S41-003

Photosynthesis and related genomics

Kurth, J., Pesaresi, P., Biehl, A., Richly, E., Weigel, M., Varotto, C., Maiwald, D. Salamini, F. and Leister, D.

Max-Planck-Institut f r Z chtungsforschung, Carl-von-Linn Weg 10, D-50829 K ln, Germany; fax: +49-2215062413, e-mail: leister@mpiz-koeln.mpg.de

Keywords: Arabidopsis, gene tagging, transcriptomics, photosystem, photosynthesis

Higher plant photosynthesis is driven by the absorption of photons in the light harvesting complexes (LHCs) of photosystem I (PSI) and II (PSII). Trapping of excitation energy in the reaction centres triggers electron and proton transport across the thylakoid membrane resulting in the synthesis of NADPH and ATP. The photosynthetic reactions are localised in the chloroplast, which derived from an ancestral cyanobacterial endosymbiont (Douglas, 1998). Best known as photosynthetic organelles, chloroplasts also have other functions, such as the synthesis of amino acids, fatty acids and lipids, plant hormones, nucleotides, vitamins and secondary metabolites, as well as the assimilation of nitrogen and sulphur. Plastids also host the transcriptional and translational machinery necessary for the expression of their own genome (the plastome). In essence, photosynthesis takes place within a complex protein network and is thus influenced by a multitude of metabolic processes. This is why the absence or modification of any of the plethora of gene products localised in the chloroplast may affect its photosynthetic function (Mascia and Robertson, 1978; Rochaix, 1992; von Wettstein *et al.*, 1971).

Arabidopsis thaliana has become a model organism for the study of the biology of flowering plants. In parallel with the sequencing of its genome, new molecular technologies allow to identify proteins, determine their expression patterns at a genome-wide scale and define their post-translational modifications and finally their functions (reviewed in: Somerville and Somerville, 1999). Here, the term functional genomics refers to the whole range of current strategies which can contribute to the assignment of functions to genes at genome-scale and which cover the fields of genomics, transcriptomics and proteomics (**Fig. 1**). The application of forward genetics, reverse genetics, microarray-based measurements of gene expression and bioinformatics for the molecular dissection of *Arabidopsis* photosynthesis will be described in the following sections.

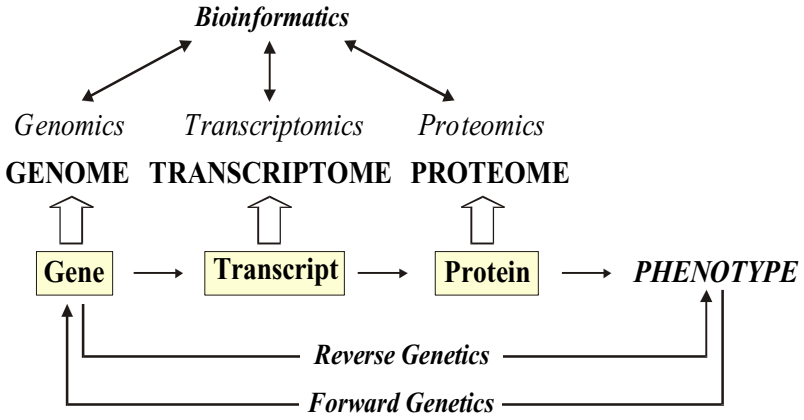

Fig.1. Overview of high-throughput procedures applied in Functional Genomics

Identification of novel photosynthetic mutations and of the genes affected

Mutants with altered photosynthetic performance are the biological material of choice to identify photosynthesis-relevant functions. Two classes of mutants are considered in our phenotypic screen of *Arabidopsis* insertion lines: mutants with an altered photosynthetic performance (PAM mutants) identified by using an automated screening system for individuals with alterations in the effective quantum yield of PSII (Φ_{II}; Varotto *et al.*, 2001) and mutants with altered pigmentation but having a wild-type Φ_{II} (ALP mutants). A total of 11,000 lines (5,500 *En*-transposon lines and 5,500 T-DNA lines) have been screened, resulting in the identification of 56 PAM and 79 ALP mutants, respectively (**Tab. 1**).

So far, we have identified 12 different mutated genes, coding for both structural and putative regulatory proteins, and of which most putative gene products are imported into the chloroplast. Nine genes confer novel photosynthesis-related functions and their molecular biology, biochemistry and physiology is currently being analysed. The most detailed characterisations have been performed for the mutation in *psaE1* (coding for the PSI-E subunit of photosystem I; Varotto *et al.*, 2000), *Prpl11* (coding for a protein of the plastid large ribosomal subunit; Pesaresi *et al.*, 2001a) and a gene coding for a putative cytoplasmatic acyl transferase.

Tab.1. Overview of identified PAM and ALP mutants

Type of mutant	***En*-population**	**T-DNA population**	**total**
	5,500 lines	5,500 lines	11,000 lines
PAM			
- normal pigmentation	6	10	16
- altered pigmentation	12	28	40
total	18	38	56
ALP			
- albino	8	26	34
- yellow or pale green	11	29	40
- others	1	4	5
total	20	59	79

Study of chloroplast functions by transcriptome analysis

Due to its limited throughput, forward genetics is a rather inefficient tool for large-scale analysis of gene functions. Reverse genetics will undoubtedly contribute more results. The vast majority of the genes relevant to photosynthesis that we have identified so far by forward genetics code, in fact, for proteins targeted to the chloroplasts. Therefore, an appropriate - straightforward and saturating -reverse genetics approach should therefore aim at the identification of mutants for all nuclear genes coding for chloroplast proteins. Such a large-scale screening effort has not yet been launched due to limitations on throughput rates.

To identify the most relevant genes affecting the photosynthetic performance of chloroplasts, we have set up a large-scale expression profiling system for - currently about 50% of - *Arabidopsis* genes predicted to encode for chloroplast proteins. In a first step, the genes encoding chloroplast proteins were identified based on their N-terminal pre-sequences (transit peptides, see next section). Between 1,900 and 2,500 *Arabidopsis* proteins with chloroplast localisation were predicted of which at least one-third were of cyanobacterial origin (Abdallah *et al.*, 2000). In a second step, about 1,800 nuclear genes whose products are predicted to be localised in chloroplasts were PCR-amplified from genomic *Arabidopsis* DNA as gene sequence-tags (GSTs) for the generation of DNA arrays (Pesaresi *et al.*, 2001b). These GST arrays (**Fig. 2**) are used for the study of the expression profiles of chloroplast genes by exploiting a variety of mRNA sources.

Currently, expression profiles in response to light stress experiments, photosynthetic mutants, and ecotypes that differ in their photosynthetic performance are analysed, demonstrating that the GST array can be very useful for monitoring differential expression of genes related to chloroplast functions. Many of the genes identified as being differentially expressed either in different environmental conditions or between the wild type and photosynthetic mutants exhibit no homology to protein sequences with known functions present in public databases. For such genes annotated as hypothetical or putative , GST array experiments confirm that these ORFs are expressed and represent functional genes. In this sense, the array allows us to carry out gene-function discovery experiments. The next upgrade of this array will include the genes for all 3,100 chloroplast proteins (Pesaresi *et al.*, 2001b) will in the long-term help us to predict or assign functions to the plethora of proteins present in higher plant chloroplasts.

On condition that similar plastid defects result also in similar transcriptional responses, the GST array can be applied for the classification of photosynthetic mutants, even if the molecular nature of their lesions is unknown. In combination with the analysis of known mutations affecting photosynthesis, this approach aims to narrow down the number of probable candidate genes thus enabling gene identification by the candidate gene approach .

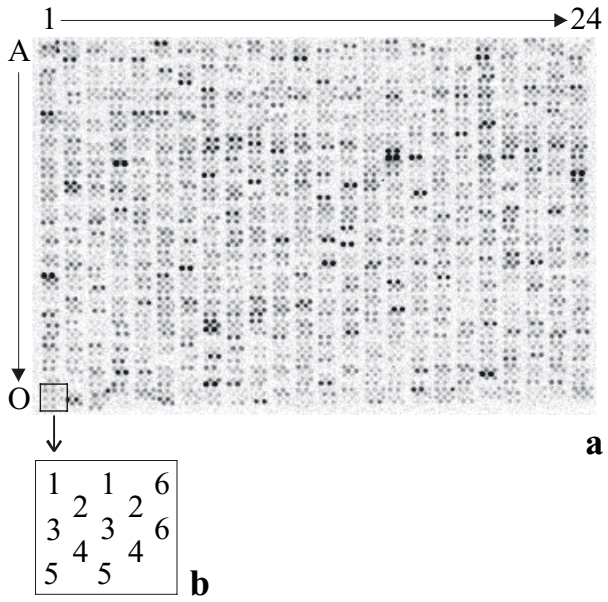

Fig.2. DNA array for plastid functions - exemplary hybridisation pattern and organisation. The DNA array used for the study of chloroplast functions contains 1,827 GSTs in duplicate, organized in 24 x 16 modules with up to 6 different GSTs or control spots each. **a**, complex ^{33}P-labelled cDNA population prepared from 3-week-old plants was hybridised to the filter; **b,** the diagram shows the order of duplicates spotted in the module O3. Control DNAs (*Arabidopsis* genomic DNA, human or fungal cDNAs) were spotted at position 6 in each module of rows A and O and were visualized by adding appropriate control cDNAs to the complex cDNA population.

Reverse Genetics for PSI subunits

The genome-scale identification and analysis of functions affecting photosynthesis must also consider the photosynthetic apparatus. The photosystems of higher plants are mosaics of plastid- and nucleus-encoded protein subunits, and some of the coding genes present in the nuclear genome of *Arabidopsis* are redundant. PSI with its relatively large fraction of nuclear encoded subunits is the object of choice for reverse genetics involving nuclear genes. When studying structure-function relationships of higher plant PSI, several questions emerge:

1. What is the biological significance of the duplication of some of the nuclear genes coding for PSI subunits?
2. Why do some subunits contain N-terminal extensions that are not present in their cyanobacterial pendants?
3. What is the function of the subunits PSI-G, -H and -N, which are specific to higher plants?
4. Which subunits are involved in cyclic electron transport, photoinhibition and regulation of PSI activity?

To address these questions and to dissect the processes leading to energy balancing, mutants for all individual subunits of PSI are required. We have adopted a medium-sized reverse genetics approach involving the ZIGIA *Arabidopsis En*-population and most publicly available *Arabidopsis* insertion mutant populations to identify mutants for these genes (**Fig. 3**). In contrast to alternative co-suppression and antisense approaches, the analysis of stable knockout mutants generated by T-DNA or transposon insertions is required to dissect the impact of each of the genes coding for PSI-E, namely *psaE1* and *psaE2*, PSI-D (*psaD1* and *psaD2*), PSI-H (*psaH1* and *psaH2*), and plastocyanin (*petE1* and *petE2*). Furthermore, stable knockout mutants are particularly useful for generating lines with multiple mutations in order to study the functional redundancy of the subunits forming PSI.

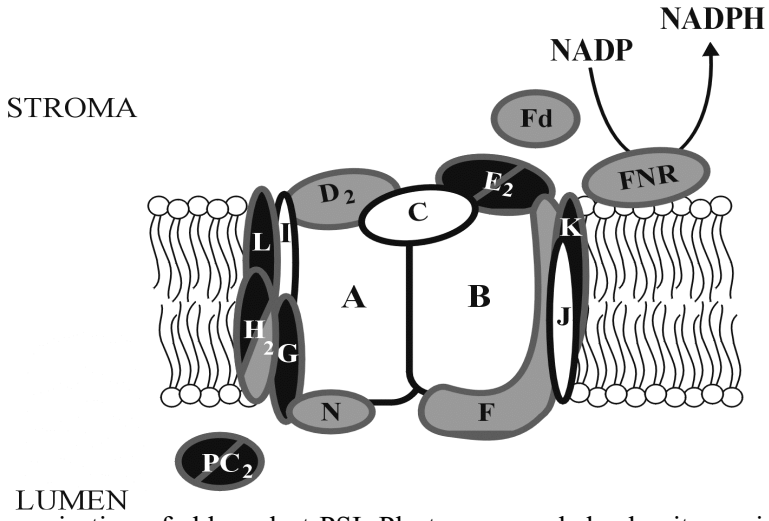

Fig.3. Molecular organisation of chloroplast PSI. Plastome-encoded subunits are indicated by white shading, whilst nucleus-encoded subunits are indicated in grey. Subunits for which insertion mutants were identified are indicated by a black filling. Subunits for which more than one gene copy exists are indicated by appropriate ciphers. PC, plastocyanin; Fd, ferredoxin; FNR, Fd-NADPH reductase

Preliminary analyses of our collection of mutants for individual subunits for PSI indicate multiple involvements in the regulation of state transition. Mutants for some subunits confer the lack of state 2 (association of the mobile pool of LHCII with PSI) as described previously (Lunde *et al.*, 2000), while other mutants exhibit the opposite case: the constitutive association of LHCII with PSI. Currently, the latter class of mutants is being characterised intensively and indicates an altered phosphorylation state of LHCII which contributes to the observed change in interphotosystemic energy distributions.

Bioinformatics

The analysis of the completed sequence of the model plant *Arabidopsis* allows to assess genome-wide structure and dynamics of gene families. To assist the generation of GST arrays (see above) we have identified the fraction of *Arabidopsis* genes putatively coding for chloroplast proteins. The recognition of chloroplast targeting signals provides a means of estimating which and how many proteins are localised in the photosynthetic organelle. The chloroplast plastome itself codes for only 87 proteins, while all other resident proteins are encoded by nuclear genes, synthesised as precursors in the cytosol and targeted to the chloroplast by their N-terminal pre-sequences (transit peptides).

An initial prediction of the number and evolutionary origin of the proteins located in the chloroplast (the chloroplast proteome) was performed at the end of 1999, when about half of the *Arabidopsis* genome had been sequenced (Abdallah *et al.*, 2000). The approach used combined a proteome-wide search for putative chloroplast transit peptides with a homology-based comparison of the chloroplast proteome with the total protein complement of a cyanobacterium (*Synechocystis*). When the total genome sequence became available end of 2000 (The *Arabidopsis* Initiative, 2000), this complete catalogue of predicted proteins of *Arabidopsis* was analysed (Pesaresi *et al.*, 2001b). By using two different prediction programs, *ChloroP* and *TargetP*, between 2,500 proteins (*ChloroP*) and 3,100 proteins (*TargetP*), respectively, with chloroplast localisation were predicted. Of those, at least 31% (*ChloroP*) and 29% (*TargetP*) are of cyanobacterial origin. Thus, in *Arabidopsis* chloroplast proteins account for up to 12% of the total proteome of *Arabidopsis* and are the proteome fraction of choice for the study of chloroplast performance and, particularly, photosynthesis by means of functional genomics.

For the generation of the chloroplast protein DNA arrays (see above), we have developed the software package *GST-PRIME* for retrieving and editing gene sequences from public databases, and then designing sets of primer pairs for use in gene amplification. Primers are designed by the program for direct amplification of GSTs from either genomic DNA or cDNA. *GST-PRIME* test runs on 2,000 randomly selected *Arabidopsis* and *Drosophila* genes were successful in about 90% of the cases. *GST-PRIME* primer pairs were tested during the amplification of the 1,900 GSTs for the set-up of the chloroplast protein DNA array: Of the primer pairs used in PCRs with genomic DNA, 95% generated the correct amplicons. Thus, *GST-PRIME* is generally suitable for strategies that aim at the large-scale or specific gene amplification from the genomes of multicellular eukaryote.

Outlook

Identification and characterisation of novel photosynthesis-related gene functions by forward genetics will continue, but the efforts to identify gene-phenotype relationships will be shifted to reverse genetics. For this purpose, suitable candidate genes will be identified by the transcriptomics approach. Currently we are establishing a second generation of DNA arrays containing 4,000 genes encoding all predicted chloroplast proteins in *Arabidopsis*. Also in the future, the analysis of energy dissipation and balancing in *Arabidopsis* photosynthesis will be followed in the context of national and international collaborations. In the long term, double and higher-grade multiple mutants will be utilised to unambiguously assign functions to specific subunits and to dissect redundant gene functions in photosynthesis. The complete sequence of the model plant *Arabidopsis* will be exploited to study aspects of photosynthesis- and chloroplast-relevant evolutionary processes. With more plant genomic sequences becoming available, intra- and interspecific analyses of genome organisation will become possible.

Acknowledgements

This work was supported by the Bundesministerium f r Bildung und Forschung, and by the Claussen-Stiftung im Stifterverband f r die Deutsche Wissenschaft (D.L.). Grateful acknowledgements are extended to Koen Dekker and Heinz Saedler.

References

Abdallah, F., F. Salamini, and D. Leister, 2000: A prediction of the size and evolutionary origin of the proteome of chloroplasts of *Arabidopsis*, *Trends Plant Sci.* **5**, 141-142.

Douglas, S. E., 1998: Plastid evolution: origins, diversity, trends, *Curr. Opin. Genet. Dev.* **8**, 655-661.

Douglas, S. E., 1998: Plastid evolution: origins, diversity, trends, *Curr. Opin. Genet. Dev.* **8**, 655-661.

Lunde, C., P. E. Jensen, A. Haldrup, J. Knoetzel, and H. V. Scheller, 2000: The PSI-H subunit of photosystem I is essential for state transitions in plant photosynthesis, *Nature,* **408**, 613-615.

Mascia, P. N., and D. S. Robertson, 1978: Studies of chloroplast development in four maize mutants defective in chlorophyll biosynthesis, *Planta* **143**, 207-211.

Pesaresi, P., C. Varotto, J. Meurer, P. Jahns, F. Salamini, and D. Leister, 2001a: Knock-out of the plastid ribosomal protein L11 in *Arabidopsis*: effects on mRNA translation and photosynthesis, *Plant J.* (in press)

Pesaresi, P., C. Varotto, E. Richly, J. Kurth, F. Salamini, and D. Leister, 2001b: Functional genomics of *Arabidopsis* photosynthesis, *Plant Physiol. Biochem.* **39**, 285-294.