the mix was incubated for 5 min at RT. The phases were separated by centrifugation (12,000×g at 4°C for 15 min) and the aqueous phase was transferred to a new tube. A 0.6× volume of isopropyl alcohol and a 0.1× volume of 3 M sodium acetate were added to this aqueous phase and incubated for 10 min at 4°C. The precipitated RNA was pelleted by centrifugation (12,000×g at 4°C for 15 min) and after the removal of the supernatant the RNA was washed twice with 70% ethanol. After drying, the RNA was resuspended in 50 μl di-ethylepyrocarbonate-treated water. The quality and quantity of the RNA was checked by agarose gel and an Agilent biochip (data not shown).

Construction of the cDNA libraries

The cDNA libraries were constructed in collaboration with the RZPD German Resource Center for Genome Research, Heidelberg. In short, the total RNA was reverse-transcribed using an oligo-(dT) primer containing a *Not*I restriction site. A *Sal*I adapter was ligated to the 5′ end of the first strand of the cDNA for second strand synthesis. The resulting cDNAs were cloned via the *Sal*I/*Not*I sites into the pENTR1A vector (Gateway, Invitrogen) by restriction and ligation reactions and subsequently transformed into *E. coli* DH10B cells.

The primary clones of the libraries were expanded using the semisolid phase method described previously (Hanahan et al. 1991; Kriegler 1990). This method was used to minimize the representational bias which can occur during the expansion of plasmid cDNA libraries. In short, around 2.5×10^6 primary clones were mixed with 1.35% low melting agarose in LB media, incubated for 1 h at 0°C and were subsequently incubated for 45 min at 37°C without agitation. The cells were collected by centrifugation ($6,000\times g$ for 20 min at RT) and the pellet resuspended in a total of 100 ml 2× LB glycerol (12.5%). The cells were filtered through cheese cloth and stored in 10-ml aliquots at −70°C.

Sequence analysis

Raw sequence data was passed through several steps, such as cloning vector removal, quality evaluation, and conversion of data formats using the Staden Package software, version 2001 (Staden et al. 2000). Sequences shorter than 80 nucleotides after vector and quality clip and sequences in good concordance to *E. coli* sequences were eliminated. The remaining EST sequences were matched to annotated *Arabidopsis* genes via BLASTX 2.2.6 (Altschul et al. 1997) using the "TIGR5 Arabidopsis protein database" including mitochondrial and plastid sequences provided by TAIR (http://www.arabidopsis.org/) as reference and the BLAST parameter "*E* value" set to 1e−10.

Reading frames and the existence of 5′-untranslated regions were calculated from the BLAST reports using a BLAST parser (BPlite, Ian Korf, Genome Sequencing Center at Washington University, St. Louis, Mo.) and scripts written in Perl. The putative clone insert length was also calculated from the BLAST results and the length of the 3′UTR was taken into account.

The sequences were clustered using d2_cluster with default parameters (Burke et al. 1999) to screen for redundancy of the clone set. Before clustering the sequences were screened for interspersed repeats and low complexity DNA sequences using the CrossMatch and RepBase database for repetitive *Arabidopsis* DNA sequence elements (Green, http://www.genome.washington.edu/UWGC/analysistools/Swat.cfm). The sequences comprising each cluster were assembled using Phrap (http://www.phrap.org/phrap.docs/phrap.html) and the assembled cluster contig sequences as well as all respective cDNA sequences were assigned to the GO-process classes (Gene Ontology, http://www.geneontology.org/) based on the BLAST hits against TIGR5 and the *Arabidopsis* GOslim annotations as provided by TAIR (ftp://tairpub:tairpub@ftp.arabidopsis.org/home/tair/Genes/Gene_Ontology).

Plasmid construction

To screen for protein-protein interaction, a LexA DNA binding domain-based vector (pBTM116-D9), adapted to the Gateway™ system (Goehler et al. 2004) was used as a bait vector. As prey vector, the pGAD10 vector (Clontech) was digested with *Eco*RI and the resulting overhangs filled in with the Klenow enzyme (NEB). Each of the three Gateway conversion cassettes was inserted into the vector by ligation, resulting in three different vectors accommodating cDNA in all possible forward frames. Correct cloning and insertion were checked by sequencing.

Shuttling of libraries

The DNA of the respective pENTR cDNA library was prepared from approximately 7×10^5 clones using the Qiagen Plasmid Maxi Kit. The DNA was transferred to the pGAD10-GW prey vectors (a mixture of the three vectors each with one Gateway cassette) by the Gateway LR reaction. In short, the prey vectors were linearized by digestion with *Bss*HI. One hundred nanograms of each prey vector were mixed with 300 ng of the respective cDNA library in the pENTR1A vector, 4 μl 5× LR reaction buffer and made up with water to 16 μl. After the addition of 4 μl Clonase mix, the reaction was incubated at 25°C for at least 20 h. Then the mix was Proteinase K treated for 10 min at 37°C and subsequently deactivated by incubation for 10 min at 80°C. Afterwards, the whole LR reaction was used to transform 50 μl DH10B cells by electroporation. The cells were then plated on 16 large plates (22×22 cm) of selective solid medium (100 μg/ml Carbenicillin). On average there were about 1×10^5 clones per plate resulting in a total of 1.6 million clones. The clones of each plate were collected in 4 ml 1× TE (+ RNase A) and the plasmid DNA was isolated using a Qiagen Plasmid Mega Kit.

Yeast two-hybrid screens

The yeast strain L40ccαU (Goehler et al. 2004) containing the bait vector was grown in YPD liquid media at 30°C for 36 h. This starter culture was used to inoculate 100 ml YPD media to a cell density of OD_{600}=0.3. The resulting culture was grown at 30°C on a shaker until OD_{600}=0.7–0.9. The cells were harvested by centrifugation (5 min at $3,000\times g$) and resuspended in 25 ml 1× TE. The cells were then collected again by centrifugation, resuspended in 1 ml Mix 1 (100 mM lithium acetate, 1 M sorbitol and 0.5× TE) and incubated at RT for 10 min. Five hundred microlitres of salmon sperm (5 mg/ml) and 50 μg of the selected prey library were mixed and added to the cells. Then 7 ml Mix 2 (100 mM lithium acetate, 45% PEG 3350, 1× TE) was added to the cells and mixed carefully. The mix was incubated for 30 min at 30°C, before 880 μl DMSO was added followed by a 14-min heat shock at 42°C. Afterwards the yeast cells were washed in 5 ml H_2O and finally resuspended in 4 ml H_2O. For the determination of transformation efficiency, 2 μl of this mix was plated on SD medium lacking leucine and tryptophane (SDII) and incubated at 30°C. After 3 days the grown colonies were counted and the total number of transformants was calculated. The rest of the cells were plated on SD media

lacking leucine, tryptophane, histidine and uracil (SDIV) on four 22×22cm plates and incubated for 7 days at 30°C. To verify putative interactions, positive clones from the screens were restreaked on SDIV media. From those strains growing after 5 days on SDIV media, prey plasmids were isolated as described previously (Hoffman and Winston 1987). To increase the plasmid DNA quantity, the DNA was transformed into *E. coli* by electroporation and subsequently reisolated (Sambrook 2001). The DNA was analyzed by restriction digest and also tested by retransformation into a yeast strain containing the respective bait plasmid. Finally, the identity of the positives clones was analyzed by sequencing.

For each bait protein a minimum of 5×10^5 clones were screened in the yeast two-hybrid system.

Results

Generation and characterization of the different *A. thaliana* cDNA libraries

Generation of cDNA libraries

The quality of a cDNA library is determined by its complexity, redundancy, and the length of the cDNAs. To maximize complexity, the source and the quality of the mRNA is crucial. Therefore, to increase the number of different mRNAs, four different cDNA libraries were generated representing a variety of specific processes (e.g. hormone response) and developmental stages (e.g. desiccation). The RNAs were derived from seedlings, hormone-treated seedlings, flowers, primary leaves and seeds at different developmental stages. Since the RNA quantities isolated from seeds and primary leaves were very low, the samples were combined and one library was generated using this material. The resulting library is referred to as the seed library hereafter. The cDNA libraries were generated in collaboration with the RZPD German Resource Center for Genome Research (Heidelberg). To yield a high number of clones in sense orientation, the RNA was reverse transcribed using the oligo-(dT) priming method. After ligation of adapters and second strand synthesis, the cDNAs were cloned into the pENTR1A vector. Primary analysis revealed that each library had an average complexity of 7×10^5 clones (Table 1). While the number of primary clones is a bit lower than for other available cDNA libraries (e.g. Stratagene, TAIR), this is compensated for by the large number of full-length clones (Table 2).

Characterization of the libraries

The quality of these libraries was evaluated by sequencing a total of 744 clones and by restriction analysis of a total of more than 70 clones from all libraries in the pENTR vector. All plasmids analyzed with either method contained an insert and of all clones sequenced only five (0.7%) were in the minus orientation (as determined by closest BLAST hit), demonstrating the efficiency of the methods used. As expected all three forward frames are present in almost equal ratios (Table 2).

To evaluate the complexity of the libraries and the representation of different transcripts within the collection, all cDNA sequences and all contig cDNA sequences resulting from cluster analysis and subsequent assembly were assigned to the GO process ontology (Gene Ontology, http://www.geneontology.org/). As shown in Fig. 1, the process ontology assignment of all sequences indicates that some GO classes, especially the protein-metabolism category, are overrepresented in our libraries in comparison to all annotated *Arabidopsis* genes (TIGR5 release). Vice versa the underrepresentation of the "transcription"-GO class is remarkable. However, overall there is no substantial difference between the clones of our cDNA libraries and the whole annotated *Arabidopsis* genome in the functional grouping. These results indicate that the cDNA libraries described here correlate well with the different genes of the functional classes of the *Arabidopsis* genome.

Shuttling into the prey vector

After quality confirmation, the cDNA libraries were transferred from the pENTR1A vector to the prey vectors via an adapted LR reaction protocol (Invitrogen). As it is gen-
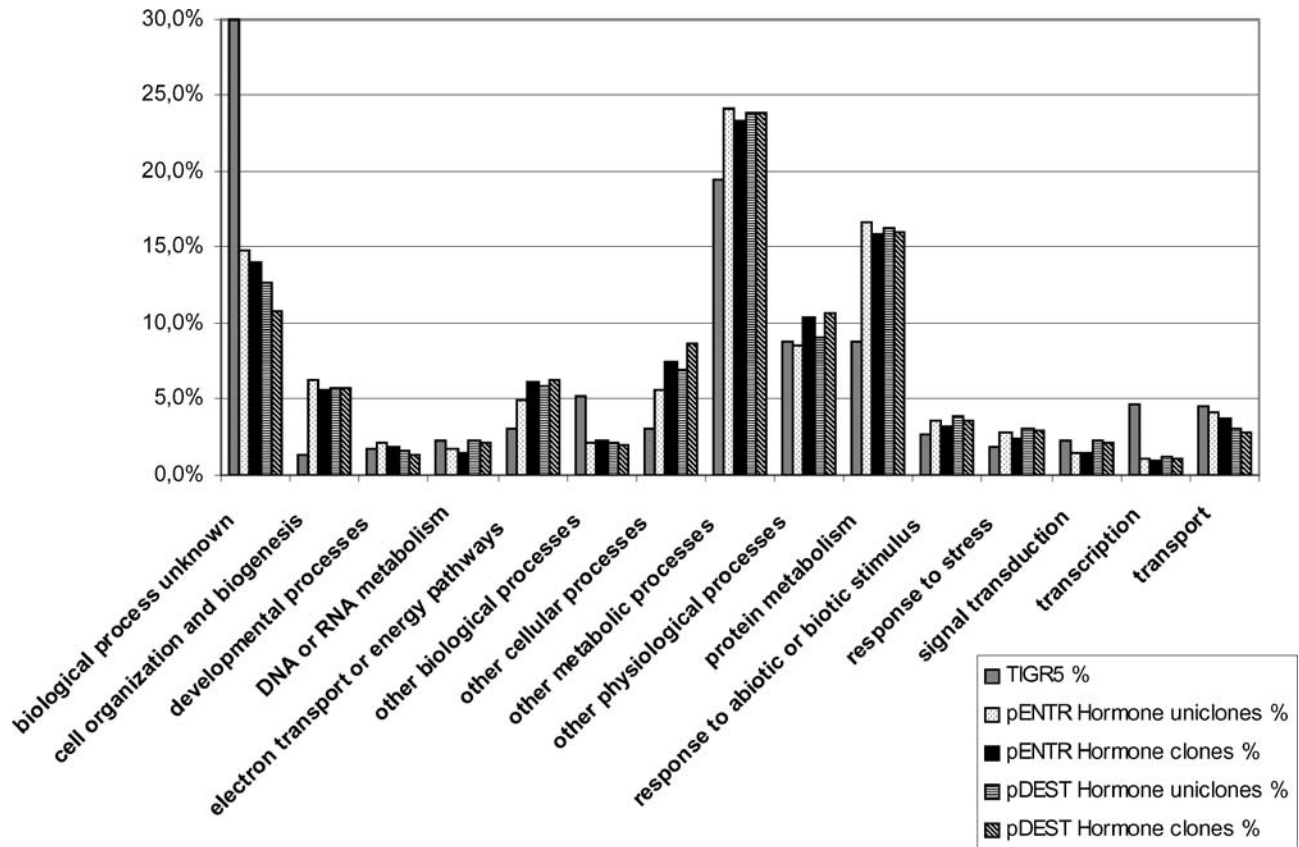
Table 1 Number of primary clones per library

Library	Total number of primary clones
Hormone-treated seedlings	776,000
Flowers	617,000
Seed	820,000
Seedlings	530,000

Table 2 Analysis of hormone library clones yielding a BLAST hit

Library	Σ sequences	Analysed sequence[a]	Sequence with TIGR5 hit	Frame (%)[b] 1	2	3	Minus	5′ UTR (%)[b]
pENTR Hormone	339	336	294	30.3	32.3	37.1	0.3	65.6
pENTR Flower	214	212	192	30.7	35.9	32.3	1.0	72.9
pENTR Seed	82	79	71	36.6	32.4	29.6	1.4	62.0
pENTR Seedling	109	105	97	35.1	29.9	35.1	0.0	72.2
pDEST Hormone	413	407	368	34.5	29.1	35.6	0.8	60.3

[a]Sequences with lengths over 80 nt after vector and quality clip and with no similarity to *E. coli* sequences
[b]Determined by BLAST result

Fig. 1 Comparison of all annotated *Arabidopsis* genes (TIGR5) and the cDNA libraries regarding their percentage content of the GO process ontology classes

erally assumed that a size bias is inherent to the in vitro recombination procedure Ohara and Temple (2001) tried to circumvent this problem by size fractioning the cDNA and shuttling different fractions separately.

In a preliminary experiment we also shuttled two size fractions separately, but could not detect any size bias compared with shuttling the whole library at once (data not shown). Therefore, we conclude that shuttling on the library scale did not have a significant size-selection bias.

To confirm these preliminary results on a larger scale, we determined the insert length of randomly picked and sequenced cDNA clones before and after the shuttling by in vitro recombination. The average insert size apparently did not change during the shuttling (Table 3) and is about 930 bp in both libraries. Interestingly, the longest cDNA clone of the library in the destination vector is longer than in the entry vector.

Redundancy of the cDNA libraries

Another problem common to the preparation of cDNA libraries is that redundancy greatly increases due to subsequent amplification steps. In our approach we tried to avoid this problem using different techniques (e.g. semisolid expansion). To validate the success of this approach, we sequenced more than 300 randomly chosen clones of the hormone cDNA library in the pENTR1A vector and of

our destination vectors, respectively. The sequences were annotated by BLAST similarity search. Sequences that were closely related were grouped together by a sequence clustering approach of the two different sequence sets—before and after the in vitro recombination of the hormone library was done. In both types of library an approximately equal number of clusters were found (Table 4). Cluster analysis merged sequences representing the same gene in one cluster. Using d2_cluster software sequences with a minimal overlap of 100 bp and at least 90% sequence identity were placed in the same cluster. Clusters containing only one or two members represented more than 90% of all cluster classes. As expected the largest clusters in both libraries consisted of proteins associated with photosynthesis. Overall the results of the cluster analysis indicated that the redundancy of the libraries was low and did not increase due to the in vitro recombination procedure (Table 4).

Table 3 Clone length before and after in vitro recombination

	pENTR hormone	pDEST hormone
Min. clone length	136	113
Max. clone length	2,730	3,655
Average clone length	947	922

Table 4 Comparison of cluster distribution before and after the in vitro recombination

Number of elements per cluster	Cluster frequency of cDNA libraries in:	
	pENTR vector	pDEST (pGAD10) vector
Singletons	276	300
2	19×	24×
3	1×	9×
5	2×[ab]	–
6	–	–
9	1×[c]	1×[c]
10	–	1×[b]
13	–	1×[a]
Sequences in total	336	407
Clusters in total/uniclones	298	336

[a] Sequences similar to genes annotated as *Arabidopsis* ribulose bisphosphate carboxylase, small subunit proteins
[b] Sequences similar to genes annotated as *Arabidopsis* chlorophyll *a/b*-binding proteins, LHCII type
[c] Sequences similar to genes annotated as *Arabidopsis* chlorophyll *a/b*-binding proteins, LHCI type

Application of the cDNA libraries in the yeast two-hybrid system

The quality of the prey libraries was tested by using several bait proteins with known interaction partners, as well as some proteins for which interactors have not yet been described. Bait proteins included some response regulators (ARR1, 2, 3 and 9) of the cytokinin signalling pathway (Grefen and Harter 2004; Heyl and Schmülling 2003). Several of these proteins are known to interact with some of the so-called *Arabidopsis* histidine phosphotransfer proteins (AHPs; Tanaka et al. 2004). For the proteins with unknown interaction partners, we tested adenine phosphoribosyl transferase 1 (APRT) and indole-3-glycerol phosphate synthase (IGPS) as bait proteins. APRT is involved in purine and cytokinin metabolism by

Table 5 Protein-protein interactions

Bait	Interacting prey (locus)	Function or homologous gene
ARR1	AHP1 (AT3G21510) AHP3 (AT5G39340)	Histidine phosphotransmitter (signal transduction)
ARR2	AHP1 (AT3G21510)	Histidine phosphotransmitter
ARR3	Unknown (At5G09540)	DNA J-like protein
ARR9	Unknown (At4G21105)	Unknown
APRT	APRT (AT1G27450)	Adenine phosphoribosyl transferase
IGPS	Unknown (At3g22440)	Hydroxyproline-rich glycoprotein

means of interconverting the corresponding free bases to the phosphorylated ribosyl (Gaillard et al. 1998; Mok and Mok 2001). IGPS is a key enzyme of tryptophane biosynthesis which is thought to be an important step in the production of the phytohormone auxin (Ouyang et al. 2000).

Using the prey library derived from the hormone-treated seedling, we were able to detect interactions with several different clones of AHPs, namely AHP1 and AHP3, when ARR1 was used as bait. AHP3 clones also interacted when ARR2 was used as bait. In total seven separate interactions could be detected using those baits (Table 5). ARR3 was found to interact with a DNA J-like protein (At5G21510). The function of At5G09540, the interaction partner of ARR9, is unknown.

Of the proteins with unknown interacting partners APRT was found to form homodimers, and IGPS interacted with a member of the hydroxyproline-rich glycoprotein family (At3g22440; Table 4).

Discussion

The generated cDNA libraries are of good quality

In most molecular experiments full-length cDNAs of the genes of interest are essential. As these are usually amplified from cDNA libraries, their quality is often crucial. The quality is defined by the complexity, as well as by the redundancy of the library and by the length and orientation of the inserted cDNAs. Complexity is especially an issue where rare transcripts are of special importance for the experiment. In some tissues the proportion of a particular mRNA can be very high, and in plants this is particularly true for photosynthetically active leaves, where the enzyme rubisco comprises up to half of the total protein (Ziegler 1998). Therefore, one must strike a careful balance between including such transcripts while trying to avoid an overrepresentation of the resulting cDNA clones within a library.

For these reasons we chose to generate several libraries and to isolate the RNA from tissues and organs in which the percentage of mRNA from the photosynthetic machinery should be relatively low. To get a good representation of the transcriptome expressed during the life cycle of *A. thaliana*, we collected samples from young and rapid developing tissue, namely seedlings and primary leaves, as well as from maturing tissues and from all stages of floral and seed development. Genes expressed only after induction with phytohormones were included by also isolating and combining RNA from seedlings treated with one of the five classical phytohormones (auxin, abscisic acid, cytokinin, ethylene and gibberellic acid). For the construction of the libraries we used the Gateway™ in vitro cloning system as this system has been proven to be very efficient in large scale approaches and is used in many functional genomic experiments (Kersten et al. 2002; Vincentelli et al. 2003; Li et al. 2004).

All sequenced clones were assigned to the GO process ontology to check if all these functional categories are represented within the library and to compare this functional distribution to the whole annotated *Arabidopsis* genome. The GO process ontology covers categories like protein metabolism, response to stress, transport, signal transduction, etc. This comparison is not exact, because of the redundancy of the libraries. Thus some clones originate from the same gene.

We tried to compensate for this fact by performing a cluster analysis and assigning the cluster sequences to the GO process classes. The comparison in Fig. 1 shows that our libraries represent the different functional groups of the *Arabidopsis* genome very well, thus demonstrating the versatility of these libraries for investigating a wide variety of biological questions. The fact that genes involved in the transcription process are underrepresented within the cDNA clones compared to the TIGR5 annotation set can be explained by the low level of transcripts, especially of transcription factors, within the cells.

In vitro recombination does not reduce the complexity of the libraries

Once a good cDNA library is at hand, it is highly desirable to use it for many different types of experiments. One way to achieve this is by generating the library in an entry vector of the Gateway system. It can then be shuttled into a large and increasing number of different expression vectors to perform further functional studies.

While the shuttling of single clones works very well, the shuttling of an entire cDNA library with the Gateway system has not yet been investigated in great detail. Of special interest is the shuttling of large cDNA clones, as it is generally believed that the efficiency of the recombination process decreases with the increase of insert size to over 2 kb (Hartley et al. 2000). One of the aims of the work presented in this paper was to test whether the in vitro recombination process leads to a bias towards smaller clones in shuttled cDNA libraries.

Therefore we sequenced more than 300 randomly picked clones before and after shuttling by in vitro recombination (Table 2). Our results indicate that there is no noticeable difference in the average size of the insert or the percentage of full-length clones and no need for size selection prior to the shuttling of a cDNA library (Table 3).

To our knowledge this is the first time the impact of the in vitro recombination reaction (LR) on the composition of a shuttled library has been analyzed in great detail. Earlier work by Ohara and Temple (2001) compared only the size fraction of cDNA plasmids before and after the LR reaction, but they also did not detect any difference in size fractions between the two libraries.

Another problem inherent to work with cDNA libraries is the overrepresentation of certain clones due to amplification steps. To examine this aspect we did a cluster analysis of the 300 sequenced clones of the hormone library in the entry vector and in the prey vectors (Fig. 1).

In both cases cluster dimension and frequency are very similar. Also those clusters that have many members belong to the same gene group. This demonstrates that no bias towards a particular clone or set of clones has occurred during the shuttling of the cDNAs from the entry vector into the destination vector.

The results of the reference yeast two-hybrid screens show the shuttled libraries are useful for protein-protein interaction studies

To evaluate the quality of our libraries for experiments in which the expression of fusion proteins is required, we performed several yeast two-hybrid screens. The proteins used as baits were chosen because most of them already had known interacting partners (for compilation see Heyl and Schmülling 2003). Indeed some of those known interactions were confirmed using our cDNA libraries. So, it had been shown in the past that ARR1 interacts with AHP1 and AHP3 (Imamura et al. 1999b; Suzuki et al. 2001). These interactions were found several times for each prey (Table 4). The same is true for ARR2, where the established interaction with AHP1 (Lohrmann et al. 2001) was verified. For ARR9 the known interactions with AHP1 and AHP3 could not be detected in our experiments. However, a new interaction with a protein of unknown function was detected. For ARR3, for which only interactions with AHPs had been previously reported in the literature (Imamura et al. 1999a), we could detect a novel interaction with a DNA J-like protein (Table 4).

Regarding the bait proteins involved in hormone metabolism (APRT, IGPS), we could also detect novel interactions, namely the homodimerization of APRT and interaction of IGPS with a hydroxyprolin-rich glycoproteins. The biological relevance of these interactions needs to be determined in further analysis.

The nature of the novel interactions found in this study is pending validation using other methods, such as affinity copurification. However, the results found in these few yeast two-hybrid screens indicate that the prey libraries used in these experiments are useful tools to confirm known and also to detect new protein interactions. Another interesting aspect of the found interaction concerns the length of the cDNAs. In general it is believed that full length clones are less likely to interact than partial clones. Since our libraries contain a high number of full length clones one could question their usefulness for yeast two-hybrid screens. However of the 11 interactions detected, 8 were full length and even contained a small 5′ UTR, meaning that at least in the interactions presented here, the full length did not interfere with the interaction.

Conclusion

With the Gateway-compatible pENTR libraries from four different tissues, we created a valuable resource for further applications and functional studies in *Arabidopsis* research

by using these libraries in the corresponding destination vectors. Furthermore, we were able to demonstrate the feasibility of the Gateway™ in vitro recombination system not only for single clones, but for entire libraries.

References

Altschul F, Madden TL, Chaffer AA, Zhang J, Zhang Z, Miller W, Lipman DL (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

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

Burke J, Davison D, Hide W (1999) d2_cluster: a validated method for clustering EST and full-length cDNA sequences. Genome Res 9:1135–1142

Bürkle L, Feilner T, Brenner W, Giavalisco P, Lehrach H, Kersten B (2003) Genomic and proteomic approaches in plant research. In: Pandalai SG (ed) Recent research developments in plant molecular biology. Research Signpost, Kerala, pp 239–263

Fang Y, Macool DJ, Xue Z, Heppard EP, Hainey CF, Tingey SV, Miao GH (2002) Development of a high-throughput yeast two-hybrid screening system to study protein-protein interactions in plants. Mol Genet Genomics 267:142–153

Gaillard C, Moffatt BA, Blacker M, Laloue M (1998) Male sterility associated with APRT deficiency in *Arabidopsis thaliana* results from a mutation in the gene *APT1*. Mol Gen Genet 257:348–353

Gibon Y, Vigeolas H, Tiessen A, Geigenberger P, Stitt M (2002) Sensitive and high throughput metabolite assays for inorganic pyrophosphate, ADPGlc, nucleotide phosphates, and glycolytic intermediates based on a novel enzymic cycling system. Plant J 30:221–235

Goehler H, Lalowski M, Stelzl U, Waelter S, Stroedicke M, Worm U, Droege A, Lindenberg KS, Knoblich M, Haenig C, Herbst M, Suopanki J, Scherzinger E, Abraham C, Bauer B, Hasenbank R, Fritzsche A, Ludewig AH, Buessow K, Coleman SH, Gutekunst CA, Landwehrmeyer BG, Lehrach H, Wanker EE (2004) A protein interaction network links GIT1, an enhancer of Huntingtin aggregation, to Huntington's disease. Mol Cell 15:853–865

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92–100

Grefen C, Harter K (2004) Plant two-component systems: principles, functions, complexity and cross talk. Planta 219:733–742

Hanahan D, Jessee J, Bloom FR (1991) Plasmid transformation of *Escherichia coli* and other bacteria. Methods Enzymol 204:63–113

Hartley JL, Temple GF, Brasch MA (2000) DNA cloning using in vitro site-specific recombination. Genome Res 10:1788–1795

Heyl A, Schmülling T (2003) Cytokinin signal perception and transduction. Curr Opin Plant Biol 6:480–488

Hilson P, Small I, Kuiper MT (2003) European consortia building integrated resources for *Arabidopsis* functional genomics. Curr Opin Plant Biol 6:426–429

Hoffman CS, Winston F (1987) A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. Gene 57:267–272

Imamura A, Hanaki N, Nakamura A, Suzuki T, Taniguchi M, Kiba T, Ueguchi C, Sugiyama T, Mizuno T (1999a) Compilation and characterization of *Arabidopsis thaliana* response regulators implicated in His-Asp phosphorelay signal transduction. Plant Cell Physiol 40:733–742

Imamura A, Hanaki N, Nakamura A, Suzuki T, Taniguchi M, Kiba T, Ueguchi C, Sugiyama T, Mizuno T (1999b) Compilation and characterization of *Arabidopsis thaliana* response regulators implicated in His-Asp phosphorelay signal transduction. Plant Cell Physiol 40:733–742

Kersten B, Bürkle L, Kuhn EJ, Giavalisco P, Konthur Z, Lueking A, Walter G, Eickhoff H, Schneider U (2002) Large-scale plant proteomics. Plant Mol Biol 48:133–141

Kitano H (2002) Systems biology: a brief overview. Science 295:1662–1664

Kriegler M (1990) Assembly of enhancers, promoters, and splice signals to control expression of transferred genes. Methods Enzymol 185:512–527

Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JD, Chesneau A, Hao T, Goldberg DS, Li N, Martinez M, Rual JF, Lamesch P, Xu L, Tewari M, Wong SL, Zhang LV, Berriz GF, Jacotot L, Vaglio P, Reboul J, Hirozane-Kishikawa T, Li Q, Gabel HW, Elewa A, Baumgartner B, Rose DJ, Yu H, Bosak S, Sequerra R, Fraser A, Mango SE, Saxton WM, Strome S, Van Den Heuvel S, Piano F, Vandenhaute J, Sardet C, Gerstein M, Doucette-Stamm L, Gunsalus KC, Harper JW, Cusick ME, Roth FP, Hill DE, Vidal M (2004) A map of the interactome network of the metazoan *C. elegans*. Science 303:540–543

Lohrmann J, Sweere U, Zabaleta E, Baurle I, Keitel C, Kozma-Bognar L, Brennicke A, Schafer E, Kudla J, Harter K (2001) The response regulator ARR2: a pollen-specific transcription factor involved in the expression of nuclear genes for components of mitochondrial complex I in *Arabidopsis*. Mol Genet Genomics 265:2–13

Mok DWS, Mok MC (2001) Cytokinin metabolism and action. Annu Rev Plant Physiol Plant Mol Biol 52:89–118

Ohara O, Temple G (2001) Directional cDNA library construction assisted by the in vitro recombination reaction. Nucleic Acids Res 29:E22

Ouyang J, Shao X, Li J (2000) Indole-3-glycerol phosphate, a branchpoint of indole-3-acetic acid biosynthesis from the tryptophan biosynthetic pathway in *Arabidopsis thaliana*. Plant J 24:327–333

Roberts JK (2002) Proteomics and a future generation of plant molecular biologists. Plant Mol Biol 48:143–154

Sambrook J (2001) Molecular cloning a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Staden R, Beal KF, Bonfield JK (2000) The Staden package, 1998. Methods Mol Biol 132:115–130

Suzuki T, Sakurai K, Ueguchi C, Mizuno T (2001) Two types of putative nuclear factors that physically interact with histidine-containing phosphotransfer (Hpt) domains, signaling mediators in His-to-Asp phosphorelay, in *Arabidopsis thaliana*. Plant Cell Physiol 42:37–45

Tanaka Y, Suzuki T, Yamashino T, Mizuno T (2004) Comparative studies of the AHP histidine-containing phosphotransmitters implicated in His-to-Asp phosphorelay in *Arabidopsis* thaliana. Biosci Biotechnol Biochem 68:462–465

Vincentelli R, Bignon C, Gruez A, Canaan S, Sulzenbacher G, Tegoni M, Campanacci V, Cambillau C (2003) Medium-scale structural genomics: strategies for protein expression and crystallization. Acc Chem Res 36:165–172

Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, Southwick AM, Wu HC, Kim C, Nguyen M, Pham P, Cheuk R, Karlin-Newmann G, Liu SX, Lam B, Sakano H, Wu T, Yu G, Miranda M, Quach HL, Tripp M, Chang CH, Lee JM, Toriumi M, Chan MM, Tang CC, Onodera CS, Deng JM, Akiyama K, Ansari Y, Arakawa T, Banh J, Banno F, Bowser L, Brooks S, Carninci P, Chao Q, Choy N, Enju A, Goldsmith AD, Gurjal M, Hansen NF, Hayashizaki Y, Johnson-Hopson C, Hsuan VW, Iida K, Karnes M, Khan S, Koesema E, Ishida J, Jiang PX, Jones T, Kawai J, Kamiya A, Meyers C, Nakajima M, Narusaka M, Seki M, Sakurai T, Satou M, Tamse R, Vaysberg M, Wallender EK, Wong C, Yamamura Y, Yuan S, Shinozaki K, Davis RW, Theologis A, Ecker JR (2003) Empirical analysis of transcriptional activity in the *Arabidopsis* genome. Science 302:842–846

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Li J, Liu Z, Qi Q, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Zhao W, Li P, Chen W, Zhang Y, Hu J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Tao M, Zhu L, Yuan L, Yang H (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296:79–92

Ziegler H (1998) Physiologie des Stoff- und Energiewechsels. In: Sitte P, Ziegler H, Ehrendorfer F, Bresinsky A (eds) Strasburger. Lehrbuch der Botanik. Fischer, Stuttgart, Germany