

**Untersuchung von Chloroplastenfunktion mit Hilfe  
differentieller Genexpressionsanalyse**

Inaugural-Dissertation  
zur Erlangung des Doktorgrades  
der Mathematisch-Naturwissenschaftlichen Fakultät  
der Universität zu Köln

Vorgelegt von  
Alexander Biehl aus Siegen

Köln, im Dezember 2004

Die vorliegende Arbeit wurde am Max-Planck-Institut für Züchtungsforschung, Köln-Vogelsang, in der Abteilung Pflanzenzüchtung und Ertragsphysiologie (Prof. Dr. F. Salamini) in der Arbeitsgruppe von Dr. D. Leister angefertigt.

Berichterstatter:

Prof. Dr. Francesco Salamini

Prof. Dr. Ulf-Ingo Flügge

Tag der mündlichen Prüfung: 01. 02. 2005

## Inhaltsverzeichnis

1. Einleitung.....	1
1.1 Das Plastom heutiger Chloroplasten.....	3
1.2 Endosymbiontischer Gentransfer.....	5
1.3 Proteinimport aus dem Cytosol.....	6
1.4 Das Proteom des Chloroplasten.....	8
1.5 Chloroplastenfunktionen.....	10
1.5.1 Photosynthese.....	11
1.6 Funktion und Herkunft des Photosynthese-Apparates.....	12
1.7 Regulation der nukleären Genexpression.....	14
1.7.1 Kommunikation vom Organell zum Zellkern.....	16
1.8 Transkriptomics des Plastiden.....	22
1.8.1 Clusteranalyse zur Gruppierung der Gene in Regulons.....	23
1.9 Ziel der Arbeit.....	24
2. Material & Methoden.....	25
2.1 Chemikalien und Enzyme.....	25
2.2 Geräte und Verbrauchsmaterial.....	26
2.3 Software.....	26
2.4 Zugang zu den Rohdaten und ergänzenden Informationen .....	27
2.5 Pflanzenmaterial.....	27
2.6 Design des 3292-GST-Makroarray.....	28
2.7 Hybridisierungsexperimente.....	30
2.3.1 Extraktion von Gesamt-RNA.....	31
2.3.2 cDNA-Erststrangsynthese.....	32
2.3.3 Hybridisierungsprozedur.....	33
2.8 Datenanalyse.....	34
2.8.1 Makroarray-Quantifizierung.....	34

2.8.2	Statistische Analyse der Makroarraydaten.....	35
2.8.3	Prozessierung und Interpretation der Expressionsdaten.....	36
2.8.3.1	Funktionelle Klassifizierung des Transkriptoms.....	36
2.8.3.2	Definition von Regulons und Transkriptomklassen.....	37
2.8.3.3	Hierarchisches Clustering.....	38
2.8.3.4	Self-Organizing-Maps (SOM-Clustering) .....	40
2.8.3.5	k-means Clustering.....	41
2.8.4	Suche nach überrepräsentierten Sequenzmotiven in der Promoterregion der Gene eines Regulons.....	41
3.	Ergebnisse.....	42
3.1	Experimenteller Aufbau zur Untersuchung des nukleären Transkriptoms des Chloroplasten.....	42
3.2	Analyse des differentiell exprimierten nukleären Trankriptoms des Chloroplasten.....	48
3.2.1	Funktionelle Klassifizierung des signifikant differentiell exprimierten Transkriptoms des Chloroplasten.....	52
3.3	Untersuchung des nukleären Transkriptoms des Plastiden durch Clusteranalyse.....	55
3.3.1	Kladogramm der 101 differentiellen Transkriptome.....	58
3.3.2	Analyse von Regulons.....	59
3.3.2.1	Vergleich von unterschiedlichen Clusteralgorithmen.....	62
3.3.2.2	Expressionsmuster der Regulons.....	63
3.3.2.3	Zusammenhang zwischen Genexpression und subzellulärer und subplastidärer Lokalisation der Genprodukte.....	66
3.3.2.4	Funktionelle Analyse der Regulons.....	68
3.3.3.5	Detailanalyse ausgewählter Regulons.....	72

4. Diskussion.....	78
4.1 Ein 3292-GST-Makroarray zur Untersuchung des plastidären Transkriptoms.....	78
4.2 Effekte von 101 unterschiedlichen Stimuli auf die Expression des nukleären Transkriptoms des Chloroplasten.....	79
4.2.1 Integration der 101 Stimuli in eine von drei Hauptklassen der transkriptionellen Antwort.....	84
4.2.2 Hinweise für einen integrativen Regulator der Transkription photosyntheserelevanter Gene.....	86
4.3 Netzwerk der Expressionsregulation des nukleären Transkriptoms des Chloroplasten.....	92
4.4 Interpretation aus evolutionäre Sicht .....	93
4.5 Ausblick .....	95
5. Zusammenfassung / Summary.....	96
5.1 Zusammenfassung.....	96
5.2 Summary.....	98
6. Literaturverzeichnis.....	100
7. Anhang.....	118
Abkürzungsverzeichnis.....	132
Erklärung.....	134
Lebenslauf.....	135
Danksagung.....	136

## 1 Einleitung

Die genetisch semiautonomen aber autoreduktiven Organellen der Pflanzenzelle, die Chloroplasten wie auch die Mitochondrien, sind die biosynthetischen beziehungsweise bioenergetischen Kompartimente der Zelle. Sequenzanalysen der Mitochondriengene und ihrer genomischen Organisation lassen darauf schließen, dass die Mitochondrien vor mehr als 1.5 Milliarden Jahren aus der Endocytobiose eines  $\alpha$ -Proteobakteriums in ein anaerobes, autotrophes Archaeobakterium entstanden sind (Gray et al., 1999). Die Chloroplasten der heutigen Pflanzen und Algen gehen auf ein zweites, singuläres endocytotisches Ereignis zurück, der Aufnahme eines Vorfahren der Cyanobakterien in eine heterotrophe, eukaryotische Wirtszelle (Douglas, 1998; Delwiche et al., 1999; Raven & Douglas, 2003).

Mit und zum Zeitpunkt der Phagozytose entstand aus Wirtszelle und Endosymbiont eine intertaxonische Chimäre aus zwei autonomen Zellen. Im Zuge ihrer Entwicklung vom ursprünglich frei lebenden Prokaryoten hin zu einem hochspezialisierten Zellorganell, verloren jedoch die Chloroplasten und die Mitochondrien ihre Autonomie. Dies spiegelt sich am deutlichsten in einer massiven Genomreduktion der Organellen im Vergleich zum Genom der heute lebenden autonomen Cyanobakterien beziehungsweise Proteobakterien wider. Zwar besitzen auch die Chloroplasten noch immer ein eigenes Genom (Plastom) und eine eigene Maschinerie zur Proteinbiosynthese über die Vorgänge der Transkription und Translation, das Plastom kodiert allerdings nur noch etwa 50 bis 200 Proteine.

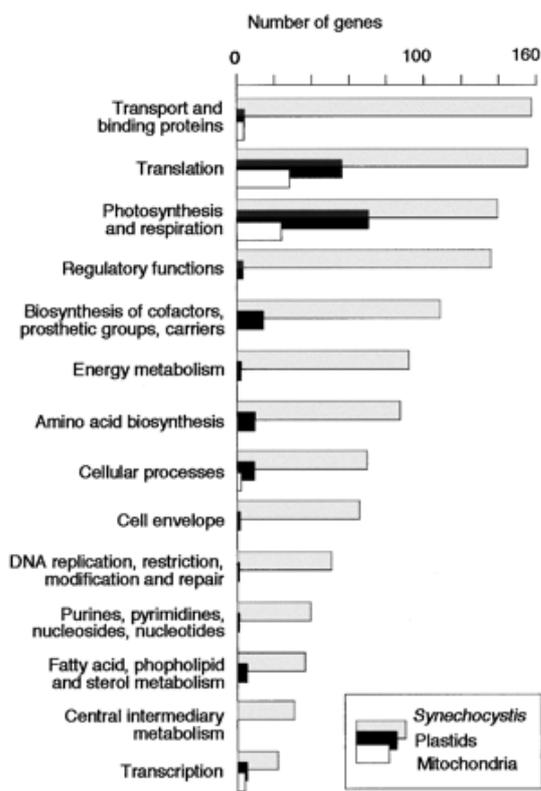
Der Verlust der Gene ging jedoch nicht zwangsläufig mit dem Verlust ihrer Funktionen einher, was sich in der prädiktierten Größe des Chloroplasten Proteoms von 2100-3600 unterschiedlicher Proteinen (Abdallah et al., 2000; The Arabidopsis Genome Initiative, 2000) ablesen lässt. Verantwortlich für die (funktionelle) Kompensation der Genomreduktion des Organells ist der Endosymbiontische Gentransfer vom Organell

zum Zellkern. Im Laufe der Evolution wurden viele der Gene des Genoms des Proto-Chloroplasten in den Zellkern transferiert und Gene mit redundanter Funktion zum Genom der Wirtszelle gingen verloren (Doolittle et al., 2003). Bedingung für eine vollständige Kompensation der Genomreduktion des Organells war der Reimport der vormalig im Chloroplasten kodierten Proteine aus dem Cytosol sowie der Import von neuen Proteinen, die für die Etablierung einer aufeinander abgestimmten Organelle-Zellkern-Interaktion benötigt wurden (Leister, 2003). Als Folge des endosymbiontischen Gentransfers wurde nicht nur genetische Information an den Zellkern abgegeben, sondern auch die transkriptionelle und translationelle Kontrolle der Expression dieser Gene und der Synthese der entsprechenden Proteine.

Die überwiegende Zahl der noch im Plastom kodierten Proteine ist entweder Teil der Transkriptions- und Translationsmaschinerie des Plastiden oder Untereinheiten des Photosynthese-Apparates in der Thylakoidmembran. Schon aus der Tatsache, dass die Multiproteinkomplexe der Photosysteme und der genetische Apparat des Plastiden in zwei verschiedenen Genomen kodiert sind, ergibt sich die Notwendigkeit der Kommunikation zwischen Organell und Zellkern [*plastid-to-nucleus signalling* (Barkan & Goldschmidt-Clermont, 2000)]. Neben den assimilatorischen Reaktionen der Photosynthese ist der Plastid an der Synthese der Aminosäuren, Fettsäuren und Lipide, der Nucleotide, Pflanzenhormone, Vitamine und sekundärer Metaboliten sowie der Fixierung anorganischen Schwefels und Stickstoffs und an dem Metabolismus vieler anderer essentieller Stoffwechselwege beteiligt (Leister, 2003). Daher mussten Signalmechanismen etabliert werden, die es dem Zellkern erlauben, den Proteinstatus des Organells zu kontrollieren und über die Expression des nukleär kodierten Transkriptoms zu regulieren (Gray et al., 2003). Die Aufschlüsselung der molekularen Natur der Signale der Kommunikation zwischen Organell und Zellkern und die Netzwerke der Regulation der Expression des nukleär kodierten Transkriptoms des Chloroplasten sind Gegenstand der aktuellen Forschung.

## 1.1 Das Plastom heutiger Chloroplasten

Durch den Transfer des Großteils der genetischen Information von der Organelle in den Zellkern, verloren die Chloroplasten ihre Autonomie, die Gesamtheit der zum Leben bzw. Überleben eines ehemals frei lebenden Cyanobakteriums benötigten Proteine durch die Vorgänge der Transkription und Translation selbst synthetisieren zu können. Das Plastom (cpDNA) kodiert nur 50 – 200 Proteine (Glockner et al., 2000), wobei die Anzahl und die Organisation der Gene je nach Art variieren kann (Martin et al., 1998). Die meisten der im Plastom verbliebenen Gene teilen sich in zwei Klassen auf: Gene, die Teile der plastidären Transkriptions- und Translationsmaschinerie kodieren (u. a. rRNAs, tRNAs, Ribosomen, RNA-Polymerase-Untereinheiten) und Gene, deren Genprodukte an den Vorgängen der Photosynthese beteiligt sind (Abb. 1.1) (Leister, 2003).



**Abb. 1.1.: Funktionelle Klassifizierung und Anzahl der proteinkodierenden Gene der Genome der Chloroplasten, Mitochondrien und Cyanobakterien.** In Bezug auf die Organellen wurde jedes Gen, das im Genom des Organells von mindestens einer von 10 verschiedenen Arten gefunden wurde, in die Berechnung mit einbezogen (Martin & Herrmann, 1998).

Die Regulation der Expression des plastidären Transkriptom unterliegt allerdings nicht mehr nur der Transkriptions- und Translationsmaschinerie der Plastiden und ist auch nicht die Miniatur eines prokaryotischen, sondern eher Teil eines eukaryotisch regulierten genetischen Systems (Hermann, 1997). Zwei DNA-abhängige RNA-Polymerasen unterschiedlichen phylogenetischen Ursprungs sind für die Transkription verantwortlich und nutzen bevorzugt unterschiedliche *cis*-Elemente im 5'-Bereich der Plastidengene. Die kernkodierte phagentypische NEP-Polymerase (*nuclear encoded polymerase*) bindet an NT-Promotoren (*nonconsensus type*) plastidärer Gene, während die PEP-Polymerase (*plastid encoded polymerase*) prokaryotischer Herkunft ist und konservierte, typisch bakterielle CT-Promotorsequenzen (*consensus type*) in der -35- (TTGACA) und der -10-Region (TATAAT) erkennt (Tonkyn & Gruissem, 1993). Einige Gene des Plastoms besitzen beide Typen von Promotoren [z.B.: *rps16* und *ycf3-psaAB* (Summer et al., 2000)], so dass sie von beiden Polymerasen transkribiert werden können. Darüber hinaus sind sechs regulierende Untereinheiten der plastomkodierten PEP-Polymerase, die  $\sigma$ -Faktoren, vom Kern kodiert (Kanamaru et al., 1999) und die Expression von drei dieser Faktoren konnte als lichtabhängig nachgewiesen werden (Isono et al., 1997).

Mittlerweile erscheint die Regulation der Expression der plastidären Gene sogar noch komplexer. Legen und Mitarbeiter (2002) zeigten, dass in PEP-defizienten Pflanzen zwar die meisten der plastomkodierten Proteine fehlten, die Transkription dieser Gene aber durch die NEP-Polymerase übernommen worden war und der Phänotyp somit auf posttranskriptionelle Prozesse zurückzuführen sein musste. Neben der NEP-Polymerase weist das Plastom im Vergleich zum prokaryotischen genetischen System weitere Neuerungen auf. Die RNA-Edierung, die ko- oder posttranskriptionelle Insertion (Rochaix, 2001) und die Deletion und/oder Substitution von Nukleotiden der mRNA (Yoshinaga et al., 1996) sind ebenso erst im Verlauf der Endosymbiose entstanden, wie die Einführung von Introns, die in Bakterien nur sehr selten zu finden sind (Ferat & Michel, 1993). Auch die Herkunft der plastidären Gene bzw. offenen Leseraster *ycf1*, *2*, *14* und *15* ist bisher nicht geklärt und sind eventuell erst in der Entwicklung vom frei lebenden Prokaryoten zum Zellorganell erworben worden (De Las Rivas et al., 2002). Berücksichtigt man, dass die DNA der heute lebenden Cyanobakterien die genetische Information für mehrere tausend Proteine trägt (Abb. 1.1) (Kaneko et al., 1996),

unterlagen die Chloroplasten über den Prozess der Endosymbiose einer drastischen Genomreduktion (Martin & Herrmann, 1998). Obwohl das Plastom des heutigen Zellorganells weniger als 10% der Proteine eines Cyanobakteriums kodiert (Abb. 1.1), gehen die Schätzungen verschiedener Arbeitsgruppen jedoch davon aus, dass das Proteom des Chloroplasten zwischen 1.000 und bis zu 5.000 Proteine umfasst (Abdallah et al., 2000; Rujan & Martin, 2000; Cavalier-Smith, 2000).

Die Diskrepanz zwischen der Anzahl der Gene des Plastoms und der Zahl der Proteine, die das Proteom des Organells bilden, wird mit den Vorgängen des „Endosymbiontischen Gentransfers“ (EGT) erklärt (Martin & Herrmann, 1998).

## 1.2 Endosymbiontischer Gentransfer

Endosymbiontischer Gentransfer ist eine spezielle und in der Natur weit verbreitete Form des lateralen Gentransfers, der zwischen den Kompartimenten eines eukaryotischen Systems unter der regulatorischen Kontrolle des Zellkerns stattfindet (Herrmann, 1997; Brown, 2001). Die Gene des ursprünglichen Organellgenoms gingen entweder verloren, wurden zum Zellkern transferiert oder sind auch heute noch Teil der cpDNA. Im Verlauf des Transfers existierte das Gen solange in doppelter Ausführung, bis das System aus Zellkern, cytosolischen Ribosomen und Organell eine Maschinerie zum gezielten Reimport des entsprechenden Genprodukts in das Organell entwickelt hatte (Heins & Soll, 1998). Die daraus resultierende genetische Redundanz führte über den Verlust der Genkopie im Organell zur Genomreduktion (Dyall et al., 2004).

Für einen erfolgreichen Transfer muss das ehemals plastomkodierte Gen im Zellkern zunächst durch einen funktionellen Promoter stabil exprimiert und das entsprechende Protein mit Hilfe einer N-terminalen Präsequenz (Transitpeptid) aus dem Cytosol in das Organell importiert werden können (Martin & Russel, 2003). Die dazu benötigte nukleär kodierte Protein-Import-Maschinerie der äußeren und inneren Chloroplastenmembran war eine Neuentwicklung, die erst nach der Endocytose des Cyanobakteriums etabliert wurde (Bölter et al., 1998; Heins & Soll, 1998; Reumann et al., 1999).

Im Falle der in den Plastiden und den Mitochondrien verbliebenen Gene ist wahrscheinlich die direkte Regulation der Gene durch die Aktivität ihrer Genprodukte in der Photosynthese und dem Elektronentransport nötig, um das empfindliche Redoxgleichgewicht im Organell zu gewährleisten und die Entstehung der hoch toxischen oxidativen Radikale zu verhindern (Allen, 2003). Da darüber hinaus einige der Genprodukte im Cytosol toxisch zu wirken scheinen [z.B. die Cytochrom-c-Oxidase der Mitochondrien, die aus dem Organell exportiert den Programmierten Zelltod (PCD) einleitet (Ludovico et al., 2002)], wäre demnach positive Selektion für die Transkription und Translation dieser Gene im Organell und für das Ausbleiben des Transfers zum Zellkern verantwortlich (Leister & Schneider, 2003).

### 1.3 Proteinimport aus dem Cytosol

Die Chloroplasten der Höheren Pflanzen besitzen drei Membransysteme. Sie sind von zwei Membranen umgeben, die auf die Plasmamembran und die äußere Membran des gram-negativen Cyanobakteriums zurückgehen und besitzen ein inneres Membransystem der Thylakoide, das das Thylakoidlumen umschließt. Viele der nukleär kodierte Proteine des Chloroplasten benötigen dazu eine Zielsequenz, um nach ihrer Synthese an den Ribosomen des Cytosols als unreife Proteinvorstufen über die Membranen in das Zellorganell importiert werden zu können. Innerhalb der Chloroplasten können die Proteine zu insgesamt sechs verschiedenen Kompartimenten transloziert werden: der äußeren (OE = *outer envelope*) und inneren Membran (IE = *inner envelope*), dem Intermembranraum und dem Stroma sowie zur Thylakoidmembran (TM) oder dem Thylakoidlumen (Dyall et al., 2004). Fast alle diese Proteine sind als Vorstufen des späteren funktionellen Proteins mit einer N-terminalen, abspaltbaren Extensionssequenz versehen, einem so genannten Chloroplasten-Transitpeptid (cTP), das als plastidspezifisches Transportsignal fungiert (Blobel et al., 1979). Die Ausnahme sind die meisten Proteine der äußeren Chloroplastenhüllmembran (OEP = *outer envelope proteins*), also auch die Proteine, die als Bestandteil eines Multiproteinkomplexes den Protein-Import-Apparat in der äußeren Chloroplasten-

membran bilden und direkt von der cytosolischen Seite in Membran inseriert werden [z.B. Toc34 (Chen & Schnell, 1997)].

Der Protein-Import in den Chloroplasten benötigt Energie aus der Hydrolyse von ATP und GTP und wird durch das Zusammenspiel des TOC-Komplex (TOC = *translocon on the outer envelope of chloroplasts*) der äußeren und dem TIC-Komplex (TIC = *translocon on the inner envelope of chloroplasts*) der inneren Hüllmembran erreicht (Jarvis & Soll, 2001). Vor dem eigentlichen Import bindet das Präprotein an einen, so genannten „Guidance-Komplex“, unter anderem bestehend aus 14-3-3-Proteinen und dem Chaperon Hsp70, das die vorzeitige Faltung des Proteins im Cytosol verhindert (Jarvis & Soll, 2001). Sobald der Guidance-Komplex mit dem gebundenen Vorstufenprotein an der äußeren Hüllmembran angelangt ist, startet der Prozess der hoch spezifischen Mustererkennung. Sie muss gewährleisten, dass trotz der kaum vorhandenen konservierten Motive der cTPs, nur Vorstufenproteine des Chloroplastenproteoms in das Stroma importiert werden (Bruce, 2001). Über die integralen Transportkanäle des TOC- und TIC-Komplexes vermitteln Chaperone (Hsp70 und Hsp100) die Passage des unreifen, ungefalteten Präproteins durch die Chloroplastenmembranen.

Für den Transport in, oder über, die TM sind vier verschiedene Translokationsmechanismen bekannt (Robinson et al., 2001). Die Komponenten und die Funktionsweise des Protein-Import-Apparates der TM sind eindeutig prokaryotischen Ursprungs (ausgenommen der spontanen Insertion, die bei Eukaryoten zu finden ist) und homologe Systeme sind deshalb bisher hauptsächlich für Bakterien beschrieben worden. Die Proteine der TM, also auch die Untereinheiten der vier Multiproteinkomplexe der beiden Photosysteme, des Cytochrom *c6/f*-Komplexes und der ATPase, werden entweder über die SRP-abhängige (*stromal recognition particle*) Translokation (z.B. Lhcb1) oder durch spontane Insertion (z.B. PsaX, PsaW, CF<sub>0</sub>-II, ELIP) in die Membran integriert. Die SRP-abhängige Translokation benötigt Energie aus der Hydrolyse von GTP und eine Zielsequenz, das ITP, die inmitten des reifen Proteins lokalisiert ist. Auch die Insertion der plastomkodierten Untereinheiten der Multiproteinkomplexe aus dem Stroma in die TM findet auf diesem Weg statt.

Die Proteine, die aus dem Stroma über die TM in das Thylakoidlumen transportiert werden müssen, besitzen nach der Abspaltung des cTPs eine zweite N-terminale

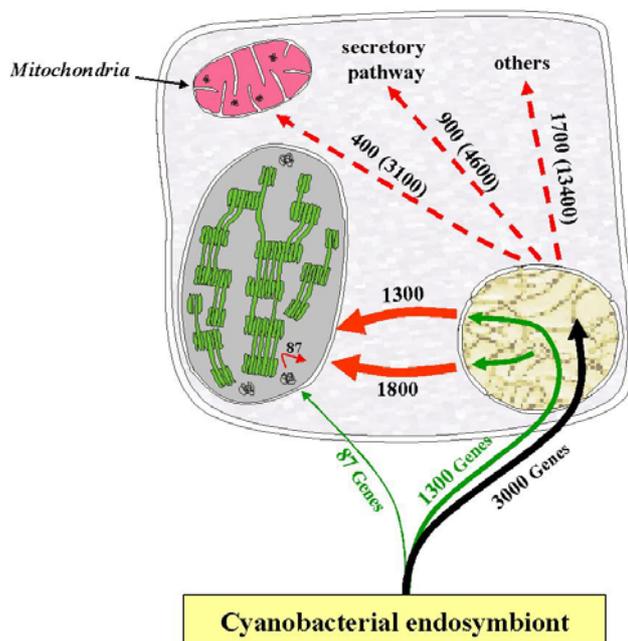
Signalsequenz (Vothknecht & Soll, 2002). Sie werden entweder über den Sec-abhängigen Transport (*Sec-dependent pathway*) oder das Tat- (*twin arginine translocation*) Translokon transloziert. In Ersterem binden die Proteine im Stroma mit ihrem ITP an eine Translokations-ATPase, ein so genanntes SecA-Protein (Mori et al., 1999). SecA scheint unter Hydrolyse von ATP in regelmäßigen Zyklen in die TM zu inserieren und sich wieder herauszulösen und so das gebundene, ungefaltete Vorstufenprotein über den eigentlichen Importkanal (SecY/E/G) durch die TM zu bewegen (Economou, 1998). Die endgültige Faltung des reifen Proteins findet erst nach der proteolytischen Abspaltung des Lumen-Transitpeptids im Thylakoidlumen statt.

## 1.4 Das Proteom des Chloroplasten

Das Proteom der Plastiden wird nicht nur in zwei verschiedenen Kompartimenten synthetisiert, sondern ist auch dualen evolutionären Ursprungs (Leister & Schneider, 2003). Das Proteom heutiger Chloroplasten ist, wie auch das Plastom, das Ergebnis aus der Endosymbiose zweier ursprünglich autonomer Zellen. In *A. thaliana* sind nur noch etwa 90 Proteine im Plastom kodiert (Leister & Schneider, 2003), im gesamten Kerngenom wurden jedoch zwischen 2100 und 3600 Gene aufgrund des Vorhandenseins eines cTP in den entsprechenden Genprodukten als Chloroplastengene vorhergesagt (Abdallah et al., 2000; The Arabidopsis Genome Initiative, 2000; Richly & Leister, 2004). Von diesen weist allerdings ein großer Teil keinerlei Homologie zu Genen der heute lebenden Cyanobakterien auf.

Andererseits kodieren von den etwa 4500 Genen im Zellkern, die aufgrund ihrer Sequenzhomologie wahrscheinlich cyanobakterieller Herkunft sind (Martin et al., 2002), nur etwa 1300 für Proteine, die ein cTP besitzen. Für den Großteil dieser Proteine (~1700) wurde zwar kein cTP gefunden, erstaunlicherweise aber wurde für etwa 10% ein Mitochondrien-Transitpeptid (mTP), beziehungsweise für ca. 20% ein Transitpeptid für das Sekretorische System, vorhergesagt (Abb. 1.2) (Martin et al., 2002). In diesen Ergebnissen spiegeln sich drei entscheidende Vorgänge wider, die im Laufe der Evolution der Pflanzenzelle zur Optimierung des genetischen und metabolischen Systems aus Zellkern und den beiden Zellorganellen beigetragen haben:

- a) eine massive Umverteilung von Proteinen des ehemaligen Cyanobakteriums in andere Kompartimente der Zelle,
- b) die Etablierung neuer Chloroplastenfunktionen, wie die des lichtsammelnden Pigment-Protein-Komplexes (LHC = *light-harvesting-complex*) der Photosysteme (Montane & Kloppstech, 2000) und die des Protein-Import-Apparates für die Proteintranslokation über oder in die innere Plastidmembran (Jarvis & Soll, 2001), sowie
- c) den Reimport eines großen Teils der Proteine in die Organelle, deren Gene im Zuge des EGT zum Zellkern transferiert wurden und nun unter der (komplexeren) regulatorischen Kontrolle eines eukaryotischen, sexuell genetischen Systems stehen.



**Abb. 1.2.: Überblick der Gene cyanobakterieller Herkunft und der subzellulären Lokalisationen der etwa 4400 korrespondierenden Proteine.** Mehr als 4000 der etwa 25700 Gene von *A. thaliana* sind cyanobakteriellen Herkunft. Interessanterweise werden die meisten der nun nukleär kodierten Proteine in andere Zellkompartimente transloziert, und mehr als die Hälfte des heutigen Proteoms des Chloroplasten stammt nicht vom cyanobakteriellen Endosymbionten (Leister & Schneider, 2003).

Die oben genannten Zahlen basieren auf struktureller Genomanalyse, in diesem Fall der Prädiktion eines cTPs. Durch aufeinander folgende spezielle Extraktions-, Fraktionierungs- und Aufreinigungsmethoden sowie anschließender Proteinanalyse (*Proteomics*), lässt sich die subzelluläre Lokalisation vieler Proteine auch experimentell

nachweisen. Durch Massenspektrometrie konnten von Kleffmann und Mitarbeitern (2004) in einer aufgereinigten Chloroplastenextraktion 690 unterschiedliche Proteine identifiziert werden. Viele der identifizierten Proteine besitzen Homologe im Proteom der Cyanobakterien, nicht aber in Hefe, was ihre Herkunft vom photosynthetischen Endosymbionten unterstreicht. Fast alle Proteine der Schlüsselfunktion im Chloroplasten, der Licht- und Dunkelreaktion der Photosynthese, wurden detektiert. Jedoch sind mehr als 30 % der gefundenen Proteine von bisher unbekannter Funktion. Weitergehende Untersuchungen hatten die Lokalisation der Proteine auf subplastidärer Ebene zum Ziel. Diese Analysen waren auf die Chloroplastenmembran (Ferro et al., 2002, 2003; Froehlich et al., 2003), das Thylakoidlumen (Peltier et al., 2002) oder die peripheren und integralen Proteine der TM (Friso et al., 2004) fokussiert. Durch die Kombination von experimentellem Nachweis und struktureller Genomanalyse über *TargetP* (Emanuelsson et al., 2000) und *LumenP* (Westerlund et al., 2003), identifizierte die Arbeitsgruppe van Wijk's insgesamt 4255 kernkodierte Chloroplastenproteine, die sich in 3387 stromale Proteine, 291 Proteine des Thylakoidlumens und 520 Proteine der TM aufteilen [*Plastid Proteome Data Base* (<http://www.ppdb.tc.cornell.edu/>)].

Mit einem anderen Ansatz testeten Richly und Leister (2004) die Vorhersagegenauigkeit verschiedener cTP-Prädiktoren an einem Set von 4500 Proteinen, deren subzelluläre Lokalisation bereits bekannt war. Sie konnten zeigen, dass eine Kombination von vier verschiedenen Computerprogrammen jedem einzelnen der cTP-Prädiktoren überlegen war und extrapolierten so, übertragen auf das gesamte Genom von *A. thaliana*, eine Zahl von 2100 kernkodierten Proteinen mit plastidärer Signalsequenz.

## 1.5 Chloroplastenfunktionen

Da das Plastom nur etwa 5% des plastidären Proteoms kodiert und 95 % der Chloroplastengene im Zellkern lokalisiert sind, kann die Morphologie und Funktion der Plastiden selbst sehr unterschiedlich sein, auch wenn jede Zelle einer Pflanze identische Plastomkopien besitzt (Leister, 2003). Bei höheren Pflanzen entwickeln sich

aus den Proplastiden junger Zellen im Zuge der Zelldifferenzierung, je nach Lichtverhältnissen und nach Gewebe, unterschiedliche Plastiden mit unterschiedlichen Funktionen. Außer den Chloroplasten des photosynthetisch aktiven Gewebes der Pflanze treten Varianten auf. Die Amyloplasten, Proteoplasten und Elaioplasten dienen als Speicherorte für Stärke, Proteine oder Lipide, während die gelben oder roten Chromoplasten Xanthophylle oder Carotinoide enthalten, die für viele Blütenfarben sowie die Farben einiger Früchte oder Wurzeln verantwortlich sind.

Die Chloroplasten sind neben der Photosynthese in viele wichtige Synthesewege der Zelle involviert. Sie sind der Ort der Assimilation von Stickstoff und Schwefel und am Metabolismus vieler essentieller Verbindungen wie Aminosäuren, Fettsäuren und Lipide, Nukleotide, Pflanzenhormone, Vitamine und verschiedener sekundärer Metabolite beteiligt. Die Photosynthese stellt jedoch die zentrale Reaktion in den Chloroplasten dar und nimmt insofern eine Sonderstellung ein, als dass das Plastom noch immer etwa die Hälfte der mindestens 70 Proteine des Photosynthese-Apparates kodiert (Abb. 1.3) (Rochaix et al., 2001; Friso et al., 2004).

### 1.5.1 Photosynthese

Bei der Photosynthese der höheren Pflanzen wird die Strahlungsenergie des Sonnenlichts in ein elektrochemisches Potential transformiert, das zur Synthese von ATP und NADPH genutzt wird, die der endergonischen Fixierung anorganischen CO<sub>2</sub> in energiereiche organische Verbindungen dienen. Die Photosynthese kann in zwei Teilreaktionen untergliedert werden:

a) bei der Lichtreaktion der Photosynthese wird die Energie aus der Absorption von Photonen durch die LHCs auf einen primären Elektronendonator der Photosysteme übertragen, ein proteingebundenes Chlorophyll *a*-Dimer, dessen Anregung zu einer gerichteten Ladungstrennung führt, wobei ein Elektron auf die stromale Seite des Komplexes transportiert und ein primärer Elektronenakzeptor reduziert wird. Unter physiologischen Bedingungen sind Photosystem II (PSII) und Photosystem I (PSI) elektrochemisch in Serie geschaltet. PSII fungiert als eine Wasser-Plastochinon-

Oxidoreduktase (Govindjee et al., 1985) und überträgt die Elektronen aus der Spaltung des Wassers über eine Elektronentransportkette aus Redoxsystemen zum Cytochrom *b6/f*-Komplex, der wiederum das Elektronendefizit des oxidierten Photosystems I ausgleicht. Nach dem Modell des modifizierten Q-Zyklus (Mitchell, 1976, Trumpower, 1990) reduziert schließlich eine Ferredoxin-NADP<sup>+</sup>-Reduktase (FNR) NADP<sup>+</sup> zu NADPH. Ergebnis des nichtzyklischen Elektronentransports ist die Spaltung zweier Wassermoleküle, die Synthese zweier NADPH-Moleküle und ein transmembraner pH-Gradient, der von der ATP-Synthetase genutzt wird, um ADP und anorganisches Phosphat (P<sub>i</sub>) zu ATP umzusetzen [Photophosphorylierung (Mitchell, 1966)].

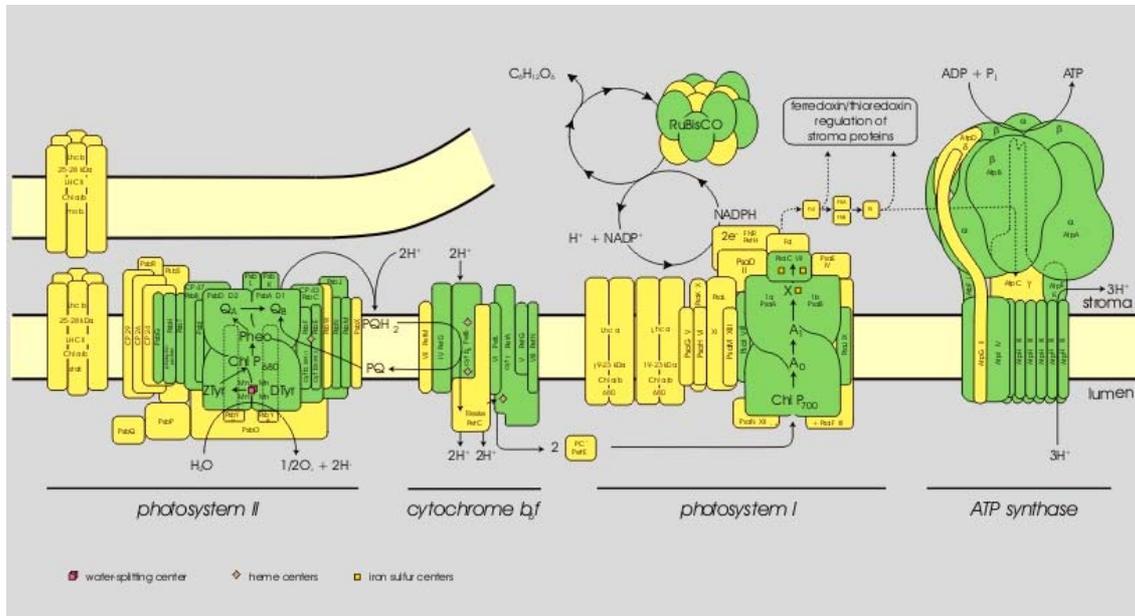
b) in der Dunkelreaktion der Photosynthese werden die in der Lichtreaktion gebildeten Reduktionsäquivalente des NADPH genutzt, um in einer endergonischen Reaktion des Calvin-Zyklus atmosphärisches CO<sub>2</sub> mit Hilfe von ATP zu energiereichen organischen Verbindungen zu reduzieren.

Der Redoxstatus der proteingebundenen Redoxsysteme (Escoubas et al., 1995), das Verhältnis von ADP zu ATP als auch der transmembrane pH-Gradient über der TM, also die direkten und indirekten Konsequenzen aus der Lichtreaktion der Photosynthese, stellen gleichzeitig auch wichtige Signale für die koordinierte Regulation der Expression der nukleär kodierten Gene des Plastiden dar (Pfannschmidt et al., 2001).

## 1.6 Funktion und Herkunft des Photosynthese-Apparates

Bei der Entwicklung der Chloroplasten bildet sich im Licht und unter dem Einfluß der Phytochrome aus der inneren Membran des Proplastiden eine komplexe Thylakoidstruktur mit Grana- und Stromapartien, deren Membranen das Lumen umschließen. Die Thylakoidmembranen der ausdifferenzierten, voll funktionsfähigen Chloroplasten sind in gestapelte (Granathylakoide) und ungestapelte (Stromathylakoide) Regionen organisiert. Die vier integralen Multi-Protein-Komplexe des Photosynthese-Apparates sind ungleichmäßig auf diese Regionen verteilt. PSII ist hauptsächlich in den Granathylakoiden zu finden, wohingegen PSI und die ATPase im Bereich der

Stromathylakoide akkumulieren, so dass NADPH und ATP nach ihrer Synthese direkt dem Calvin-Zyklus des Stroma zur Verfügung stehen. Der Cytochrom  $b_6/f$ -Komplex, der die Rolle des Mittlers im Elektronentransport zwischen PSII und PSI darstellt, ist gleichmäßig über Grana- und Stromathylakoide verteilt.



**Abb. 1.3.: Modell der Thylakoidmembran der Höheren Pflanzen.** Dargestellt sind die Zusammensetzung und die genetische Herkunft der integralen und peripheren Polypeptide der Thylakoidmembran. Plastomkodierte Proteine sind grün, die kernkodierte Proteine gelb markiert (Herrmann, 1996).

PSI, PSII, der Cytochrom  $b_6/f$ -Komplex und die plastidäre ATP-Synthase sind prokaryotischen Ursprungs (Abb. 1.3) und in ihrer Funktion in Cyanobakterien und Pflanzen weitgehend konserviert (He & Malkin, 1992). Dagegen sind die LHCs der Photosysteme erst innerhalb des eukaryotischen Systems entstanden und haben die Phycobilisomen der Cyanobakterien ersetzt. Jeder der vier großen integralen Multiprotein-Komplexe der TM setzt sich sowohl aus nukleär als auch plastomkodierten Untereinheiten zusammen.

Das PSII der höheren Pflanzen besteht aus mindestens 29 Untereinheiten, von denen 15 im Plastom kodiert sind. Das Reaktionszentrum wird von zwei integralen, ca. 34 kD großen Membranproteinen D1 (PsbA) und D2 (PsbD) gebildet, mit denen das Chlorophyll  $a$ -Dimer und die Elektronenakzeptoren Phäophytin sowie die Plastochinon-

moleküle  $Q_A$  und  $Q_B$  assoziiert sind. Zusammen mit Cytochrom  $b_{559}$  (PsbE und PsbF), CP43 (PsbC) und CP47 (PsbB) sowie einiger kleinerer integraler Proteine bilden sie den Kernkomplex des Photosystems II (Hankamer et al., 2001). PsbO, PsbP und PsbQ sind die kernkodierte Untereinheiten des Sauerstoff produzierenden Komplexes (Tsiotis et al., 1999). Weitere peripher vom Kernkomplex gelegene integrale Proteine stellen die Verbindung zu einem Multi-Protein-Pigment-Komplex, dem LHCII her. Drei kleinere (CP29, CP26 und CP24) und drei größere (Lhcb1-3) Proteine binden Chlorophyll *a*, Chlorophyll *b* sowie Xantophylle und bilden den lichtsammelnden Antennenkomplex des PSII.

Wie im PSII ist auch im Photosystem I der primäre Elektronendonator, P700\*, und die primären Elektronenakzeptoren,  $A_1$ ,  $A_2$ ,  $FeS_A$  und  $FeS_B$ , an zwei integrale plastomkodierte Proteine, PsaA und PsaB, gebunden. Zum Kernkomplex des PSI gehören außerdem die von PsaC bis PsaO benannten Untereinheiten. PsaC, PsaD und PsaE sind an der Stromaseite der TM lokalisiert und direkt oder indirekt am Transfer der Elektronen zum Ferredoxin beteiligt (Naver et al., 1996).

## 1.7 Regulation der nukleären Genexpression

Auf die Regulation der Expression plastomkodierter Proteine wurde zum Teil schon vorher eingegangen und gezeigt, dass sie nicht separiert, sondern integriert in ein genetisches System unter Kontrolle des Zellkerns ist. Der Chloroplast muss teilweise sehr schnell auf Veränderungen abiotischer Faktoren reagieren können. In Bezug auf die Photosynthese bedeutet das vor allem Reaktionen auf Veränderungen bei der Lichtqualität und Lichtquantität. Das beinhaltet Mechanismen für die kurzfristige Anpassung des Photosynthese-Apparates, einschließlich strukturellen Umbaus wie der so genannten „state transitions“ (Allen, 2001), aber auch längerfristige Adaptation über Veränderungen der Stöchiometrie der Photosysteme, etwa im Verhältnis von PSI zu PSII in der TM (Aro & Andersson, 1993).

Längerfristige Adaptation findet über die Expression von Proteinen statt und benötigt ein Signal für die Regulation der Expressionsrate. Die Transkription und Translation plastomkodierter Proteine kann über den Redoxstatus spezifischer Proteine kontrolliert

werden. Pfannschmidt und Mitarbeiter (1999) konnten zeigen, dass die Transkriptionsrate der Apoproteine von PSI und PSII über den Redoxstatus von Plastochinon, selbst Teil der Elektronentransportkette des Photosystems II, reguliert wird. Thioredoxine, deren Redoxstatus von der lichtabhängigen Reduktion von Ferredoxin, also der Aktivität des PSI abhängig ist, scheinen bei Belichtung die Rate der Proteinsynthese in den Chloroplasten 50- bis 100-fach zu erhöhen, auch wenn die Menge an Transkripten gleich bleibt (Danon & Mayfield, 1994). Die Thioredoxine vermitteln die Bindung kernkodierter Aktivatorproteine an das 5'-Ende der mRNA und damit indirekt deren Translation an den Ribosomen (Surpin et al., 2002).

Drei verschiedene Genome innerhalb einer Zelle, im Zellkern, in den Plastiden und in den Mitochondrien, erfordern die Koordination deren Expression (Gray et al., 2003). Besonders deutlich wird das im Fall der Photosynthese, die durch Vorgänge in verschiedenen Kompartimenten der Zelle beeinflusst wird und des Photosynthese-Apparates, der ein Mosaik aus kernkodierten und plastomkodierten Komponenten darstellt (Jarvis, 2001). Die Koordination der Expression dieser Genome steht unter der Kontrolle des Zellkerns und findet überwiegend auf posttranskriptionaler Ebene statt (Petracek et al., 1997, 1998; Sherameti et al., 2002), beinhaltet aber auch Signale vom Plastiden an den Zellkern, die die Expression nukleärer Chloroplastengene (Pfannschmidt et al., 2001).

Die Expression der nukleären Chloroplastengene (Kerngene, die Proteine des Chloroplasten kodieren) ist auf verschiedene Weisen reguliert. Die Photorezeptoren perzipieren Lichtqualität und -quantität und sind in der Zell- oder Gewebeentwicklung an wichtigen Photomorphosen wie der Samenkeimung, der Blühinduktion oder der Bildung von Anthocyanen und Chlorophyll beteiligt (Deng & Quail, 1999). In höheren Pflanzen sind drei verschiedene Rezeptorsysteme bekannt: Phytochrome, Cryptochrome und UV-(AB) Rezeptoren. Sie interagieren jedoch nicht direkt mit der DNA, sondern regulieren die Genexpression über eine Signaltransduktionkaskade, die Proteinkinasen, G-Proteine und das  $Ca^{2+}$ /Calmodulin-System beinhaltet (Mulligan et al., 1997). Im Gegensatz dazu stellt die Kommunikation vom Organell zum Zellkern (*plastid-to-nucleus signalling*) eine Adaption an Veränderungen der äußeren Bedingungen dar und findet zum Teil, aber nicht immer, separiert und parallel zu der cytosolischen Signaltransduktion der Lichtperzeption statt.



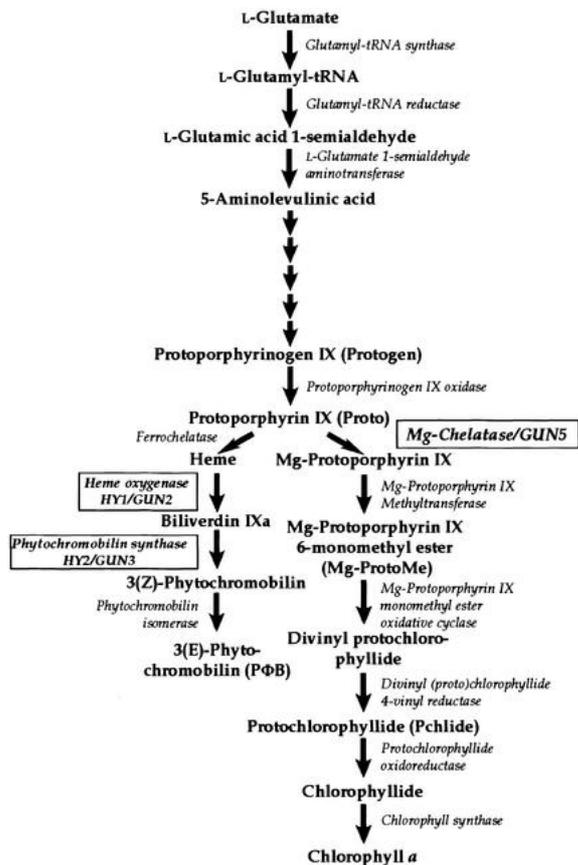
Die Photooxidation des Chlorophylls führt zur Produktion von ROS (*reactive oxygen species*), die wiederum die Expression kernkodierter Gene beeinflussen (Mullineaux & Karpinski, 2002). Carotinoide löschen die Anregungsenergie von Chlorophyllen im Triplettzustand und verhindern somit die Produktion freier Radikale und ROS. Wird die Biosynthese der Carotinoide durch das Herbizid Norflurazon gehemmt, kommt es bei intensiver Belichtung zur Bleichung der Chloroplasten. Obwohl der strukturelle Schaden durch Norflurazon auf den Chloroplasten beschränkt ist (Reiss et al., 1983), ist die Expression vieler (z.B.: *Lhcb1*, *PetG*, *E*, *H*, *Rbcs*), aber nicht aller Gene (z.B.: *Atp2*) des Kerns, die Photosynthese-relevante Proteine kodieren, deutlich reduziert. Pfannschmidt und Mitarbeiter (2001) postulierten deshalb, dass ein Redoxsignal über Thioredoxine der allgemeinen Adaption an moderate Lichtverhältnisse dient, während die stressbedingte Antwort auf extreme Lichtintensitäten von Glutathion und ROS vermittelt wird.

Weder die molekulare Natur der Komponenten dieser drei möglichen Redox-Signalketten ist bekannt, noch ob diese Signale an einem einzelnen Element der Promotoren ansprechbarer Gene zusammenlaufen oder sie die Genexpression auf verschiedenen Ebenen regulieren (Gray et al., 2003). Photosynthese-abhängige Signale scheinen auch über die Regulation der mRNA-Stabilität zu wirken (Sullivan & Gray, 2002).

Die *cue1*-Mutante [*CAB (chlorophyll a/b-binding underexpressed)*] ist nicht in der Lage, die Expression einer Reihe lichtregulierter Gene des Zellkerns nach Stimulierung durch die Phytochrome zu steigern. *CUE1* kodiert einen Phosphoenol/Phosphat Transporter der inneren Chloroplastenmembran, dessen Inaktivierung in *cue1* zu pleiotropen Effekte führt. Unter anderem ist der Shikimat-Weg beeinträchtigt und dadurch die Synthese der Flavonoide, einfacher Phenole und auch von Plastochinon (Streatfield et al., 1999). Die relative Größe des Plastochinon-Pools der *cue1*-Mutante ist 10-50 % geringer als im Wildtyp. So könnte *CUE1* letztendlich über Regulation des Redoxstatus des Plastochinon-Pools (siehe oben) auf die Genexpression Einfluss nehmen (Surpin et al., 2002). Allerdings sind durch die vielfältigen Effekte der Mutation auch andere Ursachen für die Blockade der lichtinduzierten Expression denkbar.

Die Analyse der *gun*-Mutanten [*genomes uncoupled* (Susek et al., 1993)] führte zur Identifizierung eines weiteren Signalweges der Kommunikation vom Plastid zum Zellkern. Pflanzen, die eines der bisher fünf identifizierten *gun*-Allele (*gun1-5*) besitzen, zeigen im Gegensatz zum Wildtyp keine Reduktion in der Expressionsrate von Photosynthese-relevanten Genen des Kerns, wie *RbcS* (*Rubisco small subunit*) und *Lhcb*, wenn die Entwicklung des Chloroplasten durch Photooxidation verhindert ist. In den *gun*-Mutanten ist allerdings weder die gewebs- und zellspezifische, noch die lichtabhängige oder zirkadiane Regulation der *Lhcb*-Gene beeinflusst. Vielmehr scheint spezifisch die plastidäre Regulation der Transkription der Gene im Kern beeinträchtigt zu sein (Susek et al., 1993). Schon 1984 konnten Johanningmeier und Howell zeigen, dass Mg-Protoporphyrin-IX-Monoethylester (MgProto-Me), ein Zwischenprodukt der Tetrapyrrolbiosynthese, einen wirksamen Hemmstoff für die Akkumulation von *Lhcb*-mRNA darstellt. Die vier Loci für *GUN2* bis *GUN5* konnten kloniert werden und es zeigte sich, dass die vier Gene Enzyme (*GUN2*, *GUN3* und *GUN5*) bzw. Enzymaktivatoren (*GUN4*) der Tetrapyrrolbiosynthese kodieren (Mochizuki et al., 2001; Vinti et al., 2000) (siehe Abb.1.5).

*GUN2* kodiert eine Häm-Oxygenase (Davis et al., 1999) und *GUN3* die Phytochromobilin-Synthase (Kohchi et al., 2001) für die Synthese von 3(E)-Phytochromobilin (PΦB), dem Chromophor des Phytochrome A, ausgehend von Protoporphyrin IX (Proto). Auch wenn sich die Mutationen in *gun2*- und *gun3*-Mutanten direkt nur auf die Synthese von PΦB auswirken sollten, ist jedoch auch die Biosynthese von Chlorophyll a gestört (Montgomery et al., 1999; Terry & Kendrick, 1999). *GUN5* kodiert die ChlH-Untereinheit der Mg-Chelatase. Da die ChlH-Untereinheit der Mg-Chelatase auch in Abwesenheit der Untereinheiten ChlD und ChlI Proto bindet (Karger et al., 2001), könnte die ChlH-Untereinheit selbst Teil eines Signalweges sein, die über die Bindung an Porphyrine als Konzentrationssensor fungiert und direkt oder indirekt an der Genexpression im Zellkern beteiligt ist (Mochizuki et al., 2001). *GUN4* ist als Aktivator der Mg-Chelatase nur indirekt an der Regulation der Chelatisierung von Proto zu Mg-Proto beteiligt (Larkin et al., 2003), während *GUN1* im Gegensatz zu *GUN2*, 3, 4 und 5 nicht am Tetrapyrrol-Metabolismus, sondern an einem zweiten, eventuell teilweise redundanten Signalweg beteiligt ist (Mochizuki et al., 2001; Vinti et al., 2000).



**Abb. 1.5.: Tetrapyrrol-Biosynthese des Plastiden.** Protoporphyrinogen IX entsteht, ausgehend vom Glutamat aus 5-Aminolävulinsäure, im Stroma und wird nach dem Transport in die Hüllmembran des Chloroplasten und/oder die Thylakoidmembran über Protoporphyrin IX zu Chlorophyll a oder 3(E)-Phytochromobilin umgesetzt. Die von HY1/GUN2, HY2/GUN3 und GUN5 katalysierten Reaktionen sind hervorgehoben (Surpin et al., 2002).

Zusammenfassend deutet vieles darauf hin, dass Mg-ProtoIX als ein Zwischenprodukt der Tetrapyrrolbiosynthese, ein Signalmolekül eines Kommunikationsweges von den Plastiden zum Zellkern darstellt und dessen Akkumulation zur Reduktion der Expressionsrate einer Reihe von Photosynthese-relevanten Genen führt (Abb. 1.6). Dazu passt, dass der Wildtyp unter photooxidativen Bedingungen mehrfach erhöhte Konzentrationen an Mg-ProtoIX aufweist, während die Konzentration in *gun2*- und *gun5*-Mutanten deutlich herabgesetzt ist (Strand et al., 2003).

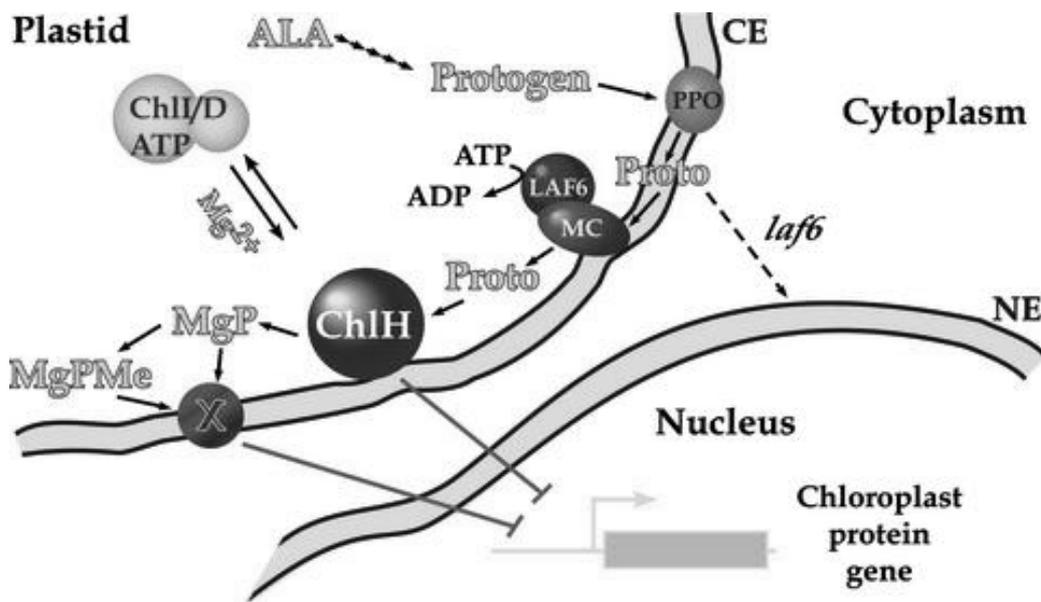
Die *laf6*-Mutante (*long after far-red6*) ist ebenfalls im Tetrapyrrolmetabolismus beeinträchtigt, wobei der ATP-abhängige Reimport von Proto in das Stroma reduziert zu sein scheint. Die daraus folgende Akkumulation von Proto im Cytosol wirkt sich

direkt oder indirekt auf einen phyA-abhängigen Signalweg aus, mit dem Ergebnis, dass bei *laf6*-Mutanten, anders als in den *gun*-Mutanten, die Expression von *Lhcb* unter photooxidativen Bedingungen reduziert ist (Moeller et al., 2001). Dies ist insofern erstaunlich, da die Konzentration von Mg-ProtoIX in den Chloroplasten wie bei den *gun*-Mutanten reduziert (die Pflanze akkumuliert 40% weniger Chlorophyll) und die Expression des *Lhcb*-Gens folglich nicht mehr unterdrückt sein sollte. Da auch die Syntheserate von PΦB nicht durch die Mutation betroffen ist, stellt anscheinend die Akkumulation von Proto im Cytoplasma das eigentliche Signal dar und hat zu Folge, dass die Induktion der Genexpression durch Licht abgeschwächt wird (Moeller et al., 2001). Das konnte dadurch bestätigt werden, indem die Behandlung von Wildtyp-Pflanzen mit Flumioxazin, einem Inhibitor der PPO (Protoporphyrinogen-IX-Oxidase), zu ähnlich hohen Konzentrationen von Proto im Cytosol und im Ergebnis zum *laf6*-Phänotyp, also reduzierter Hemmung des Hypokotylwachstums und reduzierter Expression der lichtregulierten Gene *Lhcb*, *Chs* und *FNR* führte (zum Überblick siehe: Brown et al., 2001).

Die Regulation der Tetrapyrrolsynthese steht unter der *Feedback*-Kontrolle der GLU-TR (Glutamyl-tRNA-Reduktase) durch Häm (Weinstein et al., 1993). Meskauskiene und Mitarbeiter (2002) identifizierten einen weiteren negativen Regulator der Chlorophyll-Biosynthese, das *FLU*-Protein, das unabhängig von der Aktivität von GLU-TR Einfluss auf die Syntheserate von ALA (5-Aminolävulinsäure) nimmt. *flu*-Mutanten sind nicht mehr in der Lage, die ALA-Synthese und damit die Akkumulation von Protochlorophylliden in der Dunkelheit herabzusetzen, obwohl die Häm-Konzentrationen denen des Wildtyps entsprechen (Meskauskiene et al., 2001). Sie konnten zeigen, dass das *FLU*-Protein mit der GLU-TR direkt interagiert und über die Hemmung der Aktivität des Enzyms die Chlorophyll-Biosynthese reguliert. Sollte das *FLU*-Protein Teil einer weiteren *Feedback*-Kontrolle sein, wird deutlich, wie komplex die Tetrapyrrol-Synthese und damit auch die Feinabstimmung der Kommunikation vom Plastiden zum Zellkern über die Tetrapyrrole reguliert ist (siehe Abb. 1.6).

Die Proteine, die die Elemente der oben genannten und vom Plastiden zum Zellkern gerichteten Signalwege darstellen, sind bisher kaum bekannt. Auch ist es bis jetzt nicht gelungen, Plastid- von Licht-induzierten regulatorischen *cis*-Elementen zu entkoppeln (McCormac et al., 2001), was vermuten lässt, dass beide, Licht- und Plastid-spezifische

Signale, ihren regulatorischen Einfluss über dieselben *cis*-Elementen ausüben (Kusnetsov et al., 1996). Da sich diese Signale zum großen Teil auf die Transkription der Gene im Kern auswirken, kann die differentielle Genexpressionsanalyse des gesamten nukleären Transkriptoms des Plastiden in Verbindung mit struktureller und funktioneller Genomik zur Entschlüsselung der Funktionsweise der unterschiedlichen Kommunikationswege und der Auswirkungen von Veränderungen des physiologischen Status des Organells beitragen.



**Abb. 1.6.: Modell des *Plastid-to-Nucleus Signalling* durch die Intermediate und Komponenten des Tetrapyrrolmetabolismus.** Protoporphyrinogen IX wird nach seiner Synthese im Stroma in den Membranen des Chloroplasten durch die PPO zu Protoporphyrin IX (Proto) oxidiert. Das *FLU*-Protein nimmt als Inhibitor der *GLU-TR* Einfluss auf die Syntheserate von ALA im Stroma (siehe Abb. 1.5). *LAF6*, zusammen mit einem integralem Membranprotein (*MC*), ist am ATP-abhängigen Reimport von Proto in das Stroma beteiligt. *ChlH* wird zur Chelatisierung des Proto benötigt, seine Rolle in den Vorgängen des *Plastid-to-Nucleus Signalling* ist im Text diskutiert. Die Synthese der meisten nukleär kodierten Chloroplastenproteine wird durch die Mg-Porphyrine, die Intermediate des Tetrapyrrolmetabolismus, direkt oder indirekt reprimiert. Faktor X ist möglicherweise für den Transport von Mg-Proto und Mg-ProtoME über die Chloroplastenmembran ins Cytoplasma verantwortlich. CE, Chloroplastenmembran; MgP, Mg-Proto; MgPMe, Mg-ProtoME; NE, Kernmembran (Surpin et al., 2002).

## 1.8 Transkriptomics des Plastiden

Die Transkription ist der erste Schritt im Prozess der Genexpression. Veränderungen in der Genexpression sind eine Hauptdeterminante für morphologische und phänotypische Veränderungen, sowie ein Indikator für die zelluläre Reaktion auf endogene und exogene Stimuli oder Störungen (Lockhart & Winzeler, 2000). Im Gegensatz zum Genom ist das Transkriptom hoch dynamisch und kann sich sehr schnell und dramatisch verändern. Zu wissen, wann, wo und in welchem Ausmaß ein Gen exprimiert wird, ist entscheidend, um die Aktivität und die biologische Funktion des kodierten Proteins zu verstehen (Bouchez & Höfte, 1998).

Mikro- und Makroarrays ermöglichen die Genexpressionsanalyse im großen Maßstab, entweder ausgehend vom gesamten Genom eines Organismus oder, wie in dieser Arbeit, eines ausgewählten Sets von Genen, dem nukleären Transkriptom des Plastiden. Nachdem DNA-Mikro- und Makroarrays vielfach und erfolgreich eingesetzt worden sind (DeRisi et al., 1997; Wodicka et al., 1997), um die mRNA-Abundanzen in verschiedensten Experimenten und Organismen zu messen, ist das Leistungsvermögen dieser Technik mittlerweile anerkannt und es liegt der Fokus nun auf Fragen des experimentellen Designs, der Datenanalyse und der Extraktion der biologischen Bedeutung der Ergebnisse.

Das klassische Design der differentiellen Genexpressionsanalyse ist der Vergleich der Expressionsrate eines oder mehrerer Gene unter verschiedenen Bedingungen. Ein Referenzstatus (z.B.: Wildtyp; ohne Behandlung etc.) gilt als die Kontrollbedingung und eine oder mehrere Bedingungen, in denen eine oder einige Variablen verändert sind, als Probe (z.B.: Mutante; veränderte Umweltbedingungen). In der Regel werden in einer Einzelfallstudie mehrere hundert signifikant differenziell regulierte Gene identifiziert. Die Extraktion der biologischen Bedeutung der Ergebnisse der *Transkriptomics* kann durch die Kombination mit den Gebieten der *Proteomics* und *Metabolomics* entscheidend unterstützt werden. Die Grundlage dieser Techniken bilden zum großen Teil Sequenzinformationen, erzeugt durch die vielfältigen Sequenzierungsprojekte wie dem des Arabidopsis-Genom-Projekts (The Arabidopsis Genome Initiative, 2000) sowie die Vorhersage von Genen und die Annotation der

Funktionen und subzellulären Lokalisationen der Gene bzw. Genprodukte durch bioinformatische Methoden und/oder durch den experimentellen Nachweis.

Die funktionelle Klassifizierung der Proteine hat in der letzten Zeit große Fortschritte gemacht und immer mehr Genprodukte sind in einer wachsenden Zahl von Datenbanken nach ihrer Funktion und ihrer Beteiligung an bestimmten Stoffwechselwegen kategorisiert. Als schwieriger erweist sich noch der experimentelle Nachweis der subzellulären, insbesondere der subplastidären Lokalisation, auch wenn durch die Arbeitsgruppe van Wijk's eine erste umfangreiche Datenbanken zugänglich ist (*PPDB*).

### **1.8.1 Clusteranalyse zur Gruppierung der Gene in Regulons**

Um die biologische Information herausfiltern zu können, die der Vergleich vieler Einzelfallstudien untereinander liefert, muss die unüberschaubar große Menge an Expressionsdaten reduziert und visualisiert werden. Die Clusteranalyse (*Clustering*) gehört in die Gruppe der *data mining*-Methoden und umfasst eine Vielzahl von Verfahren, die die Daten nach ihrer Ähnlichkeit bzw. Unähnlichkeit unüberwacht (*unsupervised*), also ohne Vorgabe eines Musters, in bedeutungsträchtige Gruppen klassifizieren und dadurch auf die relevanten enthaltenen Informationen reduzieren.

Clusteralgorithmen wie SOM- [*Self-Organizing-Maps* (Tamayo et al., 1999)], *k-means*- [*k-tuple means clustering* (Sherlock, 2000)] und Hierarchisches Clustering [HCL (Eisen et al., 1998; Sturn et al., 2002)] werden benutzt, um multiple Datensätze (in der vorliegenden Arbeit mehrere Expressionsprofile) miteinander zu vergleichen und (in diesem Fall) Gene nach ihrem Expressionsverhalten über eine Vielzahl von Experimenten in Gruppen gleich exprimierter Gene, in so genannte Regulons, einzuteilen. Das Ziel ist es, Cluster zu bilden, die untereinander möglichst unterschiedlich sind, das heißt eine möglichst große Distanz zueinander aufweisen, deren einzelnen Mitglieder aber wiederum sehr ähnlich sind und eine geringe Distanz zueinander aufweisen (Brazma & Vilo, 2000; Quackenbush, 2001). Es werden zwei Gruppen von Clusteralgorithmen unterschieden: Hierarchisches Clustering und *Partitioning*. Das Ergebnis des Hierarchischen Clustering ist eine geschachtelte Struktur

von Clustern, in der kleinere in größeren Clustern enthalten sind. Beim *Partitioning*, dazu gehören SOM- und *k-means*-Clustering, entsteht nur eine Ebene von Clustern, die nur Objekte und keine Cluster enthalten (siehe Kapitel 2.8.3.4; Seite 40 ff).

## 1.9 Ziel der Arbeit

Unter der Annahme, dass Gene mit ähnlichem Expressionsverhalten wahrscheinlich auch ähnliche Funktionen aufweisen, können, aufgrund der bereits annotierten und klassifizierten Gene in demselben Expressionscluster, auch Gene bisher unbekannter Funktion (dies trifft auf fast die Hälfte der Gene für das nukleäre Transkriptom des Plastiden von *A. thaliana* zu) nach dem Konzept *guilt-by-association* charakterisiert werden (Ruan et al., 1998; Quackenbush, 2003). Dieses Konzept der so genannten „*gene discovery*“ wurde schon sehr früh und mehrmals erfolgreich bei *Saccharomyces cerevisiae* angewandt (DeRisi et al., 1997; Eisen et al., 1998; Giaever et al., 2002).

Über die aufeinander abgestimmte Regulation der Genexpression von bereits klassifizierten aber funktionell unterschiedlichen Genen kann darüber hinaus ein Zusammenhang zwischen verschiedenen Genfunktionen hergestellt und die regulatorischen Netzwerke der Zelle untersucht werden. Die Identifizierung von Regulons legt die Suche nach gemeinsamen, überrepräsentierten Sequenzmotiven der genomischen DNA in der Umgebung dieser Gene nahe. Auf diese Weise können neue *cis*-regulatorischen Elemente und Transkriptionsfaktoren identifiziert und das Verständnis der regulatorischen Netzwerke der Genexpression erweitert werden.

Eine weitere Anwendung ist die Erstellung eines „transkriptionellen Fingerabdrucks“, der die Beziehung zwischen Stimulus und der transkriptionellen Antwort darstellt. Die Clusteranalyse der Expressionsprofile macht die Verwandtschaft der transkriptionellen Antworten untereinander sichtbar. Dadurch können Expressionsprofile in bereits gefundene Klassen eingeordnet (*class prediction*) oder neue Klassen gefunden werden [*class discovery* (Golub et al., 1999)]. Bei ausreichend großer Anzahl und Variation an Expressionsprofilen ist es dann möglich, unbekannte Mutationen anhand der Transkriptionsmuster bekannter Stimuli zu charakterisieren.

## 2 Material und Methoden

### 2.1 Chemikalien und Enzyme

$[\alpha\text{-}^{33}\text{P}]\text{dCTP}$	Amersham, Freiburg
Agarose	Sigma, Steinheim
Chloramphenicol	Sigma, Steinheim
Chloroform	Carl Roth Chemie, Karlsruhe
dNTP-Mix	Invitrogen, Karlsruhe
EDTA	Serva, Heidelberg
Essigsäure	Riedel-deHaën, Seelze
Ethanol	Sigma, Steinheim
Ethidiumbromid	Sigma, Steinheim
First Strand Buffer	Gibco BRL; Wiesbaden
Herings-Sperma DNA	Promega, Mannheim
Methanol	Merk, Darmstadt
Oligonukleotide	BIG Biotech, Freiburg
PCR-Puffer 10x (konzentriert)	Amersham, Freiburg
Primermix für die cDNA-Synthese	Eurogentech, Belgien
Salzsäure	J.T. Baker, Deventer (NL)
Superscript II	Invitrogen, Karlsruhe
Tris	Sigma, Steinheim
TRIzol	Invitrogen, Karlsruhe

Alle anderen Chemikalien wurden von den Firmen Roth (Karlsruhe) und Merck (Darmstadt) bezogen.

## 2.2 Geräte und Verbrauchsmaterial

Dampf-Autoklav Typ 23	Melag, Berlin
Elektrophorese-Hoefer-Kammer SE600	Amersham, Freiburg
Elektrophoresekammer DNA Sub Cell	BioRad, München
Falcon-Tubes (25 ml, 50 ml)	Becton Dickinson, Franklin Lakes (USA)
Feinwaage 1608 MP	Sartorius, Göttingen
Kühlzentrifuge Varifuge 3.0R	Heraeus Sepatech, Stuttgart
Küvetten (halb-mikro, 10x10x45)	Greiner, Frickenhausen
Magnetrührer IKAMAG RCT	IKA-Labortechnik, Staufen
Metallblock-Thermostat TCS	Labortechnik Barkey, Bielefeld
pH-Meter Modell pH537	WTW, Weilheim
Phosphor-Imager MD Storm860	Amersham, Freiburg
Phosphorscreen	Applied Biosystems, Foster City (USA)
Reaktionsgefäße (0.5 ml, 1.5 ml, 2 ml)	Eppendorf, Hamburg
SpeedVac HetoVac VR-1/120/240	Heto Lab, Wiesloch
Tischkühlzentrifuge Centrifuge 5402	Eppendorf, Hamburg
Universalschüttler KS 10	Edmund Bühler, Tübingen
UV/Visible Spectrophotometer	Amersham, Freiburg
Vortex Heidolph REAX 1R	Bender & Hobein GmbH, Freiburg

## 2.3 Software

ArrayVision™ 6.0	Imaging Research Inc, USA
ArrayStat™ 1.0	Imaging Research Inc, USA
Clone Manager Version 4.0	Scientific & Educational Software
Genesis Version 1.5.0 b1	Sturn et al., 2002
GEPAS	<a href="http://gepas.bioinfo.cnio.es/tools.html">http://gepas.bioinfo.cnio.es/tools.html</a>
GST-PRIME	Varotto et al., 2001
AraCyc	<a href="http://www.arabidopsis.org/tools/aracyc/">www.arabidopsis.org/tools/aracyc/</a>
KEGG	<a href="http://www.genome.jp/kegg/regulation.html">www.genome.jp/kegg/regulation.html</a>

## 2.4 Zugang zu den Rohdaten und ergänzenden Informationen

Alle Datensätze der 101 Expressionsprofile sind in der *GEO*-Datenbank (<http://www.ncbi.nlm.nih.gov/geo/>) unter den Zugangscodes GSE1160 bis GSE1260 öffentlich zugänglich und in die Publikation:

Biehl A, E. Richly, C. Noutsos, F. Salamini und D. Leister (2004). „Analysis of 101 nuclear transcriptomes reveals 23 distinct regulons and their relationship to metabolism, chromosomal gene distribution and co-ordination of nuclear and plastid gene expression.” *Gene* (im Druck), eingeflossen.

Ergänzende Informationen dazu sind unter <http://mpiz-koeln.mpg.de/~leister/101.html> einzusehen.

## 2.5 Pflanzenmaterial

Die 101 Expressionsprofile stellen jeweils den Vergleich des mRNA-Status von *A. thaliana*-Pflanzen zwischen einer Proben- und einer Referenzbedingung dar. Bis auf den zu untersuchenden Stimulus wurden die Pflanzen eines Experimentes unter denselben Wachstumsbedingungen (z.B.: Lichtintensität, Tag-Nacht-Rhythmus, Bewässerung etc.) kultiviert. In der Regel wurden junge Rosettenblätter von jeweils etwa 50, drei bis vier Wochen alten, Pflanzen zeitgleich, zwei Stunden nach Beginn der Tagesperiode geerntet und zu einer Probe gepoolt. Ausnahmen sind die Untersuchungen der differentiellen Expression in Stängel- und Wurzelgewebe und der Entwicklung von ein, zwei und drei Wochen alter Pflanzen. Eine detaillierte Beschreibung des Pflanzenmaterials ist in Tabelle 3.1 (Seite 47) aufgeführt.

Das Pflanzenmaterial bzw. die extrahierte RNA der meisten untersuchten Stimuli stammt überwiegend von Mutanten und/oder unter speziellen Wachstumsbedingungen herangezogenen Pflanzen des Labors der Arbeitsgruppe Leister (Ausnahmen siehe Tab. 3.1, Seite 47). Der Großteil der differentiellen Expressionsanalysen wurde von A. Biehl durchgeführt, aber auch die Ergebnisse einiger Hybridisierungs-Experimente von A. Dietzmann und J. Kurth sind mit in den Gesamtdatensatz dieser Arbeit eingeflossen.

## 2.6 Design des 3292-GST-Makroarray

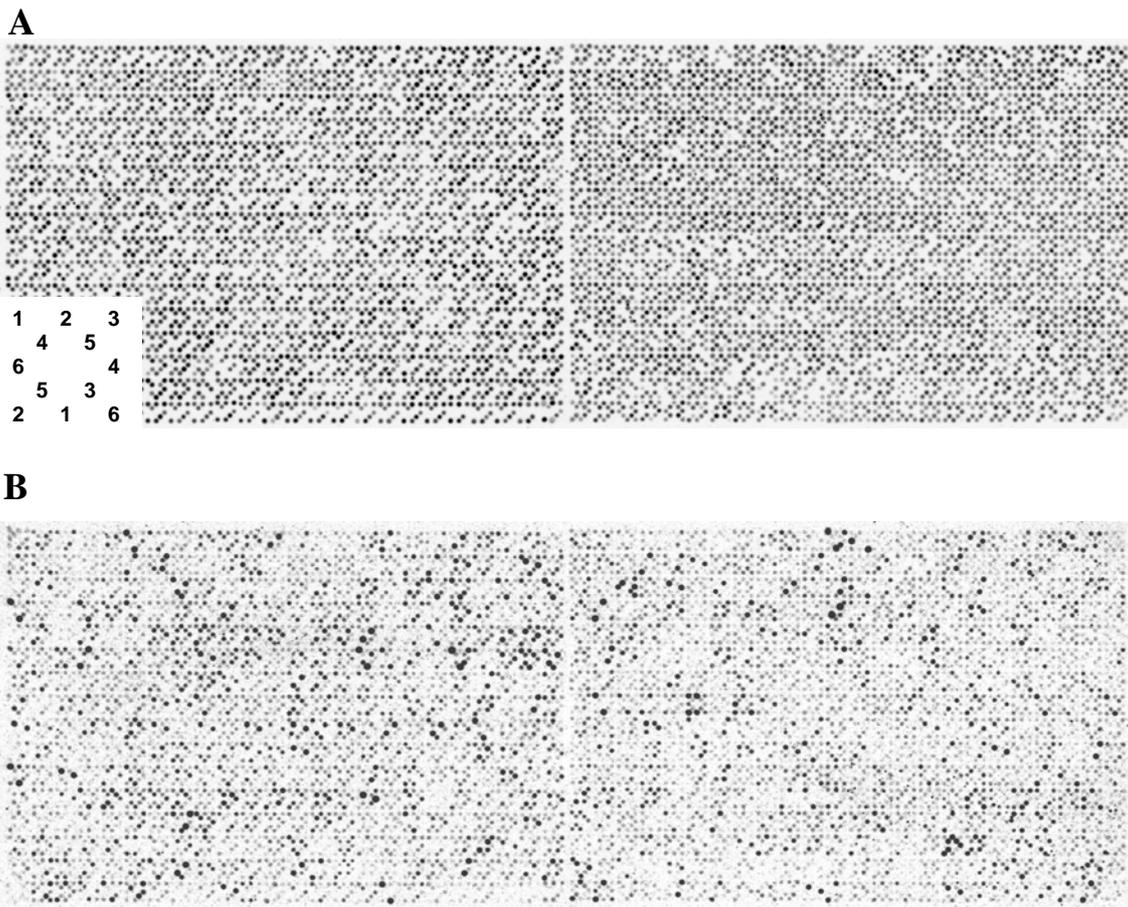
Mikro- und Makroarrays sind systematische Anordnungen hunderter bis mehrerer tausend bekannter cDNAs, die als DNA-Sonden auf einem Substrat, auf Glas oder Nylon immobilisiert worden sind. Die Identität und die Abundanz von freier markierter DNA von unbekannter Sequenz und Konzentration kann über die spezifische komplementäre Bindung (Hybridisierung) an die einzelnen cDNA-Elemente (*spots*) des Arrays bestimmt werden (Nadon et al., 2001). Während in den ersten Expressionsanalysen meist nur ESTs (*expressed sequence tags*) benutzt wurden, erlaubt die mittlerweile vollständige Sequenzinformation des Genoms verschiedener Organismen die Amplifizierung von GSTs (*gene sequence tags*) direkt von der genomischen DNA (Zhang et al., 1999).

Das Design des in dieser Arbeit verwendeten 3292-GST-Makroarrays basiert auf einem 1827-GST-Makroarray (Kurth et al., 2002), der 1999 mit der Sequenzinformation von etwa 14000 der insgesamt 25700 Genen von *Arabidopsis thaliana* von Varotto und Mitarbeitern in unserem Labor entworfen worden war. Mit der vollständigen Sequenzierung des *Arabidopsis*-Genoms konnten durch Varotto und Richly mit Hilfe von *TargetP* (Emanuelsson et al., 2000), einem Softwareprogramm für die Vorhersage N-terminaler Präsequenzen (cTPs), genomweit 2661 nukleäre Chloroplastengene identifiziert werden. Etwa 75 % der Gene des 3292-GST-Makroarrays waren bereits als exprimierte Sequenzen (ESTs) in den öffentlichen Datenbanken annotiert.

Wie von Varotto und Mitarbeitern (2001) beschrieben, wurden die GSTs der nukleären Gene des Chloroplasten und weiterer 631 Gene, deren Genprodukte nicht im Chloroplasten lokalisiert sind, mit Hilfe von *GST-PRIME* entworfen. Die Primerpaare wurden als 35-mere synthetisiert, bestehend aus einer genspezifischen 20-mer Sequenz und einer universellen 15-mer Sequenz des 5'-Endes für die Reamplifikation mit universellen Endprimern.

Die PCR-Amplifikation der 150 bis 2050 Basenpaare großen GSTs erfolgte von genomischer DNA des Ökotyps 0 (Col-0) von *A. thaliana*. Die PCR-Produkte wurden per Agarose-Gelelektrophorese analysiert und quantifiziert und als Duplikate nach dem in Abbildung 2.1 gezeigten Muster auf eine Nylonmembran aufgebracht.

Zusätzlich wurden zur Kontrolle der Qualität der Hybridisierungen genomische DNA von Col-0, humane cDNA (*CTLA4*, *PE*, *BPI*), *Inf1* und *Inf2* von *Phytophthora infestans* sowie zur alternativen Normalisierung (*Spiking*-Kontrolle) Luciferase 1 in exponentiell verschiedenen Konzentrationsstufen in das *Spotting*-Schema aufgenommen. Die Synthese der Primer, die Analyse und Quantifizierung der GST-Amplikons wie auch die Herstellung der Makroarrays wurde von Eurogentec, Seraing, Belgien durchgeführt.



**Abb. 2.1.: *Spotting*-Muster des 3292-GST-Makroarrays:** Die Abbildung **A**. zeigt das eingescannte Bild einer Hybridisierung mit  $^{33}\text{P}$ -markierten Oligonukleotiden, die komplementär an die universelle 15-mer Sequenz jedes GSTs gebunden haben. Dadurch konnte die Qualität der gespotteten Amplikons verifiziert werden. Der Makroarray ist aus einem Raster von 16 x 48 Untereinheiten aufgebaut. Maximal sechs verschiedene GSTs oder Kontrollen sind pro Untereinheit in Duplikaten aufgetragen. Unten rechts ist das *Spotting*-Muster der GSTs der Untereinheiten des Rasters schematisch dargestellt. In **B**. ist das Ergebnis der Hybridisierung einer komplexen und durch  $^{33}\text{P}$ -markierten cDNA-Probe abgebildet. Die Probe hat hier sequenzspezifisch an die jeweils korrespondierende Hybridisierungssonde gebunden.

Das automatisierte Design der Primer mit Hilfe des Softwareprogramm *GST-PRIME* gewährleistete sowohl die Genspezifität der Primer als auch die PCR-Amplifizierung von überwiegend Exonbereichen der genomischen DNA in einer vorherrschenden Größe von 450-850 bp für mehr als 90% der GSTs. Zudem wurde von Richly und Varotto eine *BLASTN*-Analyse durchgeführt, um die Möglichkeit von Überkreuzhybridisierungen zu untersuchen. Nur für 90 der 3292 GSTs wurde eine Sequenzhomologie zu irgendeinem anderen offenen Leserahmen von *A. thaliana* gefunden und im Vergleich der 3292 GSTs untereinander besaßen nur 14 GSTs einen homologen Sequenz von mehr als 100 bp (Richly et al., 2003).

Durch die universelle 15-mer Sequenz am 5'-Ende der für die Amplifikation der GSTs benutzten Primer, konnte die Qualität des *Spotting*s der DNA-Amplikons sehr einfach durch die Hybridisierung komplementärer und durch <sup>33</sup>P-markierter Oligonukleotide überprüft werden (Abb. 2.1, A).

### 2.7 Hybridisierungsexperimente

DNA-Arrays zeichnen sich durch eine Konstruktion aus, bei der auf einer festen Matrix die genspezifischen Sequenzen immobilisiert werden und mit markierter freier cDNA hybridisieren können. In diesem Sinne kann man DNA-Arrays als eine Umkehrung des Northern-Blots betrachten (Freeman, 2000), bei dem die Probe immobilisiert wird und mit freien, markierten Sonden hybridisieren kann. Hintergrund für die Genexpressionsanalyse mit DNA-Arrays ist die Annahme, dass, je stärker ein Gen exprimiert ist, desto größer die Menge markierter RNA und folglich das zu detektierende Signal ist.

Allerdings können nur relative Transkriptionsniveaus und keine absoluten Mengen oder gar die Kopienzahl bestimmt und miteinander verglichen werden. Bei der Erststrang-Synthese handelt es sich um 1:1 Umschreibungen. Ebenso besteht zwischen der Menge an radioaktiv markierter Probe und der detektierten Signalintensität ein linearer Zusammenhang (Freeman, 2000).

### 2.7.1 Extraktion von Gesamt-RNA

1 bis 5 g Blattmaterial (bzw. im speziellen Fall auch Wurzel- oder Stengelgewebe) wurden geerntet, das Pflanzengewebe in flüssigem N<sub>2</sub> aufgebrochen, in ein 50 ml Falcon-Reaktionsgefäß überführt und zur Extraktion der Gesamt-RNA 20 ml TRIzol (Life Technologies), ein monophasischen Reagen mit Guanidin-Thiocyanat und Phenol, hinzu gegeben. Das Blattmaterial wurde unter gelegentlichem Schwenken für 5 min bei Raumtemperatur (RT) in TRIzol inkubiert und daran anschließend für 5 min bei 4500 rpm zentrifugiert. Das Pellet wurde verworfen und zu dem Überstand 0.2 Volumen Chloroform addiert. Die Inkubation (5 min bei RT) erfolgte unter regelmäßigem Schwenken. Danach wurde die Lösung für 10 min bei 4500 rpm zentrifugiert, die obere, wässrige Phase abgenommen und in ein neues 50 ml Reaktionsgefäß überführt.

Je ein halbes Volumen Isopropanol und 0.8 M Natriumcitrat/1.2 M NaCl wurden hinzu gegeben, der Ansatz wiederum für 10 min geschwenkt und daran anschließend für 10 min bei 4°C und 4500 rpm zentrifugiert. Der Überstand wurde verworfen und das RNA-Pellet zweimal in kaltem Ethanol (70 %) gewaschen, kurz getrocknet und in 250 µl ddH<sub>2</sub>O unter Zugabe von RNaseH-Inhibitor (Invitrogen) aufgenommen und bei 55-60°C für 10 min resuspendiert.

Die Absorptionsmessung zur Bestimmung der Konzentration und des Reinheitsgrades der RNA in wässriger Lösung erfolgte in einem Spektralfotometer mittels Quarzglasküvette (1000 µl) bei 260 nm und 280 nm. Die RNA wurde 1:100 verdünnt und die Absorption bei einer Wellenlänge von 260 nm ( $A_{260}$ ) fotometrisch bestimmt. Die Konzentration im Messvolumen der Küvette (1000 µl) berechnet sich als Quotient des gemessenen Extinktionswertes ( $A_{260nm}$ ), des  $A_{260}$ -Faktors und des Verdünnungsfaktors. Demnach ist die Konzentration  $c$  (µg/µl):  $c = A_{260} \times 40 \times 100$ .

Bei 260 nm und einer optischen Dichte (OD) von 1 beträgt die Konzentration an einzelsträngiger RNA 40 µg/ml. Aus dem Quotienten  $A_{260nm} / A_{280nm}$  lässt sich der Reinheitsgrad der RNA bestimmen. Für RNA sollte der Quotient optimalerweise zwischen 1.5 und 2.0 liegen. Für die nachfolgende cDNA-Erststrangsynthese wurde die Ausgangskonzentration an RNA auf 2-3 µg/µl in ddH<sub>2</sub>O eingestellt.

### 2.7.2 cDNA-Erststrangsynthese

20-30 µg der aus Pflanzenmaterial isolierten Gesamt-RNA wurden durch reverse Transkription in cDNA umgeschrieben und gleichzeitig durch den Einbau von [ $\alpha$ - $^{33}\text{P}$ ]dCTP radioaktiv markiert.

Dazu wurden 10 µl der Gesamt-RNA mit 5 µl einer Antisense-Primer-Mischung von Oligonukleotiden (je 2.5 pmol) versetzt. Die Antisense-Primer-Mischung setzte sich aus den Primerpaaren zusammen, die schon für die Amplifizierung der GSTs des Makroarray verwendet wurden und somit passend für die Bindung in Antisense-Richtung an jedes der als GST auf dem Makroarray aufgebrauchten 3292 Gene waren. Für die Bindung der Primer an die RNA (*Annealing*) wurde das Reaktionsvolumen für 10 min bei 70 °C erhitzt, anschließend kurz auf Eis gekühlt und bei 43 °C äquibriert. Der Reaktionsmix für die reverse Transkription (6 µl 5-fach Puffer, 3 µl DTT, 1,5 µl Nukleotid-Mix (je 16 mM dAGT, dGTP, dTTP, 100 µM dCTP), 50 µCi [ $\alpha$ - $^{33}\text{P}$ ]dCTP (>2500 Ci/mmol, Amersham Pharmacia Biotech) sowie 1 µl Reverse Transkriptase (SuperScriptII, 200 U/µl, Life Technologies) wurden zugegeben und der Reaktionsansatz für 1 h bei 43 °C inkubiert. Die Reaktion wurde mit 1 µl 0.5 M EDTA abgestoppt und die RNA durch Zugabe von 3 µl 3 M NaOH und 1 µl 1 % SDS für 30 min bei 65 °C hydrolisiert.

Nach weiteren 15 min bei RT wurde die Lösung mit 3 µl 2 M HCl und 10 µl Tris/HCl pH 8.0 neutralisiert. Die cDNA wurde anschließend durch Zugabe von 5 µl tRNA (10 mg/ml), 5 µl Na-Acetat pH 5,3 und 60 µl Isopropanol für 30 min bei -20 °C gefällt, um nicht eingebaute Nukleotide zu entfernen. Der Niederschlag wurde pelletiert, getrocknet und in 100 µl ddH<sub>2</sub>O aufgenommen.

Zur Bestimmung der Effizienz der Reaktion wurde die Einbaurate der markierten Nukleotide bestimmt. Dazu wurde vor der Isopropanolfällung ein Aliquot entnommen, in einem Fällungsmix (7 µl ddH<sub>2</sub>O; 7 µl 0.05 M EDTA pH 8.0; 5 µl 10 mg/ml tRNA) aufgenommen und je zu Hälfte auf zwei Glasfibrerfiltern aufgebracht. Ein Filter wurde dreimal 5 min in eiskalter TCA-Lösung (10 % TCA, 1 % Na-Pyrophosphat) gewaschen, kurz mit EtOH abgespült und getrocknet. Beide Filter wurden mit 5 ml Szintillationscocktail versetzt und die Radioaktivität in einem Szintillationszähler bestimmt.

Die Einbaurate errechnet sich aus: Einbau in % = (gewaschen/ungewaschen) \* 100. Ab einer Einbaurate von 30 % konnten die Proben für die Hybridisierungen verwendet werden.

### 2.7.3 Hybridisierungsprozedur

Die Hybridisierungen wurden in 300 ml Glasröhren im Rotationsofen ausgeführt. Die GST-Makroarrays wurden für mindestens 30 min bei 65 °C in 30 ml Hybridisierungspuffer (0,5 M Na<sub>2</sub>HPO<sub>4</sub>, pH 7.2; 7 % SDS; 1 mM EDTA pH 8.0) und zusätzlichen 300 µl denaturierten Heringssperma prähybridisiert, um unspezifische Bindungen zu verhindern. Danach wurde die markierte cDNA Probe hinzu gegeben und die Makroarrays bei 65 °C über Nacht hybridisiert. Daran anschließend wurde der Hybridisierungspuffer durch Waschpuffer (2 x SSPE, 0.1 % SDS) ausgetauscht und die Membranen für jeweils zweimal 30 min bei 65 °C gewaschen. Im abschließenden Waschschrift wurden die Filter für 15 min bei Raumtemperatur in 50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 7.2; 0.2 % SDS geschwenkt. Die Membranen wurden kurz angetrocknet, in Haushaltsfolie luftblasenfrei eingepackt und auf Phosphoimagerplatten (Molecular-Dynamics) für drei Tage exponiert.

Die DNA-Makroarrays konnten fünf- bis achtmal wieder verwendet werden. Da die zu vergleichenden Wachstumsbedingungen auf ein und dasselbe Filterset hybridisiert wurden, musste die gebundene cDNA eines vorangegangenen Hybridisierungsexperiments vor jedem erneuten Durchlauf zunächst entfernt werden (*Stripping*). Die Makroarrays wurden in einer Schale für 10 min mit einer 85 °C heißen Lösung von 0.1% SDS in 1 Liter ddH<sub>2</sub>O überschichtet und auf einen Schüttler gestellt. Mit einem Oberflächen-Geigerzählerrohr wurde die Membranen auf verbliebene Radioaktivität kontrolliert und bei +4 °C in einem Plastikbeutel eingeschweißt im Kühlschrank aufbewahrt.

## 2.8 Datenanalyse

Jedes der 101 Expressionsprofile basiert auf dem Vergleich einer Kontrollbedingung (z. B.: Wildtypstamm, Zeitpunkt 0 einer Zeitreihe, Medium ohne Zusatz) mit einer Bedingung, deren Änderungen relativ zur Kontrollbedingung von Interesse ist (z. B. Mutantenstamm, Zeitpunkte nach einer Behandlung, Medium mit Zusatz). Für beide Vergleichsbedingungen wurden jeweils mindestens drei unabhängige Hybridisierungsexperimente auf drei verschiedenen Makroarrays durchgeführt. Das bedeutet, dass um den Einfluss von Variationen in der Qualität der Makroarrays untereinander zu minimieren, für jede Wachstumsbedingung drei unabhängige komplexe Proben hergestellt und beide Bedingungen reziprok auf dieselben Makroarrays hybridisiert wurden. Aus der Hybridisierung einer komplexen Probe ging ein separater Datensatz mit jeweils zwei Datenpunkten pro GST hervor. Ein Expressionsprofil stellt den Vergleich von mindestens je sechs Expressionsdaten eines jeden der 3292 GSTs pro Vergleichsbedingung dar.

### 2.8.1 Makroarray-Quantifizierung

Die Phosphoimagerplatten wurden mit einem Phosphorimagersystem (Storm 860, Molecular Dynamics) bei einer Auflösung von 50  $\mu\text{m}/\text{Pixel}$  eingelesen und als Graustufenbild mit 16bit Farbtiefe, d. h. in 65536 Graustufen gespeichert. Die Zuordnung der einzelnen Positionen auf dem Makroarray mit den entsprechenden GST-Identifikatoren, die Detektion artifizieller Signale, sowie die Quantifizierung der Signalintensitäten nach Subtraktion des lokalen Hintergrunds und Normalisierung, erfolgte mit der Software *ArrayVision* (Version 6.0, Imaging Research Inc.). In *ArrayVision* wurde zunächst ein ideales Raster definiert, das danach sowohl automatisch als auch manuell dem eingelesenen Bild des Makroarrays angepasst werden konnte. Der durchschnittliche Grauwert innerhalb jedes Rasterelementes wurde berechnet, der umgebende lokale Hintergrund subtrahiert und die Signalintensitäten jeder einzelnen Position des Rasters gegen die Signalintensitäten aller Positionen

normalisiert (Kurth et al., 2002). Das Ergebnis einer Hybridisierung, eines unabhängigen Experiments, ist eine von *ArrayVision* rearrangierte und exportierte Tabelle, in der das normalisierte und, um den lokalen Hintergrund korrigierte, Intensitätsvolumen der Duplikate eines GSTs in zwei Spalten einer Reihe dem entsprechenden GST-Identifikator zugeordnet ist.

### **2.8.2 Statistische Analyse der Makroarraydaten**

Der Quantifizierung des Makroarrays schließt sich die statistische Analyse und Qualitätskontrolle der Expressionsdaten an. Für die Erstellung eines Expressionsprofils wurden die Datensätze von mindestens drei unabhängigen Experimenten pro Vergleichsbedingung in einer *Excel*-Tabelle zusammengestellt und in das Software-Programm *ArrayStat* (Version 1.0, Imaging Research Inc.) importiert. *ArrayStat* (Version 1.0, Imaging Research Inc.) ermöglicht auch bei relativ kleiner Anzahl von Datenpunkten pro GST und Bedingung die Berechnung von Signifikanzintervallen (*p*-Werten) und Standardfehlern, kontrolliert die Fehlerrate der „Falsch-Positiven“ sowie detektiert und eliminiert statistische Ausreißer innerhalb der Datensätze. Verwendet wurde ein proportionales statistisches Modell, in dem die Daten zur Basis 10 logarithmiert werden, der Standardfehler wurde gepoolt (*Pooled Common Error*), um der geringen Stichprobenzahl pro Datenpunkt Rechnung zu tragen und ein *z-Test* durchgeführt, um signifikante Unterschiede zwischen den Datensätzen zweier verschiedener Vergleichsbedingungen zu identifizieren. Signifikante Unterschiede zwischen den Datensätzen heißt im Falle der Expressionsanalyse die Transkripte herauszufiltern, die einen eindeutigen Unterschied in ihrer Expressionsrate unter zwei verschiedenen Wachstumsbedingungen zeigen. Zumindest zwei bis drei unabhängige Datensätze pro Vergleichsbedingung sind nötig, um unter Verwendung spezieller statistischer Algorithmen, die der geringen Stichprobenzahl pro Datenpunkt Rechnung tragen, verlässliche Ergebnisse zu erzielen (Lee et al., 2000). Mindestens zwei Drittel der Datenpunkte pro GST und Bedingung (z.B.: 4/6 oder 6/8 Datenpunkte bei drei bzw. 4 unabhängigen Hybridisierungen pro Wachstumsbedingung) mussten nach der

Bereinigung der Datensätze von artifiziellen Signalen (in *ArrayVision*) und statistischen Ausreißern in die Berechnung der durchschnittlichen Expressionsrate eingehen, um nicht von der weiteren statistischen Analyse ausgeschlossen zu werden. Für die Berechnung der Falsch-Positiv-Fehlerrate wurde ein nominaler Alpha-Wert von 0.05 und zur Korrektur der Falsch-Positiven die *False Discovery Rate* gewählt.

Das Resultat der statistischen Analyse durch *ArrayStat* ist eine Tabelle mit der durchschnittlichen logarithmierten Expression jedes der 3292 GSTs für jeweils eine der beiden Vergleichsbedingungen, aus der sich die relative Expression jedes Gens zwischen der Proben- und Referenzbedingung errechnen lässt. Ein Signifikanzintervall (*p*-Wert) beschreibt die Wahrscheinlichkeit der differentiell (zwischen zwei Vergleichsbedingungen) unterschiedlichen Expression eines Gens.

### **2.8.3 Prozessierung und Interpretation der Expressionsdaten**

#### **2.8.3.1 Funktionelle Klassifizierung des Transkriptoms**

Die Genprodukte der 3292 Gene des Makroarrays wurden durch den Abgleich der Einträge in öffentlich zugängigen Datenbanken funktionell klassifiziert. Hauptsächlich beruht die Definition der funktionellen Klassen und die Einteilung der Gene auf den Annotationen der entsprechenden Genprodukte in der *KEGG PATHWAY Database* (<http://www.genome.jp/kegg/pathway.html>) und des *AraCyc Server* (<http://www.arabidopsis.org/tools/aracyc/>). Die entsprechenden Proteine der GSTs wurden in 15 funktionelle Klassen unterteilt (amino acid metabolism; carbohydrate metabolism; other; other metabolic functions; photosynthesis/dark reaction; photosynthesis/light reaction; protein modification and fate; protein phosphorylation; protein synthesis; secondary metabolism; sensing, signalling and cellular communication; stress response; transcription; transport; unclassified).

Darüber hinaus konnten 147 von 3292 Genprodukten anhand der *Plastid Proteome Data Base* [*PPDB* (<http://www.ppdb.tc.cornell.edu/>)] nach ihrer subplastidären Lokalisation charakterisiert werden.

### 2.8.3.2 Definition von Regulons und Transkriptomklassen

Es wurde Hierarchisches Clustering (HCL) für die Klassifizierung der 101 Expressionsprofile (*Class Discovery* durch *Class Prediction*) nach ihrer Ähnlichkeit und zur Ermittlung der Anzahl der zu erwartenden Regulons, abhängig von der Ähnlichkeit der Transkriptionsrate der 3292 Gene in 101 differentiellen Expressionsprofilen, benutzt.

Unterschiedliche Clusteralgorithmen haben verschiedene Vor- und Nachteile. Deshalb wurden in der vorliegenden Arbeit die Datensätze sowohl mit HCL als auch mit den Verfahren des *Partitioning* – SOM- und k-means-Clustering - untersucht und so die Ergebnisse der drei Clustermethoden gegeneinander abgesichert. Aufgrund der Überlegenheit bei der Filterung von *Noise* und in der Erkennung von Datenmustern (Zschunke, 2003) wurden zur Definition der Regulons letztendlich die Ergebnisse des SOM-Clustering benutzt. Da die Methoden des *Partitioning* keine Information über Distanzen zwischen den Regulons liefern, wurden die Software *GEPAS* (<http://gepas.bioinfo.cnio.es/tools.html>) benutzt, um die durch *Self-Organizing-Maps* definierten Regulons nochmals hierarchisch zu clustern (siehe Abb. 3.11, Seite 64).

Die Clusteranalysen mit den Algorithmen von HCL, SOM- und *k-means*-Clustering sowie die Visualisierung der Ergebnisse wurden mit dem *Genesis*-Softwarepaket [Version 1.3.0 – 1.5.0 (Sturn et al., 2002)] durchgeführt. Grundsätzlich ist jede dieser Methoden dazu geeignet, die vorherrschenden Strukturen in einem Datensatz aufzuspüren (Golub, 2001).

Mit *Genesis* wurde zuerst eine Expressionsmatrix erstellt. In einer Eingabetabelle wurde die relative Expressionsrate ( $x_{ij}$ ) jedes der 3292 Gene gegen die 101 Wachstumsbedingungen aufgetragen. Hochregulierte Gene werden von einem Bereich von  $1 < x_{ij} < +\infty$ , herunterregulierte Gene von einem Bereich von  $0 \leq x_{ij} < 1$  repräsentiert. Deshalb müssen die Expressionsraten logarithmiert werden, so dass hochregulierten Genen positive Werte und herunterregulierten Gene negative Werte zugeordnet werden. Benutzt wurde der Logarithmus zur Basis 2 ( $\log_2$ ), der eine offensichtlichere Unterscheidung im Sinne von „doppelt“ und „halb“ zulässt.

Die Qualität der Regulons und deren biologische Aussagekraft sind von verschiedenen Parametern abhängig. Dazu zählt neben der Clusteranzahl und des Clusteralgorithmus auch das Distanzmaß. Je nach Algorithmus muss eine Distanzfunktion zur Bestimmung des Abstands zweier Elemente ( $d(x_i, y_i)$ , zum Beispiel die euklidische Distanz) und/oder eine Methode zur Berechnung des Mittelpunktes oder des Zentroiden eines Clusters bekannt sein. Die Anzahl fehlender Werte in der Datenmatrix muss vor der Distanzberechnung reduziert werden. Fehlende Werte können durch Effekte während des Spotting des Makroarrays, der Hybridisierung und der Datenanalyse auftreten. Ist die Distanzberechnung ( $d$ ) nicht unabhängig von der Anzahl der Vektorelemente  $n$ , so wie bei der Euklidischen Distanz ( $d_E$ ):

$$d_E = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad , \text{ werden Vektoren mit fehlenden Werten im Vergleich zu Vektoren}$$

ohne fehlende Werte anders und unter Umständen falsch gewichtet (Sturn, 2000). Wenige fehlende Werte sind in der Regel kein Problem, so dass auf die Expressionsdaten ein Filter angewendet wurde, der jedes der 3292 GSTs, für das nicht in mindestens 95 von 101 Expressionsprofilen ein Wert vorhanden war, von der weiteren Clusteranalyse ausgeschlossen wurde. Nach der Filterung zur Reduzierung verblieben 1616 von 3292 Genen. Bei Mikro- bzw. Makroarraydaten sind die Merkmalsräume stetig, so dass als metrisches Distanzmaß die Euklidische Distanz gewählt wurde. Die Daten wurden durch das Quadratische Mittel (*root mean square*) der Gesamtheit der Gene und der Experimente dividiert, um schwache Signale im Vergleich zu stärkeren zu amplifizieren.

### 2.8.3.3 Hierarchisches Clustering

Die modifizierte Genexpressionsmatrix wurde auf zwei verschiedene Weisen untersucht. Es wurden a) die Expressionsraten der 1616 Gene über die 101 Experimente, also die Datenreihen der Expressionsmatrix mit jeweils 101 Expressionsdaten pro Zeile, miteinander verglichen, und

b) die Expressionsmuster der 101 Vergleichsbedingungen, also die 101 Spalten der Expressionsmatrix mit jeweils 1616 Expressionsdaten pro Spalte.

Werden zwei Gene (Reihen) ähnlich exprimiert, kann vermutet werden, dass die Gene ko-reguliert und vielleicht auch funktionell miteinander verbunden sind [*guilt-by-association* (D'haeseleer et al., 2000)]. Aus dem Vergleich der Experimente lassen sich Klassen bilden (*Class Discovery*), die den Effekt eines Stimulus mit dem Muster der transkriptionellen Antwort in Beziehung setzen (Sturn, 2000).

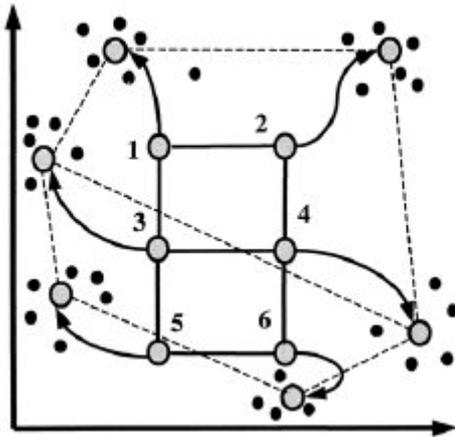
Ein Clusteralgorithmus soll Cluster definieren, in denen die Variabilität auf Intraclusterebene gering ist, dessen Gene sich aber in Bezug auf ihr Expressionsverhalten von Genen anderer Cluster unterscheiden. Hierarchisches Clustering wird dann angewandt, wenn in einer Datenmenge  $S$  keine offensichtliche Partition in separierte Datengruppen/Cluster vorliegt. Im Zuge des HCL wird ein Kladogramm erstellt, dessen Knoten Teilmengen der Datenmenge  $S$  darstellen. Dabei ist  $S$  die Wurzel des Baumes, die einzelnen Objekte in  $S$  die Blätter und die inneren Knoten stellen die Vereinigungsmenge der jeweiligen Kindknoten dar. Die Länge der einzelnen Äste sind ein Maß für die Verwandtschaft, in der sich die Ähnlichkeit der Elemente zueinander widerspiegelt. Der verwendete agglomerativer Algorithmus beinhaltet mehrere aufeinander folgende Schritte:

- 1) Zuerst wird eine Distanzmatrix, abhängig von der Ähnlichkeit der Objekte, also der Gene und/oder Experimente untereinander berechnet. Im ersten Durchlauf besteht jeder  $n$  Cluster demnach nur aus einem Gen.
- 2) Es werden die zwei Cluster mit der kleinsten Distanz zueinander gesucht.
- 3) Diese werden zu einem neuen Cluster vereinigt und die Distanzen aller neu gebildeten Cluster untereinander neu berechnet.
- 4) Im Weiteren werden die Schritte 2 und 3 so lange wiederholt, bis nur noch ein Cluster übrig bleibt ( $n \rightarrow 1$ ).

Die Distanz zwischen zwei Clustern wurde mit der *average linkage*-Methode bestimmt. In diesem Verfahren wird die Distanz zwischen zwei Clustern aus der durchschnittlichen Distanz zwischen allen Objekten beider Cluster miteinander errechnet. Der Vorteil dieses Verfahrens ist, dass kein Vorwissen über die Gene oder spezielle Einstellungen notwendig sind. Von Nachteil ist, dass der Benutzer selbst die Cluster identifizieren und markieren muss. Da aber durch das HCL die Partition der

Datenmenge in separierte Datengruppen offensichtlich wird, die Ähnlichkeit der Elemente zueinander visualisiert werden kann und sich die Anzahl der Cluster bestimmen lässt, wird Hierarchisches Clustering meist vor, und zusammen mit, den Verfahren des *Partitioning* eingesetzt.

### 2.8.3.4 Self-Organizing-Maps (SOM-Clustering)



**Abb. 2.2.: Prinzip des SOM-Algorithmus.** Dargestellt anhand eines 3 x 2-Gitters. Die Referenzvektoren (1-6) bilden die Gitterknoten und nähern sich mit jedem Durchlauf des Algorithmus den Eingabevektoren, also den Datenpunkten (schwarze Punkte) an. Die Pfeile zeigen die mögliche Richtung der Annäherung der Referenzvektoren an die Eingabevektoren.

Das neuronale Netzwerkmodell der *Self-Organizing-Maps* ist eine populäre Methode der Clusteranalyse und dem *k-means*-Clustering sehr ähnlich. Beim SOM-Clustering wird der  $d$ -dimensionale Eingaberaum auf ein ein- oder zweidimensionales Gitter projiziert. Die Anzahl der Cluster muss vorgegeben werden und entspricht der Anzahl der Knoten des Gitters, die als Referenzvektoren bezeichnet werden. Das Neuronennetz wirkt als topologische Merkmalskarte. Wird nun ein Signal an diese Karte herangeführt, so werden nur diejenigen Gebiete der Karte erregt, die dem Signal ähnlich sind. Während des Algorithmus werden die Referenzvektoren in einem iterativen Prozess in Richtung der Eingabevektoren dirigiert. Dies wird solange wiederholt, bis die Cluster optimal getrennt sind.

Der größte Nachteil des Verfahrens liegt darin, dass die Anzahl der Cluster bereits vor dem Beginn der Berechnung abgeschätzt werden muss und die Verwandtschaft der Objekte eines Clusters untereinander nicht deutlich wird. Deshalb wurde die durch HCL bestimmte Anzahl der Regulons (25) für die *Partitioning*-Verfahren übernommen.

### 2.8.3.5 k-means Clustering

Beim k-means-Algorithmus ist eine gewünschte Anzahl k von Clustern und eine Funktion zur Bestimmung des Mittelpunktes eines Clusters bekannt. Der Algorithmus läuft folgendermaßen ab:

1. Die Initialisierung: Die Auswahl von k (muss vorgegeben werden) Clusterzentren.
2. Zuordnung: Jedes Objekt wird dem ihm am nächsten liegenden Clusterzentrum zugeordnet.
3. Neuberechnung: Die Clusterzentren werden für jedes Cluster neu berechnet.
4. Wiederholung: Ändert sich nun die Zuordnung der Objekte, wiederholt der Algorithmus von Schritt 2 an, ansonsten Abbruch.

### 2.8.4 Suche nach überrepräsentierten Sequenzmotiven in der Promoterregion der Gene eines Regulons

Die Software *MotifFinder* [TAIR (<http://www.arabidopsis.org/tools/bulk/motiffinder/index.jsp>)] erkennt und vergleicht die Häufigkeit von 6-mer „Wörtern“ 500 oder 1000 Basenpaare vor jedem Genlocus eines Datensatzes von Gensequenzen mit der Häufigkeit desselben Sequenzmotivs innerhalb der Promoterregion der Gene des gesamten Genoms. Nach der Eingabe einer Liste mit den AGI-Codes der Gene eines Regulons und des zu untersuchenden Sequenzbereiches, in diesem Fall 500 Basenpaare vor dem ATG-Translationinitiations-Codon, extrahiert *MotifFinder* automatisch die entsprechenden DNA-Sequenzen.

Die Ausgabetablelle gibt unter anderem die Häufigkeit des Sequenzmotivs im Datensatz

der Anfrage, die Häufigkeit des Sequenzmotivs in der vollständigen Genomsequenz und den p-Wert der Signifikanz an und listet die Gene auf, die das entsprechende Sequenzmotiv beinhalten.

Vorrangiges Ziel der Analyse waren die Regulons, die sich nach der Clusteranalyse, funktioneller Klassifizierung und Identifizierung der subplastidären Lokalisation als interessanteste Kandidaten erwiesen hatten. Nur die überrepräsentierten Sequenzmotive mit einem p-Wert  $< 10^{-5}$  (nach Hudson & Quail, 2003), also der höchsten Signifikanz, wurden sowohl gegen öffentliche Datenbanken für bereits beschriebene Promotermotive geblastet, als auch in einschlägigen Veröffentlichungen gesucht.

## 3 Ergebnisse

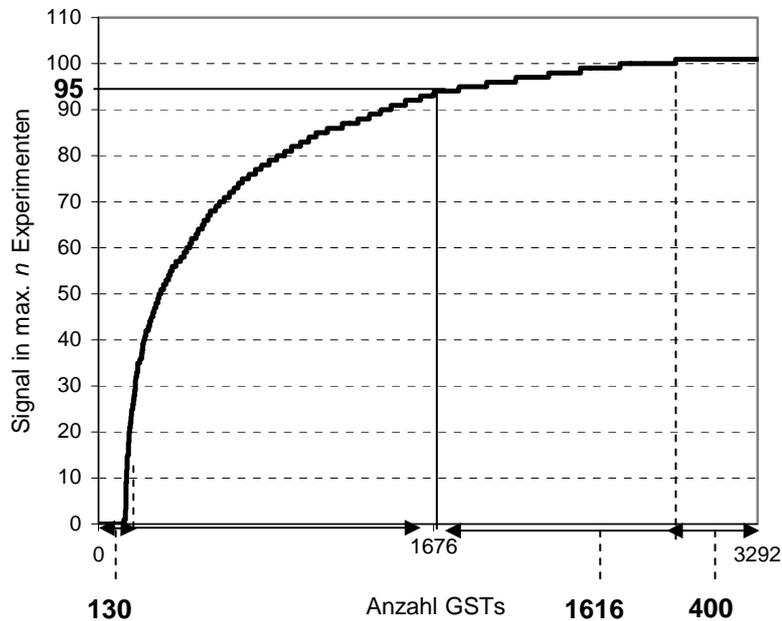
### 3.1 Experimenteller Aufbau zur Untersuchung des nukleären Transkriptoms des Chloroplasten

Zu Beginn des Projekts wurde die Verlässlichkeit der Expressionsanalyse durch den Makroarray stichprobenartig, sowohl durch QRT-PCR in unserem Labor [Kurth, Lindhout, Richly und Biehl (Kurth et al., 2002)] als auch durch Northern-Blotting in einer Kolaboration mit Paul Jarvis' Labor (Kubis et al., 2002) erfolgreich bestätigt.

Neben der speziellen Fragestellung hinter jeder einzeln betrachteten differentiellen Expressionsstudie [u.a. *prpl11* (Kurth et al., 2002); *ppi1* (Kubis et al., 2003); *mak3* (Pesaresi et al., 2003); *atpc* (Maiwald et al., 2003); (Bosco et al., 2004); *psad1*, *psad2* (Ihnatowicz et al., 2004); *toc159* (Kubis et al., 2004); *aap3* (Okumoto et al., 2004)] war bei der Auswahl der getesteten Bedingungen von besonderem Interesse, wie sich verschiedene Wachstumsbedingungen (mit unterschiedlichsten Stimuli für die Funktionen des Chloroplasten) auf die Expression des nukleäre Transkriptom des Chloroplasten auswirken. Andererseits sollten sowohl möglichst ähnliche (z.B.: Mutanten der Photosystem-Untereinheiten), als auch möglichst gegensätzliche Stimuli (z.B.: *gun* und *cue1*) (siehe Tab. 3.1), anhand des jeweils resultierenden Transkriptoms miteinander verglichen werden.

In zumindest 10 der 101 Expressionsprofilen wurden 3162 Gene, das entspricht mehr als 96 % aller 3292 Gene des durch den Makroarray repräsentierten Transkriptoms des Plastiden, sowohl unter dem Einfluss der speziellen Wachstums- als auch der Referenzbedingung, exprimiert (Abb. 3.1). Jedoch konnte nur für etwa die Hälfte aller Gene in mindestens 95 der 101 Bedingungen differentielle Expression nachgewiesen

werden. Von diesen sind etwa 400 Gene unter jedem der untersuchten Stimuli exprimiert, beispielsweise auch in den Plastiden des nichtphotosynthetischen Wurzelgewebes.



**Abb. 3.1.: Differentielle Genexpression in 101 Wachstumsbedingungen:** Nur für 130 der 3292 GSTs konnte in weniger als 10 der untersuchten 101 Wachstumsbedingungen ein Signal detektiert werden. 1616 Gene wurden in 95 % aller Stimuli transkribiert (bzw. 1676 Gene nur in weniger als 95 von 101 Bedingungen), davon wiederum nur etwa 400 Gene in allen 101 Bedingungen.

Daraus lässt sich schließen, dass auch die meisten der offenen Leserahmen, die bisher nicht als ESTs annotiert worden waren, zumindest unter speziellen Stimuli exprimiert werden. Gleichzeitig scheint ein Großteil der Gene, die das plastidäre Proteom kodieren, nicht fortwährend transkribiert zu werden. Das würde bedeuten, dass die Expression der Gene nicht konstitutiv ist und nicht nur quantitativ reguliert wird, sondern die Gene unter bestimmten Bedingungen selektiv ein- oder ausgeschaltet werden.

Bezeichnung Experiment	Genotyp oder Behandlung	Haupteffekt auf die Chloroplastenfunktionen
<b>Behandlungen</b>		
1200pm	Lichtadaptierte Pflanzen geerntet um 12.00 pm vs. dunkel adaptierte Pflanzen geerntet um 7.30 am nach 30min Licht	Zirkadiane Rhythmik, Photosynthese und andere Funktionen
1600pm	Lichtadaptierte Pflanzen geerntet um 16.00 pm vs. dunkel adaptierte Pflanzen geerntet um 7.30 am nach 30min Licht	Zirkadiane Rhythmik, Photosynthese und andere Funktionen
1830pm	Lichtadaptierte Pflanzen geerntet um 18.30 pm vs. dunkel adaptierte Pflanzen geerntet um 7.30 am nach 30min Licht	Zirkadiane Rhythmik, Photosynthese und andere Funktionen
1h_D	Lichtadaptiert nach 1h Dunkelheit vs. lichtadaptiert	Photosynthese und andere Funktionen
1week	1 Woche alte Pflanzen vs. 4 Wochen alte Pflanzen	Entwicklung des Chloroplasten
2h_rec	1h Licht nach 1h Dunkelheit vorher lichtadaptierter Pflanzen vs. lichtadaptiert	Photosynthese und andere Funktionen
2weeks	2 Wochen alte Pflanzen vs. 4 Wochen alte Pflanzen	Entwicklung des Chloroplasten
3weeks	3 Wochen alte Pflanzen vs. 4 Wochen alte Pflanzen	Entwicklung des Chloroplasten
blue200micE	48 Stunden bei 20 microE Blaulicht vs. 48 Stunden bei 20 microE Weißlicht <sup>a</sup>	Photosynthese und andere Funktionen
blue20micE	48 Stunden bei 200 microE Blaulicht vs. 48 Stunden bei 20 microE Weißlicht <sup>a</sup>	Photosynthese und andere Funktionen
HL_1h	1 h hohe Lichtintensität <sup>b</sup> vs. normale Lichtintensität <sup>c</sup>	Photosynthese und andere Funktionen
HLrec_2d	48 h Erholung <sup>b</sup> nach 1 h Lichtstress <sup>c</sup> vs. bevor Lichtstress <sup>b</sup>	Verschiedene Funktionen
HLrec_2h	2 h Erholung <sup>b</sup> nach 1 h Lichtstress <sup>c</sup> vs. bevor Lichtstress <sup>b</sup>	Verschiedene Funktionen
L_1h	1 h normale Lichtintensität <sup>b</sup> vs. über Nacht dunkel adaptiert	Photosynthese und andere Funktionen
L30min	30 min normale Lichtintensität <sup>b</sup> vs. dunkel adaptiert	Photosynthese und andere Funktionen
mCO2_2d	Niedrig-CO <sub>2</sub> -Stress: 2 d mit 0.003 % (v/v) CO <sub>2</sub> vs. normale CO <sub>2</sub> -Konzentrationen	Calvin Zyklus, Respiration
mCO2_4d	Niedrig-CO <sub>2</sub> -Stress: 4 d mit 0.003 % (v/v) CO <sub>2</sub> vs. normale CO <sub>2</sub> -Konzentrationen	Calvin Zyklus, Respiration
mec_wou	2 Tage nach mechanischer Verletzung vs. ohne Verletzung	Induzierte Seneszenz, Verwundungsstress
ML_4C_rec_24h_I	24 h mittlere Lichtintensität <sup>d</sup> bei 4 °C vs. 20 °C <sup>e</sup>	Photoinhibition der Photosysteme
ML_4C_rec_24h_II	24 h Erholung nach 24 h mittlerer Lichtintensität <sup>d</sup> bei 4°C vs. 20°C <sup>e</sup>	Photoinhibition der Photosysteme
ML_4C	24 h Erholung nach 24 h mittlerer Lichtintensität <sup>d</sup> bei 4°C vs. 4°C <sup>e</sup>	Photoinhibition der Photosysteme
pAsc	4 h H <sub>2</sub> O + 4 h 25 mM Ascorbat (pH 5,5) vs. 8 h H <sub>2</sub> O <sup>f</sup>	Verschiedene Funktionen
pBro	Behandlung mit dem Herbizid Bromoxynil vs. unbehandelt	Photosynthese und andere Funktionen
pBroL	Behandlung mit dem Herbizid Bromoxynil im Licht vs. Behandlung mit Bromoxynil in Dunkelheit	Photosynthese und andere Funktionen
pCK	Mit Cytokinin behandelte (2 h) Zellkultur vs. unbehandelte Zellkultur <sup>g</sup>	Verschiedene Funktionen
pCO2_10d	Hoch-CO <sub>2</sub> -Stress: 10 d mit 1 % (v/v) CO <sub>2</sub> vs. normale CO <sub>2</sub> -Konzentrationen	Calvin Zyklus, Respiration
pCO2_1d	Hoch-CO <sub>2</sub> -Stress: 1 d mit 1 % (v/v) CO <sub>2</sub> vs. normale CO <sub>2</sub> -Konzentrationen	Calvin Zyklus, Respiration
pCO2_6d	Hoch-CO <sub>2</sub> -Stress: 6 d mit 1 % (v/v) CO <sub>2</sub> vs. normale CO <sub>2</sub> -Konzentrationen	Calvin Zyklus, Respiration
pFe	Eisen-Stress: Blätter besprüht mit einer Eisenlösung (0.06% Fe <sup>2+</sup> -Chelat-Lösung) vs. ohne Behandlung	Oxidativer Stress
pmildew	2 Tage nach Mehltau-Infektion vs. ohne Infektion	Induzierte Seneszenz, Verwundungsstress
pPar	Behandlung mit Herbizid Paraquat vs. unbehandelt	Photosynthese und andere Funktionen
pParL	Behandlung mit Herbizid Paraquat im Licht vs. Behandlung mit Paraquat in Dunkelheit	Photosynthese und andere Funktionen
pPro_2d	48 h 100mM Prolin vs. ohne Behandlung <sup>h</sup>	Erhöhte (toxische) Konzentrationen von zellulärem Prolin
pPro_30min	30' mit 100 mM Prolin-Behandlung vs. H <sub>2</sub> O-Behandlung <sup>h</sup>	Erhöhte (toxische) Konzentrationen von zellulärem Prolin
pPro_3h	3 h mit 100 mM Prolin-Behandlung vs. H <sub>2</sub> O-Behandlung <sup>h</sup>	Erhöhte (toxische) Konzentrationen von zellulärem Prolin
pPro_6h	6 h mit 100 mM Prolin-Behandlung vs. H <sub>2</sub> O-Behandlung <sup>h</sup>	Erhöhte (toxische) Konzentrationen von zellulärem Prolin
PSI_DCMU	Wachstum unter PSI-spezifischem Licht und behandelt mit dem Elektronentransport-Inhibitor DCMU vs. PSII-Licht <sup>i</sup>	Photosynthese, Redoxstatus des Plastiden
PSII_DCMU	Wachstum unter PSII-spezifischem Licht und behandelt mit dem Elektronentransport-Inhibitor DCMU vs. PSI- Licht <sup>i</sup>	Photosynthese, Redoxstatus des Plastiden
PSII	Wachstum unter PSII-spezifischem Licht und behandelt mit dem Elektronentransport-Inhibitor DCMU vs. PSII- Licht <sup>i</sup>	Photosynthese, Redoxstatus des Plastiden
pStau	4 h H <sub>2</sub> O + 4 h 25 mM Staurosporin (pH 5,5) vs. 8 h H <sub>2</sub> O <sup>f</sup>	Photosynthese und andere Funktionen

Fortsetzung Tabelle 3.1. →

pStaupAsc	4 h 2µM Staurosporin + 4 h 25 mM Ascorbat (pH 5,5) vs. 4 h 25 mM Ascorbat <sup>f</sup>	Photosynthese und andere Funktionen
pSuc	MS-Medium mit 2% Saccharose vs. ohne Behandlung	Photosynthese und andere Funktionen
pSuc_30min	30 min Saccharose-Behandlung <sup>k</sup> vs. H <sub>2</sub> O Behandlung <sup>h</sup>	Photosynthese und andere Funktionen
pSuc_3h	3 h Saccharose-Behandlung <sup>k</sup> vs. H <sub>2</sub> O Behandlung <sup>h</sup>	Photosynthese und andere Funktionen
pSuc_6h	6 h Saccharose-Behandlung <sup>k</sup> vs. H <sub>2</sub> O Behandlung <sup>h</sup>	Photosynthese und andere Funktionen
red200micE	48 h bei 20 microE Rotlicht vs. 48 Stunden bei 20 microE Weißlicht <sup>a</sup>	Photosynthese und andere Funktionen
red20micE	48 h bei 200 microE Rotlicht vs. 48 Stunden bei 20 microE Weißlicht <sup>a</sup>	Photosynthese und andere Funktionen
root	Wurzel vs. Rosettenblätter	Unterschiedlich Pflanzengewebe
stem	Stängel vs. Rosettenblätter	Unterschiedlich Pflanzengewebe

**Mutanten**

aap3	<i>aap3</i> vs. WT <sup>h</sup>	Fehlt AAP3-Protein der Wurzel: beeinträchtigt Aminosäuren-Import im Wurzelgewebe
aap6	<i>aap6</i> vs. WT <sup>h</sup>	Fehlt AAP6-Protein des Xylems: beeinträchtigte Aufnahme von Aminosäuren im Xylem
atc6	<i>atc6</i> <sup>++</sup> vs. WT	Fehlt Cytochrom c6 Protein atc6 (Elektronentransport)
atpc	<i>atpc1-1</i> <sup>m</sup> vs. WT (24 h Licht)	Fehlt ATP Synthase Untereinheit: Photophosphorylierung
atpd	<i>atpd-1</i> <sup>++</sup> Maiwald et al., 2003) vs. WT (Tag-Nacht-Rhythmus)	Fehlt ATP Synthase Untereinheit: Photophosphorylierung
cue1	<i>cue1-1</i> <sup>a</sup> (Streatfield et al., 1999) vs. WT	Fehlt PEP Translokator: Austausch von Metaboliten, Plastid-Signaling
fluD	<i>flu</i> (Meskauskiene et al., 2001), dunkel adaptiert vs. WT, dunkel adaptiert	Fehlt FLU, Ein Regulator der Chlorophyll Biosynthese
fluL	<i>flu</i> (Meskauskiene et al., 2001) Licht-adaptiert vs. WT, dunkel adaptiert	Fehlt FLU, Ein Regulator der Chlorophyll Biosynthese
gun1	<i>gun1-1</i> (Susek et al., 1993) vs. WT	Unbekannt, Plastid-Signaling
gun5	<i>gun5</i> (Susek et al., 1993; Mochizuki et al., 2001) vs. WT	Fehlt die ChlH-Untereinheit der Mg-Chelatase: Plastid-Signaling
hcf145	<i>hcf145</i> <sup>m</sup> vs. WT	Aggregation des Photosystems I gestört: Chloroplastenen Biogenese
kn09	<i>kn09</i> <sup>h</sup> vs. WT	Blockiert im mitochondrialen Prolin-Katabolismus
kn09pPro_2d	Hyersensitive Prolin-utante <i>kn09</i> <sup>h</sup> mit 100 mM Prolin behandelt vs. unbehandelter WT	Erhöhte (toxische) Konzentrationen von zellulärem Prolin
laf6_D	Dunkeladaptierte Arabidopsis <i>laf6</i> -Mutante vs. dunkel adaptierter WT (landsberg)	Fehlt atABC1-Transporter, Signaltransduktion der Phytochrome, Licht-regulierten Genexpression
laf6_L_1h	1h im Licht adaptierte Arabidopsis <i>laf6</i> -Mutante vs. dunkel adaptierter WT (landsberg)	Fehlt atABC1-Transporter, Signaltransduktion der Phytochrome, Licht-regulierten Genexpression
mak3	<i>atmak3-1</i> <sup>++</sup> vs. WT	Defekt in der cytoplasmatischen N-Acetyltransferase: verschiedene Funktionen
pam46	<i>pam46</i> <sup>++</sup> vs. WT	Defekt eines unbekannt Proteins, alternierende Photosynthese
pam48	<i>pam48</i> <sup>++</sup> vs. WT	Defekt eines unbekannt Proteins, alternierende Photosynthese
petc	<i>petC</i> <sup>++</sup> (Maiwald et al., 2003) vs. WT	Fehlt Rieske-Protein petC (Elektronentransport)
pete1	<i>petE1</i> <sup>++</sup> auf MS-Medium gewachsen vs. WT	Fehlt Plastocyanin-Protein PETE1 (Elektronentransport)
pete1_pete2	<i>petE1_pete2</i> <sup>++</sup> auf MS-Medium gewachsen vs. WT	Fehlen die Plastocyanin-Proteine PETE1 und petE2 (Elektronentransport)
Pete1_soil	<i>petE1</i> <sup>++</sup> auf Erde gewachsen vs. WT	Fehlt Plastocyanin-Protein PETE1 (Elektronentransport)
pete2	<i>petE2</i> <sup>++</sup> auf MS-Medium gewachsen vs. WT	Fehlt Plastocyanin-Protein PETE2 (Elektronentransport)
Pete2_soil	<i>petE2</i> <sup>++</sup> auf Erde gewachsen vs. WT	Fehlt Plastocyanin-Protein PETE2 (Elektronentransport)
ppi1	<i>ppi1</i> <sup>P</sup> (Jarvis et al., 1998) vs. WT	Defekt im Import von Chloroplasten-Vorläuferproteinen

Fortsetzung Tabelle 3.1. →

ppi2	<i>ppi2</i> <sup>p</sup> vs. WT	Defekt im Import von Chloroplasten-Vorläuferproteinen
prpl11	<i>prpl11-1</i> <sup>++</sup> (Pesaresi et al., 2001) vs. WT	Fehlt das plastidäre Ribosomen-Protein PRPL11: plastidäre Proteinsynthese
prs1	Prolyl-tRNA-Synthetase Mutante <sup>++</sup> vs. WT	Fehlt Prolyl-tRNA-Synthetase
prs1_D	Prolyl-tRNA-Synthetase Mutante <sup>++</sup> , dunkel adaptiert vs. WT, dunkel adaptiert	Fehlt Prolyl-tRNA-Synthetase
prs2	Prolyl-tRNA-Synthetase Mutante <sup>++</sup> vs. WT	Fehlt Prolyl-tRNA-Synthetase
prs2_D	Prolyl-tRNA-Synthetase Mutante <sup>++</sup> , dunkel adaptiert vs. WT, dunkel adaptiert	Fehlt Prolyl-tRNA-Synthetase
psad1	<i>psad1-1</i> <sup>++</sup> vs. WT	Fehlt Photosystem I Protein PSI-D1: Photosynthese
psae1	<i>psae1-1</i> <sup>++</sup> (Varotto et al., 2000) vs. WT	Fehlt Photosystem I Protein PSI-E1: Photosynthese
psae2	<i>psae2</i> <sup>++</sup> vs. WT	Fehlt Photosystem I Protein PSI-E2: Photosynthese
psaf	<i>psaf</i> <sup>e</sup> vs. WT	Weniger als 20 % des Photosystem I Proteins PSI-F nachweisbar (Photosynthese)
psag	<i>psag</i> <sup>++</sup> vs. WT	Fehlt Photosystem I Protein PSI-G: Photosynthese
psah2	<i>psah2</i> <sup>++</sup> vs. WT	Fehlt Photosystem I Protein PSI-H2: Photosynthese
psah2k	<i>psak_pсах2</i> <sup>++</sup> vs. WT	Fehlt Photosystem I Proteins PSI-K und PSI-H2: Photosynthese
psak	<i>psak</i> <sup>++</sup> vs. WT	Fehlt Photosystem I Protein PSI-K: Photosynthese
psan	<i>psan-1</i> <sup>++</sup> vs. WT	Fehlt Photosystem I Protein PSI-N: Photosynthese
psao	<i>psao-1</i> <sup>++</sup> vs. WT	Fehlt Photosystem I Protein PSI-O: Photosynthese
psbs	<i>Psbs</i> <sup>++</sup> vs. WT	Fehlt Photosystem I Protein PSII-S: Photosynthese; Photoinhibition
sut2	<i>sut2</i> <sup>h</sup> vs. WT	Fehlt Saccharose-Sensor/Transporter SUT2; Kohlenhydrattransport
tata	<i>tataA</i> <sup>p</sup> vs. WT	Beeinträchtiger Protein-Import über die innere Chloroplastenmembran
tic40	<i>tic40</i> <sup>p</sup> vs. WT	Beeinträchtiger Protein-Import über die innere Chloroplastenmembran
toc120	<i>toc120</i> <sup>p</sup> vs. WT	Beeinträchtiger Protein-Import über die äußere Chloroplastenmembran
toc132	<i>toc132</i> <sup>p</sup> vs. WT	Beeinträchtiger Protein-Import über die äußere Chloroplastenmembran
toc160	<i>toc160</i> <sup>p</sup> vs. WT	Beeinträchtiger Protein-Import über die äußere Chloroplastenmembran
toc34	<i>toc34</i> <sup>p</sup> vs. WT	Beeinträchtiger Protein-Import über die äußere Chloroplastenmembran
toc75	<i>toc75</i> <sup>p</sup> vs. WT	Beeinträchtiger Protein-Import über die äußere Chloroplastenmembran
toc89	<i>toc89</i> <sup>p</sup> vs. WT	Beeinträchtiger Protein-Import über die äußere Chloroplastenmembran
tpt1	<i>tpt1</i> <sup>q</sup> vs. WT	Triosephosphate (TP)-spezifischer Transport ist um 95 % reduziert

**Tab. 3.1.: Auflistung der untersuchten Wachstumsbedingungen und Genotypen.** Jedes Expressionsprofil (siehe „Bezeichnung Experiment“) ist der differentielle Vergleich (siehe „Genotyp oder Behandlung“) der Expression des Transkriptoms des Chloroplasten in einer Wachstumsbedingung zu einer Referenzbedingung oder eines genomischen Status zum Wildtyp. Der vermutete Haupteffekt auf die Funktionen des Chloroplasten ist beschrieben.<sup>a</sup> RNA von J. Chory's Labor; <sup>b</sup> 2000  $\mu\text{mol Photonen m}^{-2}\text{s}^{-1}$ ; <sup>c</sup> 80  $\mu\text{mol Photonen m}^{-2}\text{s}^{-1}$ ; <sup>d</sup> 100  $\mu\text{mol Photonen m}^{-2}\text{s}^{-1}$ ; <sup>e</sup> RNA von A. Haldrup; <sup>f</sup> RNA von M. Baier's Labor; <sup>g</sup> RNA von T.Schmülling; <sup>h</sup> RNA von W.B. Frommer's Labor; <sup>i</sup> RNA von T. Pfanschmidt; <sup>k</sup> 60 mM Saccharose in MS-Medium; <sup>m</sup> RNA von J. Meurer's Labor; <sup>p</sup> RNA von P. Jarvis' Labor; <sup>q</sup> RNA von U.I. Flügges' Labor; <sup>++</sup> in unserem Labor isolierte Mutante. WT, Wildtyp.

### 3.2 Analyse des differentiell exprimierten nukleären Transkriptoms des Chloroplasten

Die Expression der Gene des nukleär kodierten Transkriptoms des Chloroplasten wurde in 101 differentiellen Vergleichen von jeweils einer Wachstumsbedingung oder eines genetischen Status zu einer entsprechenden Kontrollbedingung analysiert. Die untersuchten, unterschiedlichen Stimuli reichen von Null-Mutationen, die sich auf die Photosynthese, den Proteinimport in die Plastiden, die Kommunikation zwischen Organelle und Zellkern oder die Transkriptions- und Translationsmaschinerie des Chloroplasten auswirken über Wachstumsbedingungen, in denen Wildtyppflanzen unterschiedlichen Lichtqualitäten- oder Quantitäten, Veränderungen des atmosphärischen CO<sub>2</sub>-Gehalts, verschiedenen Herbiziden oder Stressbedingungen ausgesetzt wurden bis zum Vergleich verschiedener Pflanzengewebe und Entwicklungsstufen des Chloroplasten (Tab. 3.1).

Die Abbildung 3.2 zeigt das signifikant differentiell regulierte Transkriptom von *A. thaliana* über 101 getestete Wachstumsbedingungen. Mit Hilfe statistischer Methoden wurde die Anzahl signifikant differentiell exprimierter Gene bestimmt und zwischen Induktion und Repression der Genexpression in Abhängigkeit des gegebenen Stimulus unterschieden. Zur Bestimmung der signifikant differentiell exprimierten Gene wurde ein proportionales statistisches Modell mit gepoolter Berechnung des Standardfehlers, ein *z*-Test, ein nominales Alpha von 0.05 zur Berechnung des Signifikanzniveaus und die *False Discovery Rate* zur Korrektur der Rate der Falsch-Positiven gewählt.

Die absolute Anzahl der signifikant differentiell regulierten Gene variiert sehr stark innerhalb der 101 Expressionsprofile. Während sich die Behandlung von *A. thaliana* mit Bromoxynil (57 Gene) oder Cytokinin (172 Gene) nur auf weniger als 2 % bzw. 5 % des untersuchten nukleären plastidären Transkriptoms auswirkt, führt die Behandlung mit Saccharose (30 min) zu signifikanten Veränderungen der Expression von annähernden 80 % (2603 von 3292 Genen) aller Gene.

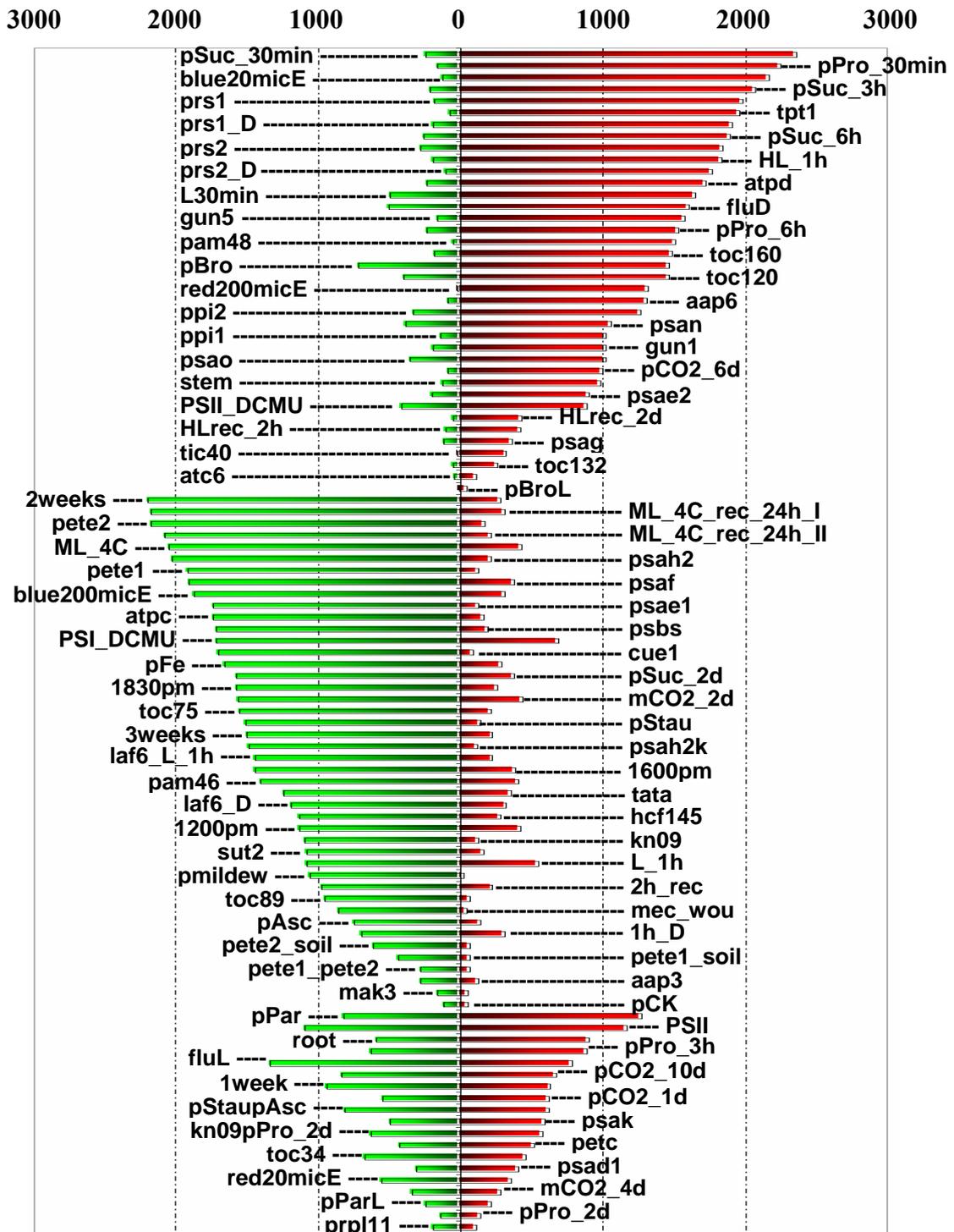
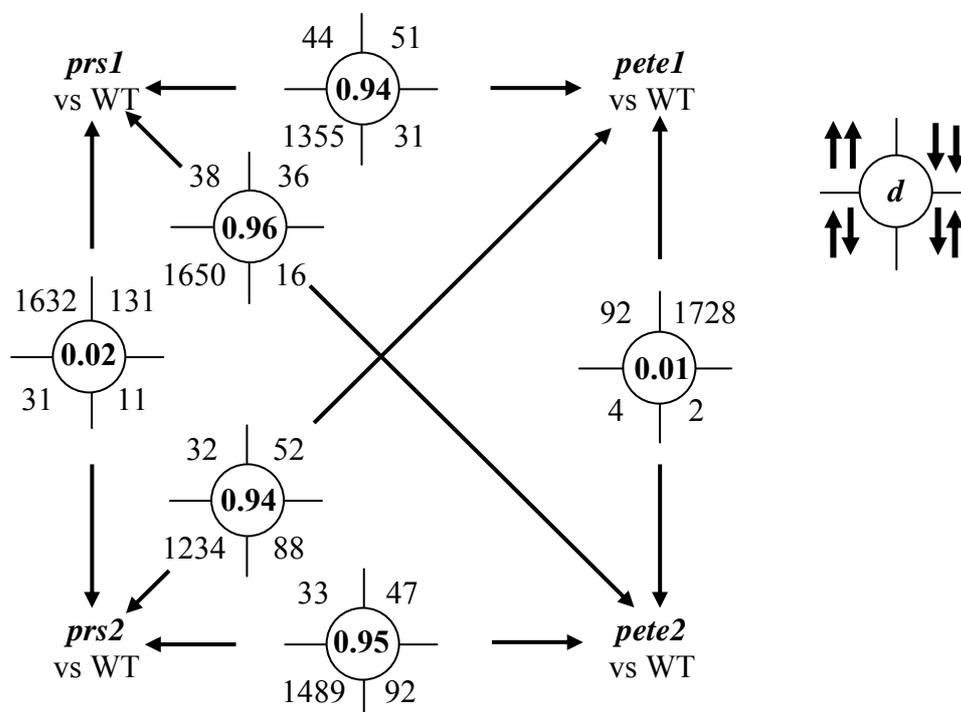


Abb.3.2.: Signifikant differentielle Expression des nukleären plastidären Transkriptoms unter 101 verschiedenen Bedingungen. Zwei Hauptklassen der transkriptionellen Antwort sind erkennen: in 38 Expressionsprofilen sind mehr als 75 % der signifikant differentiell exprimierten Gene hoch-, in 44 Expressionsprofilen dagegen herunterreguliert. Nur 19 Expressionsprofile zeigen ein +/- ausgewogenes Verhältnis von hoch- und herunterregulierten Genen. Die Säulen stellen die differentiell exprimierten Gene dar (grün für herunter-, rot für hochregulierte Gene). Die Abkürzungen ergeben sich aus Tab. 3.1.

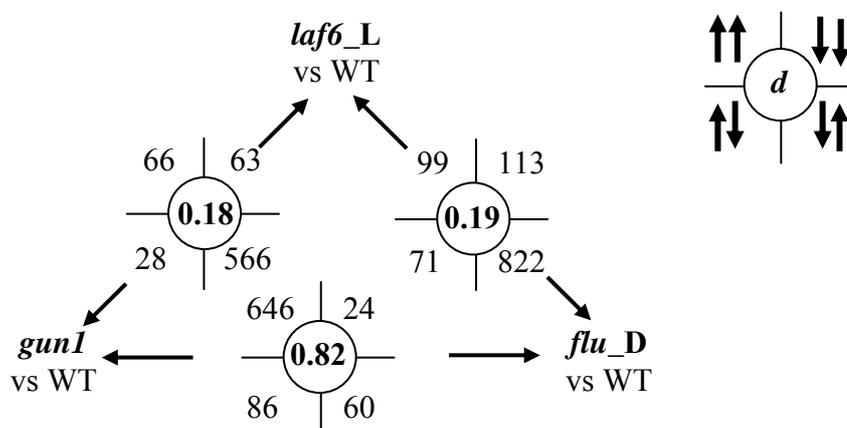
Erstaunlicherweise zeigt die Expression des nukleären Transkriptom des Chloroplasten in 82 von 101 differentiellen Expressionsanalysen einen eindeutigen Trend. Die Expression von jeweils mindestens 75 % aller signifikant differentiell exprimierten Gene ist entweder induziert (in 38 der Expressionsprofile) oder reprimiert (in 44 der Expressionsprofile). Nur 19 Expressionsprofile zeigen ein mehr oder minder ausgewogenes Verhältnis von hoch- zu herunterregulierten Genen. Das impliziert, dass das nukleäre Transkriptom des Chloroplasten tendenziell als Ganzes nach einem „Entweder- (hochreguliert) Oder- (herunterreguliert)“ Prinzip reguliert ist. Diese Vermutung wird im Kapitel 4.2.1 (Seite 84) ausführlich diskutiert.



**Abb. 3.3.:** Quervergleich der signifikant differentiell regulierten Gene des nukleären Transkriptom des Chloroplasten von jeweils zwei allelischen Mutanten. Die jeweils gemeinsamen signifikant differentiell regulierten Gene der allelischen Mutanten *prs1* und *prs2* sowie *pete1* und *pete2* wurden paarweise miteinander verglichen. Für jeden Vergleich ist die Anzahl der gemeinsamen identischen Gene gleichen (hoch/hoch oder runter/runter) oder entgegengesetzten Trends der Expression (hoch/runter oder runter/hoch) angegeben. Aus dem Quotienten der gleich regulierten Gene zur Gesamtanzahl der gemeinsamen Gene errechnet sich ein Distanzmaß  $d$  zweier Expressionsprofile. Je größer das Distanzmaß  $d$  (max. 1), desto unterschiedlicher sind die Datensätze.

Die Auswirkungen der verschiedenen Stimuli auf die Expression des Transkriptoms sind größtenteils nur sehr schwer vorherzusagen. Zumindest aber ist zu erwarten, dass sehr ähnliche Stimuli ein sehr ähnliches Expressionsmuster der Gene zeigen. Unter dieser Prämisse wurde die Reproduzierbarkeit und die Verlässlichkeit der Ergebnisse der differentiellen Expressionsanalyse durch den Vergleich einiger der Expressionsprofile untereinander bestätigt. Schon bei der Erstellung des signifikant differentiell regulierten Transkriptoms ist zu erkennen, dass allelische Mutationen (z.B.: *prs1* und *prs2*, *prs1\_D* und *prs2\_D*, *pete1* und *pete2*, *pete1\_soil* und *pete2\_soil*) eine sehr ähnliche transkriptionelle Antwort zur Folge haben (Abb. 3.2).

Mehr noch zeigt sich im direkten Quervergleich der Expressionsprofile, dass nicht nur die absolute Anzahl und der vorherrschende Trend (Induktion oder Repression) der differentiell regulierten Gene übereinstimmen, sondern auch dieselben Gene überwiegend gleich reguliert sind (Abb. 3.3). Aus dem paarweise durchgeführten Vergleich der vier Expressionsprofile entsteht eine jeweils etwa 1400 bis 1850 Genen große Schnittmenge der gemeinsamen, signifikant differentiell regulierten Gene. Ein identischer Stimulus durch die allelischen Mutationen führt zu einer hohen Übereinstimmung (und einer Distanz  $d \rightarrow 0$ ) der Transkriptome.

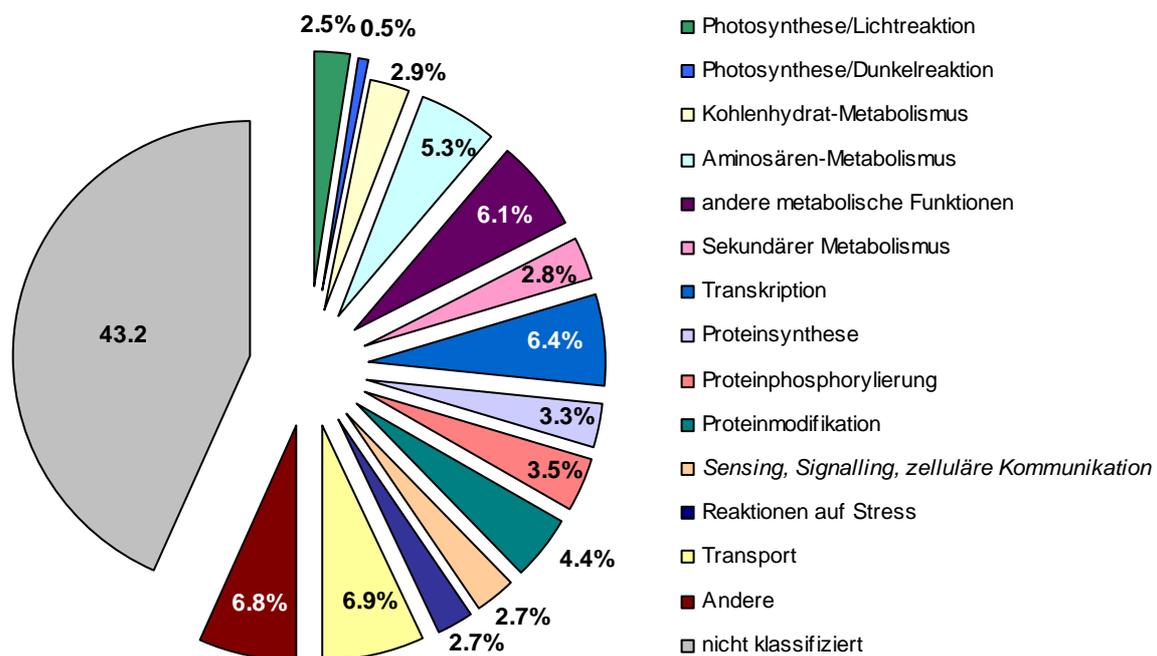


**Abb. 3.4.: Quervergleich der signifikant differentiell regulierten Gene von verschiedenen Mutanten mit Einfluss auf den Tetrapyrrolmetabolismus.** Die jeweils gemeinsamen signifikant differentiell regulierten Gene der Mutanten des Tetrapyrrolmetabolismus *gun1*, *laf6* und *flu\_D* wurden paarweise miteinander verglichen. Für die Erklärung der Abbildung siehe Abb. 3.3.

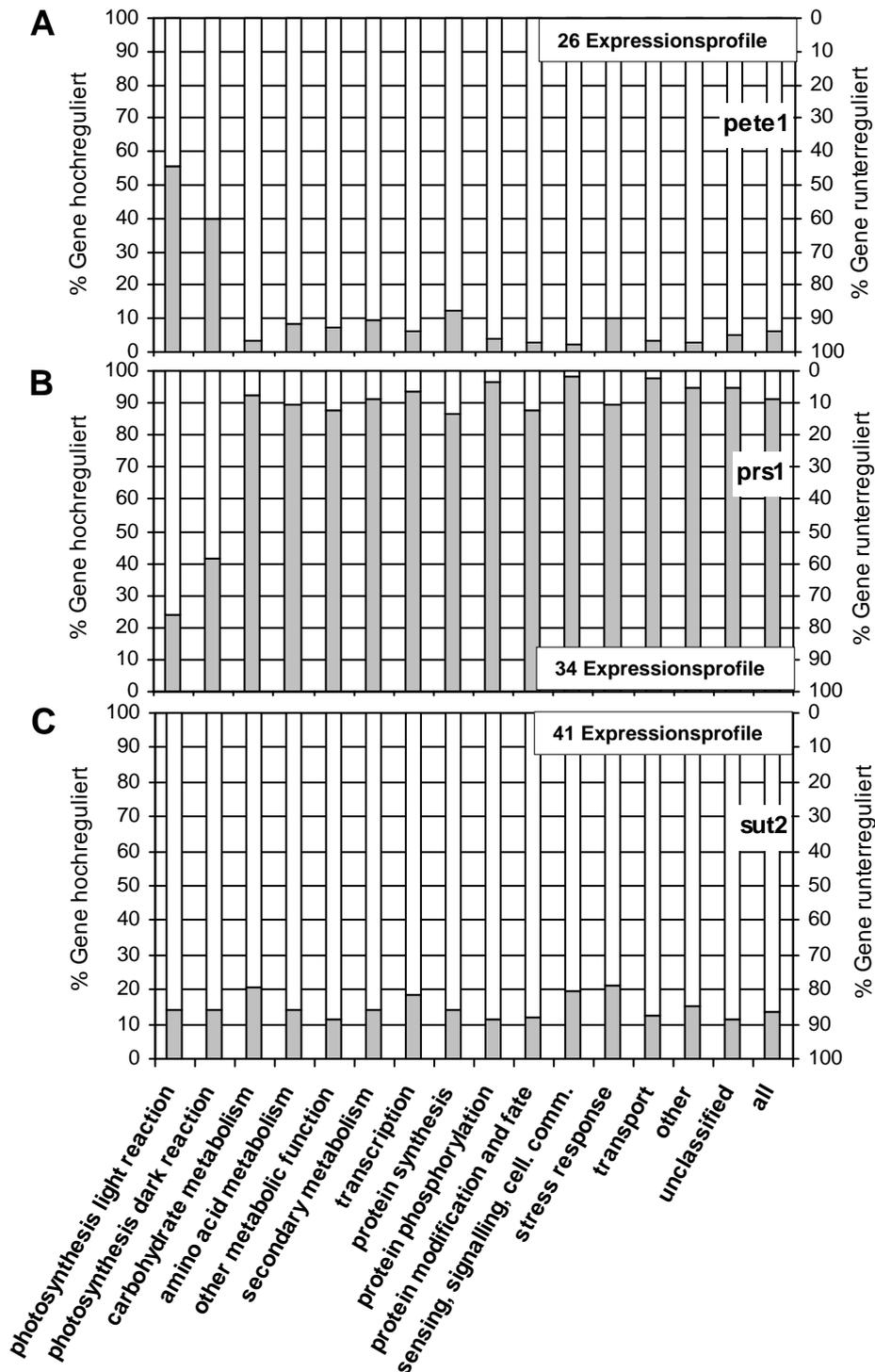
Ein Distanzmaß  $d$  von annähernd 1, so wie zwischen den Mutationen des Plastocyanin auf der einen und der Prolyl-tRNA-Synthetase auf der anderen Seite, zeigt dagegen die jeweils fast vollständig entgegengesetzte Expression der gemeinsamen Gene an.

Das gleiche Phänomen, die hohe Übereinstimmung der Transkriptome von ähnlichen Stimuli und umgekehrt, die großen Distanzen zwischen den Expressionsprofilen von entgegen wirkenden Stimuli, ist auch im Vergleich der Expressionsprofile der Mutanten mit beeinträchtigtem Tetrapyrrolmetabolismus [*gun1*, *laf6* und *flu\_D* (Abb. 3.4)] und der *Signalling*-Mutanten [*gun1*, *gun5* und *cue1* (Daten nicht gezeigt)] untereinander deutlich zu erkennen. Diese Beobachtungen konnten auch im Kapitel 3.3.1 (Seite 58) anhand der Ergebnisse der Clusteranalyse der 101 Transkriptome bestätigt werden.

### 3.2.1 Funktionelle Klassifizierung des signifikant differentiell exprimierten Transkriptoms des Chloroplasten



**Abb. 3.5.: Funktionelle Klassifizierung des nukleären Transkriptoms des Chloroplasten.** Die Gene des nukleären Transkriptoms des Chloroplasten, das durch den 3292-GST-Makroarray repräsentiert ist, wurden in 15 funktionelle Klassen eingeteilt. Die Zahlen innerhalb der Diagrammsegmente geben den prozentualen Anteil einer funktionellen Klasse vom gesamten nukleären Transkriptom wieder.



**Abb. 3.6.: Drei Klassen der Antwort des funktionell klassifizierten und differenziell exprimierten Transkriptoms des Chloroplasten.** Anhand der Expressionsprofile *pete1*, *prs1* und *sut2* sind die drei Hauptklassen der Antwort des funktionell klassifizierten und differenziell exprimierten Transkriptoms des Chloroplasten dargestellt. In 26 Expressionsprofilen sind die Gene der Photosynthese gegen den Trend des gesamten Transkriptoms hochreguliert und in 34 Expressionsprofilen gegen den Trend herunterreguliert. Nur unter 41 der 101 Stimuli zeigt das gesamte Transkriptom eine +/- einheitliche Expression.

Zunächst wurde die Gesamtheit der korrespondierenden Gene bzw. Genprodukte des 3292-GST-Makroarrays anhand ihrer Annotationen in den öffentlich zugängigen Datenbanken *AraCyc* und *KEGG PATHWAY* jeweils einer von 15 funktionellen Klassen zugeordnet. Die funktionelle Klassifizierung des annähernd vollständigen nukleären Transkriptoms des Chloroplasten zeigt, dass die Funktion von mehr als 40 % der Gene bisher nicht bekannt ist (Abb. 3.5). Die nächst größeren, funktionell definierten Klassen beinhalten Genprodukte, die bei den Vorgängen von Transport, Transkription und dem Metabolismus der Aminosäuren der Zelle eine Rolle spielen. Gene des Kerngenoms, die Proteine der Licht- und Dunkelreaktion der Photosynthese kodieren, machen hingegen insgesamt nur etwa 3 % des bekannten Transkriptoms des Chloroplasten aus.

Jedes einzelne der 101 Experimente wurde getrennt betrachtet und das jeweils (unterschiedliche) signifikant differentiell regulierte Transkriptom zur Verdeutlichung des Einflusses eines Stimulus auf die Genexpression funktionell klassifiziert. Generell sind die meisten funktionellen Genklassen bezüglich ihrer Expression jeweils mit dem Trend des gesamten nukleären Transkriptoms reguliert (Abb. 3.6, C).

Die Ausnahme stellen die Gene der Licht- und Dunkelreaktion der Photosynthese dar (Abb. 3.6, A und B). In insgesamt 60 der 101 differentiellen Expressionsanalysen wurden die Gene, deren Genprodukte in die Vorgänge der Photosynthese involviert sind, gegen den Trend des gesamten Transkriptoms entweder hoch- oder herunterreguliert.

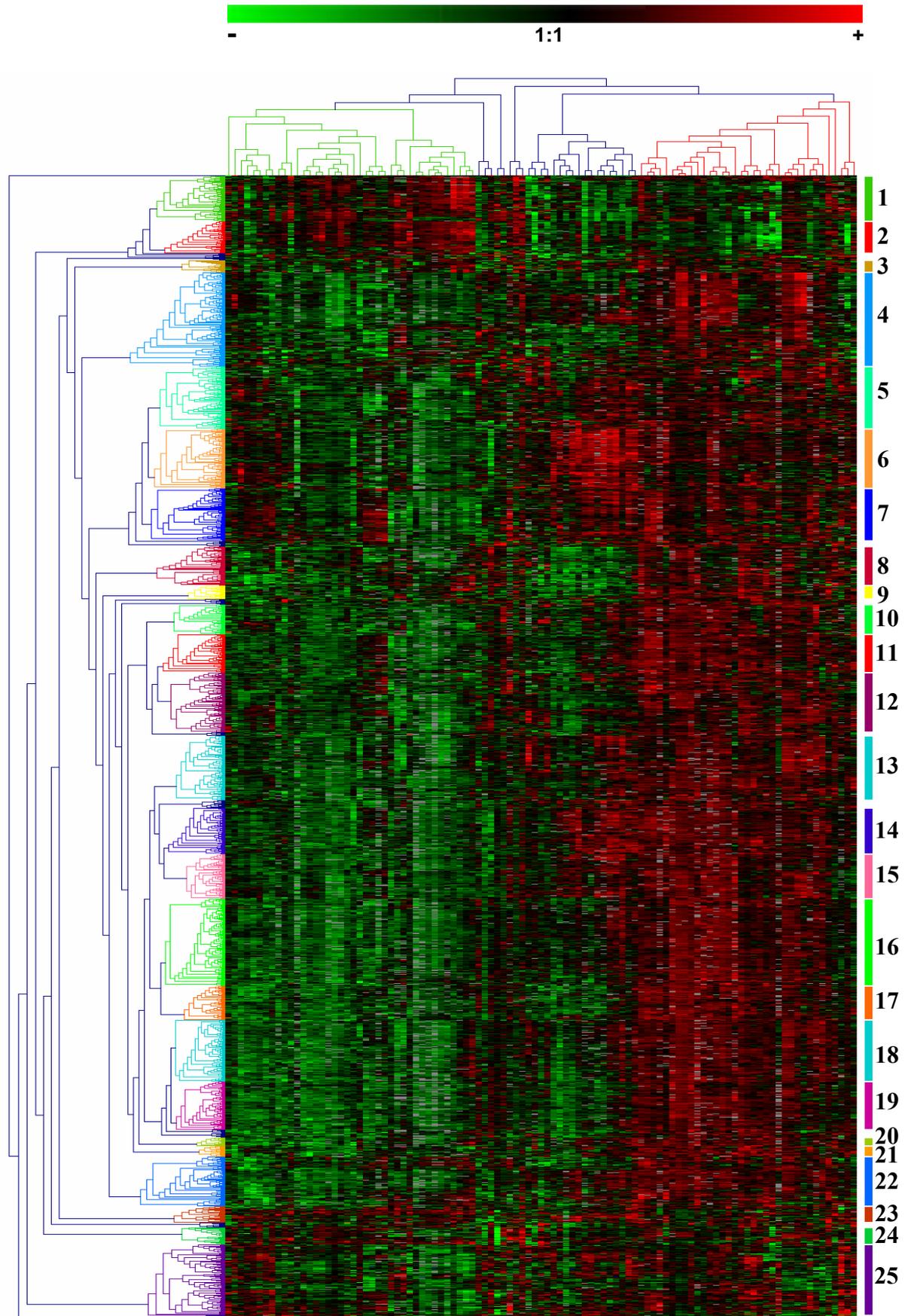
Um die Transkriptionsrate der Gene in den 101 Expressionsprofilen und die Expressionsprofile untereinander detaillierter vergleichen zu können, wurde eine Clusteranalyse aller Expressionsdaten durchgeführt.

### 3.3 Untersuchung des nukleären Transkriptom des Plastiden durch Clusteranalyse

Im Gegensatz zu den oben beschriebenen qualitativen Analysen, wird im Clustering nicht nur zwischen Induktion und Repression der Genexpression unterschieden, sondern auch die Expressionsrate mit einbezogen (quantitativer Ansatz). Für den Vergleich der 101 Expressionsprofile und für die Analyse des Expressionsverhaltens der Gene wurden drei verschiedene Methoden der Clusteranalyse eingesetzt. Hierarchisches Clustering, SOM- und k-means-Clustering benutzen unterschiedliche Algorithmen und wurden nebeneinander eingesetzt, um so die Ergebnisse der drei Verfahren gegeneinander abzusichern.

Auf die Expressionsdaten wurde zuerst Hierarchisches Clustering angewandt, um die Anzahl der zu erwartenden Cluster für die Methoden des *Partitioning* zu bestimmen (Abb. 3.7). Als Distanzmaß wurde die Euklidische Distanz verwendet. Um verlässliche Ergebnisse zu erzielen, musste die Anzahl fehlender Messwerte in der Distanzmatrix reduziert werden. Folglich wurden die Expressionsdaten einem Filter unterzogen, so dass nur die 1616 Gene, die in mindestens 95 von 101 Vergleichen von Proben- und Referenzbedingung exprimiert worden waren, für das Clustering berücksichtigt wurden. Das Hierarchische Clustering bietet gegenüber den *Partitioning* den Vorteil, dass die zu erwartende Anzahl an Regulons nicht schon vorab festgelegt werden muss, die Zahl der Regulons aber nachher für das *Partitioning* genutzt werden kann (Sturn, 2000). In der Abbildung 3.7 wird deutlich, dass sich aus dem Clustering der Expressionsdaten der Gene unterschiedlich große Gruppen ähnlich regulierter Gene ergeben. Die Distanz der Gene untereinander wird durch ein Kladogramm wiedergegeben und lässt in diesem Fall die Unterscheidung in 25 unterschiedlich exprimierte Regulons zu.

Hierarchisches Clustern ermöglicht es, neben den Genen auch die Verwandtschaft der Experimente darzustellen. Ist der Datensatz der Experimente ausreichend groß, kann ein Stimulus durch die Ähnlichkeit bzw. Unähnlichkeit der beobachteten transkriptionellen Antwort zu den nächstgelegenen bzw. entferntesten Nachbarn im Kladogramm charakterisiert werden (Abb. 3.8).

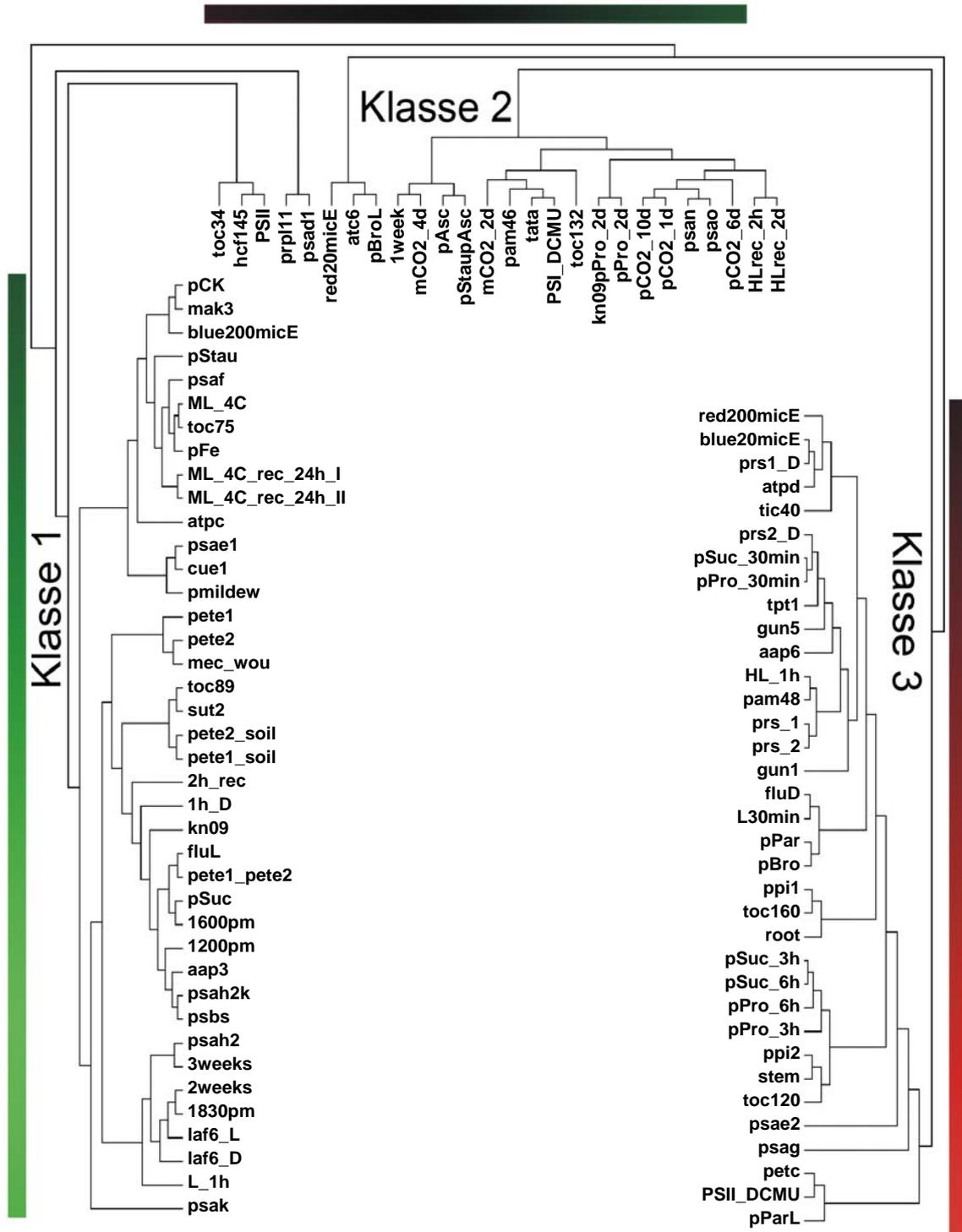


**Abb. 3.7: Hierarchisches Clustering der Expressionsdaten von 1616 Genen über 101 Bedingungen.** Dargestellt ist das *average linkage* Clustering von 101 Probenbedingungen und 1616 Genen, die in mindestens 95 von 101 Bedingungen differentiell exprimiert worden waren. Dem jeweiligem differentiellen Expressionswert jedes Gens ist in einer durchgehenden Zeile, unterteilt in die 101 Spalten der Wachstumsbedingungen, ein Farbcode von grün für herunterregulierte bis rot für hochregulierte Expression zugeordnet. Das obere Kladogramm gibt die Verwandtschaft der Expressionsprofile wieder, das Kladogramm auf der linken Seite die Ähnlichkeit des Expressionsverhaltens der Gene. Anhand der Distanzunterschiede im Kladogramm der Gene wurde das Transkriptom in 25 Regulons unterteilt und die Regulons farblich gekennzeichnet und nummeriert. Im Kladogramm der Bedingungen sind die Expressionsprofile mit überwiegend herunterreguliertem Transkriptom in grün, die Expressionsprofile mit vorherrschend hochreguliertem Transkriptom in rot markiert. Für eine detaillierte Darstellung des Kladogramms der Expressionsprofile siehe Abbildung 3.8.

### 3.3.1 Kladogramm der 101 differentiellen Transkriptome

Durch das Hierarchische Clustering der 101 Experimente ergibt sich eine Art Karte der transkriptionellen Antworten auf die unterschiedlichsten Stimuli (Abb. 3.8). Wie schon in Abbildung 3.2 gezeigt wurde, führten der überwiegende Teil der Wachstumsbedingungen und Mutationen zu einem entweder vorherrschend hochreguliertem oder herunterreguliertem nukleären Transkriptom des Chloroplasten.

Obwohl nun nur noch das gemeinsam differentiell exprimierte Transkriptom von 1616 Genen in das Hierarchische Clustering einging, werden beide Klassen der transkriptionellen Antworten wieder gefunden. Die deutliche Trennung ist im Kladogramm der Expressionsprofile durch die geringen Distanzen zwischen den Experimenten innerhalb einer Klasse und die großen Distanzen zwischen den Experimenten unterschiedlicher Klassen erkennbar. In einer dritten, kleineren Gruppe von 26 Expressionsprofilen sind die Gene des Transkriptoms in einem ausgeglichen Verhältnis sowohl hoch- wie auch herunterreguliert. So wie bereits im Überkreuzvergleich der gemeinsamen, signifikant differentiell exprimierten Gene der allelischen Mutanten (Abb. 3.3), wird die Verwandtschaft der Expressionsmuster von identischen Stimuli auch im Hierarchischen Clustering nochmals bestätigt (z.B.: *prs1* und *prs2*, *prs1\_D* und *prs2\_D*, *pete1* und *pete2*, *pete1\_soil* und *pete2\_soil*).



**Abb. 3.8.: Kladogramm der 101 Bedingungen.** Zur besseren Übersicht ist das Kladogramm der Expressionsprofile aus Abbildung 3.7 modifiziert dargestellt. Die Reihenfolge der Expressionsprofile von links unten nach rechts unten ist entsprechend der Reihenfolge von links nach rechts in Abb. 3.7. Das Clustering führt zu einem in drei Klassen aufgeteiltes Kladogramm: Expressionsprofile mit überwiegend induziertem Transkriptom sind in grün (Klasse 1), mit reprimierten Transkriptom in rot (3) markiert. In einer weiteren, deutlich abgegrenzten Klasse (2), ist das Transkriptom ausgewogen exprimiert.

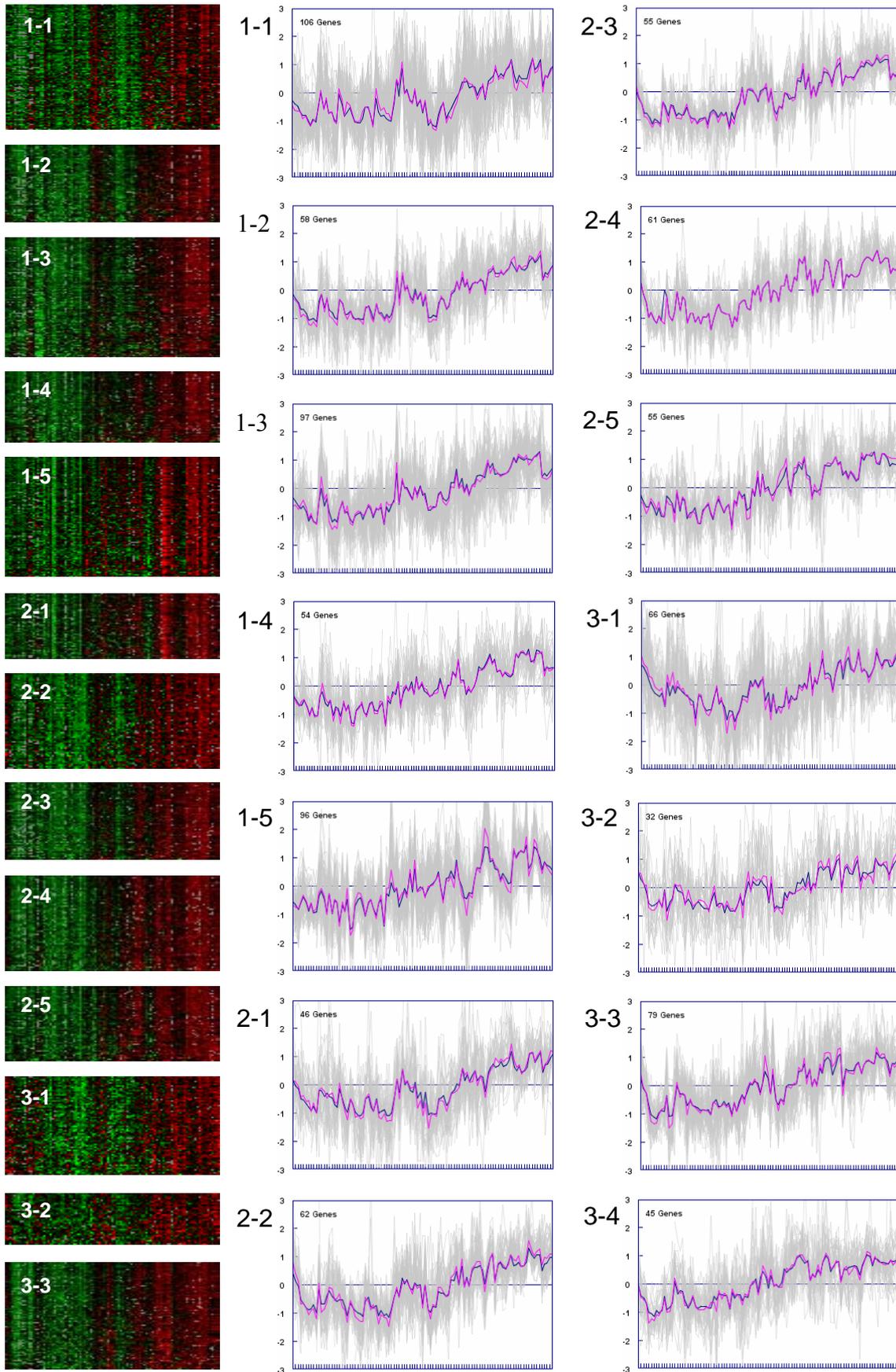
### 3.3.2 Analyse von Regulons

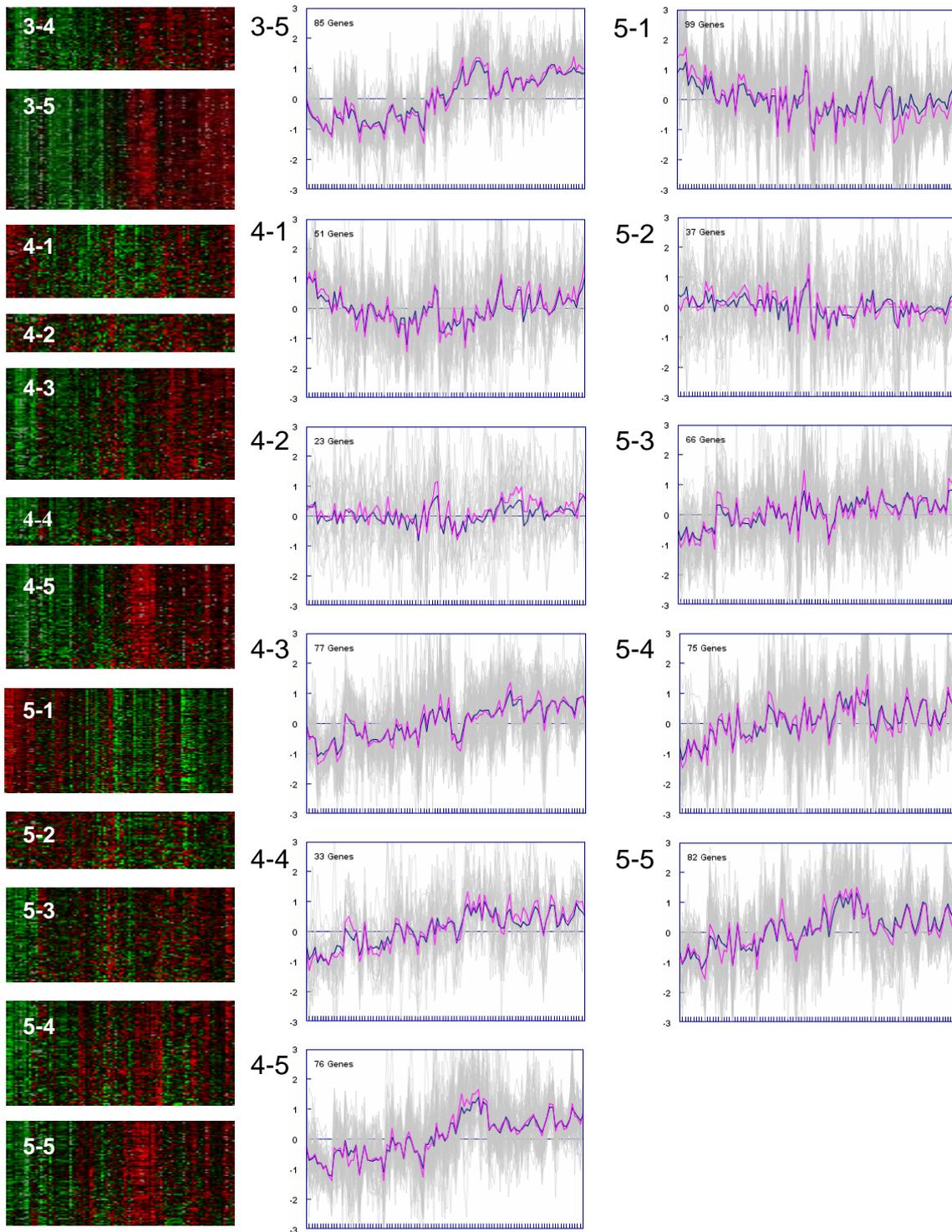
Hierarchisches Clustering wurde sowohl für die Experimente als auch für die Gene angewandt (Abb. 3.7). Im Kladogramm der Gene konnten 25 unterschiedlich exprimierte und unterschiedlich große Regulons definiert werden. Für die Unterscheidung zwischen getrennt voneinander regulierte Genclustern gibt es im Hierarchischen Clustering keine festgelegten Grundregeln (Sturn, 2000). Die Kriterien für die Identifizierung der Regulons sind die Distanzen (visualisiert im Kladogramm der Gene) auf Inter- und Intraclusterebene sowie eine vorher festzulegende Minimalgröße für die Anzahl der Gene eines Regulons.

Die Einteilung der Gene nach ihrem Expressionsverhalten in ähnlich regulierte Gengruppen unterliegt im Hierarchischen Clustering der subjektiven Betrachtung des Experimentators. Deshalb wurde für die endgültige Definition der Regulons SOM-Clustering durchgeführt und dabei die vorher bestimmte Anzahl von 25 zu erwartenden Regulons verwendet. Das Ergebnis des SOM-Clustering ist nur eine Ebene von Clustern, die nur Objekte und keine weiteren Cluster enthalten.

Die Regulons entstanden durch 1.616.000 Iterationen, also jeweils 1000 Wiederholungen des SOM-Algorithmus für jeden der 1616 Genvektoren der 5x5-dimensionalen Matrix (siehe Kapitel 2.8.3.4, Seite 40). Neben den farblich kodierten Expressionsraten der Mitglieder jedes Clusters, gibt der Median der Diagramme das Expressionsverhalten der Gesamtheit der Gene wieder (Abb. 3.9). Jeder der Regulons zeigt ein ganz spezielles Expressionsmuster und unterscheidet sich dadurch von jedem anderen Cluster der SOM-Matrix. Die Distanzen der Regulons untereinander sind in Abbildung 3.11 durch Hierarchisches Clustering dargestellt.

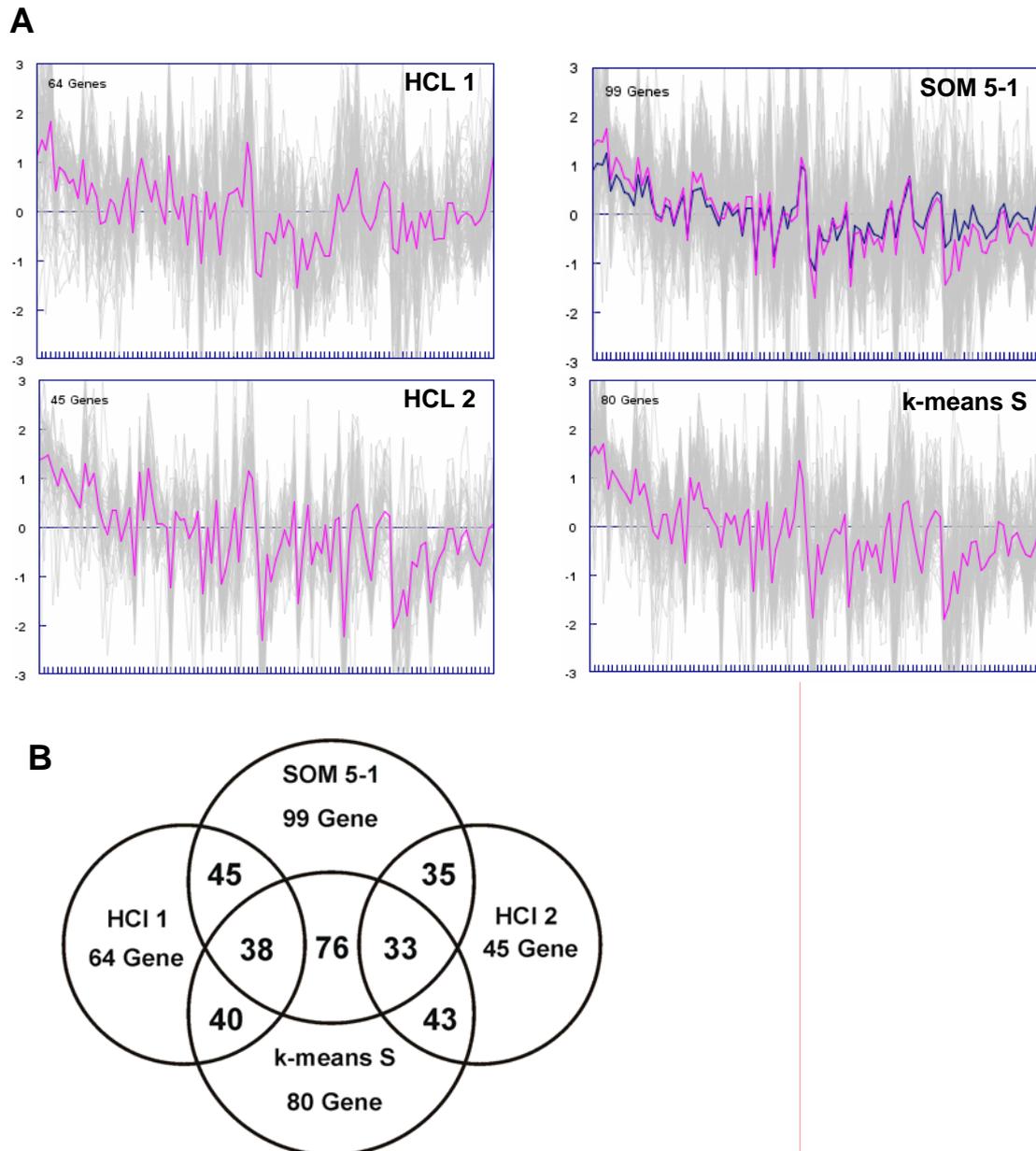
Da das SOM-Clustering einen anderen, iterativen Algorithmus benutzt, findet im Vergleich zum Hierarchischen Clustering eine teilweise Umverteilung der Gene zwischen den Regulons statt. Zwar unterscheiden sich die Regulons des SOM-Clustering in der Anzahl der Gene von den Regulons des Hierarchischen Clustering und des ebenfalls durchgeführten k-means-Clustering (siehe auch Abb. A2; Seite 119 ff.), generell werden aber über jede der drei Clustermethoden dieselben, deutlich voneinander getrennten Expressionsmuster wieder gefunden (Abb. 3.10 A).





**Abb. 3.9.:** Ergebnis des SOM-Clustering von 1616 Genen. Dargestellt sind die farblich visualisierten Expressionsdaten der durch SOM-Clustering definierten 25 Regulons (links) und die dazu korrespondierenden Expressionsdiagramme (rechts) mit SOM-Vektor (blau) und Median (rosa) des Clusters. Verwendet wurde eine 5x5-dimensionale SOM-Matrix und 1.616.000 (1000 x 1616) Iterationen. Die Reihenfolge der Bedingungen ergibt sich aus Abb. 3.8.

## 3.3.2.1 Vergleich von unterschiedlichen Clusteralgorithmen



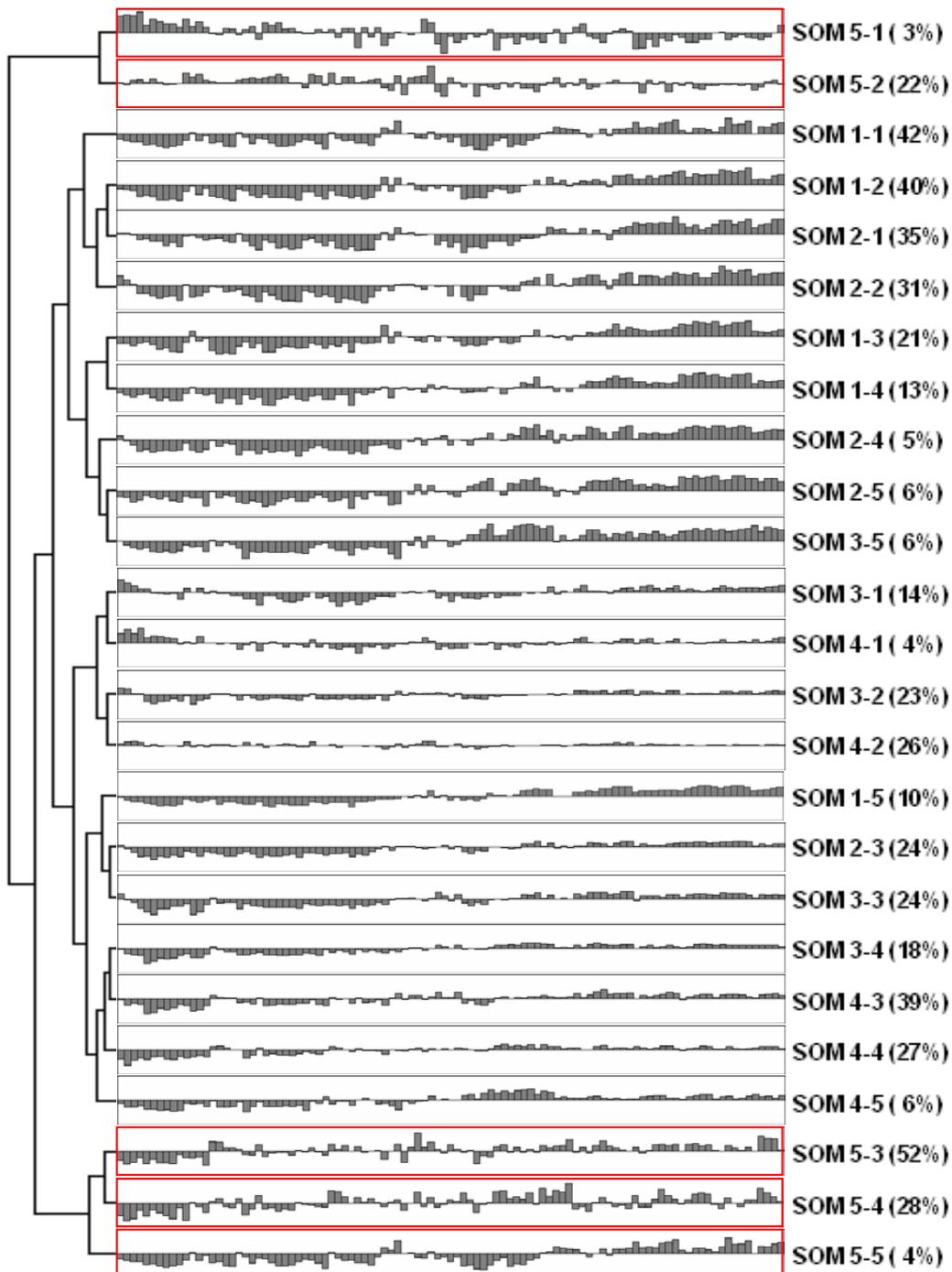
**Abb. 3.10.: Vergleich von Hierarchischem, k-means- und SOM-Clustering.** Die hohe Übereinstimmung der Ergebnisse der drei Clusteralgorithmen ist exemplarisch anhand von Regulon 1 und 2 aus dem Hierarchischem Clustering gezeigt. Schon in **A.** wird deutlich, dass die *Partitioning*-Methoden sehr ähnliche Expressionsmuster geclustert haben. Die Mediane des SOM-Clusters 5-1 und des k-means-Clusters S sind annähernd deckungsgleich. Dieses Expressionsmuster findet sich auch in den Clustern 1 und 2 des Hierarchischen Clustering wieder. In **B.** ist zu erkennen, dass die Gene des HCL-Regulon 1 und 2 durch die Clusteralgorithmen des *Partitioning* größtenteils zu einem Cluster vereinigt worden sind.

Mit dem Vergleich verschiedener Clusteralgorithmen sollten die Ergebnisse der Clusteranalyse verifiziert werden. In der Abbildung 3.10 sind die Ergebnisse der verschiedenen Clusteralgorithmen anhand des subjektiv erstellten Regulons 1 und Regulons 2 aus dem Hierarchischen Clustering exemplarisch dargestellt und miteinander verglichen worden. Es ist zu erkennen, dass 86% der Gene des SOM-Clusters 5-1 bzw. 98 % des Cluster k-means S aus den Regulons 1 und 2 des HCL stammen. Unter Vorgabe der 25 zu erwartenden Regulons aus dem HCL wurden durch die Algorithmen des *Partitioning* die meisten der Gene der HCL-Regulons 1 und 2 zu einem Regulon vereinigt. Ein Vergleich der Methoden des *Partitioning* untereinander zeigt sogar noch eine deutlichere Übereinstimmung. 95 % der Gene des k-means-Regulons S sind auch durch den SOM-Algorithmus miteinander geclustert worden.

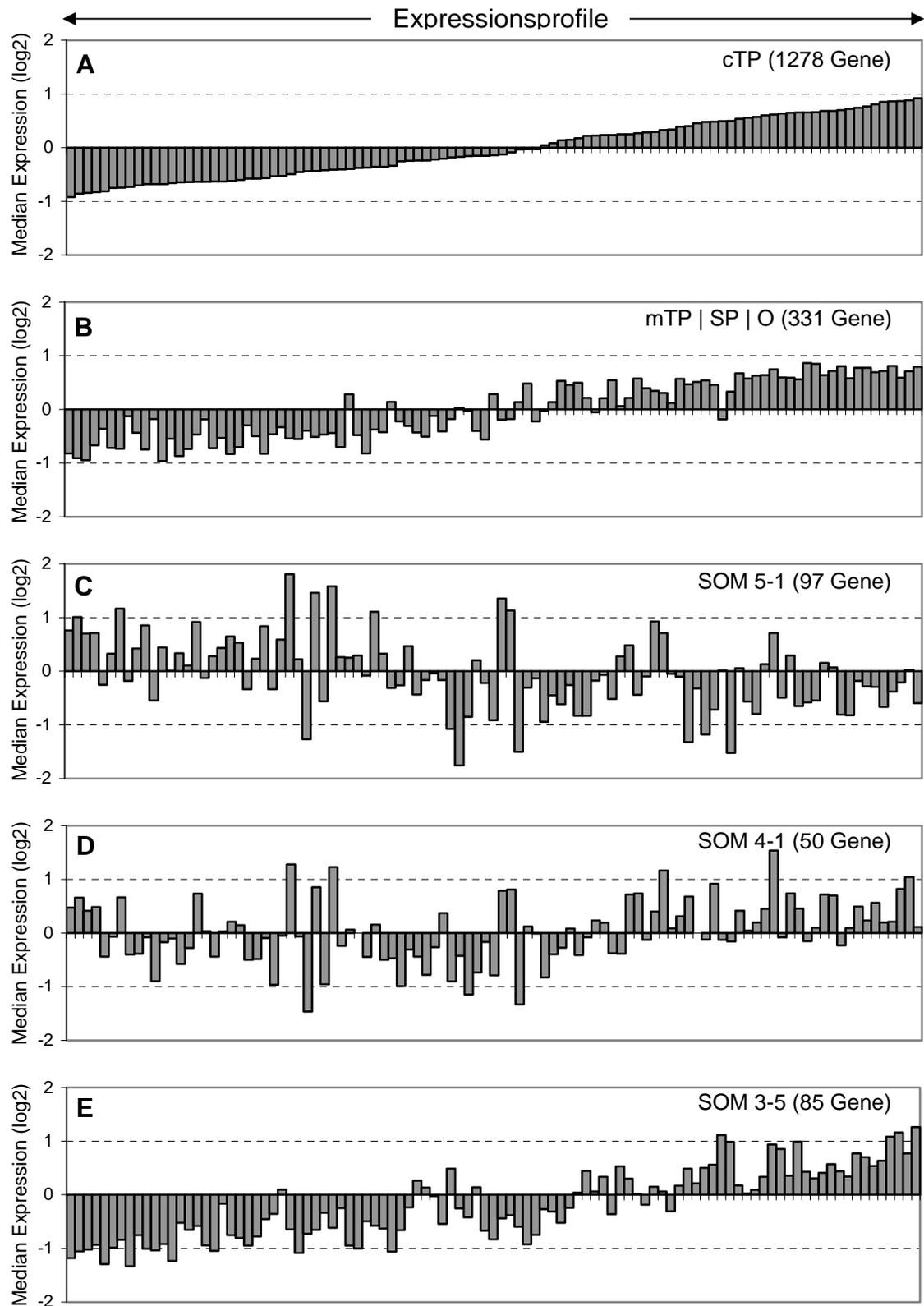
### 3.3.2.2 Expressionsmuster der Regulons

Da durch SOM-Clustering nur eine Ebene von Clustern entsteht und der Algorithmus kein Distanzmaß für die Verwandtschaft der Cluster untereinander angibt, wurden die SOM-Cluster durch die Software *GEPAS* nochmals hierarchisch geclustert (Abb. 3.11). Das Ergebnis ist ein Kladogramm der Interclusterdistanzen. Dadurch lassen sich die Regulons identifizieren, die entweder sehr ähnlich, oder noch wichtiger, sehr unähnlich zum Rest des Transkriptoms reguliert sind. Je größer die Distanz eines Regulons zum Rest des Transkriptoms ist, desto eher stellt er ein Zielobjekt für die Suche nach überrepräsentierten regulatorischen Elementen der Genexpression und funktionellen Genklassen dar.

Im Kladogramm ist zu erkennen, dass die Gesamtheit der Gene von fünf (SOM 5-1 und SOM 5-2 sowie SOM 5-3, SOM 5-4 und SOM 5-5) der 25 Regulons ein sehr spezielles Expressionsverhalten zeigen, das sich deutlich vom allgemeinen Trend des übrigen Transkriptoms unterscheidet. Da die Gene von Regulon 1 und 2 aus dem Hierarchischen Clustering zum größten Teil in Regulon SOM 5-1 wieder zu finden sind, lässt sich dies im Kladogramm des Hierarchischen Clustering auch visuell nachvollziehen (siehe Abb. 3.7). Zusätzlich ist für jedes Regulons der prozentuale Anteil der Gene angegeben, die nicht für Proteine des Plastiden kodieren.



**Abb.3.11. Hierarchisches Kladogramm der SOM-Regulons.** Das Kladogramm gibt die Distanzen der Regulons untereinander wieder. Die Gene der Regulons der X-Dimension 5 der SOM-Matrix sind im Vergleich zum Rest des Transkriptoms sehr speziell exprimiert (markiert in rot). Die Säulen der Expressionsdiagramme geben den Median der Expressionsrate aller Gene eines Regulons für jede der Bedingungen wieder (entsprechend Abb. 3.9). Die Prozentangaben beziehen sich auf den Anteil der Gene eines Regulons, die nicht plastidäre Proteine kodieren (siehe nachfolgendes Kapitel 3.3.2.4). Die Reihenfolge der Bedingungen ergibt sich aus Abb. 3.8.



**Abb. 3.12.:** Vergleich der Expression der Gene mit cTP gegenüber Genen anderer subzellulärer Lokalisationen bzw. der Expression der Gene von ausgewählten Regulons. Die Reihenfolge der 101 Expressionsprofile wurde nach dem Median der Expression aller Gene mit prädiziertem cTP (A) sortiert.

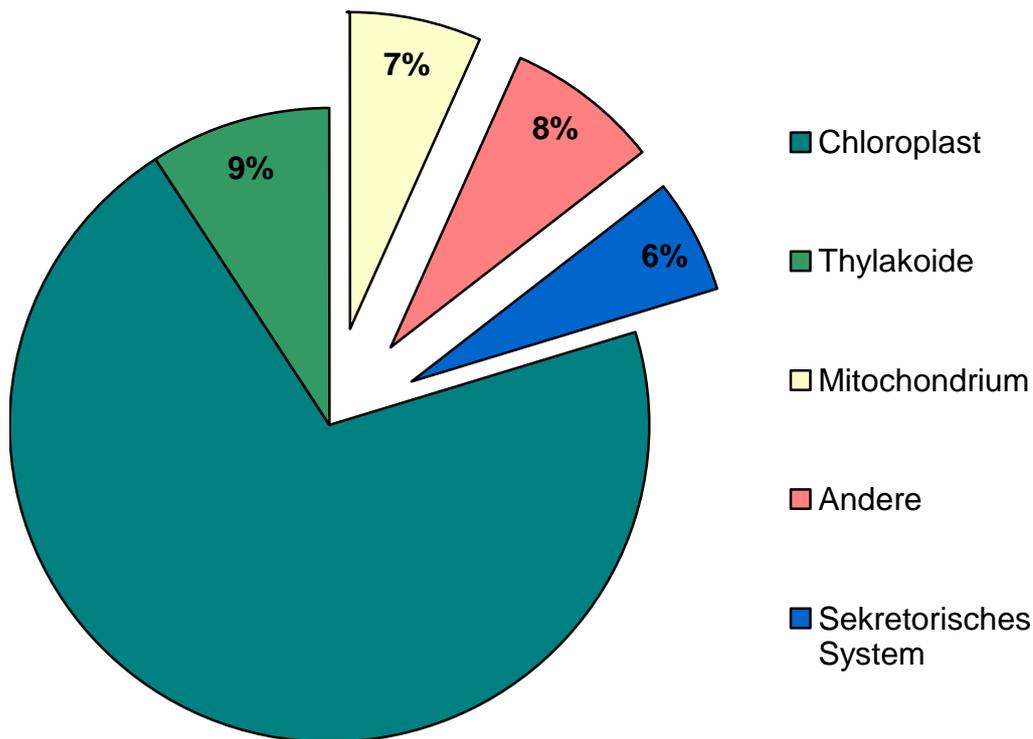
In Abbildung 3.12 sind die Expressionsprofile aufsteigend nach dem Median der Expression aller Gene mit prädiktiertem cTP sortiert worden. Es ist deutlich zu erkennen, dass es keine signifikant unterschiedliche Regulation der Expression zwischen den Genen gibt, die Proteine des Chloroplasten kodieren und denen, deren Genprodukte anderswo in der Zelle lokalisiert sind (siehe Abb. 3.12, A und B).

Im Gegensatz dazu sind die Gene des Regulons SOM 4-1 (D) sehr häufig, die Gene des Regulons SOM 5-1 (C) sogar überwiegend, entgegen dem Trend der Expression aller Gene des durch den Makroarray repräsentierten Transkriptoms exprimiert. Exemplarisch für den Grossteil der übrigen Regulons zeigt sich in Regulon SOM 3-5 (E), dass selbst in den Genclustern, deren Mitglieder sich wie im Falle der beiden oben genannten Regulons fast ausschließlich aus Chloroplastengen rekrutieren, keine deutlich unterschiedliche Expression im Vergleich zu den Genen festgestellt werden kann, die Proteine anderer subzellulärer Lokalisation kodieren.

### **3.3.2.3 Zusammenhang zwischen Genexpression und subzellulärer und subplastidärer Lokalisation der Genprodukte**

Das Design des Makroarrays basiert auf der computerunterstützten Prädiktion und/oder dem experimentelle Nachweis der subzellulären und zum Teil auch der subplastidären Lokalisationen, der vom nukleären Genom von *A. thaliana* kodierten, Genprodukte. 631 nukleäre Gene, die zusätzlich zu den 2661 Chloroplastengen des Kerns in das Set der Hybridisierungssonden des Arrays aufgenommen wurden, kodieren Proteine des Mitochondriums (7 %), des Sekretorischen Systems (6 %) oder anderer Lokalisationen (8 %) (Abb. 3.13). Dieses Design des Makroarrays erlaubt es, einen möglichen Zusammenhang zwischen Genexpression und subzellulärer Lokalisation des Genprodukts zu untersuchen.

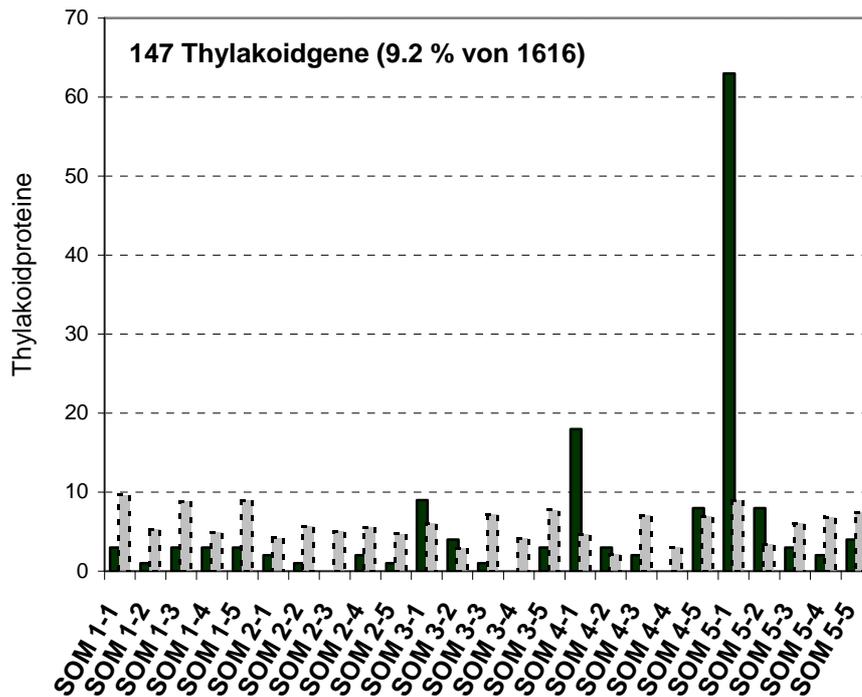
Der Anteil der Gene eines Regulons, die nicht für plastidäre Proteine kodieren, variiert zwischen den Genclustern sehr stark (Abb. 3.11). Während in einigen Regulons fast ausschließlich nukleäre Chloroplastengene zu finden sind (z.B.: SOM 5-1, SOM 4-1), kodieren mehr als die Hälfte der Gene des Regulons SOM 5-3 Proteine anderer subzellulärer Lokalisation.



**Abb. 3.13.: Subzelluläre und subplastidäre Lokalisationen der Genprodukte des 3292-GST-Makroarrays.** Das vorhergesagte nukleäre Transkriptom des Chloroplasten beinhaltet 2661 Gene (Richly et al., 2003) und macht damit etwa 80 % des durch den 3292-GST-Makroarray abgebildeten Transkriptoms aus. 9 % dieser Gene kodieren Proteine der Thylakoide, während durch 7 %, 6 % und 8 % aller gespotteten GSTs Teile des nukleären Transkriptoms des Mitochondriums, des Sekretorischen Systems und anderer subzellulärer Lokalisationen abgebildet werden (Design des DNA-Arrays und *Blast*-Analyse der subzellulären Lokalisationen der Genprodukte von Varotto und Richly).

Ein noch deutlicheres Bild ergibt sich bei der Betrachtung der subplastidären Lokalisationen. Für 147 Genprodukte der 1616 Gene des Gesamtdatensatzes konnte bisher der experimentelle Nachweis für die Lokalisation integral oder peripher der Thylakoidmembran oder im Lumen der Thylakoide erbracht werden (*PPBD*).

Eindeutig sind Gene für Thylakoidproteine in Regulon SOM 5-1 überrepräsentiert und machen etwa 65 % aller Gene des Clusters aus (Abb. 3.14). In Regulon SOM 4-1 kodieren annähernde 40 % der Gene und in Regulon SOM 5-2 immerhin 21 % der Gene für Proteine, die in den Thylakoiden lokalisiert sind.

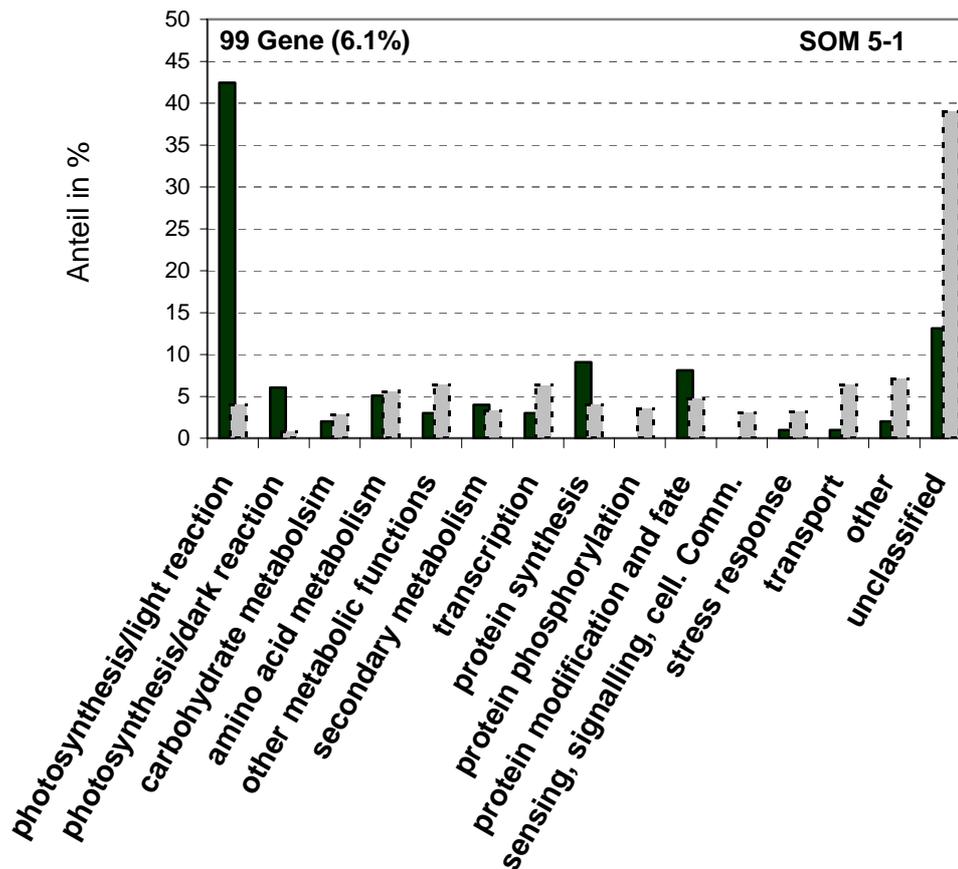


**Abb. 3.14:** Verteilung der 147 in den Thylakoiden lokalisierten, nukleär kodierten Genprodukte auf die Regulons. Für 147 (9,2 %) der 1616 Genprodukte des nukleären Transkriptoms des Chloroplasten ist der experimentelle Nachweis für die Lokalisation im Thylakoidlumen oder peripher oder integral in der Thylakoidmembran erbracht. Der tatsächlichen Anzahl der Thylakoidgene (schwarze Säulen) eines Regulons ist die zu erwartende Anzahl (graue Säulen) bei zufälliger Verteilung gegenüber gestellt.

### 3.3.2.4 Funktionelle Analyse der Regulons

Um die zufällig zu erwartende relative Verteilung mit der tatsächlichen Verteilung der funktionellen Genklassen innerhalb eines Regulons vergleichen zu können, wurden die 1616 Gene des gefilterten, differentiell regulierten Transkriptoms bezüglich der Zugehörigkeit zu Regulons funktionell klassifiziert.

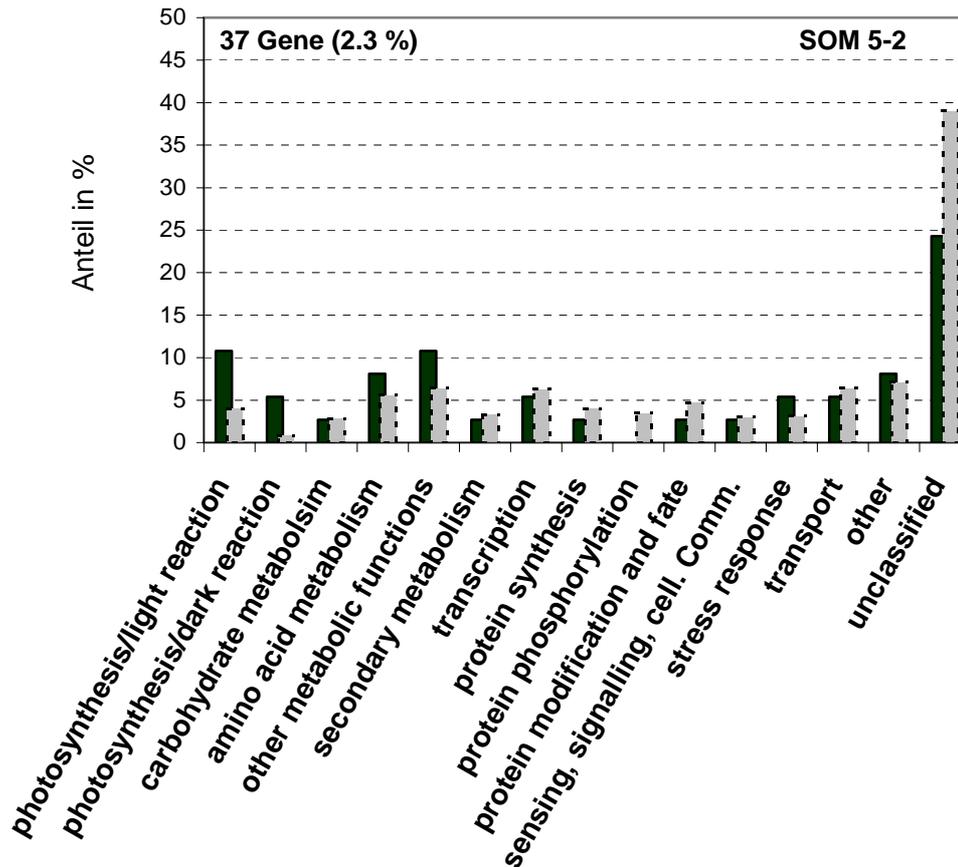
Drei Regulons konnten identifiziert werden, die eine deutliche Akkumulation einer oder mehrerer funktioneller Genklassen aufweisen. Im Regulon SOM 5-1, der, wie aus dem Hierarchischen Clustering der Regulons hervorgeht, zusammen mit Regulon SOM 5-2 sehr speziell reguliert zu sein scheint (Abb. 3.11), ist die Klasse der Photosynthesegene eindeutig überrepräsentiert (Abb. 3.15).



**Abb. 3.15.: Funktionelle Klassifizierung der Gene des Regulons SOM 5-1.** Die Gene des Regulons wurden in 15 funktionelle Klassen unterteilt und die relativen Abundanzen der funktionellen Genklassen (schwarze Säulen) im Regulon mit der Verteilung der funktionellen Genklassen im gesamten Ausgangsdatensatz des 1616 Gene umfassenden Transkriptom verglichen (graue Säulen). Regulon SOM 5-1 beinhaltet 6.1 % (99 Gene) der Gene des Gesamtdatensatzes.

Fast die Hälfte der Gene des Regulons kodieren Proteine, die in die Licht- oder Dunkelreaktion der Photosynthese involviert sind. Anders ausgedrückt, finden sich über 65 % der Photosynthesegene des gesamten Datensatzes von 1616 Genen in einem Cluster wieder, der nur etwa 6 % des gefilterten, differentiell exprimierten nukleären Transkriptoms des Chloroplasten ausmacht (siehe auch Abb. A1, Seite 118). Nahezu alle nukleär kodierten Untereinheiten der Photosysteme PSI und PSII und der LHCs sind in diesem Regulon zu finden (siehe Tab. 3.3).

Auch Regulon SOM 5-2 weist eine Akkumulation von Genen auf, die in Beziehung zu den Vorgängen der Photosynthese stehen (Abb. 3.16).

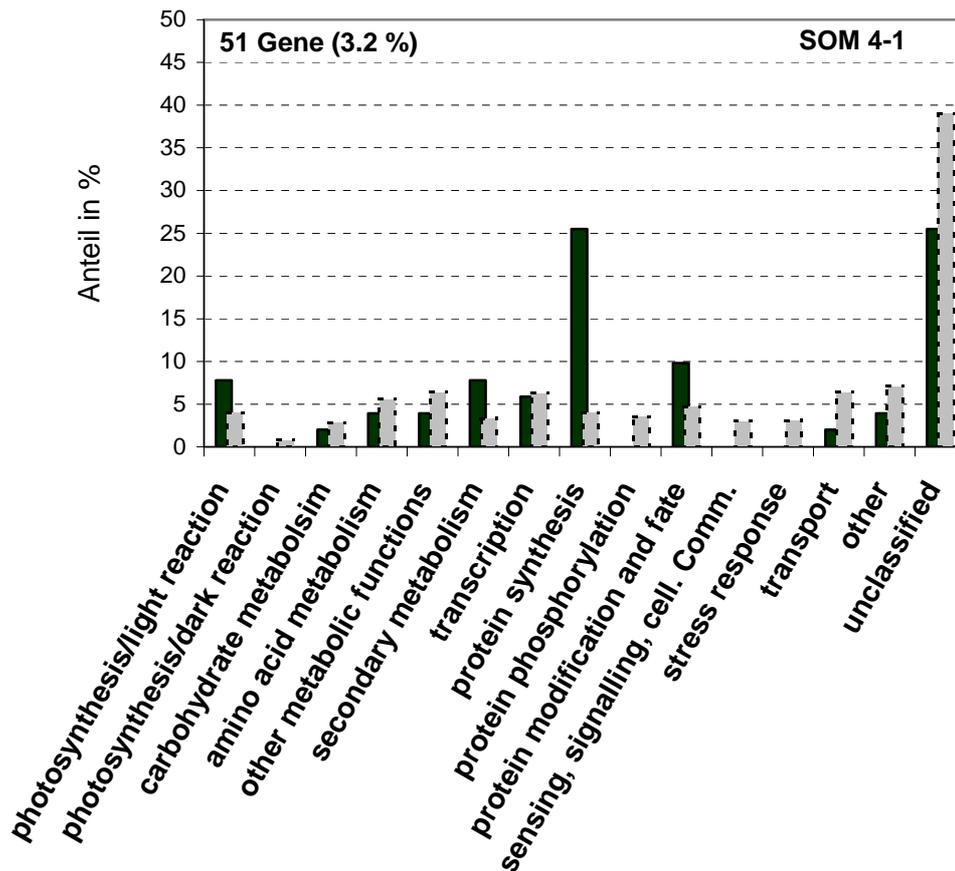


**Abb. 3.16.: Funktionelle Klassifizierung der Gene des Regulons SOM 5-2.** Die Gene des Regulons wurden in 15 funktionelle Klassen unterteilt und die relativen Abundanzen der funktionellen Genklassen (schwarze Säulen) im Regulon mit der Verteilung der funktionellen Genklassen im gesamten Ausgangsdatensatz des 1616 Gene umfassenden Transkriptoms verglichen (graue Säulen). Regulon SOM 5-2 beinhaltet 2.3 % (37 Gene) der Gene des Gesamtdatensatzes.

Erstaunlicherweise ist dieser Regulons jedoch weit weniger eindeutig funktionell definiert als Regulon SOM 5-1, obwohl beide Gene mit sehr ähnlichen Expressionsverhalten beinhalten.

Das zeigt, dass zwischen zwei untereinander sehr ähnlich, im Vergleich zum restlichen Transkriptom aber sehr speziell regulierten Regulons, eine eindeutige Trennung nach der Genfunktion stattgefunden hat. Andererseits unterstreicht dieses Ergebnis das Potential des SOM-Clustering, unterschiedliche Expressionsmuster zu erkennen und voneinander zu trennen. Ein weiteres Regulon zeigt eine interessante Verteilung der funktionellen Genklassen:

bei Regulon SOM 4-1 sind die Gene der Proteinsynthese deutlich überrepräsentiert. 20% der zur Klasse der Proteinsynthese zugeordneten Gene des Gesamtdatensatzes fallen in diesen Gencluster (siehe Abb. A1, Seite 118). Innerhalb der funktionellen Klasse der Proteinsynthese sind es hier hauptsächlich die Gene der nukleär kodierten Ribosomenproteine des Chloroplasten, die bei 101 Experimenten ähnlich reguliert waren (siehe Tab. 3.4).



**Abb. 3.17.: Funktionelle Klassifizierung der Gene des Regulons SOM 4-1.** Die Gene des Regulons wurden in 15 funktionelle Klassen unterteilt und die relativen Abundanzen der funktionellen Genklassen (schwarze Säulen) im Regulon mit der Verteilung der funktionellen Genklassen im gesamten Ausgangsdatsatz des 1616 Gene umfassenden Transkriptom verglichen (graue Säulen). Regulon SOM 4-1 beinhaltet 3.2 % (51 Gene) der Gene des Gesamtdatensatzes.

Die funktionelle Klassifizierung aller anderen Regulons ist in Abbildung A2 (Seite 119) dargestellt. Durch die weiterführende Unterteilung in kleinere und enger gefasste funktionelle Klassen könnten vielleicht auch diese Regulons funktionell definiert werden.

### 3.3.3.5 Detailanalyse ausgewählter Regulons

Anhand der Ergebnisse aus der funktionellen Klassifizierung und der Bestimmung der subzellulären und subplastidären Lokalisationen (Abb. 3.14) der Gene jedes Regulons sowie der im hierarchischen Kladogramm (Abb. 3.11) dargestellten Distanzen der Regulons zueinander, wurden die Regulons SOM 5-1 und SOM 4-1 als die prädestinierten Kandidaten für eine tiefer gehende Analyse identifiziert.

Mittels *MotifFinder* wurde die Häufigkeit von 6-mer Sequenzmotiven in der Promoterregion (bis 500 Basenpaare vor dem Startcodon) jedes der Gene dieser Regulons mit der Frequenz im gesamten Genom verglichen. Der p-Wert wurde durch die Binominalverteilung bestimmt und nur Sequenzmotive mit einem p-Wert  $< 10^{-5}$  genauer klassifiziert (nach Hudson & Quail, 2003).

Nach diesen Kriterien konnten ausschließlich in Regulon SOM 5-1 überrepräsentierte Sequenzmotive identifiziert werden (Tab. 3.2): Die G-Box-Motive ACGTGG/CCACGT (Uno et al., 2000), GCCACG/CGTGGC (Hudson & Quail, 2003), CACGTG (Toledo-Ortiz et al., 2003; Huq et al., 2004) und GGATAA/TTATCC (Hudson & Quail, 2003) sowie das I-Box-Motiv GATAAG/CTTATC (Escobar et al., 2004) wurden bereits als *cis*-regulatorische Sequenzmotive des Netzwerkes der von Phytochrom A regulierten Genexpression in der Literatur beschrieben. Ein weiteres überrepräsentiertes Sequenzmotiv (ATGTGG/TACACC) ist jedoch bisher nicht mit der lichtregulierten Expressionskontrolle in Verbindung gebracht worden.

Die funktionelle Klassifizierung des differentiell regulierten Transkriptoms erlaubte es, in der großen Datenmenge der 1616 Genen, die sich nach ihrem Expressionsmuster in 25 deutlich voneinander verschiedene Regulons einteilen ließen, übergeordnete Muster mit wichtiger biologischer Information zu extrahieren. Um eine mögliche funktionelle Beziehung einiger oder mehrerer Gene eines Regulons untereinander zu analysieren, müssen die Mitglieder eines potentiell interessanten Regulons genauer betrachtet werden. In den Tabellen 3.3 und 3.4 sind die Gene der Regulons SOM 5-1 und SOM 4-1 zusammen mit den Ergebnissen der funktionellen Klassifizierung der Genprodukte, der Bestimmung der subzellulären und subplastidären Lokalisationen sowie der Analyse überrepräsentierter Sequenzmotive in der Promoterregion der Gene aufgelistet.

	oligoMer	Sequenzen mit mind. 1 oligoMer	Alle Sequenzen des Genoms mit mind. 1 oM	p-Wert	Typ	Signalling-Typ
1	ACGTGG	42/97	3817/28088	6.88E-13	ABR G-Box	phyA-Signalling/ ABA-Signalling
	CCACGT	42/97	3818/28088	6.94E-13	ABR G-Box	phyA-Signalling/ ABA-Signalling
2	GCCACG	32/97	2291/28088	2.59E-12	G-Box	phyA-Signalling (zirkadiane Rhythmik)
	CGTGGC	32/97	2291/28088	2.59E-12	G-Box	phyA-Signalling (zirkadiane Rhythmik)
3	CACGTG	33/97	2819/28088	1.13E-10	G-Box	phyA-Signalling
4	GGATAA	48/97	6337/28088	4.22E-09	G-Box / Gata	phyA-Signalling
	TTATCC	48/97	6337/28088	4.22E-09	G-Box / Gata	phyA-Signalling
5	GATAAG	48/97	6488/28088	9.29E-09	I-Box	phyA-Signalling
	CTTATC	48/97	6489/28088	9.34E-09	I-Box	phyA-Signalling
6	ATGTGG	36/97	5128/28088	6.00E-06	?	?
	TACACC	36/97	5128/28088	6.00E-06	?	?

**Tab. 3.2.: Überrepräsentierte 6-mer Sequenzmotive in der Promoterregion der Gene des Regulons SOM 5-1.** Fünf der sechs überrepräsentierten 6-mer Sequenzmotive in der Promoterregion (500 bp vor dem Startcodon) der Gene des Regulons SOM 5-1 sind G-Box-, GATA- oder I-Box-Motive, die bereits als regulatorische Elemente des *phyA-Signallings* identifiziert worden waren. Die Funktion eines weiteren Motivs (ATGTGG/TACACC) ist bisher nicht beschrieben worden.

Bei genauerer Betrachtung der Gene des Regulons SOM 5-1 fällt auf, dass außer den funktionell in Beziehung stehenden Untereinheiten der Photosysteme und der Lichtsammelkomplexe weitere funktionelle Subklassen gefunden werden. Die drei Glycerinaldehyd-3-Phosphate-Dehydrogenasen (*GAP*) des Datensets sind ebenso ähnlich exprimiert, wie die kleinen nukleär kodierten Untereinheiten und die plastomkodierte große Untereinheit der RuBisCo.

Neben einer Vielzahl von Untereinheiten der plastidären Ribosomen sind auch die Thioredoxine deutlich überrepräsentiert. Interessanterweise finden sich mit der Mg-protoporphyrinIX-Chelatase (CHLI) und der Mg-ProtoME-Cyclase zwei Enzyme des Tetrapyrrolmetabolismus wieder, dessen Intermediate direkten Einfluss auf die Expression photosyntheserelevanter Gene im Kern zu haben scheinen (Surpin et al., 2002).

UNIQUID	PPDB	1	2	3	4	5	6	Fkt. Genklasse	Annotation
At3g26650	x				o	o		amino acid metabolism	glyceraldehyde 3-phosphate dehydrogenase A subunit (GapA)
At1g12900							o	amino acid metabolism	glyceraldehyde 3-phosphate dehydrogenase
At1g42970	x				o	o	o	amino acid metabolism	glyceraldehyde-3-phosphate dehydrogenase
At2g14750		o		o			o	amino acid metabolism	adenosine phosphosulfate kinase
At3g54050		o	o	o	o	o	o	carbohydrate metabolism	fructose-bisphosphatase precursor
At4g38970					o			carbohydrate metabolism	putative fructose-bisphosphate aldolase
At3g01500	x				o	o		other	carbonic anhydrase, location in thylakoid
At3g62030					o			other	peptidylprolyl isomerase ROC4
At3g11170								other metabolic functions	omega-3 fatty acid desaturase
At5g61410								other metabolic functions	ribulose-5-phosphate-3-epimerase
At5g05580								other metabolic functions	omega-3 fatty acid desaturase
At1g56190						o	o	photosynthesis/dark reaction	phosphoglycerate kinase, putative
At1g32060					o	o	o	photosynthesis/dark reaction	phosphoribulokinase precursor
rbcl	x							photosynthesis/dark reaction	large subunit of RuBisCO
At5g38430						o	o	photosynthesis/dark reaction	RuBisCO small chain 1b precursor
At5g38420		o	o	o	o	o	o	photosynthesis/dark reaction	RuBisCO small chain 2b precursor
At1g67090	x	o	o	o		o		photosynthesis/dark reaction	RuBisCO small unit, putative
At4g09650	x	o	o		o			photosynthesis/light reaction	AtpD; ATP synthase delta subunit
At3g61470	x	o	o	o	o	o	o	photosynthesis/light reaction	Lhca2
At1g61520	x	o	o	o	o	o		photosynthesis/light reaction	Lhca3
At3g47470	x		o					photosynthesis/light reaction	Lhca4
At1g29920	x	o		o	o			photosynthesis/light reaction	Lhcb1.1 (LHCP AB 160)
At1g29910	x					o	o	photosynthesis/light reaction	Lhcb1.2 (LHCP AB 180)
At2g34430	x						o	photosynthesis/light reaction	Lhcb1.4
At2g34420	x						o	photosynthesis/light reaction	Lhcb1.5
At2g05100	x						o	photosynthesis/light reaction	Lhcb2.1
At2g05070	x							photosynthesis/light reaction	Lhcb2.2
At5g54270	x				o			photosynthesis/light reaction	Lhcb3 chlorophyll a/b binding protein
At5g01530	x	o	o	o		o		photosynthesis/light reaction	Lhcb4.1 (CP29)
At3g08940	x						o	photosynthesis/light reaction	Lhcb4.2 (CP29)
At2g40100	x						o	photosynthesis/light reaction	Lhcb4.3 (CP29)
At4g10340	x				o		o	photosynthesis/light reaction	Lhcb5 (CP26)
At1g15820	x	o	o		o	o	o	photosynthesis/light reaction	Lhcb6 (CP24)
At1g76100	x	o		o		o		photosynthesis/light reaction	PetE1 (plastocyanin)
At1g20340	x	o		o		o		photosynthesis/light reaction	PetE2 (plastocyanin)
At1g60950		o	o	o			o	photosynthesis/light reaction	PetF1; ferredoxin precursor
At1g10960		o	o	o			o	photosynthesis/light reaction	PetF2; ferredoxin precursor isolog
At2g26500	x							photosynthesis/light reaction	PetM
atpB	x							photosynthesis/light reaction	plastid gene product; ATP synthase
At1g03130	x					o	o	photosynthesis/light reaction	PsaD2
At4g28750	x	o	o	o	o		o	photosynthesis/light reaction	PsaE1
At2g20260	x	o	o	o	o			photosynthesis/light reaction	PsaE2
At1g31330	x					o	o	photosynthesis/light reaction	PsaF
At1g55670	x	o	o	o			o	photosynthesis/light reaction	PsaG
At3g16140	x	o	o	o	o	o		photosynthesis/light reaction	PsaH1
At1g52230	x	o	o	o			o	photosynthesis/light reaction	PsaH2
At1g30380	x				o	o		photosynthesis/light reaction	PsaK
At5g64040	x	o			o	o	o	photosynthesis/light reaction	PsaN
At5g66570	x	o	o	o	o	o		photosynthesis/light reaction	PsbO1
At3g50820	x	o	o	o	o		o	photosynthesis/light reaction	PsbO2
At2g30790	x	o	o	o	o	o		photosynthesis/light reaction	PsbP2
At4g21280	x	o	o	o		o	o	photosynthesis/light reaction	PsbQ1
At4g05180	x	o	o	o		o	o	photosynthesis/light reaction	PsbQ2

Fortsetzung Tabelle 3.1. →

At1g79040	x	o	o	o			photosynthesis/light reaction	PsbR
At1g44575	x	o				o	photosynthesis/light reaction	PsbS protein
At3g21055	x	o	o	o	o	o	photosynthesis/light reaction	PsbT1
At1g51400	x	o	o	o	o	o	photosynthesis/light reaction	PsbT2
At2g06520	x	o		o	o		photosynthesis/light reaction	PsbX
At1g67740	x				o	o	photosynthesis/light reaction	PsbY
At2g04700		o	o		o	o	photosynthesis/light reaction	FTR; ferredoxin-thioredoxin reductase
At2g39470	x				o	o	photosynthesis/light reaction	thylakoid lumen protein, PsbP domain
At1g02560	x		o				protein modification and fate	Clp protease proteolytic subunit (nClpP1)
At5g50920	x	o	o				protein modification and fate	ClpC-V
At4g03520	x						protein modification and fate	m-type thioredoxin
At2g30950	x					o	protein modification and fate	putative ftsH chloroplast protease (VAR2)
At1g03680	x				o		protein modification and fate	putative thioredoxin-m
At2g28800	x	o	o				protein modification and fate	chloroplast membrane protein ALBINO3
At3g02730					o		protein modification and fate	thioredoxin f1
At3g15360	x	o		o	o		protein modification and fate	thioredoxin m4
At2g37220	x					o	protein synthesis	RNA-binding protein cp29
At3g25920						o	protein synthesis	chloroplast ribosomal protein L15
At1g35680					o		protein synthesis	chloroplast ribosomal protein L21
At5g54600					o		protein synthesis	chloroplast ribosomal protein L24
At3g62410		o			o	o	protein synthesis	CP12 protein precursor-like protein
At3g27850	x				o	o	protein synthesis	chloroplast 50S ribosomal protein L12-C
At2g47400		o	o	o		o	protein synthesis	putative chloroplast protein CP12
At2g33450						o	protein synthesis	chloroplast ribosomal protein L28
At5g14320						o	protein synthesis	chloroplast ribosomal protein S13
At4g34050							secondary metabolism	caffeoyl-CoA O-methyltransferase - like
At4g18480					o	o	secondary metabolism	Mg-protoporphyrin IX chelatase (CHL1)
At3g56940	x	o	o	o		o	secondary metabolism	Mg-ProtoME cyclase
At5g54770		o	o			o	secondary metabolism	thiazole biosynthetic enzyme (ARA6)
At5g67030					o	o	secondary metabolism	zeaxanthin epoxidase precursor (ABA1)
At4g34190	x						stress response	stress encanced protein 1 (SEP1), Lil4
At1g09340					o	o	transcription	putative RNA-binding protein
At4g24770	x				o	o	transcription	RNA-binding protein RNP-T precursor
At5g52440	x				o		transport	HCF106
At2g35370	x						unclassified	glycine decarboxylase complex H protein
At1g16030							unclassified	heat shock protein 70, Hsp70
At1g12090					o		unclassified	hypothetical protein
At3g12120							unclassified	hypothetical protein
At4g01050	x	o	o				unclassified	hypothetical protein
At4g01150	x					o	unclassified	hypothetical protein
At2g45180						o	unclassified	lipid transfer protein (LTP)
At3g59400						o	unclassified	putative protein
At4g22890	x	o	o			o	unclassified	putative protein
At4g34350					o	o	unclassified	putative protein
At1g54780	x	o	o	o	o	o	unclassified	thylakoid lumen protein
At1g21500					o	o	unclassified	unknown protein

**Tab. 3.3.: Gene des Regulons SOM 5-1.** Die Gene des Regulons sind nach ihrer Zugehörigkeit zu den funktionellen Genklassen sortiert. Die mit 1, 2, 3, 4, 5 und 6 überschriebenen Spalten stehen für die *phyA*-regulierten *cis*-Elemente 1, ACGTGG/CCACGT; 2, GCCACG/CGTGGC; 3, CACGTG; 4, GGATAA/TTATCC und 5, GATAAG/CTTATC sowie ein bisher nicht klassifiziertes Sequenzmotiv 6, ATGTGG/TACACC. Die Spalte PPDB gibt den experimentellen Nachweis für die Lokalisation des Genprodukts in den Thylakoiden wieder. Für At4g34050 wurde keine N-terminale Zielsequenz, für At2g35370 ein mTP und für At2g45180 ein Transitpeptid für das Sekretorische System prädiiziert.

UNIQUID	TP	PPDB	Fkt. Genklasse	Annotation
At4g08870	C(2)		amino acid metabolism	putative arginase
At5g04590	C(2)		amino acid metabolism	sulphite reductase
At5g17230	C(4)		carbohydrate metabolism	phytoene synthase
At2g42220	C(1)		other	hypothetical protein
At3g54900	C(4)		other	putative protein
At4g15560	C(1)		other metabolic functions	DEF (CLA1) protein
At2g37660	C(1)		other metabolic functions	unknown protein
At4g04640	C(1)	x	photosynthesis/light reaction	AtpC1; ATP synthase gamma subunit
At4g32260	C(3)	x	photosynthesis/light reaction	AtpG; ATP synthase beta' chain
At4g17600	C(3)	x	photosynthesis/light reaction	Li3 protein (similar to LHC group of proteins)
At5g47110	C(2)		photosynthesis/light reaction	Li3 protein (similar to LHC group of proteins)
At4g12800	C(5)	x	photosynthesis/light reaction	PsaL
At1g50250	C(3)	x	protein modification and fate	chloroplast FtsH protease, putative VAR2
At1g12410	C(5)	x	protein modification and fate	ClpP-protease
At4g24280	C(3)	x	protein modification and fate	heat shock protein 70, putative / HSP70
At3g01480	C(2)	x	protein modification and fate	peptidyl-prolyl cis-trans isomerase, TLP 40 rotamase
At1g54500	C(3)	x	protein modification and fate	rubredoxin, putative
At1g50320	C(5)		protein modification and fate	thioredoxin, putative
At1g32990	C(1)		protein synthesis	chloroplast ribosomal protein L11
At1g79850	C(3)		protein synthesis	chloroplast ribosomal protein S17; HCF60 (maize)
At1g68590	C(1)	x	protein synthesis	plastid-specific 30S ribosomal protein 3 / PSRP-3
At4g34110	C(5)		protein synthesis	poly(A)-binding protein
At1g64510	C(1)	x	protein synthesis	chloroplast plastid ribosomal protein S6
At5g13510	C(5)		protein synthesis	chloroplast ribosomal protein L10
At3g54210	C(1)		protein synthesis	chloroplast ribosomal protein L17
At1g48350	C(1)		protein synthesis	chloroplast ribosomal protein L18
At2g43030	C(5)		protein synthesis	chloroplast ribosomal protein L3
At2g24090	C(4)		protein synthesis	chloroplast ribosomal protein L35
At1g05190	C(5)		protein synthesis	chloroplast ribosomal protein L6
At5g15760	C(3)		protein synthesis	ribosomal protein PSRP3
At4g20360	C(4)	x	protein synthesis	translation elongation factor EF-Tu precursor, chloroplast
At3g55120	C(4)		secondary metabolism	chalcone isomerase
At1g44446	C(3)		secondary metabolism	chlorophyll a oxygenase, chlorophyll b synthase
At3g51820	C(3)	x	secondary metabolism	chlorophyll synthetase
At2g20860	M(3)		secondary metabolism	lipoic acid synthase (LIP1)
rpoC1		x	transcription	plastid gene product; RNA polymerase
At5g24120	C(5)		transcription	RNA polymerase sigma subunit SigE (sigE)
At3g53460	C(3)	x	transcription	RNA-binding protein cp29 protein
At4g25050	C(5)		transport	acyl carrier - like protein
At2g23670	C(4)	x	unclassified	hypothetical protein
At1g75340	C(5)		unclassified	hypothetical protein
At3g22690	C(4)		unclassified	hypothetical protein
At4g12520	SP(1)		unclassified	lipid transfer protein (LTP) family protein
At5g14910	C(3)	x	unclassified	putative protein
At4g21210	C(3)		unclassified	putative protein
At4g30620	C(4)		unclassified	putative protein
At4g31560	C(1)		unclassified	putative protein
At2g46820	C(3)	x	unclassified	unknown protein
At1g15980	C(5)		unclassified	unknown protein
At2g20890	C(5)		unclassified	unknown protein
At3g22970	C(1)		unclassified	unknown protein

**Tab. 3.4.: Gene des Regulons SOM 4-1.** Siehe Tab. 3.3. Die subzelluläre Lokalisation wurde durch die Prädiktion möglicher Transitpeptide (TP) vorgenommen. C, Chloroplast; O, andere; SP, Sekretorisches System; M, Mitochondrium.

In Regulon SOM 4-1 clustern elf ribosomale Untereinheiten, ein Elongationsfaktor der Translation und zwei RNA-bindende Proteine. Darüber hinaus finden sich mit der plastomkodierte RNA-Polymerase und dem nukleär kodierte sigma-Faktor ein Großteil des nukleär kodierte genetischen Apparates des Chloroplasten in diesem Regulon wieder. Bemerkenswert ist, dass zwar die  $\gamma$ - und  $\beta$ -Untereinheiten der ATPase ähnlich exprimiert sind, die  $\delta$ -Untereinheit der ATPase allerdings im „Photosynthese-Regulon“ SOM 5-1 wieder zu finden ist. Diese und weitere Beobachtungen sind in Kapitel 4.3 (Seite 92) ausführlich diskutiert.

## 4 Diskussion

### 4.1 Ein 3292-GST-Makroarray zur Untersuchung des plastidären Transkriptoms

Dem Design des 3292-GST-Makroarrays liegt die mittlerweile vollständige Sequenz des *Arabidopsis*-Genoms zugrunde, so dass ein Set von 1827 nukleär kodierten Chloroplastengen (Kurth et al., 2002), noch basierend auf der partiellen genomischen Sequenzinformation, von Varotto und Richly durch cTP-Prädiktion mittels *TargetP* auf 2661 GSTs erweitert werden konnte. Neben 631 zusätzlich gewählten Genen anderer subzellulärer Lokalisationen bilden die korrespondierenden mRNAs das nukleär kodierte Transkriptom des Chloroplasten fast vollständig ab (Richly et al., 2003).

Die große Anzahl von Genen, die Existenz von Introns und die zu Beginn der Sequenzierungsprojekte nur unvollständig verfügbaren Genomsequenzen der Höheren Eukaryoten, ließ bis vor kurzem die direkte Amplifizierung von genspezifischen Sonden (GSTs) von der genomischen DNA nur bei Prokaryoten und einfachen Eukaryoten (Hefe) zu. Stattdessen wurden und werden teilweise nur die Sequenzen von ESTs für die Expressionsanalyse genutzt (Wang et al., 2000; Schaffer et al., 2001). ESTs repräsentieren einen großen Teil des transkribierten Genoms und sind zudem frei von Intronsequenzen. Da aber nur etwa 75% des durch *TargetP* prädizierten nukleär kodierten Transkriptoms als ESTs annotiert ist, besteht der große Vorteil des verwendeten GST-Makroarrays in der Abbildung des gesamten potentiellen Transkriptoms.

Dies konnte gezeigt werden, indem 96 % der korrespondierenden Gene des 3292-GST-Makroarrays in mehr als 10 von 101 Expressionsprofilen differentiell und zumindest in einem der 101 Experimente, signifikant differentiell verschieden exprimiert wurden

(siehe Abb. 3.1, Seite 43). Andererseits sind nur etwa 400 Gene, das entspricht 10 Prozent des abgebildeten Transkriptoms, in jedem (101/101) der Experimente als differentiell exprimiert nachgewiesen worden. Diese Ergebnisse bestätigen zum einen, dass auch der Großteil der nicht als ESTs annotierten Gene transkribiert wurde und unterstreicht damit die Vorteile der GST-Methode gegenüber einem SONDENSATZ von ausschließlich ESTs (nur 75 % der 3292 GSTs sind als ESTs annotiert). Darüber hinaus wird deutlich, dass die Expression der meisten Gene nicht konstitutiv ist und ein Großteil der Gene nur auf einen speziellen Stimulus hin exprimiert wird.

Das generelle Problem der statistischen Analyse der Makroarraydaten ist, dass die Anzahl  $n$  der Datenpunkte pro Gen und Experiment klein ist. Dem gegenüber steht die sehr große Anzahl  $p$  der in einem Experiment untersuchten Parameter (Gene). In der Literatur wird vom „large  $p$ , small  $n$ “-Problem gesprochen. Durch den Einsatz von *ArrayStat* konnten die Ergebnisse der Expressionsanalysen statistisch abgesichert werden. Wurden in früheren Arbeiten noch unflexible Schwellenwerte für die Unterscheidung von differentiell exprimierten Genen festgelegt [z. B.: eine Expressionsrate zwischen Probe und Referenz von  $>1.5$  bzw.  $<0.67$  (Chen, 2001b);  $>2.5$  bzw.  $<0.4$  (Kurth et al., 2002);  $>4$  bzw.  $<0.25$  (Notterman, 2001)], wurde in dieser Arbeit für jeden Vergleich ein individuelles Signifikanzniveau in Abhängigkeit der Qualität und der statistischen Verteilung der Daten berechnet (siehe Kapitel 2.8.2, Seite 35).

Die Reproduzierbarkeit der Ergebnisse durch die Expressionsanalyse mit dem verwendeten GST-Makroarray konnten sowohl durch die Methoden der QRT-PCR (Kurth et al., 2002) als auch mit Northern-Blotting (Kubis et al., 2002) bestätigt werden.

## **4.2 Effekte von 101 unterschiedlichen Stimuli auf die Expression des nukleären Transkriptoms des Chloroplasten**

Die *Class Discovery* versucht verschiedenen Proben aufgrund von Ähnlichkeiten eines oder mehrerer Parameter in Gruppen einzuteilen, ohne dabei eine möglicherweise bereits bestehende Klassenzuordnung der Proben in Bezug auf andere Merkmale zu beachten. In dieser Arbeit wurde das Expressionsverhalten des nukleär kodierten Transkriptoms des Chloroplasten von *A. thaliana* als Antwort auf 101 unterschiedliche

Stimuli untersucht. Von den Stimuli - unterschiedlichen genetischen Status und/oder speziellen Wachstumsbedingungen der *Arabidopsis*-Pflanzen - war bekannt, dass sie sich auf unterschiedliche Chloroplastenfunktionen, darunter die Photosynthese, den Proteinimport in die Plastiden, die Kommunikation zwischen Organelle und Zellkern oder die Transkriptions- und Translationsmaschinerie und viele weitere Vorgänge im Organell und in der Zelle auswirken (Tab 3.1, Seite 47). Das Ziel war, eine Beziehung zwischen den größtenteils bekannten physiologischen Effekten der Stimuli anhand des daraus resultierenden Expressionsmusters des Transkriptoms des Chloroplasten herzustellen (*Class Prediction*). Die zweite Möglichkeit, die Analyse der Gene und der Regulation der Expression des Transkriptoms, ist wiederum mehr oder weniger unabhängig vom Wissen über die Auswirkungen eines einzelnen Stimulus und benötigt Information über die Gene (*Gene Prediction*), unter anderem auch, um bisher nicht klassifizierten Genen eine Funktionen zuordnen zu können (*Gene Discovery*). Dieser grundsätzlich andere Ansatz wird in Kapitel 4.2.2 ausführlich diskutiert.

Das Transkriptom des Chloroplasten scheint sehr dynamisch zu sein. So sind je nach Stimulus und bezogen auf die Referenzbedingung zwischen 2 % (pBroL) und 80 % (pSuc\_30') der Gene signifikant unterschiedlich reguliert (siehe Abb. 3.2; Seite 49). Für den Großteil aller Stimuli gilt, dass sie sich entweder fast ausschließlich induzierend oder aber reprimierend auf die Expressionsrate der Gene auswirken. Nur in etwa 25 % der Bedingungen ist ein mehr oder weniger ausgeglichenes differentiell reguliertes Transkriptom zu beobachten (siehe Abb. 3.2; Abb. 3.7 und Abb. 3.8, Seiten 49, 56 und 58).

Der direkte, paarweise durchgeführte Vergleich von einigen ausgewählten Stimuli untereinander zeigt, dass diese drei Klassen nicht nur durch den allgemeinen Trend der Expression des Transkriptoms gebildet werden, sondern auch größtenteils dieselben Gene entweder gleichsam reguliert sind oder im Vergleich sehr unterschiedlicher Stimuli entgegen gesetzt hoch- oder herunterreguliert werden (siehe Abb. 3.3; Seite 50). So haben sehr ähnliche Stimuli, wie im Falle allelischer Mutanten, eine im hohen Maße qualitativ übereinstimmende transkriptionelle Antwort zur Folge. Gleiches gilt in den meisten Fällen auch für die Bedingungen, denen schon vorher durch die *Class Prediction* (siehe Tab. 3.1, Seite 47) ein sehr ähnlicher physiologischer Effekt zugewiesen worden war. Beispielsweise ist das Expressionsverhalten des Transkriptoms

der dunkeladaptierten *flu*-Mutante der *gun1*-Mutante sehr viel ähnlicher, als dem der lichtadaptierten *laf6*-Mutante (siehe Abb. 3.4, Seite 51; siehe dazu auch Kapitel 1.7.1, Seite 16).

Für den Vergleich aller 101 Stimuli untereinander und auf der Basis des differentiellen Expressionsverhältnis einer gemeinsam exprimierten Schnittmenge von 1616 Genen wurde ein hierarchisches Clustering durchgeführt (siehe Abb. 3.7, Seite 56 und Abb. 3.8, Seite 58). Auch hier finden sich die drei oben beschriebenen Hauptklassen der transkriptionellen Antwort wieder. 40 der untersuchten Stimuli bilden die Transkriptomklasse 1, die aufgrund der Expressionsrate der 1616 gemeinsam exprimierten Gene deutlich von den Stimuli separiert sind, die in ein Transkriptom der Klasse 2 oder Klasse 3 resultieren. Die Transkriptome der Klasse 1 und Klasse 3 (35 Bedingungen) weisen wiederum einander entgegen gesetzt entweder vorherrschend Induktion (Klasse 3) oder aber Repression (Klasse 1) der Genexpression auf.

In Tabelle 3.1 wurde versucht, die möglichen Effekte jeder der getesteten Bedingungen auf die Funktionen des Chloroplasten vorherzusagen. Die Grundlage für die Interpretation der Ergebnisse aus dem Hierarchischen Clustering der 101 differentiell exprimierten Transkriptome bildet zu einem großen Teil die *Class Prediction*, das Wissen über den physiologischen Effekt einer speziellen Wachstumsbedingung oder eines genomischen Status. Schon aufgrund möglicher pleiotropher Effekte ist die Interpretation des Kladogramms der Expressionsprofile schwierig.

Wie zu erwarten war und wodurch die Aussagekraft der Transkriptionskarte (siehe Abb. 3.8, Seite 58) bestätigt wird (deren Grundlage nicht nur die qualitativen, sondern auch die quantitativen Expressionsdaten sind), sind die Expressionsprofile allelischer Mutanten aufgrund ihrer Ähnlichkeit direkte Nachbarn (z.B.: *prs1* und *prs2*, *prs1\_D* und *prs2\_D*, *pete1* und *pete2*, *pete1\_soil* und *pete2\_soil*). Viele weitere physiologisch verwandte Stimuli clustern in jeweils einer der Hauptklassen der transkriptionellen Antworten. So ist *psah2* zusammen mit *psak* und der Doppelmutante *psah2k* sowie weiteren Mutanten der Photosystemuntereinheiten (*psaf*, *psae1* und *psbs*) zusammen in einer Klasse (Klasse 1) wieder zu finden. In dieser Klasse clustern mit *cue1*, *pete* und pStau [wobei Staurosporin die vom Redoxstatus von PSI abhängige und durch Peroxiredoxin vermittelte Expression vieler photosyntheserelevanter Gene unterdrückt (Baier et al., 2004)] darüber hinaus weitere Bedingungen, deren Effekt sich auf die

nukleäre Transkription ebenfalls über die Störung des Redoxstatus der Thylakoidmembran auswirken sollten (diskutiert für *cue1* in Surpin et al., 2002). Neben *cue1* und pStau ist mit *laf6* eine Mutante in Klasse 1 vertreten, die eine Störung in der Tetrapyrrol-biosynthese aufweist und als Folge dessen, die Expression einiger nukleärer Chloroplastengene durch erhöhte Konzentrationen von Protoporphyrinogen IX im Cytosol herabgesetzt sein sollte (siehe auch Kapitel 1.7.1, Seite 16 ff.).

Konsequenterweise finden sich dagegen die *gun1*, *gun5* und die dunkeladaptierte *flu*-Mutante in der Transkriptomklasse 3 wieder (siehe Abb. 3.8). Auch sie besitzen eine Mutation im Tetrapyrrolmetabolismus, allerdings sind in diesen Fällen die Aufhebung der transkriptionshemmenden Effekte durch Mg-ProtoME [*gun1*, *gun5* (Strand et al., 2003) bzw. die ungehemmte Synthese von 5-Aminolävulinsäure 3 [*flu* (Meskauskiene et al., 2002) Auslöser für induzierte Genexpression im Zellkern.

Andererseits wird aber auch deutlich, dass teilweise vermeintlich ähnliche aber nicht identische Stimuli einen ganz unterschiedlichen Trend der Expression des Transkriptoms zeigen können. Beispielsweise clustern einige der PSI-Mutanten (*psad1*, *psag*, *psan*, *psao*) in der jeweils entgegen gesetzten Transkriptomklasse, obwohl sie Untereinheiten desselben Multiproteinkomplexes sind. Das gleiche trifft auch auf und Mutanten der Untereinheiten des Proteinimport-Apparates [*toc75* (Klasse 1); *toc34*, *toc134* (Klasse 2); *toc120*, *toc160*, *tic40*, *ppi1*, *ppi2* (Klasse 3)] zu (siehe Abb. 3.8).

Diese Ergebnisse sind zunächst unerwartet, allerdings ist bei der Interpretation zu berücksichtigen, dass teilweise die Mutanten, die man im Zuge der *Class Prediction* (siehe Tab. 3.1, Seite 47) in einer gemeinsamen Transkriptomklasse erwarten würde (beispielsweise die Photosynthese-Mutanten *psao*, *psan*, *psad* und *psae*), zwar erstaunlicherweise unterschiedliche Expressionsprofile, andererseits aber wiederum auch unterschiedliche Phänotypen aufweisen [siehe dazu: *psae* (Varotto et al., 2000); *psad* (Ihnatowicz et al., 2004); *psao* (Jensen et al., 2004)]. Im Falle der beiden ATPase-Mutanten *atpc* und *atpd* ist zu berücksichtigen, dass sie unter sehr unterschiedlichen Wachstums- und insbesondere Lichtbedingungen gewachsen sind [*atpc*, Tag-Nacht-Rhythmik; *atpd*, 24 h konstantes Licht (siehe Tab. 3.1). Hier scheint der Faktor Licht den Effekt der Mutation in Bezug auf den Einfluss auf die Transkription der nukleären Chloroplastengene vollständig zu überlagern.

Auch können Peptiduntereinheiten ein und desselben Proteinkomplexes unterschiedliche Funktionen besitzen. So wird für die *ChlH*-Untereinheit (*gun5*) der Mg-Chelatase die Funktion eines Konzentrationssensors diskutiert, die allerdings nur zusammen mit der Untereinheiten *ChlD* und *ChlI* Proto zu Mg-ProtoIX umsetzen kann (Karger et al., 2001).

Zudem ist die Regulation der Transkription nicht die einzige Möglichkeit, die Funktion des Proteoms als Antwort auf einen Stimulus zu modifizieren. Die Koordination der Genexpression findet zu einem großen Teil auf posttranskriptionaler Ebene statt (Petracek et al., 1997, 1998; Sherameti et al., 2002) und auch translationale und posttranslationale Prozesse (Zerges et al., 2002) können der Anpassung des Proteoms an unterschiedliche Bedingungen dienen. So nimmt beispielsweise die Photosynthese über den Redoxstatus der TM zwar Einfluss auf die Transkription (Karpinski et al., 1999; Oswald et al., 2001), allerdings werden wohl auch die Stabilität der Transkripte (Petracek et al., 1998), die Beladung der Ribosomen (Tang et al., 2003), die Initiation der Translation (Shen et al., 2001) und die Protein-Akkumulation (Yang et al., 2001) über Redox-Signale modifiziert.

Zumindest auf der Ebene des nukleär kodierten plastidären Transkriptomts scheinen die zum Teil sehr verschiedenen Stimuli in letztendlich zwei vorherrschende transkriptionelle Antworten integriert zu werden und entweder in überwiegend Induktion (Klasse 3) oder Repression (Klasse 1) der Genexpression des Transkriptomts zu resultieren. Dies alles spricht für einen auf transkriptioneller Ebene arbeitenden binären Hauptschalter, so wie er schon zuvor im Clustering der Expressionsprofile von 35 unterschiedlichen Stimuli beobachtet und beschrieben werden konnte (Richly et al., 2003) und so wie dessen Existenz hier durch die Erweiterung des Probensatzes auf 101 Bedingungen eindrucksvoll bestätigt werden konnte. Eine weitergehende Klassifizierung des nukleären Transkriptomts zeigt zudem, dass verschiedene, und zum Teil entgegengesetzte, Regulationsmechanismen der nukleären Genexpression vorliegen (siehe Kapitel 4.2.2).

### 4.2.1 Integration der 101 Stimuli in eine von drei Hauptklassen der transkriptionellen Antwort

Einer der wichtigsten abiotischen Parameter, über den die Genexpression im Zellkern reguliert wird, ist die Lichtqualität und Quantität. Die Entwicklung des Chloroplasten wird über eine Signalkaskade gesteuert, an deren Anfang die Lichtperzeption durch die Photorezeptoren des Cytosols steht, die Phytochrome und Cryptochrome (Moeller et al., 2001). Die Pflanzenzelle reguliert und koordiniert auf diesem Wege unter anderem die Expression von den nukleären und plastidären Genen, die die Komponenten des Photosynthese-Apparates des Chloroplasten kodieren (Strand et al., 2003).

Aus der in dieser Arbeit durchgeführten differentiellen Expressionsanalyse geht hervor, dass bei den meisten getesteten Wachstumsbedingungen eine große Anzahl (zumeist mehr als 1000) von Genen des nukleären Transkriptoms des Chloroplasten

- i) als Antwort auf den jeweiligen Stimulus hin eine signifikant veränderte Expressionsrate im Vergleich zur entsprechenden Referenzbedingung aufweisen (siehe Abb. 3.2, Seite 49) und
- ii) dass 75 % aller Stimuli in einen eindeutigen Trend der Expression des annähernd gesamten Transkriptoms resultieren, entweder vorherrschend Induktion oder vorherrschend Repression.

Diese beiden klar voneinander getrennten Klassen werden durch die bereits bekannten Signalling-Mutanten *gun1*, *gun5* und die dunkeladaptierte *flu*-Mutante in der Transkriptomklasse 3 und durch die infolge der *Class Prediction* als entgegen gesetzten vermuteten Stimuli der lichtadaptierte *flu*-Mutante, der *laf6*- und der *cue1*-Mutante sowie der Behandlung mit Staurosporin (pStau) in Transkriptomklasse 1 charakterisiert (Abb. 3.8, Seite 58).

Neben den Mutanten der Tetrapyrrolbiosynthese sind im Satz der 101 Bedingungen viele zusätzliche Stimuli enthalten, die Veränderungen in einem der vier weiteren Parameter verursachen sollten, die nach heutiger Ansicht als retrogrades Signal von den Plastiden zum Zellkern gerichtet Einfluss auf die nukleäre Genexpression nehmen [z.B.: plastidäre Genexpression (*prp111*); ROS (HL\_1h); Redoxstatus der Thylakoidmembran

und der Thioedoxine (PSI\_DCMU, PSII\_DCMU, pPar, pBro, pFe, Photosystem-Mutanten)].

Allerdings unterliegt nicht das vollständige nukleäre Transkriptom des Chloroplasten der Regulation durch den oben beschriebenen, übergeordneten und binär arbeitenden Hauptschalter (Richly et al., 2003). Schon bei der Betrachtung jedes einzelnen der 101 Expressionsprofile wurde deutlich, dass es sehr häufig neben dem allgemeinen Trend der transkriptionellen Antwort zu entweder vorherrschend Induktion oder Repression des Großteils des nukleären Transkriptoms einige Gene gab, die dem Trend genau entgegen gesetzt exprimiert worden waren (Abb. 3.5, Seite 53). Nach der funktionellen Klassifizierung des durch den 3292-GST-Makroarray abgebildeten Transkriptoms (Abb. 3.5, Seite 52) fallen diese Gene hauptsächlich in die funktionellen Genklassen der Licht- und Dunkelreaktion der Photosynthese sowie mit Abstrichen in die Klasse der Proteinsynthese.

Da aber speziell die photosyntheserelevanten Gene Ziel der retrograden Koordination der Genexpression sind (Surpin et al., 2002; Gray et al., 2003; Strand et al., 2003), scheint das Expressionsverhalten des restlichen (und überwiegenden) Teils des nukleären Transkriptoms des Chloroplasten eingebettet zu sein in die Regulation des gesamten nukleären Transkriptoms aller subzellulärer Lokalisationen. Diese Vermutung ist nur aufgrund des speziellen Designs des verwendeten Makroarrays möglich. Ein Vergleich der Expressionsraten der Gene, die das nukleäre Transkriptom des Chloroplasten repräsentieren mit den Genen, die Proteine anderer subzellulärer Lokalisationen kodieren, zeigt keine eindeutig unterschiedliche Regulation (siehe Abb. 3.12, A und B, Seite 65).

Differentiell unterschiedliche Transkription hat zwar stattgefunden, allerdings ist anzunehmen, dass die durch den Makroarray abgebildeten Teil-Transkriptome des Sekretorischen Systems, des nukleären Transkriptoms der Mitochondrien und anderer subzellulärer Lokalisationen (siehe Abb. 3.11, Seite 64) den allgemeinen Trend des gesamten Transkriptoms der Zelle wiedergeben. Demnach wäre ein allgemeiner, binär arbeitender Hauptschalter dafür verantwortlich, dass bestimmte, bisher nicht identifizierte Signale, die durchaus vom Chloroplasten kommen könnten, als Folge der meisten der untersuchten Stimuli zu entweder vorherrschend Induktion oder vorherrschend Repression der Genexpression im Zellkern und auch des Großteils des

nukleären Transkriptom des Chloroplasten führen. Die koordinierte Expression einer großen Anzahl von Genen als Antwort auf spezielle Stimuli konnte beispielsweise schon in Hefe nachgewiesen werden, wo 900 Gene des Transkriptoms aufgrund von Stressbedingungen gleichsam aktiviert werden konnten (Gasch et al., 2000).

Ein zweiter integrativer Regulator, vermeintlich arbeitend in der Art, wie er von Gray und Mitarbeitern (2003) beschrieben worden ist (Abb. 4.1), wäre jedoch diesem allgemeinen binären Hauptschalter wiederum übergeordnet und dient der speziellen Koordination der Genexpression photosyntheserelevanter Gene aufgrund verschiedener retrograder Signale vom Chloroplasten an den Zellkern. Diese Vermutung wird im Weiteren ausführlich diskutiert.

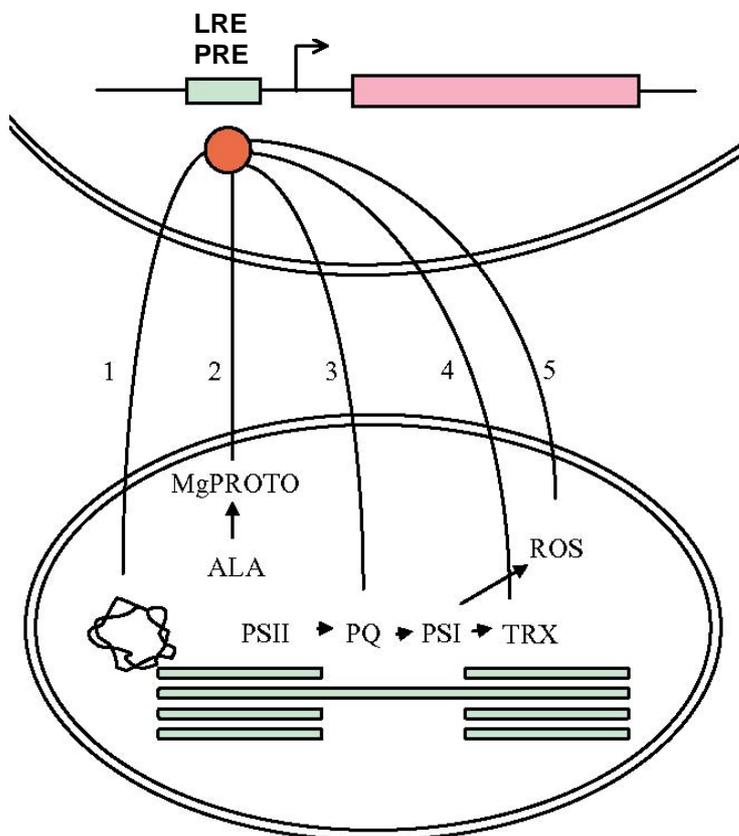
#### **4.2.2 Hinweise für einen integrativen Regulator der Transkription photosyntheserelevanter Gene**

Es gibt mittlerweile Hinweise, dass zumindest drei verschiedene Signalwege existieren, die vom Chloroplasten zum Zellkern gerichtet sind und ebenso wie die Photorezeptoren Einfluss auf die nukleäre Genexpression nehmen (Mochizuki et al., 2001; Surpin et al., 2002). Gray und Mitarbeiter (2003) gehen sogar von insgesamt fünf verschiedenen Prozessen innerhalb des Plastiden aus, durch deren Vorgänge Signale generiert werden, die die Expression vor allem von photosyntheserelevanten Genen im Zellkern direkt oder indirekt beeinflussen (Abb. 4.1).

Dazu gehören neben den Signalen ausgehend von den Intermediaten des Tetrapyrrolmetabolismus (Mochizuki et al., 2001), dem Redoxstatus der Thylakoidmembran und dem Einfluss von ROS (Pfannschmidt et al., 2001), auch die Synthese plastomkodierter Proteine und der Redoxstatus der Thiothione.

Gray und Mitarbeiter (2003) diskutieren einen integrativen, bisher unbekanntem Regulator der nukleären Transkription, über den sich diese Signale auswirken und der über ein spezielles Promotorelement der ansprechbaren Gene die nukleäre Expression entscheidend beeinflusst. Dabei interagiert der integrative Regulator mit demselben

Promotorelement, über das auch als *light response element* (LRE) die lichtabhängige Expressionskontrolle vermittelt wird, über das aber im speziellen Fall der Plastidsignale als *plastid response element* (PRE) Einfluss auf die Genexpression genommen wird.



**Abb. 4.1.: Modell der vom Plastiden zum Zellkern gerichteten Signalwege und der Interaktion mit den Promotoren photosyntheserelevanter, nukleärer Gene.** Die Regulation der nukleären Genexpression ist vermutlich von zumindest fünf verschiedenen Prozessen innerhalb der Plastiden beeinflusst. Die Signalwege sind abhängig von: (1) Expression des Plastoms; (2) Tetrapyrrolsynthese; (3) Redoxstatus des Plastochinon-Pools; (4) Redoxstatus der Thioredoxine; (5) ROS. Im oberen Teil der Abbildung ist ein nukleäres Gen schematisch dargestellt, das ein Promotorelement besitzt, das sowohl über lichtabhängige (LRE) als auch über Signale des Plastiden (PRE) ansprechbar ist. Die kodierende Region des Gens für ein photosyntheserelevantes Protein ist in Rosa gekennzeichnet. Der rote Kreis symbolisiert einen allgemeinen, bisher nicht identifizierten Faktor und Regulator der Transkriptionsmaschinerie, der die Signale verschiedener Signalwege integriert. ALA, 5-Aminolävulinsäure; LRE (*light response element*); MgPROTO, Mg-Protoporphyrin; PSI, Photosystem I, PSII, Photosystem II; PRE (*plastid response element*); PQ, Plastochinon; ROS (*reactive oxygen species*); TRX, Thioredoxin (modifiziert von Gray et al., 2003).

Wie schon in Kapitel 1.7.1 (Seite 16) beschrieben wurde, basiert diese Vermutung auf den vielfältigen Expressionsanalysen mit Carotinoid-defizienten und mit dem Herbizid Norflurazon behandelten Wildtyp-Pflanzen (Mayfield & Taylor, 1984; Batschauer et al., 1986; Oelmüller et al., 1986; Burgess & Taylor, 1987; Sagar et al., 1988).

Die Ergebnisse lassen darauf schließen, dass ein retrogrades Signalsystem, gerichtet vom Chloroplasten zum Zellkern, für die Regulation der Transkription der zumeist photosyntheserelevanten Gene verantwortlich ist. Allerdings war es für keines dieser Gene möglich, ein regulatorisches Promoterelement - ansprechbar auf Behandlung mit Norflurazon - von den durch Licht regulierten und als LRE bekannten Promoterelementen zu separieren. Tatsächlich zeigten transgene Tabakpflanzen, die das GUS-Reportergen unter der Kontrolle eines Promoters mit vielfachen Kopien von LREs trugen, deutlich herabgesetzte Expression nach Behandlung mit Norflurazon (Martinez-Hernandez et al., 2002).

Die ersten Hinweise für den Einfluss der plastidären Genexpression auf die Expressionsrate nukleär kodierter Gene ergaben sich aus der Analyse der Enzyme des Calvin-Zyklus (Bradbeer et al., 1979). Gerstenpflanzen, denen plastidäre Ribosomen fehlten, zeigten deutlich reduzierte Konzentrationen von Phosphoribulokinase und Glyceraldehyd-3-Phosphat-Dehydrogenase (*GAP*). Hess und Mitarbeiter (1991, 1994) konnten in anschließenden Untersuchungen zeigen, dass in diesen Pflanzen auch andere Komponenten des Calvin-Zyklus, der LHCs und Enzyme der Photorespiration auf transkriptioneller Ebene reprimiert waren.

Durch Expressionsanalysen von Wildtyp-Pflanzen, in denen die plastidäre Translation durch chloroplastenspezifische Inhibitoren wie Chloramphenicol oder Streptomycin gehemmt war, konnte gezeigt werden, dass auch hier die Expression einer Reihe von nukleären, photosyntheserelevanten Genen (darunter wieder *Lhcb1* und *RbcS*) herabgesetzt war (Sullivan & Gray, 1999).

Zusammen mit den bereits dargestellten Plastidsignalen, vermittelt durch die Zwischenprodukte aus der Tetrapyrrolbiosynthese und durch den Redoxstatus der TM (Kapitel 1.7.1, Seite 16), wird die Rolle der Thioredoxine als Regulatoren der nukleären Transkription noch konträr diskutiert (Gray et al., 2003). Jedoch deutet vieles darauf hin, dass einige der direkt durch die Vorgänge der Photosynthese im Chloroplasten generierten Signale Einfluss auf die nukleäre Transkription von hauptsächlich - aber

nicht nur - der Gene nehmen, die photosyntheserelevante Proteine kodieren. Dabei scheint sich der retrograde Signalweg vom Chloroplasten zum Zellkern derselben regulatorischen Promoterelemente zu bedienen, durch den auch die lichtabhängige Genexpression reguliert wird (Puente et al., 1996).

Im Datensatz der 101 Expressionsprofile lieferte die funktionelle Klassifizierung des signifikant, differentiell unterschiedlich exprimierten Transkriptoms des Chloroplasten einen ersten Hinweis für die Existenz zweier unterschiedlicher Regulationsmechanismen. Das in der Abbildung 3.3 (Seite 50) anhand der Expressionsprofile von *pete1* und *prs1* gezeigte Phänomen von Induktion der Photosynthesegene bei gleichzeitiger Repression des restlichen Transkriptoms beziehungsweise Repression der Photosynthesegene bei gleichzeitiger Induktion des restlichen Transkriptoms war ähnlich auch für die Signalling-Mutanten *gun1*, *gun5* und die dunkeladaptierte *flu*-Mutante (Schema *pete1*) sowie für die lichtadaptierte *flu*, *laf6*- und *cue1*-Mutante (Schema *prs1*) zu beobachten (Daten nicht gezeigt).

Durch die zuvor schon in Bezug auf die Probenbedingungen besprochene Clusteranalyse ließ sich auch das Expressionsverhalten der 1616 Gene des in 101 Bedingungen gemeinsam differentiell exprimierten, nukleären Transkriptoms des Chloroplasten genauer untersuchen. Unter Vorgabe der 25 zu erwartenden Genclustern aus dem Hierarchischen Clustering wurden durch SOM-Clustering 25 Regulons definiert (Abb. 3.7, Seite 56), die dann für die weiteren Analysen benutzt wurden.

Schon im Hierarchischen Clustering ist erkennbar, dass sich neben dem Großteil der entweder vorherrschend induzierten oder überwiegend reprimierten Gene in den Transkriptomklassen 1 bzw. 3 auch Regulons finden, deren Gene sich genau entgegengesetzt zum allgemeinen Trend der Expression verhalten (Abb. 3.7, Regulon 1 und 2, 23 und 24). Die speziellen Expressionsmuster und die Gene dieser Regulons wurden auch durch die Methode des SOM-Clustering wieder gefunden (Abb. 3.10, Seite 62). Aus dem Hierarchischen Clustering der Regulons geht hervor, dass die Gene von fünf der 25 Regulons sehr unterschiedlich zum Rest des Transkriptoms reguliert zu sein scheinen (Abb. 3.11, Seite 64).

Bei Betrachtung der subzellulären (Abb. 3.11) und subplastidären (Abb. 3.14, Seite 68) Lokalisationen der Gene jedes Regulons fällt auf, dass insbesondere in den Regulons

SOM 5-1 und SOM 4-1 eine Akkumulation von fast ausschließlich Chloroplastengen gefunden hat und dass auf subplastidärer Ebene die Gene, die Proteine der Thylakoide kodieren, eindeutig überrepräsentiert sind.

Die funktionelle Klassifizierung der Gene dieser Regulons definiert das Regulon SOM 5-1 als einen Gencluster, in dem hauptsächlich Gene der Licht- und Dunkelreaktion der Photosynthese sowie eine überdurchschnittlich hohe Zahl an Genen der Proteinsynthese und der Proteinmodifikation miteinander geclustert haben (Abb. 3.15, Seite 69), während in Regulon SOM 4-1 vorherrschend Gene der Proteinsynthese, des Photosynthese-Apparates und (mit Abstrichen) der Proteinmodifikation wieder zu finden sind (Abb. 3.17, Seite 71). Beide Regulons erweisen sich durch die subzellulären und subplastidären Lokalisationen und die Verteilung der Gene auf die funktionellen Genklassen, sowie durch ein zum Rest des Transkriptom sehr spezielles Expressionsverhalten (siehe Abb. 3.11, Seite 64), als erste Kandidaten für die Suche nach einem speziellen Modus der Regulation der Genexpression.

Im Promoterbereich der Gene des Regulons SOM 5-1 sind auch bei strikter Auswahl des Signifikanzintervalls [p-Wert  $< 10^{-5}$  (Hudson & Quail, 2003)] eindeutig überrepräsentierte Sequenzbereiche auszumachen. Die vier G-Box Elemente und das gefundene I-Box Element sind schon zuvor als *cis*-regulatorische Sequenzmotive einer durch Phytochrom A regulierten, lichtabhängigen Signalkaskade beschrieben worden (Hudson & Quail, 2003; Toledo-Ortiz et al., 2003; Huq et al., 2004; Escobar et al., 2004).

Ein detaillierter Blick auf die Gene der Regulons SOM 5-1 und SOM 4-1 zeigt, dass in beiden Regulons in weitere funktionelle Subklassen unter der Ebene von Photosynthese und Proteinsynthese unterschieden werden kann. So kodieren über 85 % der Photosynthesegene des Regulons SOM 5-1 für eine der Peptiduntereinheiten der vier integralen Multiproteinkomplexe in der TM (Tab 3.3, Seite 75). Fast für jedes dieser Gene konnte mittlerweile nachgewiesen werden, dass die Genexpression unter anderem auch durch die retrograden Signalwege des *plastid-to-nucleus* Signalling reguliert wird [z.B.: *psbO*, *psbP*, *petC* und *fed-1* (Reiss et al., 1983), *atpC*, *atpD*, *petE* und *psaF* (Bolle et al., 1994, 1996), *lhcb1* und *rbcS* (Giuliano et al., 1988)].

Auch viele der übrigen Gene dieser Regulons wurden bereits zuvor als Ziel der Expressionsregulation über das *plastid-to-nucleus* Signalling identifiziert. Hess und

Mitarbeiter (1991, 1994) beobachteten, dass die Transkription von verschiedenen nukleär kodierten Enzymen des Calvin-Zyklus, der Photorespiration und auch die Glyceraldehyd-3-Phosphate-Dehydrogenase (*GAP*) durch Inhibition der Genexpression im Plastom herabgesetzt werden konnte. Außer der Phosphoglycerat-Kinase und den Peptiduntereinheiten der RuBisCo (Calvin-Zyklus) finden sich in Regulon SOM 5-1 auch alle drei *gap*-Gene des Datensatzes sowie der Glycin-Decarboxylase-Komplex H (involviert in Photorespiration) wieder. Mehr noch konnten Marri und Mitarbeiter (2004) nachweisen, dass *GAP* und die Phosphoribulokinase (*PRK*) über die Bindung des Peptids *CPI2* miteinander interagieren. Die Bildung dieses Supramolekular-komplexes tritt zusammen mit der lichtabhängigen Modulation der Aktivität von *GAP* und *PRK in vivo* auf. Konsequenterweise sind außer den drei Genkopien der *GAP* auch die *PRK* und *CPI2* im Regulon vertreten (Tab. 3.3).

Neben den Genen des Photosynthese-Apparates finden sich mit *hcf106* (Mori & Cline, 2002) und *alb3* (High et al., 1997; Bellafiore et al., 2002) zwei Gene, deren Genprodukte als Bestandteil des Protein-Import-Apparates der TM direkt an der funktionellen Aggregation der Photosysteme beteiligt sind. Darüber hinaus clustern im selben Regulon weitere Gene, die für Enzyme des proteolytischen Umsatzes der Proteinkomplexe der TM kodieren [*ClpC-V*, *nClpP1* und *VAR2* (Sakamoto et al., 2002, 2003)] und dadurch direkt mit den Vorgängen der Photosynthese in Verbindung stehen. Interessanterweise sind viele der Gene des Regulons SOM 5-1 direkt mit den retrograden Kommunikationswegen zwischen Organelle und Zellkern in Verbindung zu bringen. Der Zusammenhang zwischen der Translationsaktivität an den plastidären Ribosomen und der Transkriptionsrate nukleär kodierter Chloroplastengene (Hess et al., 1991, 1994) wurde schon am Beispiel von *GAP* gezeigt. Gleiches gilt für die Peptiduntereinheiten des Photosynthese-Apparates und ihre Bedeutung für den Redoxstatus der TM. Der Elektronenfluss über die in Serie geschalteten PSII und PSI bestimmt darüber hinaus über die Ferredoxin-Thioredoxin-Reduktase direkt den Redoxstatus der Thioredoxine, denen von Pfannschmidt und Mitarbeitern (2001) ebenfalls eine Rolle in der Kommunikation zwischen Chloroplast und Zellkern zugeschrieben wurde.

Die plastidären Ribosomen und die Thioredoxine clustern ebenso in Regulons SOM 5-1 wie zwei der Schlüsselenzyme des letzten bekannten retrograden Signalweges, der

durch die Intermediate des Tetrapyrrolmetabolismus vermittelt wird. Die *CHLI*-Untereinheit der Mg-Protoporphyrin IX Chelatase ist als eine von drei Peptiduntereinheiten am Umsatz von Protoporphyrin IX zu Mg-Protoporphyrin IX beteiligt (siehe auch Abb. 1.6, Seite 21 und Kapitel 1.7.1, Seite 16 ff.). Mg-Protoporphyrin IX und Mg-Protoporphyrin IX 6-Monomethylester (MgProtoMe) gelten als retrograde Signale, die im Cytoplasma integriert werden [siehe Abb. 4.1 (Gray et al., 2003)] und zu einer Inhibition der nukleären Transkription photosyntheserelevanter Gene führen (Surpin et al., 2002).

Erst kürzlich wurde eine Untereinheit der MgProtoMe-Cyclase (*CHL27*) identifiziert. *CHL27* ist an der Umsetzung von MgProtoMe zu Divinyl-Protochlorophyllid beteiligt. Antisense-Linien von *A. thaliana* mit reduziertem Gehalt an *CHL27* weisen eine erhöhte Konzentration an MgProtoMe auf und besitzen chlorotische Blätter mit reduzierter Abundanz fast aller Proteinuntereinheiten der Photosysteme (Tottey et al., 2003). Diese Ergebnisse bekräftigen die Hypothese, nach der MgProtoMe ein hemmendes retrogrades Signal für die nukleäre Transkription photosyntheserelevanter Gene darstellt. Diese beiden Gene des Tetrapyrrolmetabolismus clustern zusammen mit *hsp70*, dessen lichtabhängige Expressionsinduktion in *CHLI*-defizienten, nicht aber in Null-Mutanten der *CHLH*-Untereinheit der Mg-Protoporphyrin IX Chelatase deutlich gestört ist (Brusslan & Peterson, 2002). Dazu passt, dass das *chlH*-Gen in Regulon SOM 1-4 zu finden ist und zumindest auf transkriptioneller Ebene ganz anders reguliert zu sein scheint, als die *CHLI*-Untereinheit.

### **4.3 Netzwerk der Expressionsregulation des nukleären Transkriptoms des Chloroplasten**

Es fällt auf, dass im Regulon SOM 5-1 viele der Gene clustern, die sowohl als Ziel der bisher bekannten retrograden Signalwege diskutiert werden, als auch deren Genprodukte eine Beteiligung an der Generierung des Signals zugesprochen wird. Die überrepräsentierten Sequenzmotive in der Promoterregion der Gene dieses Regulons lassen darüber hinaus schlussfolgern, dass, wie schon von Gray und Mitarbeitern (2003)

postuliert wurde, die retrograde Signaltransduktion die Genexpression über dieselben *cis*-Elemente reguliert, über die auch die lichtabhängige Signalkaskaden der Phytochrome und Cryptochrome im Zuge der Chloroplastenentwicklung wirken.

Durch die Zusammenstellung der Gene des Regulons SOM 4-1 bestätigt sich, dass mit der Photosynthese und der Proteinsynthese gerade die Chloroplastenfunktionen eine spezielle Regulation der Genexpression aufweisen, deren Proteine noch immer in zwei verschiedenen Genomen der Zelle kodiert sind. Außer den Genen des genetischen Apparates des Chloroplasten sind auch in diesem Regulon weitere Gene zu finden, deren Genprodukte am proteolytischen Umsatz der Proteine der TM beteiligt sind oder die durch die Aktivität des Photosynthese-Apparates direkt beeinflusst werden (Thioredoxine, Rubredoxin).

Es erscheint logisch, dass ein Mechanismus existiert, der, gesteuert über retrograde Signale vom Chloroplasten zum Zellkern, vor allem der koordinierten Genexpression von plastomkodierten und kernkodierten Proteinen dient. Nur durch eine spezielle Art der Transkriptionsregulation ist es möglich, dass die Expression der kernkodierten Photosystem-Untereinheiten und des nukleären genetischen Apparates des Chloroplasten der allgemeinen transkriptionellen Reaktion durch den binären Hauptschalter entkommt. Nach den Ergebnissen der hier vorgestellten Transkriptom-analyse, erscheint eine Steuerung der koordinierten Genexpression über die nukleäre Kontrolle der Ribosomen-Abundanz und/oder die Aktivität der plastomkodierten RNA-Polymerase durch die kernkodierten  $\sigma$ -Faktoren und/oder eine direkte *Feedback*-Kontrolle über die retrograden Signale des *plastid-to-nucleus* Signalling wahrscheinlich.

#### 4.4 Interpretation aus evolutionärer Sicht

Der Existenz eines allgemeinen Hauptschalter der nukleären Transkription und eines speziellen Regulationsmechanismus für die Transkription der kernkodierten Proteinsynthese- und Photosynthesegene, dürften die Folgen aus dem Endosymbiontischen Gentransfer und dem Erwerb neuer Funktionen des Chloroplasten im Verlauf der Endosymbiose sein. Ein Großteil der nukleär kodierten Genprodukte des heutigen

Proteoms des Chloroplasten ist erst im Zuge der Endosymbiose des ehemaligen Cyanobakteriums in die eukaryotische Wirtszelle neu erworben worden [so wie beispielsweise der Protein-Import-Apparat der äußeren und inneren Plastidmembran (Jarvis & Soll, 2001)]. Für etwa 1800 Proteine des plastidären Proteoms finden sich folglich keine Homologe in den heute noch frei lebenden Cyanobakterien (Martin et al., 2002). Etwa 1300 Gene des Transkriptoms des Chloroplasten sind im Verlauf des der Endosymbiose durch die Vorgänge des EGT vom Organell an den Zellkern transferiert worden, während etwa 90 Proteine noch immer im Plastom kodiert sind (Leister & Schneider, 2003). Der Flaschenhals des EGT ist die stabile Genexpression eines transferierten Gens im Kern durch den Erwerb eines funktionellen Promoters (Martin & Herrmann, 1998). Dadurch wurden sowohl das heute nukleär kodierte Transkriptom cyanobakteriellen Ursprungs, als auch die neu erworbenen Genfunktionen, Teil eines eukaryotischen Systems unter der regulatorischen Kontrolle des Zellkerns (Herrmann, 1997).

Die Gene des Photosynthese-Apparates und der plastomkodierten genetischen Maschinerie zur Transkription dieser Gene und der Translation der Proteine sind dagegen wohl aufgrund positiver Selektion im Plastom verblieben (Ludovico et al., 2002; Leister & Schneider, 2003). Entscheidend dafür scheint die Notwendigkeit der direkten Regulation dieser Gene durch die Aktivität ihrer Genprodukte in der Photosynthese und dem Elektronentransport zu sein, die das empfindliche Redoxgleichgewicht im Organell zu gewährleisten und die Entstehung von hoch toxischen oxidativen Radikale zu verhindern (Allen, 2003).

Für den überwiegenden Teil des nukleären Transkriptoms des Chloroplasten scheint, zumindest auf Ebene der Transkription, keine spezielle Regulation der Genexpression vorzuliegen und auch nicht benötigt zu werden. Ein Hinweis dafür ist die hier gezeigte Existenz eines transkriptionellen Hauptschalters, der sowohl die Genexpression dieser Chloroplastengene, als auch der Gene anderer subzellulärer Lokalisationen, gleichsam reguliert.

Im Falle des Photosynthese-Apparates und der genetischen Maschinerie des Chloroplasten war es dagegen aufgrund des dualen genetischen Ursprung nötig, die Expression dieser Gene aufeinander abzustimmen, so dass beispielsweise die Stöchiometrie der Multiproteinkomplexe in der TM gewährleistet bleibt.

## 4.5 Ausblick

Um die hier gemachten Beobachtungen noch genauer untersuchen und die Schlussfolgerungen untermauern zu können, wäre es sehr hilfreich, parallel zur Expression des kernkodierte, auch das plastomkodierte Transkriptom des Chloroplasten zu untersuchen. Drei der vier Plastomgene des Datensatzes clustern in Regulon SOM 5-1 bzw. Regulon SOM 4-1, jedoch ist eine statistische Aussage nicht möglich.

Von besonderem Interesse sind die bisher nicht klassifizierten Gene dieser beiden Regulons. Sie stellen ausgewiesene Kandidaten dar, um durch das *guilt-by-association*-Prinzip, ebenso wie der Großteil der bereits funktionell klassifizierten Gene der Regulons, direkt oder indirekt an den Vorgängen der Photosynthese oder Proteinsynthese beteiligt zu sein und/oder sogar ebenfalls eine Rolle in den Signalwegen der Kommunikation zwischen den Chloroplasten und dem Zellkern zu spielen. Aktuell versuchen wir in der Arbeitsgruppe von Dario Leister, diese Frage durch reverse Genetik zu beantworten.

## 5 Zusammenfassung / Summary

### 5.1 Zusammenfassung

Die Chloroplasten haben als Folge ihrer Entwicklung vom ehemals frei lebenden Cyanobakterium hin zum biosynthetischen Zellorganell einer eukaryotischen Zelle die Fähigkeit verloren, alle zum Leben beziehungsweise zum Überleben benötigten Proteine selbstständig zu synthetisieren. Die postendosymbiontische Evolution des Proto-Chloroplasten war geprägt von einem massiven Gentransfer in den Zellkern, so dass mehr als 90 % des Proteoms heutiger Chloroplasten kernkodiert und dadurch auch die Regulation der Chloroplastenfunktionen größtenteils auf den Zellkern übergegangen ist.

Die Entwicklung eines vollständig funktionellen Chloroplasten ist somit abhängig von der koordinierten Expression von kernkodierten und plastomkodierten Genen als Antwort auf sowohl entwicklungspezifische als auch umgebungsbedingte Signale. Die Regulation der Transkription dieser Gene ist eine effektive Methode, um Einfluss auf die Funktionen des Chloroplasten zu nehmen. Die Transkriptionsregulation einer Reihe von Genen des Kerns, insbesondere des zu annähernd gleichen Teilen im Kern und Plastom kodierten photosynthetischen Apparates, ist abhängig von der Präsenz funktioneller Plastiden in der Zelle und beinhaltet retrograde Signale, die vom Chloroplasten zum Zellkern gerichtet sind.

In dieser Arbeit wurde durch einen Makroarray die differentielle Expression von 3292 nukleären Genen, die größtenteils Proteine des Plastiden kodieren und fast das vollständige kernkodierte Transkriptom des Chloroplasten darstellen, unter 101 verschiedenen Bedingungen untersucht. Eine gemeinsame Schnittmenge von 1616 konstitutiv exprimierten Genen zeigte nur drei übergeordnete Hauptklassen der transkriptionellen Antwort. Die meisten der getesteten Stimuli resultieren in entweder

vorherrschend induzierte Transkription oder aber vorherrschend reprimierte Transkription der meisten und überwiegend jeweils identischen Gene dieses Teils des plastidären Transkriptoms. Anscheinend existiert ein integrativer Hauptschalter, der das nukleäre Transkriptom des Chloroplasten als ein Ganzes reguliert.

Anhand ihres Expressionsverhaltens ließen sich die 1616 Gene in 25 voneinander separierte, ko-exprimierte Gruppen (Regulons) einteilen. Ausgenommen der Regulons SOM 4-1 und SOM 5-1 sind alle anderen Regulons heterogen und beinhalten Gene, die Proteine unterschiedlicher subzellulärer Lokalisationen und verschiedener biochemischer Funktionen kodieren. Das lässt darauf schließen, dass die verschiedenen Zellorganellen und/oder verschiedenste Stoffwechselwege der Zelle zumindest auf der Ebene der Transkription im Zellkern koordiniert sind.

Die Ausnahme sind die Gene des Kerngenoms, die die Untereinheiten des Photosynthese-Apparates oder aber Proteine, die beteiligt sind an der Transkriptions- und Translationsmaschinerie der plastomkodierten Gene (insbesondere Ribosomen-Untereinheiten) (Regulon SOM 4-1 und Regulon SOM 5-1) kodieren. Die Ko-Expression dieser Gene deutet darauf hin, dass ein bisher unbekannter Mechanismus, möglicherweise über die nukleäre Kontrolle der Ribosomen-Abundanz im Chloroplasten, die Genexpression gerade des Teils des Proteoms koordiniert, der noch immer sowohl im Kern als auch im Plastom kodiert ist. Die Ergebnisse aus den Expressionsanalysen unterstützen zudem die Vermutung, dass die retrograden Signale vom Organell zum Zellkern an der Regulation der nukleären Transkription maßgeblich beteiligt sind.

Die Ko-Regulation der Gene der Photosynthese und der ribosomalen Proteine des Chloroplasten entzieht sich der oben dargestellten generellen Expressionsregulation von nukleären Chloroplastengen, die kontrolliert wird durch einen kürzlich beschriebenen transkriptionellen Hauptschalter. Durch den Vergleich zu anderen Chloroplastenfunktionen zeigt diese Beobachtung die Existenz eines speziellen Modus der transkriptionellen Regulation der Photosynthese an. Aus der evolutionären Sicht deuten diese Ergebnisse darauf hin, dass die funktionelle Integration des Proto-Chloroplasten in die eukaryotische Wirtszelle mit der Etablierung von verschiedenen Ebenen der nukleären Kontrolle der Transkription einherging.

## 5.2 Summary

Along with their development from a once free-living cyanobacteria to the biosynthetic organelles of the eucaryotic cell, chloroplasts have lost their ability to synthesize all the proteins needed to live and survive on their own. Post-endosymbiotic evolution of the proto-chloroplast was characterized by gene transfer to the nucleus. Hence, today more than 90 % of the chloroplast proteins are nuclear-encoded and the regulation of chloroplast functions includes nuclear transcriptional control.

The development of a fully functional chloroplast therefore depends on the coordinate expression of nuclear and chloroplast genes in response to both developmental and environmental signals. Control of transcript abundance of these genes is an effective strategy for regulating chloroplast functions. Transcriptional Regulation of a whole string of nuclear genes, especially those coding for the photosynthetic apparatus which is encoded almost equally in both the nuclear and the plastid genome, depends on the presence in the cell of functional plastids and involves retrograde signals that originate in the chloroplast acting to regulate transcription of nucleus-encoded photosynthetic genes.

In this work, by means of a GST macroarray the expression profiles of 3292 nuclear *Arabidopsis* genes, most of them encoding chloroplast proteins and representing almost the whole nuclear plastid transcriptome, were determined from 101 different conditions. A common intersection of 1616 most-constitutive expressed genes only showed three main classes of response from the nuclear chloroplast transcriptome. Most of the 101 environmental and genetic conditions tested resulted in either predominantly induction of gene expression or prevalent repression of transcription of always highly identical subsets of genes indicating the existence of a major switch that regulates the nuclear chloroplast transcriptome as a whole.

Due to their expression, the 1616 genes fell into 25 distinct groups of co-regulated genes (regulons). Except regulons SOM 4-1 and SOM 5-1, regulons are heterogeneous and consist of genes coding for proteins with different subcellular locations or contributing to several biochemical functions. This implies that different organelles and/or metabolic pathways are co-ordinated at the nuclear transcriptional level.

The co-expression of nuclear genes coding for subunits of the photosystems or encoding proteins involved in the transcription/translation of plastome genes (particularly ribosome polypeptides) (regulons SOM 4-1 and SOM 5-1, respectively) implies the existence of a novel mechanism that co-ordinates plastid and nuclear gene expression and involves nuclear control of plastid ribosome abundance. The results from the expression profiling also supports the conjecture that the retrograde signals from the chloroplasts play an operative role in the regulation of nuclear transcription.

The co-regulation of genes for photosystem and plastid ribosome proteins escapes a previously described general control of nuclear chloroplast proteins imposed by a transcriptional master switch, highlighting a mode of transcriptional regulation of photosynthesis which is different compared to other chloroplast functions. From the evolutionary standpoint, the results provided indicate that functional integration of the proto-chloroplast into the eucaryotic cell was associated with the establishment of different layers of nuclear transcriptional control.

## 6 Literatur

- Abdallah, F., F. Salamini und D. Leister (2000). A prediction of the size and evolutionary origin of the proteome of chloroplasts of Arabidopsis. Trends Plant Sci **5**(4): 141-2.
- Allen, J. F. (2003). The function of genomes in bioenergetic organelles. Philos Trans R Soc Lond B Biol Sci **358**(1429): 19-37; discussion 37-8.
- Allen, J. F. und J. Forsberg (2001). Molecular recognition in thylakoid structure and function. Trends Plant Sci **6**(7): 317-26.
- Alon, U., N. Barkai, D. A. Notterman, K. Gish, S. Ybarra, D. Mack und A. J. Levine (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc Natl Acad Sci U S A **96**(12): 6745-50.
- Aro, E. M., I. Virgin und B. Andersson (1993). Photoinhibition of Photosystem II. Inactivation, protein damage and turnover. Biochim Biophys Acta **1143**(2): 113-34.
- Baier, M., E. Stroher und K. J. Dietz (2004). The acceptor availability at photosystem I and ABA control nuclear expression of 2-Cys peroxiredoxin-A in Arabidopsis thaliana. Plant Cell Physiol **45**(8): 997-1006.
- Barkan, A. und M. Goldschmidt-Clermont (2000). Participation of nuclear genes in chloroplast gene expression. Biochimie **82**(6-7): 559-72.
- Batschauer, A., E. Mosinger, K. Kreuz, I. Dorr und K. Apel (1986). The implication of a plastid-derived factor in the transcriptional control of nuclear genes encoding the light-harvesting chlorophyll a/b protein. Eur J Biochem **154**(3): 625-34.
- Bellafiore, S., P. Ferris, H. Naver, V. Gohre und J. D. Rochaix (2002). Loss of Albino3 leads to specific depletion of the light-harvesting system. Plant Cell **14**(9): 2303-14.

- Biehl A, E. Richly, C. Noutsos, F. Salamini und D. Leister (2004). Analysis of 101 nuclear transcriptomes reveals 23 distinct regulons and their relationship to metabolism, chromosomal gene distribution and co-ordination of nuclear and plastid gene expression. Gene (im Druck).
- Blobel, G., P. Walter, C. N. Chang, B. M. Goldman, A. H. Erickson und V. R. Lingappa (1979). Translocation of proteins across membranes: the signal hypothesis and beyond. Symp Soc Exp Biol **33**: 9-36.
- Bolle, C., V. V. Kusnetsov, R. G. Herrmann und R. Oelmüller (1996). The spinach AtpC and AtpD genes contain elements for light-regulated, plastid-dependent and organ-specific expression in the vicinity of the transcription start sites. Plant J **9**(1): 21-30.
- Bolle, C., S. Sopory, T. Lubberstedt, R. G. Herrmann und R. Oelmüller (1994). Segments encoding 5'-untranslated leaders of genes for thylakoid proteins contain cis-elements essential for transcription. Plant J **6**(4): 513-23.
- Bolter, B., J. Soll, A. Schulz, S. Hinnah und R. Wagner (1998). Origin of a chloroplast protein importer. Proc Natl Acad Sci U S A **95**(26): 15831-6.
- Bosco, C. D., L. Lezhneva, A. Biehl, D. Leister, H. Strotmann, G. Wanner und J. Meurer (2004). Inactivation of the chloroplast ATP synthase gamma subunit results in high non-photochemical fluorescence quenching and altered nuclear gene expression in *Arabidopsis thaliana*. J Biol Chem **279**(2): 1060-9.
- Bouchez, D. und H. Hofte (1998). Functional genomics in plants. Plant Physiol **118**(3): 725-32.
- Bradbeer, J.W., Y.E. Atkinson, T. Börner, R. Hagemann (1979). Cytoplasmic synthesis of plastid polypeptides may be controlled by plastid-synthesized RNA. Nature **279**, 816-17.
- Brazma, A. und J. Vilo (2000). Gene expression data analysis. FEBS Lett **480**(1): 17-24.
- Brown, D. E., A. M. Rashotte, A. S. Murphy, J. Normanly, B. W. Tague, W. A. Peer, L. Taiz und G. K. Muday (2001). Flavonoids act as negative regulators of auxin transport in vivo in *arabidopsis*. Plant Physiol **126**(2): 524-35.
- Brown, E. C., A. Somanchi und S. P. Mayfield (2001). Interorganellar crosstalk: new

- perspectives on signaling from the chloroplast to the nucleus. Genome Biol **2**(8): 1021.
- Bruce, B. D. (2001). The paradox of plastid transit peptides: conservation of function despite divergence in primary structure. Biochim Biophys Acta **1541**(1-2): 2-21.
- Brusslan, J.A. Peterson, M.P. (2002). Tetrapyrrole regulation of nuclear gene expression. Photosynthesis Research **71**: 185-94.
- Burgess, C. D., R. W. Siebers, G. L. Purdie, C. Taylor und T. J. Maling (1987). The relationship between the QT interval and plasma amiodarone concentration in patients on long-term therapy. Eur J Clin Pharmacol **33**(2): 115-8.
- Cavalier-Smith, T. (2000). Membrane heredity and early chloroplast evolution. Trends Plant Sci **5**(4): 174-82.
- Chen, B. P., Y. S. Li, Y. Zhao, K. D. Chen, S. Li, J. Lao, S. Yuan, J. Y. Shyy und S. Chien (2001). DNA microarray analysis of gene expression in endothelial cells in response to 24-h shear stress. Physiol Genomics **7**(1): 55-63.
- Chen, D. und D. J. Schnell (1997). Insertion of the 34-kDa chloroplast protein import component, IAP34, into the chloroplast outer membrane is dependent on its intrinsic GTP-binding capacity. J Biol Chem **272**(10): 6614-20.
- Chin, J. Y., K. J. Leister und C. E. Wenner (2001). Density differential responses of embryonic fibroblasts. In Vitro Cell Dev Biol Anim **37**(5): 263-5.
- Danon, A. und S. P. Mayfield (1994). Light-regulated translation of chloroplast messenger RNAs through redox potential. Science **266**(5191): 1717-9.
- Davis, S. J., J. Kurepa und R. D. Vierstra (1999). The *Arabidopsis thaliana* HY1 locus, required for phytochrome-chromophore biosynthesis, encodes a protein related to heme oxygenases. Proc Natl Acad Sci U S A **96**(11): 6541-6.
- De Las Rivas, J., J. J. Lozano und A. R. Ortiz (2002). Comparative analysis of chloroplast genomes: functional annotation, genome-based phylogeny, and deduced evolutionary patterns. Genome Res **12**(4): 567-83.
- Delwiche, C. F. (1999). Tracing the Thread of Plastid Diversity through the Tapestry of Life. Am Nat **154**(S4): S164-S177.

- Deng, X. W. und P. H. Quail (1999). Signalling in light-controlled development. Semin Cell Dev Biol **10**(2): 121-9.
- DeRisi, J. L., V. R. Iyer und P. O. Brown (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. Science **278**(5338): 680-6.
- Deuschle, K., D. Funck, G. Forlani, H. Stransky, A. Biehl, D. Leister, E. van der Graaff, R. Kunze und W. B. Frommer (2004). The Role of  $\delta^1$ -Pyrroline-5-Carboxylate Dehydrogenase in Proline Degradation. Plant Cell **16**(12):3413-25.
- D'Haeseleer, P., S. Liang und R. Somogyi (2000). Genetic network inference: from co-expression clustering to reverse engineering. Bioinformatics **16**(8): 707-26.
- Doolittle, W. F., Y. Boucher, C. L. Nesbo, C. J. Douady, J. O. Andersson und A. J. Roger (2003). How big is the iceberg of which organellar genes in nuclear genomes are but the tip? Philos Trans R Soc Lond B Biol Sci **358**(1429): 39-57; discussion 57-8.
- Douglas, A. E. und J. A. Raven (2003). Genomes at the interface between bacteria and organelles. Philos Trans R Soc Lond B Biol Sci **358**(1429): 5-17; discussion 517-8.
- Douglas, S. E. (1998). Plastid evolution: origins, diversity, trends. Curr Opin Genet Dev **8**(6): 655-61.
- Dyall, S. D., M. T. Brown und P. J. Johnson (2004). Ancient invasions: from endosymbionts to organelles. Science **304**(5668): 253-7.
- Economou, A. (1998). Bacterial preprotein translocase: mechanism and conformational dynamics of a processive enzyme. Mol Microbiol **27**(3): 511-8.
- Eisen, M. B., P. T. Spellman, P. O. Brown und D. Botstein (1998). Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A **95**(25): 14863-8.
- Emanuelsson, O., H. Nielsen, S. Brunak und G. von Heijne (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol **300**(4): 1005-16.
- Escobar, M. A., K. A. Franklin, A. S. Svensson, M. G. Salter, G. C. Whitelam und A. G. Rasmusson (2004). Light regulation of the Arabidopsis respiratory chain. multiple discrete photoreceptor responses contribute to induction of type II NAD(P)H

- dehydrogenase genes. Plant Physiol **136**(1): 2710-21.
- Escoubas, J. M., M. Lomas, J. LaRoche und P. G. Falkowski (1995). Light intensity regulation of cab gene transcription is signaled by the redox state of the plastoquinone pool. Proc Natl Acad Sci U S A **92**(22): 10237-41.
- Ferat, J. L. und F. Michel (1993). Group II self-splicing introns in bacteria. Nature **364**(6435): 358-61.
- Ferro, M., D. Salvi, S. Brugiere, S. Miras, S. Kowalski, M. Louwagie, J. Garin, J. Joyard und N. Rolland (2003). Proteomics of the chloroplast envelope membranes from *Arabidopsis thaliana*. Mol Cell Proteomics.
- Ferro, M., D. Salvi, H. Riviere-Rolland, T. Vermat, D. Seigneurin-Berny, D. Grunwald, J. Garin, J. Joyard und N. Rolland (2002). Integral membrane proteins of the chloroplast envelope: identification and subcellular localization of new transporters. Proc Natl Acad Sci U S A **99**(17): 11487-92.
- Freeman, T. (2003). Platform technologies for microarray analysis. Brief Funct Genomic Proteomic **2**(1): 4-6.
- Friso, G., L. Giacomelli, A. J. Ytterberg, J. B. Peltier, A. Rudella, Q. Sun und K. J. Wijk (2004). In-depth analysis of the thylakoid membrane proteome of *Arabidopsis thaliana* chloroplasts: new proteins, new functions, and a plastid proteome database. Plant Cell **16**(2): 478-99.
- Froehlich, J. E., C. G. Wilkerson, W. K. Ray, R. S. McAndrew, K. W. Osteryoung, D. A. Gage und B. S. Phinney (2003). Proteomic study of the *Arabidopsis thaliana* chloroplastic envelope membrane utilizing alternatives to traditional two-dimensional electrophoresis. J Proteome Res **2**(4): 413-25.
- Gasch, A. P., P. T. Spellman, C. M. Kao, O. Carmel-Harel, M. B. Eisen, G. Storz, D. Botstein und P. O. Brown (2000). Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell **11**(12): 4241-57.
- Giaever, G., A. M. Chu, L. Ni, C. Connelly, L. Riles, S. Veronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. Andre, A. P. Arkin, A. Astromoff, M. El-Bakkoury, R. Bangham, R. Benito, S. Brachat, S. Campanaro, M. Curtiss, K. Davis, A.

- Deutschbauer, K. D. Entian, P. Flaherty, F. Foury, D. J. Garfinkel, M. Gerstein, D. Gotte, U. Guldener, J. H. Hegemann, S. Hempel, Z. Herman, D. F. Jaramillo, D. E. Kelly, S. L. Kelly, P. Kotter, D. LaBonte, D. C. Lamb, N. Lan, H. Liang, H. Liao, L. Liu, C. Luo, M. Lussier, R. Mao, P. Menard, S. L. Ooi, J. L. Revuelta, C. J. Roberts, M. Rose, P. Ross-Macdonald, B. Scherens, G. Schimmack, B. Shafer, D. D. Shoemaker, S. Sookhai-Mahadeo, R. K. Storms, J. N. Strathern, G. Valle, M. Voet, G. Volckaert, C. Y. Wang, T. R. Ward, J. Wilhelmy, E. A. Winzeler, Y. Yang, G. Yen, E. Youngman, K. Yu, H. Bussey, J. D. Boeke, M. Snyder, P. Philippsen, R. W. Davis und M. Johnston (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. Nature **418**(6896): 387-91.
- Giuliano, G., N. E. Hoffman, K. Ko, P. A. Scolnik und A. R. Cashmore (1988). A light-entrained circadian clock controls transcription of several plant genes. Embo J **7**(12): 3635-42.
- Glockner, G., A. Rosenthal und K. Valentin (2000). The structure and gene repertoire of an ancient red algal plastid genome. J Mol Evol **51**(4): 382-90.
- Golub, T. R., D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield und E. S. Lander (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science **286**(5439): 531-7.
- Govindjee, T. Kambara und W. Coleman (1985). The electron donor side of photosystem II: the oxygen evolving complex. Photochem Photobiol **42**(2): 187-210.
- Gray, J. C. (2003). Chloroplast-to-nucleus signalling: a role for Mg-protoporphyrin. Trends Genet **19**(10): 526-9.
- Gray, J. C., J. A. Sullivan, J. H. Wang, C. A. Jerome und D. MacLean (2003). Coordination of plastid and nuclear gene expression. Philos Trans R Soc Lond B Biol Sci **358**(1429): 135-44; discussion 144-5.
- Gray, M. W., G. Burger und B. F. Lang (1999). Mitochondrial evolution. Science **283**(5407): 1476-81.
- Hankamer, B., E. Morris, J. Nield, A. Carne und J. Barber (2001). Subunit positioning and

- transmembrane helix organisation in the core dimer of photosystem II. FEBS Lett **504**(3): 142-51.
- He, W. Z. und R. Malkin (1992). Specific release of a 9-kDa extrinsic polypeptide of photosystem I from spinach chloroplasts by salt washing. FEBS Lett **308**(3): 298-300.
- Heins, L. und J. Soll (1998). Chloroplast biogenesis: mixing the prokaryotic and the eukaryotic? Curr Biol **8**(6): R215-7.
- Herrmann, R. G., R. Martin, W. Busch, G. Wanner und U. Hohmann (1996). Physical and topographical mapping among Triticeae chromosomes. Symp Soc Exp Biol **50**: 25-30.
- Herrmann, R.G. (1997) Eukaryotism, towards a new interpretation. In Schenk,H.E.A., Herrmann,R.G., Jeon,K.W., Müller,N.E. und Schwemmler,W. (eds), *Eukaryotism and Symbiosis*. Springer, Berlin, Germany, pp. 73–118.
- Hess, W.R., R. Schendel, T. Börner, W. Rüdiger, (1991) Reduction of mRNA level for two nuclear encoded light regulated genes in the barley mutant albobrians is not correlated with phytochrome content and activity. J. Plant Physiol **138**: 292-98.
- Hess, W.R., A. Müller, F. Nagy, T. Börner (1994) Ribosome-deficient plastids affect transcription of light-induced nuclear genes: genetic evidence for a plastid-derived signal. Mol Gen Genet **242**: 305-12.
- High, S., R. Henry, R. M. Mould, Q. Valent, S. Meacock, K. Cline, J. C. Gray und J. Luirink (1997). Chloroplast SRP54 interacts with a specific subset of thylakoid precursor proteins. J Biol Chem **272**(17): 11622-8.
- Hudson, M. E. und P. H. Quail (2003). Identification of promoter motifs involved in the network of phytochrome A-regulated gene expression by combined analysis of genomic sequence and microarray data. Plant Physiol **133**(4): 1605-16.
- Huq, E., B. Al-Sady, M. Hudson, C. Kim, K. Apel und P. H. Quail (2004). Phytochrome-interacting factor 1 is a critical bHLH regulator of chlorophyll biosynthesis. Science **305**(5692): 1937-41.
- Isono, K., Y. Niwa, K. Satoh und H. Kobayashi (1997). Evidence for transcriptional

- regulation of plastid photosynthesis genes in *Arabidopsis thaliana* roots. Plant Physiol **114**(2): 623-30.
- Jarvis, P. (2001). Intracellular signalling: the chloroplast talks! Curr Biol **11**(8): R307-10.
- Jarvis, P., L. J. Chen, H. Li, C. A. Peto, C. Fankhauser und J. Chory (1998). An *Arabidopsis* mutant defective in the plastid general protein import apparatus. Science **282**(5386): 100-3.
- Jarvis, P. und J. Soll (2001). Toc, Tic, and chloroplast protein import. Biochim Biophys Acta **1541**(1-2): 64-79.
- Johanningmeier, U. und S. H. Howell (1984). Regulation of light-harvesting chlorophyll-binding protein mRNA accumulation in *Chlamydomonas reinhardtii*. Possible involvement of chlorophyll synthesis precursors. J Biol Chem **259**(21): 13541-9.
- Kanamaru, K., M. Fujiwara, M. Seki, T. Katagiri, M. Nakamura, N. Mochizuki, A. Nagatani, K. Shinozaki, K. Tanaka und H. Takahashi (1999). Plastidic RNA polymerase sigma factors in *Arabidopsis*. Plant Cell Physiol **40**(8): 832-42.
- Kaneko, T., S. Sato, H. Kotani, A. Tanaka, E. Asamizu, Y. Nakamura, N. Miyajima, M. Hirose, M. Sugiura, S. Sasamoto, T. Kimura, T. Hosouchi, A. Matsuno, A. Muraki, N. Nakazaki, K. Naruo, S. Okumura, S. Shimpo, C. Takeuchi, T. Wada, A. Watanabe, M. Yamada, M. Yasuda und S. Tabata (1996). Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions (supplement). DNA Res **3**(3): 185-209.
- Karger, G. A., J. D. Reid und C. N. Hunter (2001). Characterization of the binding of deuteroporphyrin IX to the magnesium chelatase H subunit and spectroscopic properties of the complex. Biochemistry **40**(31): 9291-9.
- Karpinski, S., H. Reynolds, B. Karpinska, G. Wingsle, G. Creissen und P. Mullineaux (1999). Systemic signaling and acclimation in response to excess excitation energy in *Arabidopsis*. Science **284**(5414): 654-7.
- Kleffmann, T., D. Russenberger, A. von Zychlinski, W. Christopher, K. Sjolander, W. Gruissem und S. Baginsky (2004). The *Arabidopsis thaliana* chloroplast proteome

- reveals pathway abundance and novel protein functions. Curr Biol **14**(5): 354-62.
- Kohchi, T., K. Mukougawa, N. Frankenberg, M. Masuda, A. Yokota und J. C. Lagarias (2001). The Arabidopsis HY2 gene encodes phytochromobilin synthase, a ferredoxin-dependent biliverdin reductase. Plant Cell **13**(2): 425-36.
- Kubis, S., R. Patel, J. Combe, J. Bedard, S. Kovacheva, K. Lilley, A. Biehl, D. Leister, G. Rios, C. Koncz und P. Jarvis (2004). Functional specialization amongst the Arabidopsis Toc159 family of chloroplast protein import receptors. Plant Cell **16**(8): 2059-77.
- Kurth, J., C. Varotto, P. Pesaresi, A. Biehl, E. Richly, F. Salamini und D. Leister (2002). Gene-sequence-tag expression analyses of 1,800 genes related to chloroplast functions. Planta **215**(1): 101-9.
- Kusnetsov, V., C. Bolle, T. Lubberstedt, S. Sopory, R. G. Herrmann und R. Oelmüller (1996). Evidence that the plastid signal and light operate via the same cis-acting elements in the promoters of nuclear genes for plastid proteins. Mol Gen Genet **252**(6): 631-9.
- Larkin, R. M., J. M. Alonso, J. R. Ecker und J. Chory (2003). GUN4, a regulator of chlorophyll synthesis and intracellular signaling. Science **299**(5608): 902-6.
- Lee, M. L., F. C. Kuo, G. A. Whitmore und J. Sklar (2000). Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. Proc Natl Acad Sci U S A **97**(18): 9834-9.
- Legen, J., S. Kemp, K. Krause, B. Profanter, R. G. Herrmann und R. M. Maier (2002). Comparative analysis of plastid transcription profiles of entire plastid chromosomes from tobacco attributed to wild-type and PEP-deficient transcription machineries. Plant J **31**(2): 171-88.
- Leister, D. (2003). Chloroplast research in the genomic age. Trends Genet **19**(1): 47-56.
- Leister, D. und A. Schneider (2003). From genes to photosynthesis in Arabidopsis thaliana. Int Rev Cytol **228**: 31-83.
- Lockhart, D. J. und E. A. Winzler (2000). Genomics, gene expression and DNA arrays. Nature **405**(6788): 827-36.

- Ludovico, P., F. Rodrigues, A. Almeida, M. T. Silva, A. Barrientos und M. Corte-Real (2002). Cytochrome c release and mitochondria involvement in programmed cell death induced by acetic acid in *Saccharomyces cerevisiae*. Mol Biol Cell **13**(8): 2598-606.
- Maiwald, D., A. Dietzmann, P. Jahns, P. Pesaresi, P. Joliot, A. Joliot, J. Z. Levin, F. Salamini und D. Leister (2003). Knock-out of the genes coding for the Rieske protein and the ATP-synthase delta-subunit of Arabidopsis. Effects on photosynthesis, thylakoid protein composition, and nuclear chloroplast gene expression. Plant Physiol **133**(1): 191-202.
- Martin, W. und R. G. Herrmann (1998). Gene transfer from organelles to the nucleus: how much, what happens, and Why? Plant Physiol **118**(1): 9-17.
- Martin, W., T. Rujan, E. Richly, A. Hansen, S. Cornelsen, T. Lins, D. Leister, B. Stoebe, M. Hasegawa und D. Penny (2002). Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc Natl Acad Sci U S A **99**(19): 12246-51.
- Martin, W. und M. J. Russell (2003). On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. Philos Trans R Soc Lond B Biol Sci **358**(1429): 59-83; discussion 83-5.
- Martinez-Hernandez, A., L. Lopez-Ochoa, G. Arguello-Astorga und L. Herrera-Estrella (2002). Functional properties and regulatory complexity of a minimal RBCS light-responsive unit activated by phytochrome, cryptochrome, and plastid signals. Plant Physiol **128**(4): 1223-33.
- Maxwell, D. P., D. E. Laudenbach und N. Huner (1995). Redox Regulation of Light-Harvesting Complex II and cab mRNA Abundance in *Dunaliella salina*. Plant Physiol **109**(3): 787-795.
- Mayfield, S. P. und W. C. Taylor (1984). Carotenoid-deficient maize seedlings fail to accumulate light-harvesting chlorophyll a/b binding protein (LHCP) mRNA. Eur J Biochem **144**(1): 79-84.

- McCormac, D. J., H. Litz, J. Wang, P. D. Gollnick und J. O. Berry (2001). Light-associated and processing-dependent protein binding to 5' regions of rbcL mRNA in the chloroplasts of a C4 plant. J Biol Chem **276**(5): 3476-83.
- Meskauskiene, R., M. Nater, D. Goslings, F. Kessler, R. op den Camp und K. Apel (2001). FLU: a negative regulator of chlorophyll biosynthesis in *Arabidopsis thaliana*. Proc Natl Acad Sci U S A **98**(22): 12826-31.
- Mitchell, P. (1976). Vectorial chemistry and the molecular mechanics of chemiosmotic coupling: power transmission by proticity. Biochem Soc Trans **4**(3): 399-430.
- Mochizuki, N., J. A. Brusslan, R. Larkin, A. Nagatani und J. Chory (2001). Arabidopsis genomes uncoupled 5 (GUN5) mutant reveals the involvement of Mg-chelatase H subunit in plastid-to-nucleus signal transduction. Proc Natl Acad Sci U S A **98**(4): 2053-8.
- Moller, S. G., T. Kunkel und N. H. Chua (2001). A plastidic ABC protein involved in intercompartmental communication of light signaling. Genes Dev **15**(1): 90-103.
- Montane, M. H. und K. Kloppstech (2000). The family of light-harvesting-related proteins (LHCs, ELIPs, HLIPs): was the harvesting of light their primary function? Gene **258**(1-2): 1-8.
- Montgomery, B. L., K. C. Yeh, M. W. Crepeau und J. C. Lagarias (1999). Modification of distinct aspects of photomorphogenesis via targeted expression of mammalian biliverdin reductase in transgenic Arabidopsis plants. Plant Physiol **121**(2): 629-39.
- Mori, H. und K. Cline (2002). A twin arginine signal peptide and the pH gradient trigger reversible assembly of the thylakoid [Delta]pH/Tat translocase. J Cell Biol **157**(2): 205-10.
- Mori, H., E. J. Summer, X. Ma und K. Cline (1999). Component specificity for the thylakoidal Sec and Delta pH-dependent protein transport pathways. J Cell Biol **146**(1): 45-56.
- Mulligan, R. M., J. Chory und J. R. Ecker (1997). Signaling in plants. Proc Natl Acad Sci U S A **94**(7): 2793-5.
- Mullineaux, P. und S. Karpinski (2002). Signal transduction in response to excess light:

- getting out of the chloroplast. Curr Opin Plant Biol **5**(1): 43-8.
- Nadon, R. und J. Shoemaker (2002). Statistical issues with microarrays: processing and analysis. Trends Genet **18**(5): 265-71.
- Naver, H., M. P. Scott, J. H. Golbeck, B. L. Moller und H. V. Scheller (1996). Reconstitution of barley photosystem I with modified PSI-C allows identification of domains interacting with PSI-D and PSI-A/B. J Biol Chem **271**(15): 8996-9001.
- Notterman, D. A., U. Alon, A. J. Sierk und A. J. Levine (2001). Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. Cancer Res **61**(7): 3124-30.
- Oelmuller, R., Levitan, I., Bergfeld, R., Rajasekhar, V.K., Mohr, H. (1986). Expression of nuclear genes as affected by treatments acting on the plastids. Planta **168**: 482-492
- Oelmuller, R., R. E. Kendrick und W. R. Briggs (1989). Blue-light mediated accumulation of nuclear-encoded transcripts coding for proteins of the thylakoid membrane is absent in the phytochrome-deficient aurea mutant of tomato. Plant Mol Biol **13**(2): 223-32.
- Okumoto, S., W. Koch, M. Tegeder, W. N. Fischer, A. Biehl, D. Leister, Y. D. Stierhof und W. B. Frommer (2004). Root phloem-specific expression of the plasma membrane amino acid proton co-transporter AAP3. J Exp Bot **55**(406): 2155-68.
- Peltier, J. B., O. Emanuelsson, D. E. Kalume, J. Ytterberg, G. Friso, A. Rudella, D. A. Liberles, L. Soderberg, P. Roepstorff, G. von Heijne und K. J. van Wijk (2002). Central functions of the luminal and peripheral thylakoid proteome of Arabidopsis determined by experimentation and genome-wide prediction. Plant Cell **14**(1): 211-36.
- Pesaresi, P., N. A. Gardner, S. Masiero, A. Dietzmann, L. Eichacker, R. Wickner, F. Salamini und D. Leister (2003). Cytoplasmic N-terminal protein acetylation is required for efficient photosynthesis in Arabidopsis. Plant Cell **15**(8): 1817-32.
- Pesaresi, P., C. Lunde, P. Jahns, D. Tarantino, J. Meurer, C. Varotto, R. D. Hirtz, C. Soave, H. V. Scheller, F. Salamini und D. Leister (2002). A stable LHCII-PSI aggregate and suppression of photosynthetic state transitions in the psae1-1 mutant of

- Arabidopsis thaliana. Planta **215**(6): 940-8.
- Pesaresi, P., C. Varotto, J. Meurer, P. Jahns, F. Salamini und D. Leister (2001). Knock-out of the plastid ribosomal protein L11 in Arabidopsis: effects on mRNA translation and photosynthesis. Plant J **27**(3): 179-89.
- Pesaresi, P., C. Varotto, E. Richly, A. Lessnick, F. Salamini und D. Leister (2003). Protein-protein and protein-function relationships in Arabidopsis photosystem I: cluster analysis of PSI polypeptide levels and photosynthetic parameters in PSI mutants. J Plant Physiol **160**(1): 17-22.
- Petracek, M. E., L. F. Dickey, S. C. Huber und W. F. Thompson (1997). Light-regulated changes in abundance and polyribosome association of ferredoxin mRNA are dependent on photosynthesis. Plant Cell **9**(12): 2291-300.
- Petracek, M. E., L. F. Dickey, T. T. Nguyen, C. Gatz, D. A. Sowinski, G. C. Allen und W. F. Thompson (1998). Ferredoxin-1 mRNA is destabilized by changes in photosynthetic electron transport. Proc Natl Acad Sci U S A **95**(15): 9009-13.
- Pfannschmidt, T., A. Nilsson, A. Tullberg, G. Link und J. F. Allen (1999). Direct transcriptional control of the chloroplast genes psbA and psaAB adjusts photosynthesis to light energy distribution in plants. IUBMB Life **48**(3): 271-6.
- Pfannschmidt, T., K. Schutze, M. Brost und R. Oelmüller (2001). A novel mechanism of nuclear photosynthesis gene regulation by redox signals from the chloroplast during photosystem stoichiometry adjustment. J Biol Chem **276**(39): 36125-30.
- Puente, P., N. Wei und X. W. Deng (1996). Combinatorial interplay of promoter elements constitutes the minimal determinants for light and developmental control of gene expression in Arabidopsis. Embo J **15**(14): 3732-43.
- Quackenbush, J. (2001). Computational analysis of microarray data. Nat Rev Genet **2**(6): 418-27.
- Quackenbush, J. (2003). Genomics. Microarrays--guilt by association. Science **302**(5643): 240-1.

- Reiß, T., Bergfeld, R., Link, G., Thien, W. und Mohr, H. (1983). Photooxidative destruction of chloroplasts and its consequences for cytosolic enzyme levels and plant development. Planta **159**, 518-28.
- Reumann, S., J. Davila-Aponte und K. Keegstra (1999). The evolutionary origin of the protein-translocating channel of chloroplastic envelope membranes: identification of a cyanobacterial homolog. Proc Natl Acad Sci U S A **96**(2): 784-9.
- Richly, E., A. Dietzmann, A. Biehl, J. Kurth, C. Laloi, K. Apel, F. Salamini und D. Leister (2003). Covariations in the nuclear chloroplast transcriptome reveal a regulatory master-switch. EMBO Rep **4**(5): 491-8.
- Richly, E. und D. Leister (2004). An improved prediction of chloroplast proteins reveals diversities and commonalities in the chloroplast proteomes of Arabidopsis and rice. Gene **329**: 11-6.
- Robinson, C., S. J. Thompson und C. Woolhead (2001). Multiple pathways used for the targeting of thylakoid proteins in chloroplasts. Traffic **2**(4): 245-51.
- Rochaix, J. D. (2001). Assembly, function, and dynamics of the photosynthetic machinery in *Chlamydomonas reinhardtii*. Plant Physiol **127**(4): 1394-8.
- Rodermel, S. (2001). Pathways of plastid-to-nucleus signaling. Trends Plant Sci **6**(10): 471-8.
- Rujan, T. und W. Martin (2001). How many genes in Arabidopsis come from cyanobacteria? An estimate from 386 protein phylogenies. Trends Genet **17**(3): 113-20.
- Sagar, A.D., Briggs, W.R. und Thompson, W.F. (1988). Nuclear-cytoplasmic partitioning of phytochrome-regulated transcripts in *Pisum sativum*. Plant Physiol **88** 1397-1402.
- Sakamoto, W., T. Tamura, Y. Hanba-Tomita und M. Murata (2002). The VAR1 locus of Arabidopsis encodes a chloroplastic FtsH and is responsible for leaf variegation in the mutant alleles. Genes Cells **7**(8): 769-80.
- Sakamoto, W., A. Zaltsman, Z. Adam und Y. Takahashi (2003). Coordinated regulation and complex formation of yellow variegated1 and yellow variegated2, chloroplastic

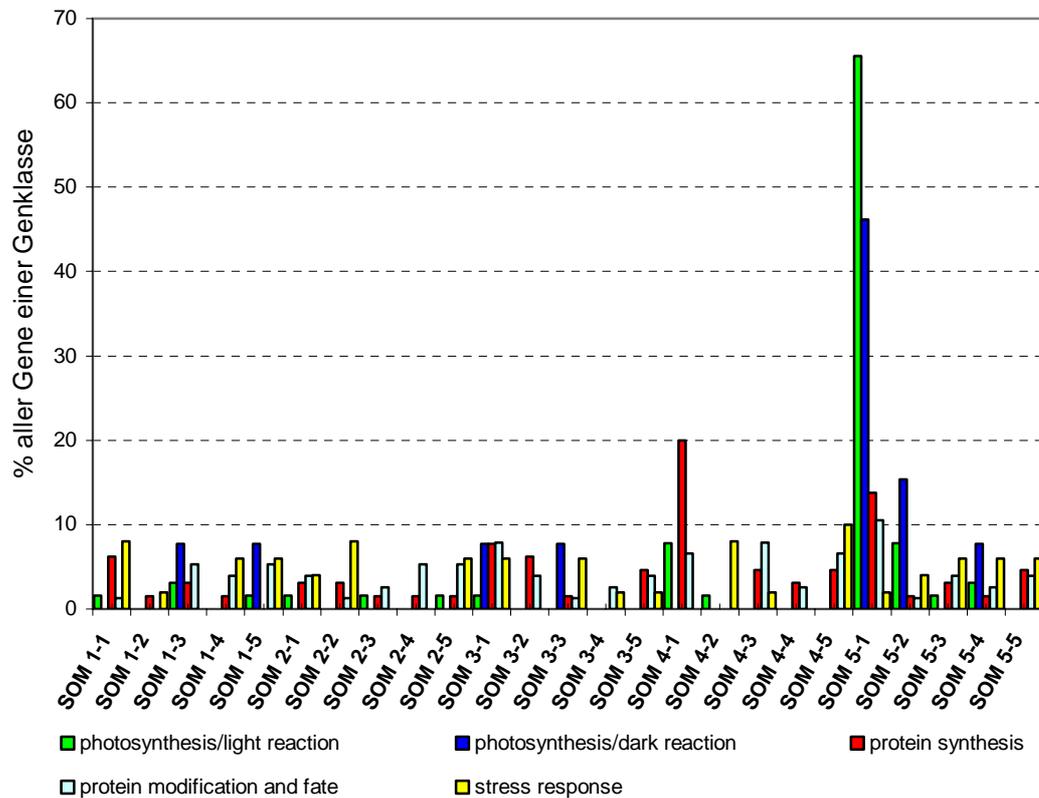
- FtsH metalloproteases involved in the repair cycle of photosystem II in Arabidopsis thylakoid membranes. Plant Cell **15**(12): 2843-55.
- Schaffer, R., J. Landgraf, M. Accerbi, V. Simon, M. Larson und E. Wisman (2001). Microarray analysis of diurnal and circadian-regulated genes in Arabidopsis. Plant Cell **13**(1): 113-23.
- Shen, Z. J., S. K. Kim, O. S. Kwon, Y. S. Lee und B. J. Moon (2001). Specific inhibition of transforming growth factor-beta2 expression in human osteoblast cells by antisense phosphorothioate oligonucleotides. Eur J Biochem **268**(8): 2331-7.
- Sherameti, I., M. Nakamura, Y. Y. Yamamoto, T. Pfannschmidt, J. Obokata und R. Oelmüller (2002). Polyribosome loading of spinach mRNAs for photosystem I subunits is controlled by photosynthetic electron transport. Plant J **32**(5): 631-9.
- Sherlock, G. (2000). Analysis of large-scale gene expression data. Curr Opin Immunol **12**(2): 201-5.
- Strand, A., T. Asami, J. Alonso, J. R. Ecker und J. Chory (2003). Chloroplast to nucleus communication triggered by accumulation of Mg-protoporphyrinIX. Nature **421**(6918): 79-83.
- Streatfield, S. J., A. Weber, E. A. Kinsman, R. E. Hausler, J. Li, D. Post-Beittenmiller, W. M. Kaiser, K. A. Pyke, U. I. Flugge und J. Chory (1999). The phosphoenolpyruvate/phosphate translocator is required for phenolic metabolism, palisade cell development, and plastid-dependent nuclear gene expression. Plant Cell **11**(9): 1609-22.
- Sturn, A. (2000). Cluster analysis for large scale gene expression studies. Master's thesis, TU Graz, [www.genome.tugraz.at](http://www.genome.tugraz.at)
- Sturn, A., J. Quackenbush und Z. Trajanoski (2002). Genesis: cluster analysis of microarray data. Bioinformatics **18**(1): 207-8.
- Sullivan, J. A. und J. C. Gray (1999). Plastid translation is required for the expression of nuclear photosynthesis genes in the dark and in roots of the pea lip1 mutant. Plant Cell **11**(5): 901-10.
- Sullivan, J. A. und J. C. Gray (2002). Multiple plastid signals regulate the expression of the

- pea plastocyanin gene in pea and transgenic tobacco plants. Plant J **32**(5): 763-74.
- Summer, H., T. Pfannschmidt und G. Link (2000). Transcripts and sequence elements suggest differential promoter usage within the *ycf3-psaAB* gene cluster on mustard (*Sinapis alba* L.) chloroplast DNA. Curr Genet **37**(1): 45-52.
- Surpin, M., R. M. Larkin und J. Chory (2002). Signal transduction between the chloroplast and the nucleus. Plant Cell **14 Suppl**: S327-38.
- Susek, R. E., F. M. Ausubel und J. Chory (1993). Signal transduction mutants of *Arabidopsis* uncouple nuclear CAB and RBCS gene expression from chloroplast development. Cell **74**(5): 787-99.
- Tamayo, P., D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E. S. Lander und T. R. Golub (1999). Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc Natl Acad Sci U S A **96**(6): 2907-12.
- Tang, L., S. Bhat und M. E. Petracek (2003). Light control of nuclear gene mRNA abundance and translation in tobacco. Plant Physiol **133**(4): 1979-90.
- Terry, M. J. und R. E. Kendrick (1999). Feedback inhibition of chlorophyll synthesis in the phytochrome chromophore-deficient *aurea* and *yellow-green-2* mutants of tomato. Plant Physiol **119**(1): 143-52.
- The *Arabidopsis* Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408**(6814): 796-815.
- Toledo-Ortiz, G., E. Huq und P. H. Quail (2003). The *Arabidopsis* basic/helix-loop-helix transcription factor family. Plant Cell **15**(8): 1749-70.
- Tonkyn, J. C. und W. Gruissem (1993). Differential expression of the partially duplicated chloroplast S10 ribosomal protein operon. Mol Gen Genet **241**(1-2): 141-52.
- Totter, S., M. A. Block, M. Allen, T. Westergren, C. Albrieux, H. V. Scheller, S. Merchant und P. E. Jensen (2003). *Arabidopsis* CHL27, located in both envelope and thylakoid membranes, is required for the synthesis of protochlorophyllide. Proc Natl Acad Sci U S A **100**(26): 16119-24.
- Trumpower, B. L. (1990). The protonmotive Q cycle. Energy transduction by coupling of

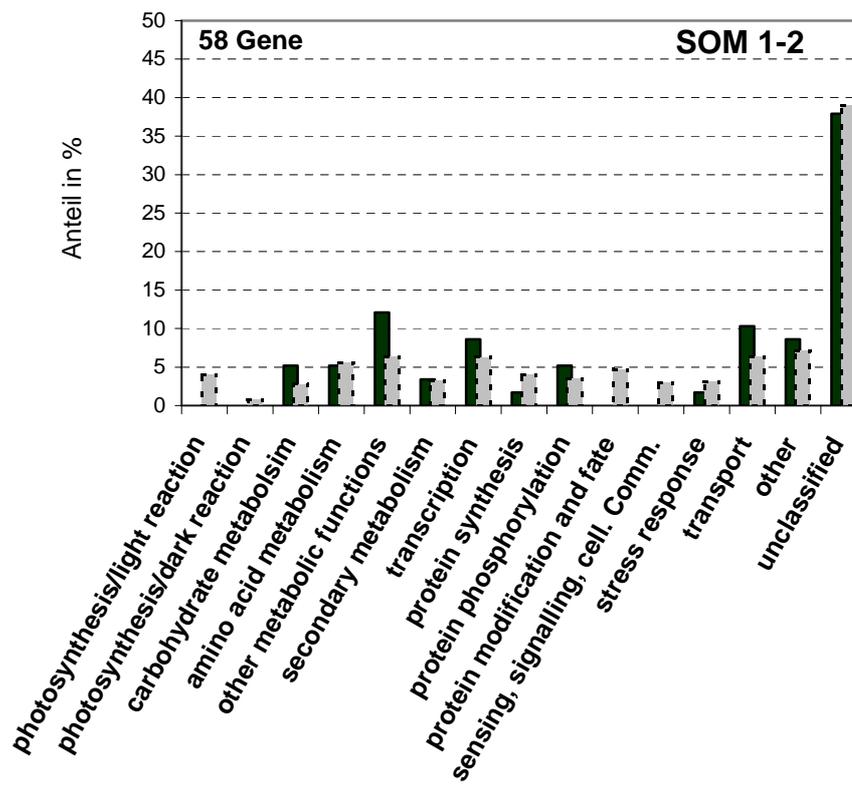
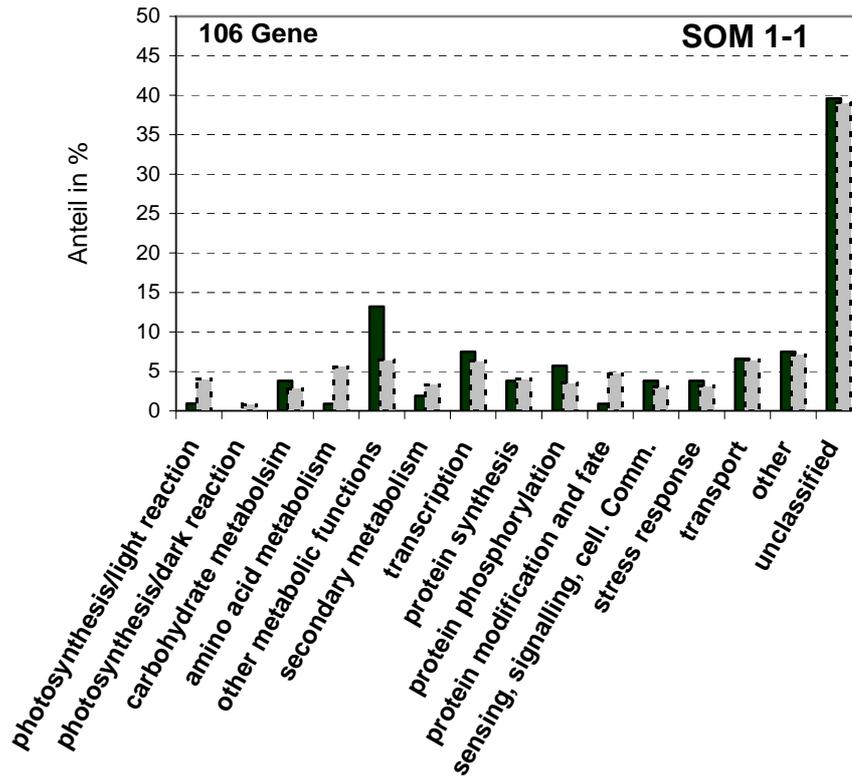
- proton translocation to electron transfer by the cytochrome bc1 complex. J Biol Chem **265**(20): 11409-12.
- Uno, Y., T. Furihata, H. Abe, R. Yoshida, K. Shinozaki und K. Yamaguchi-Shinozaki (2000). Arabidopsis basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. Proc Natl Acad Sci U S A **97**(21): 11632-7.
- Varotto, C., D. Maiwald, P. Pesaresi, P. Jahns, F. Salamini und D. Leister (2002). The metal ion transporter IRT1 is necessary for iron homeostasis and efficient photosynthesis in Arabidopsis thaliana. Plant J **31**(5): 589-99.
- Varotto, C., P. Pesaresi, P. Jahns, A. Lessnick, M. Tizzano, F. Schiavon, F. Salamini und D. Leister (2002). Single and double knockouts of the genes for photosystem I subunits G, K, and H of Arabidopsis. Effects on photosystem I composition, photosynthetic electron flow, and state transitions. Plant Physiol **129**(2): 616-24.
- Varotto, C., P. Pesaresi, J. Meurer, R. Oelmüller, S. Steiner-Lange, F. Salamini und D. Leister (2000). Disruption of the Arabidopsis photosystem I gene psaE1 affects photosynthesis and impairs growth. Plant J **22**(2): 115-24.
- Varotto, C., E. Richly, F. Salamini und D. Leister (2001). GST-PRIME: a genome-wide primer design software for the generation of gene sequence tags. Nucleic Acids Res **29**(21): 4373-7.
- Vinti, G., A. Hills, S. Campbell, J. R. Bowyer, N. Mochizuki, J. Chory und E. Lopez-Juez (2000). Interactions between hy1 and gun mutants of Arabidopsis, and their implications for plastid/nuclear signalling. Plant J **24**(6): 883-94.
- Wang, C., R. Francis, S. Harirchian, D. Battle, B. Mayhew, M. Bassett, W. E. Rainey und R. G. Pestell (2000). The application of high density microarray for analysis of mitogenic signaling and cell-cycle in the adrenal. Endocr Res **26**(4): 807-23.
- Weigel, M., P. Pesaresi und D. Leister (2003). Tracking the function of the cytochrome c6-like protein in higher plants. Trends Plant Sci **8**(11): 513-7.
- Weigel, M., C. Varotto, P. Pesaresi, G. Finazzi, F. Rappaport, F. Salamini und D. Leister (2003). Plastocyanin is indispensable for photosynthetic electron flow in

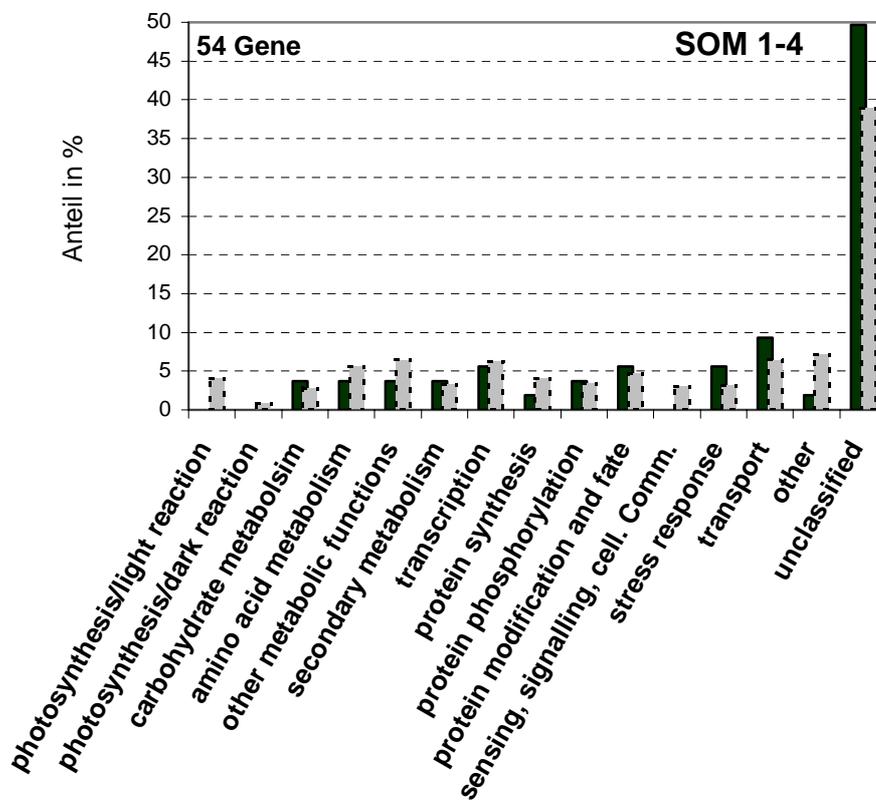
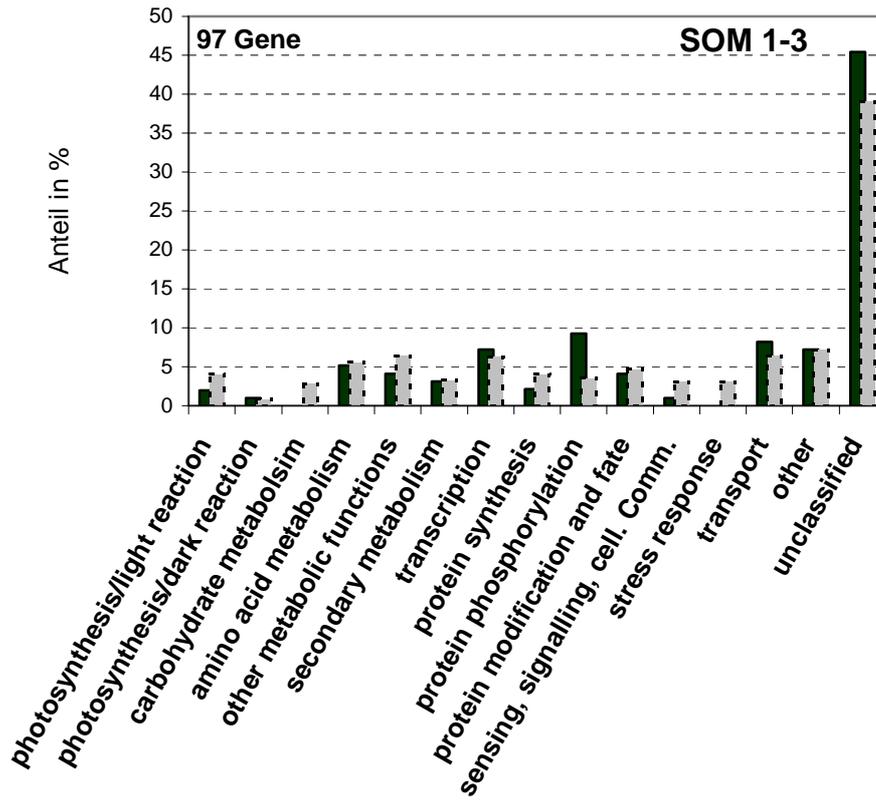
- Arabidopsis thaliana*. J Biol Chem **278**(33): 31286-9.
- Weinstein, J. D., R. W. Howell, R. D. Leverette, S. Y. Grooms, P. S. Brignola, S. M. Mayer und S. I. Beale (1993). Heme Inhibition of [delta]-Aminolevulinic Acid Synthesis Is Enhanced by Glutathione in Cell-Free Extracts of *Chlorella*. Plant Physiol **101**(2): 657-65.
- Westerlund, I., G. Von Heijne und O. Emanuelsson (2003). LumenP--a neural network predictor for protein localization in the thylakoid lumen. Protein Sci **12**(10): 2360-6.
- Westphal, S., J. Soll und U. C. Vothknecht (2001). A vesicle transport system inside chloroplasts. FEBS Lett **506**(3): 257-61.
- Wodicka, L., H. Dong, M. Mittmann, M. H. Ho und D. J. Lockhart (1997). Genome-wide expression monitoring in *Saccharomyces cerevisiae*. Nat Biotechnol **15**(13): 1359-67.
- Yang, D.H., Andersson, B., Aro, E.-M. und Ohad, I. (2001) The redox state of the plastoquinone pool controls the level of the light-harvesting chlorophyll a/b binding protein complex II (LHCII) during photoacclimation. - Photos Res **68**: 163-74.
- Yoshinaga, K., H. Inuma, T. Masuzawa und K. Uedal (1996). Extensive RNA editing of U to C in addition to C to U substitution in the *rbcL* transcripts of hornwort chloroplasts and the origin of RNA editing in green plants. Nucleic Acids Res **24**(6): 1008-14.
- Zerges, W., S. Wang und J. D. Rochaix (2002). Light activates binding of membrane proteins to chloroplast RNAs in *Chlamydomonas reinhardtii*. Plant Mol Biol **50**(3): 573-85.
- Zhang, M. Q. (1999). Large-scale gene expression data analysis: a new challenge to computational biologists. Genome Res **9**(8): 681-8.

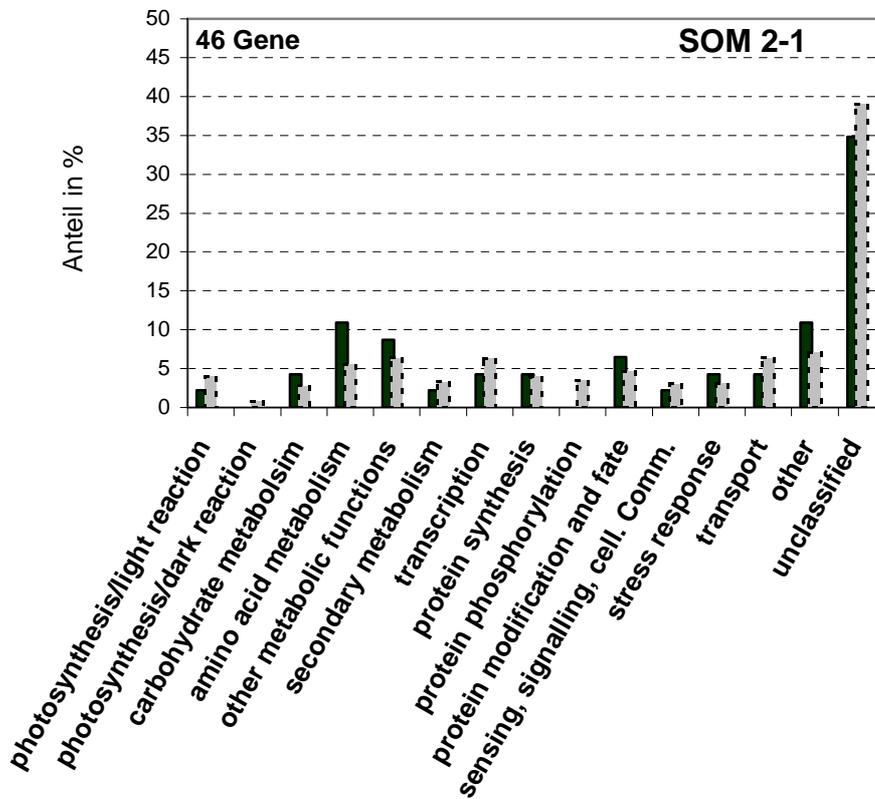
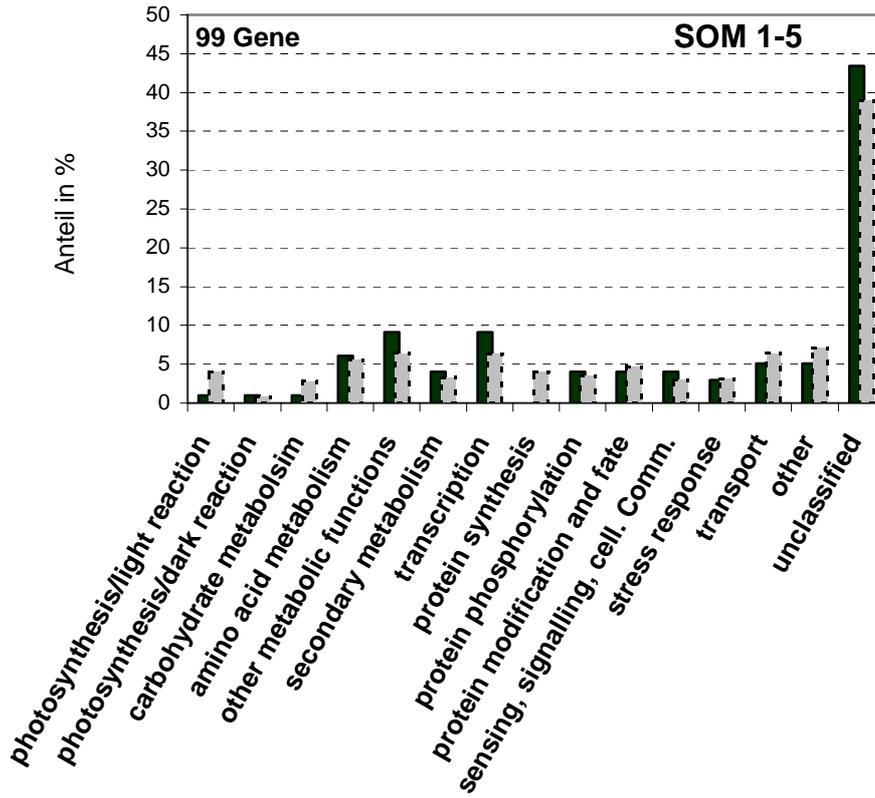
## 7. Anhang

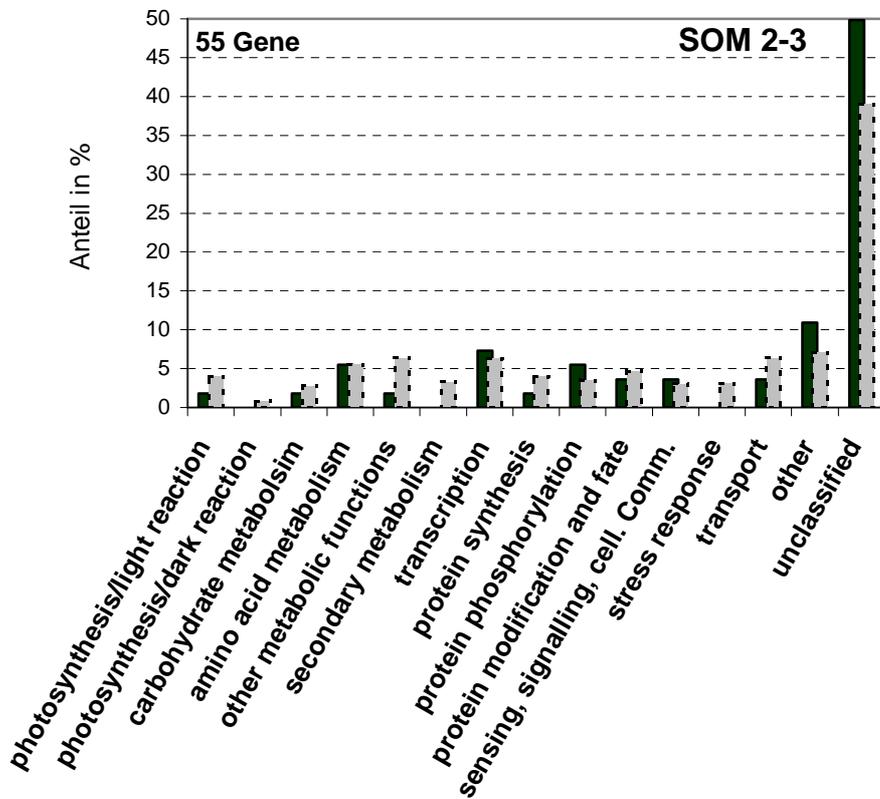
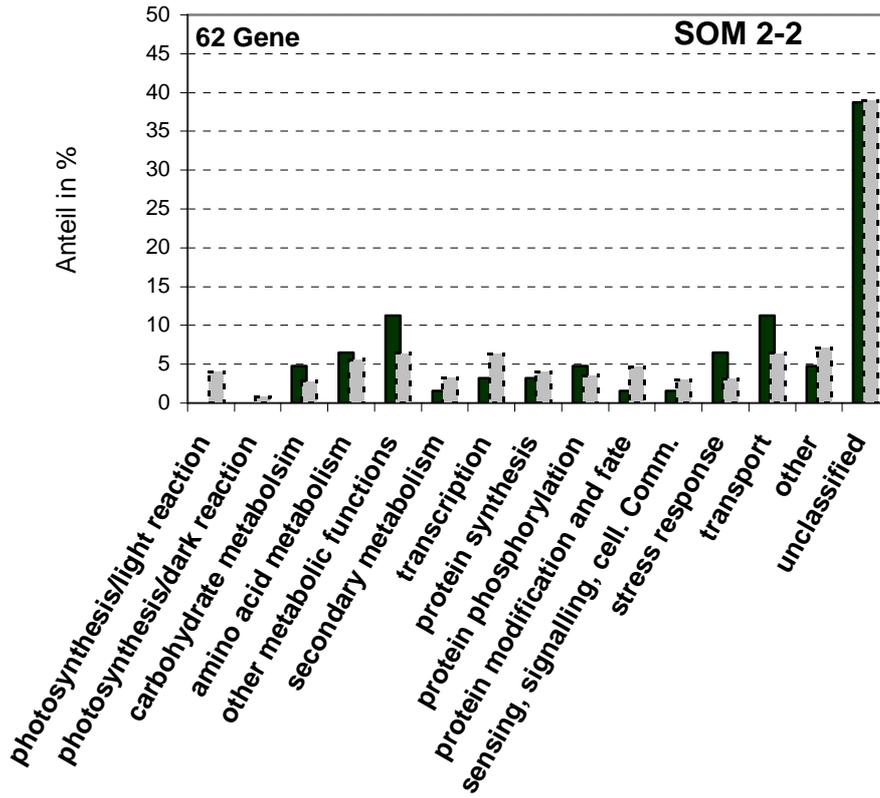


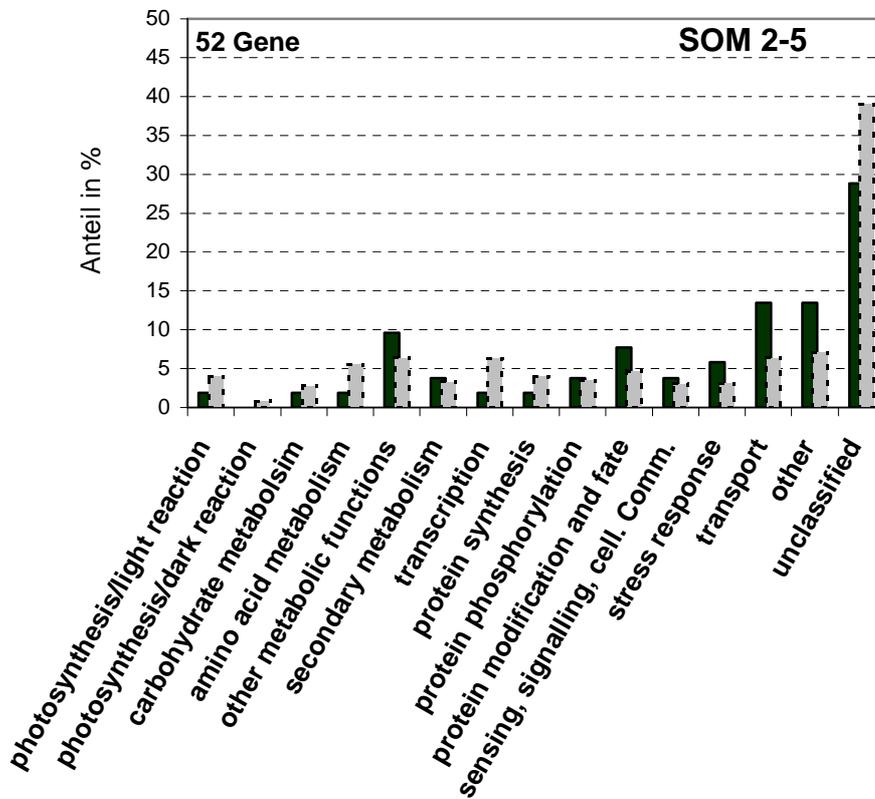
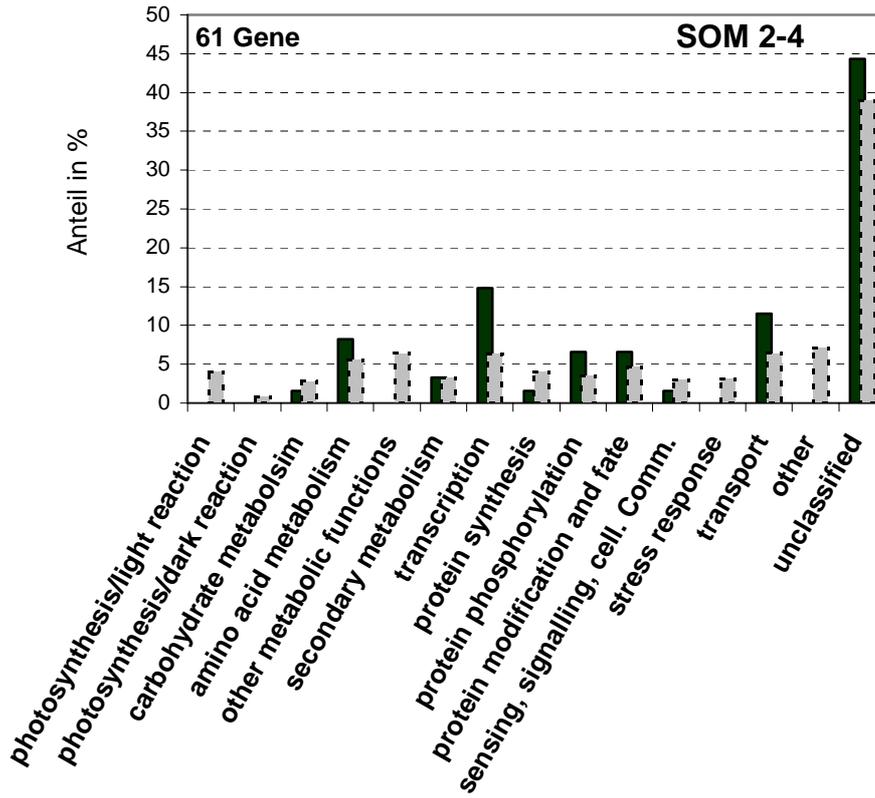
**Abb. A 1.: Verteilung der funktionellen Genklassen auf die Regulons.** Dargestellt ist die Verteilung der Gene von 5 der 15 funktionellen Genklassen auf die 25 Regulons. Angegeben ist der prozentuale Anteil der Gene einer Genklasse in jedem der Regulons an der Gesamtzahl aller Gene der Genklasse im Gesamtdatensatz. Beispielsweise sind in Regulon SOM 4-1 20 % aller Gene der Proteinsynthese zu finden, die wiederum mehr als 25 % der Gene des Regulon repräsentieren (siehe Abb. A2 und Abb. 3.17, Seite 71).

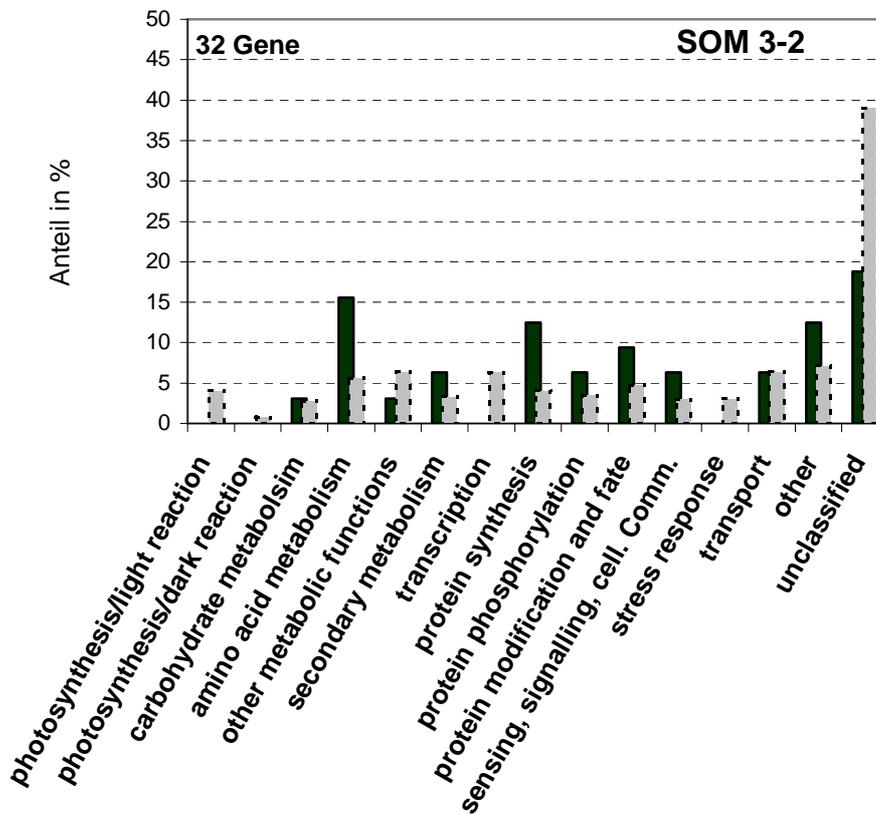
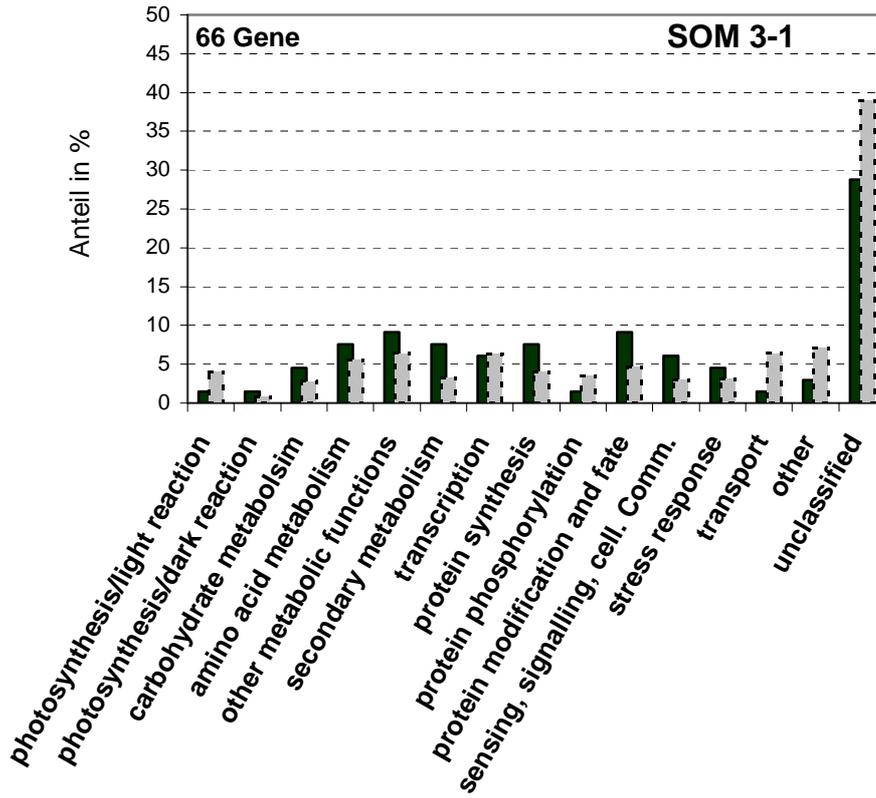


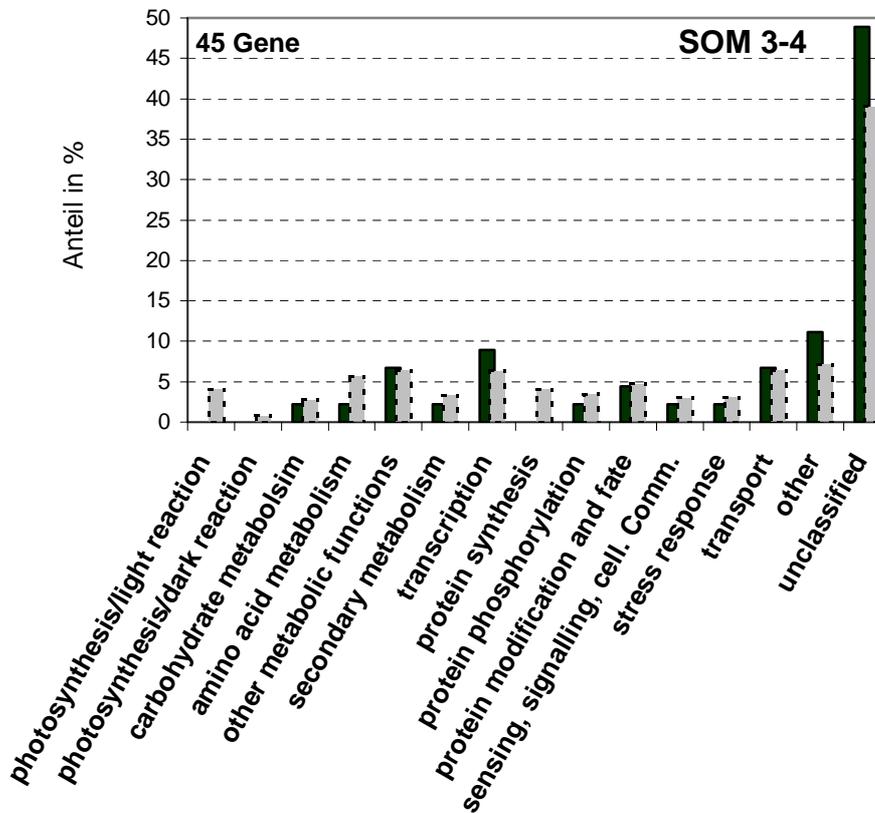
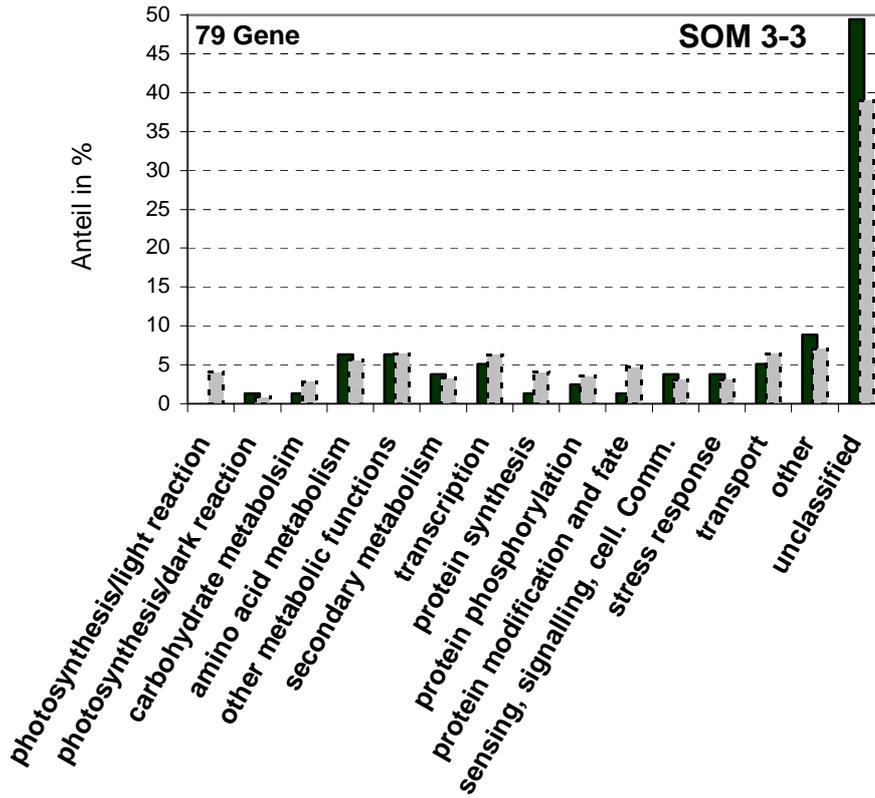


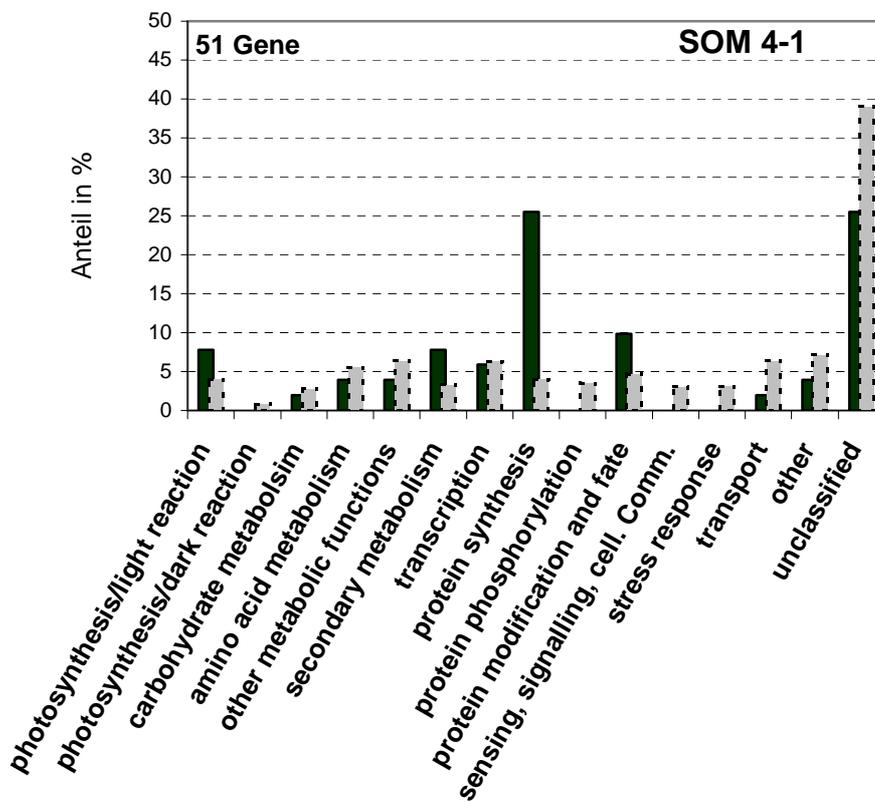
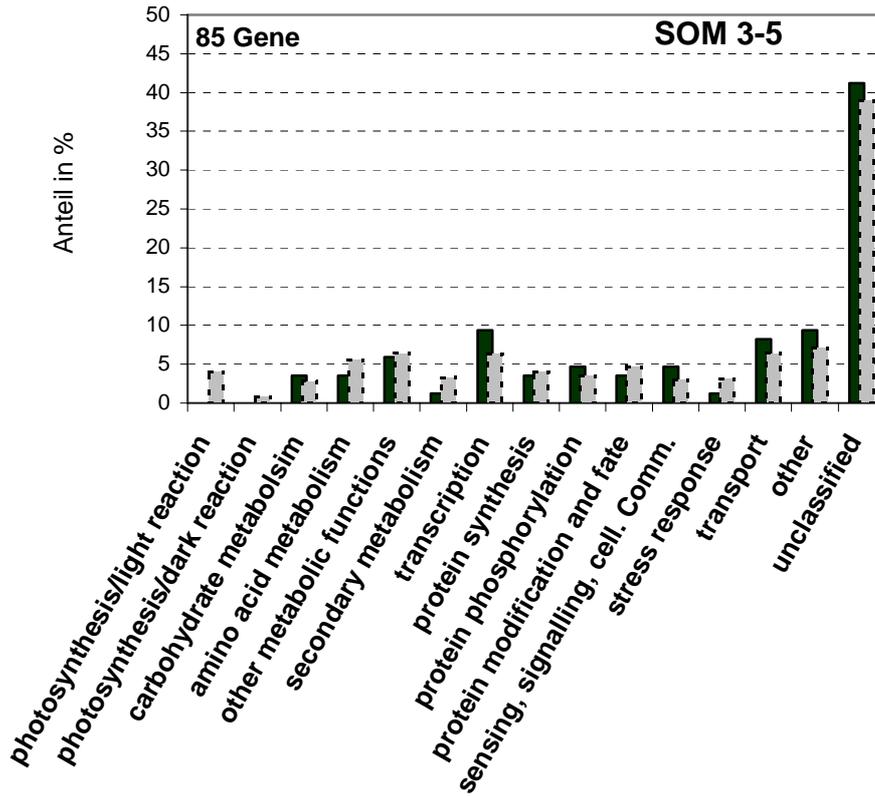


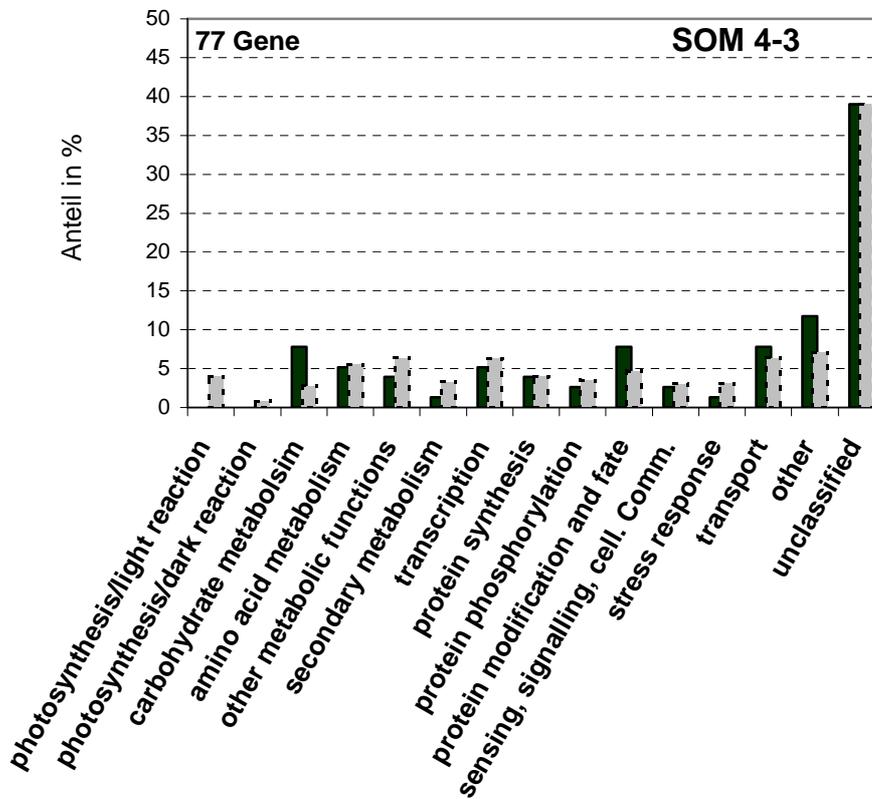
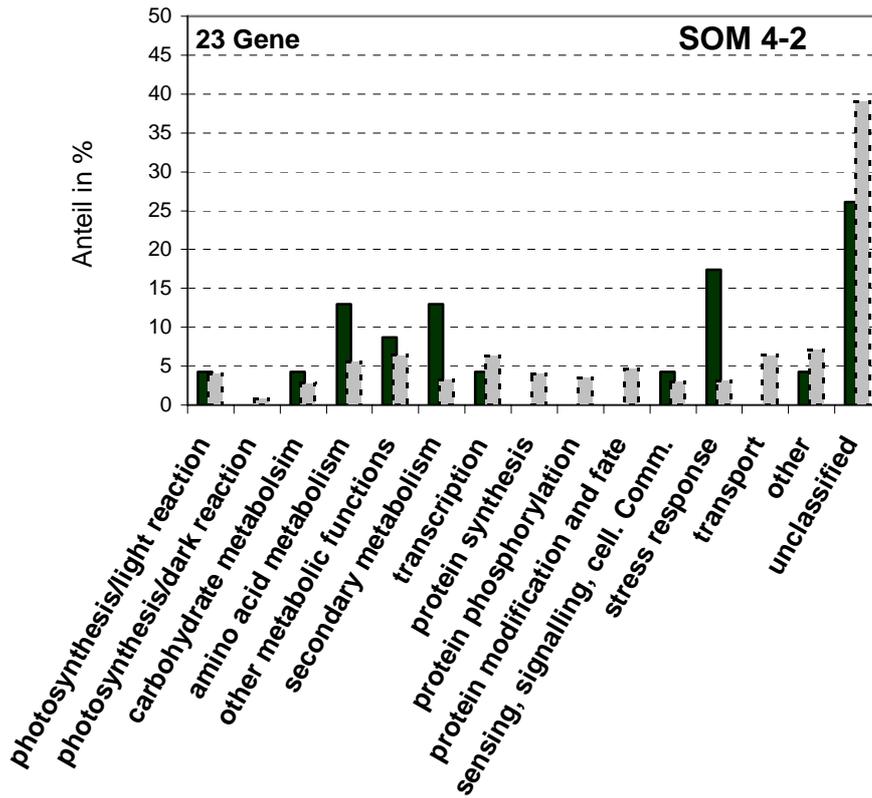


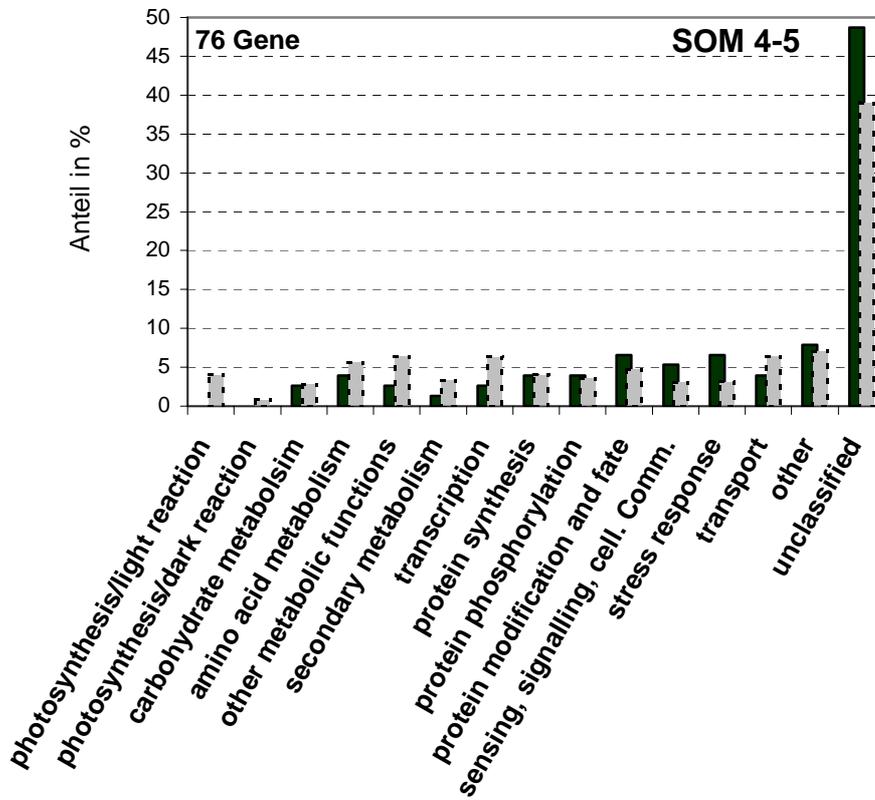
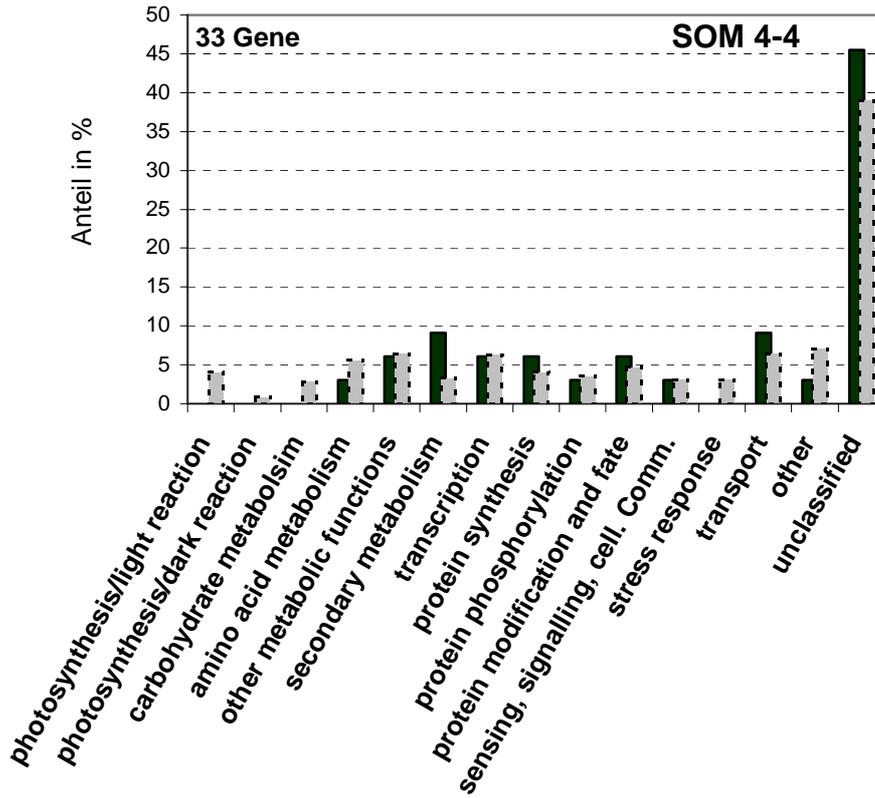


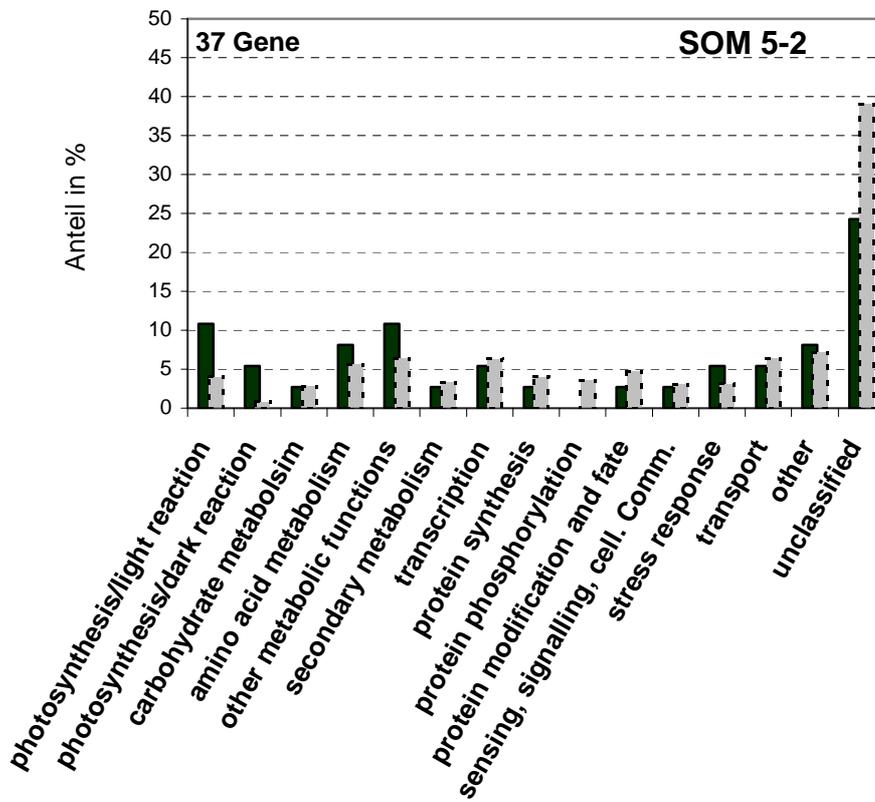
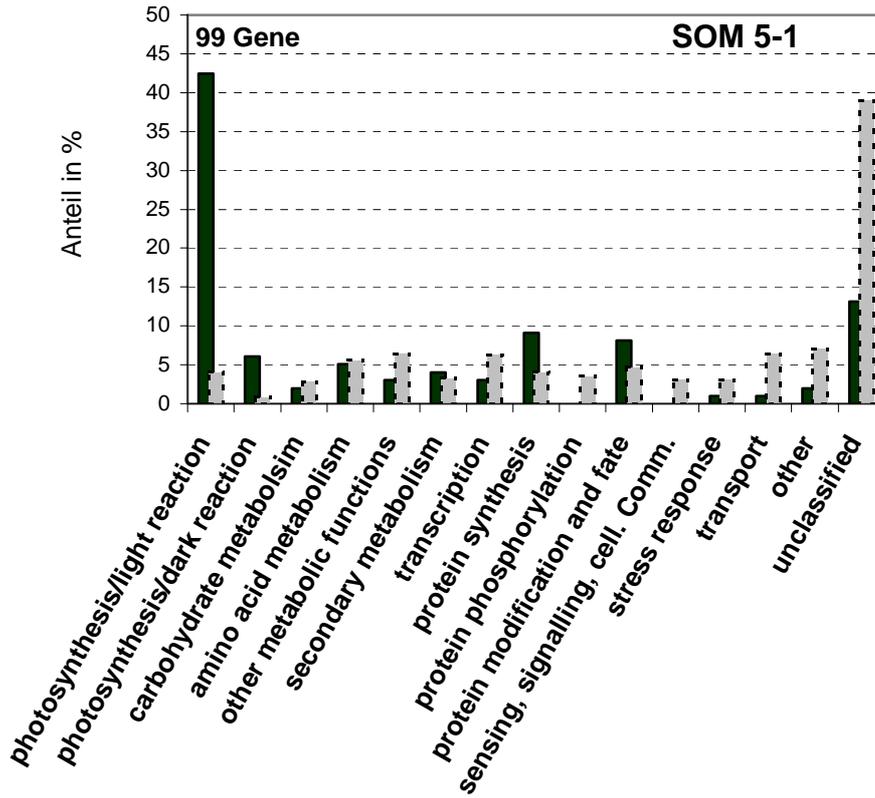


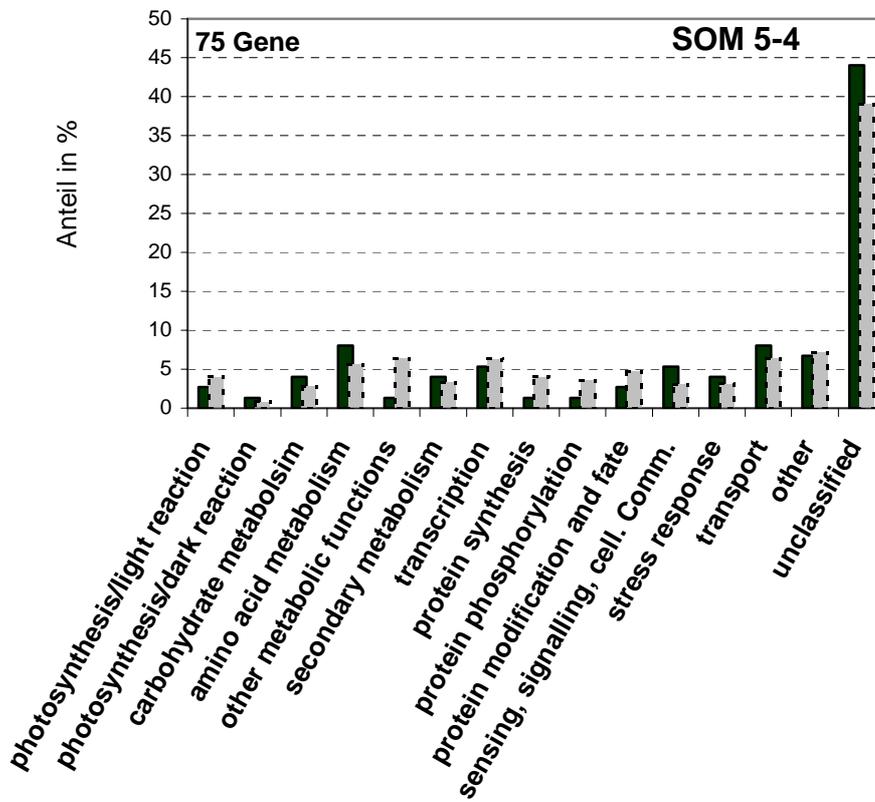
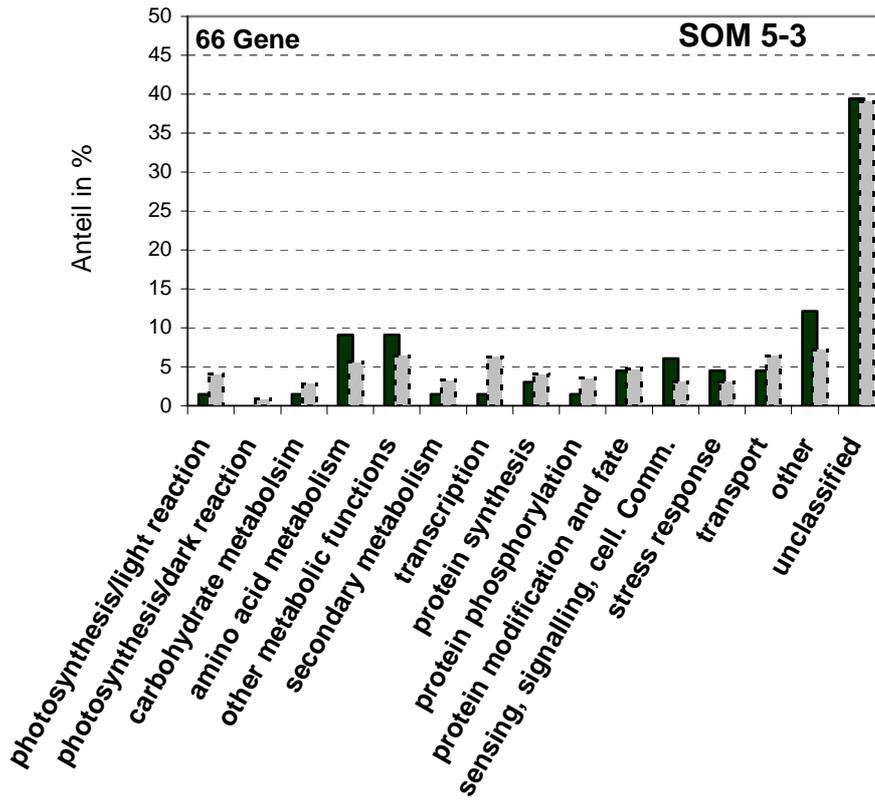


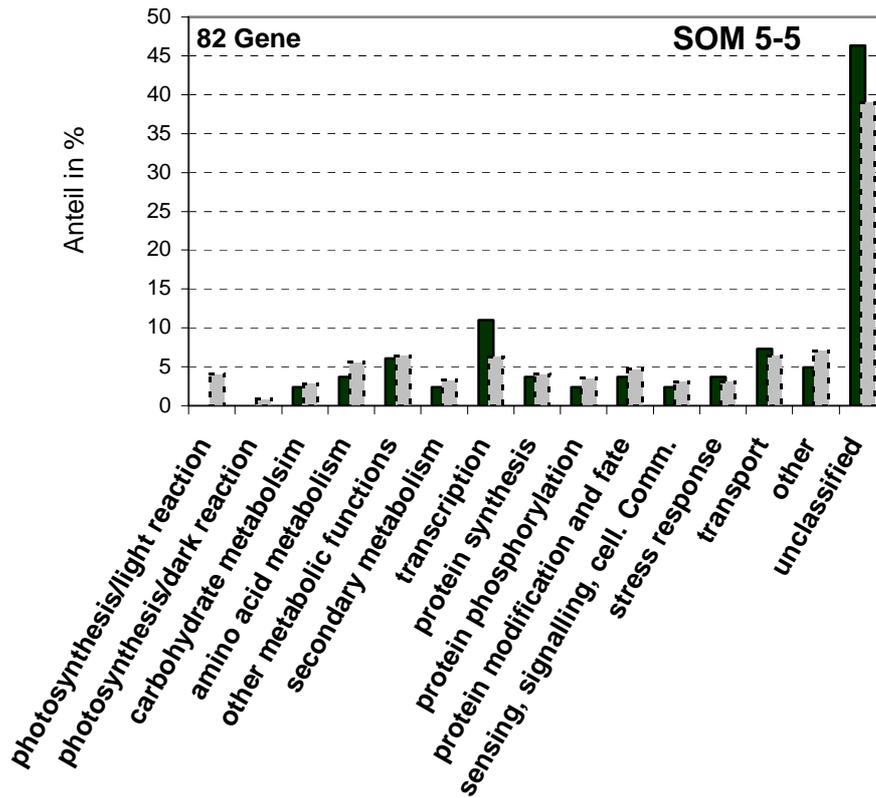












**Abb. A 2.: Funktionelle Klassifizierung der Gene aller 25 Regulons.** Die Gene der Regulons wurden in 15 funktionelle Klassen unterteilt und die relativen Abundanzen der funktionellen Genklassen (schwarze Säulen) im Regulon mit der Verteilung der funktionellen Genklassen im gesamten Ausgangsdatensatz des 1616 Gene umfassenden Transkriptoms verglichen (graue Säulen).

## Abkürzungsverzeichnis

°C	Grad Celsius
μ~	Mikro
Abb.	Abbildung
ALA	5-Aminolävulinsäure
ATP	Adenosin-5'-triphosphat
bp	Basenpaare
CAB	Chlorophyll a/b bindendes Protein
cDNA	revers transkribierte mRNA (complementary DNA)
Ci	Curie
Col-0	Arabidopsis Ökotyp Columbia
cTP	Chloroplasten-Transitpeptid
cue	<i>chlorophyll a/b-binding underexpressed</i>
Cyt	Cytochrom
dATP	2'-Desoxyadenosin-5'-triphosphat
dCTP	2'-Desoxycytidin-5'-triphosphat
ddH <sub>2</sub> O	doppelt destilliertes Wasser
dGTP	2'-Desoxyguanosin-5'-triphosphat
DNA	Desoxyribonukleinsäure
dNTP	2'-Desoxyribonukleosid-5'-triphosphat
dTTP	2'-Desoxythymidin-5'-triphosphat
DTT	Dithiothreitol
EDTA	Ethylendiamintetraessigsäure
EGT	Endosymbiotischer Genetransfer
EST	<i>expressed sequence tag</i>
FD	Ferredoxin
FNR	Ferredoxin-NADP <sup>+</sup> -Reduktase
GLU-TR	Glutamyl-tRNA-Reduktase
GST	<i>gene sequence tag</i>
GTP	Guanosin-5'-triphosphat
gun	<i>genomes uncoupled</i>
h	Stunde
HCl	Chlorwasserstoff, Salzsäure
IE	innere Chloroplastenmembrane
Kb	Kilobasenpaare
KCl	Kaliumchlorid
l	Liter
laf	<i>long after far red</i>
LHC	lichtsammelnder Antennenkomplex der Photosysteme ( <i>light harvesting complex</i> )
M	Molar

m~	Milli~
mtDNA	mitochondriale DNA
MgProto	Mg-Protoporphyrin IX
MgProtoME	Mg-Protoporphyrin IX Monomethylester
min	Minute
mRNA	kodierende RNA ( <i>messenger RNA</i> )
NaCl	Natriumchlorid
NaOH	Natriumhydroxid
OM	äußere Chloroplastenmembran
ORF	offener Leserahmen ( <i>open reading frame</i> )
p~	Piko
PΦB	Phytochromobilin
PC	Plastocyanin
PCR	Polymerasekettenreaktion ( <i>polymerase chain reaction</i> )
PPDB	<i>Plastid Proteome Data Base</i>
PPO	Protoporphyrinogen-IX-Oxidase
PPT	Phosphoenolpyruvat/Phosphat Translokator
Proto	Protoporphyrinogen IX
PS	Photosystem
RNA	Ribonukleinsäure
rpm	Umdrehungen pro Minute ( <i>rounds per minute</i> )
RT	Raumtemperatur
ROS	reaktive oxigene Zwischenprodukte ( <i>reactiv oxygene species</i> )
SDS	Natriumdodecylsulfat
sek	Sekunde
SOM	<i>Self-Organizing-Maps</i>
SPP	stromale Prozessierungs-Peptidase
SRP	<i>signal recognition particle</i>
Tab.	Tabelle
TIC	Protein-Import-Apparat der inneren Chloroplastenmembran ( <i>translocase of the inner membrane of chloroplasts</i> )
TM	Thylakoidmembran
TOC	Protein-Import-Apparat der äußeren Chloroplastenmembran ( <i>translocase of the outer membrane of chloroplasts</i> )
Tris	Trishydroxymethyl-aminomethan
tRNA	transfer RNA
U	Einheit ( <i>unit</i> )
UTP	Uridin-5'-triphosphat
UV	ultraviolett
WT	Wildtyp

Ich versichere, daß ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; daß diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; daß sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, daß ich eine solche Veröffentlichung vor Abschluß des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Herrn Prof. Dr. Francesco Salamini betreut worden.

Köln, den 12. Dezember 2004

Alexander Biehl

Teilpublikationen:

- Biehl A, E. Richly, C. Noutsos, F. Salamini und D. Leister (2004). Analysis of 101 nuclear transcriptomes reveals 23 distinct regulons and their relationship to metabolism, chromosomal gene distribution and co-ordination of nuclear and plastid gene expression. Gene (im Druck).
- Okumoto, S., W. Koch, M. Tegeder, W. N. Fischer, A. Biehl, D. Leister, Y. D. Stierhof und W. B. Frommer (2004). Root phloem-specific expression of the plasma membrane amino acid proton co-transporter AAP3. J Exp Bot **55**(406): 2155-68.
- Kubis, S., R. Patel, J. Combe, J. Bedard, S. Kovacheva, K. Lilley, A. Biehl, D. Leister, G. Rios, C. Koncz und P. Jarvis (2004). Functional specialization amongst the Arabidopsis Toc159 family of chloroplast protein import receptors. Plant Cell **16**(8): 2059-77.
- Deuschle, K., D. Funck, G. Forlani, H. Stransky, A. Biehl, D. Leister, E. van der Graaff, R. Kunze und W. B. Frommer (2004). The Role of  $\Delta^1$ -Pyrroline-5-Carboxylate Dehydrogenase in Proline Degradation. Plant Cell **16**(12):3413-25.
- Bosco, C. D., L. Lezhneva, A. Biehl, D. Leister, H. Strotmann, G. Wanner und J. Meurer (2004). Inactivation of the chloroplast ATP synthase gamma subunit results in high non-photochemical fluorescence quenching and altered nuclear gene expression in Arabidopsis thaliana. J Biol Chem **279**(2): 1060-9.
- Richly, E., A. Dietzmann, A. Biehl, J. Kurth, C. Laloi, K. Apel, F. Salamini und D. Leister (2003). Covariations in the nuclear chloroplast transcriptome reveal a regulatory master-switch. EMBO Rep. **4**: 491-8.
- Kurth, J., C. Varotto, P. Pesaresi, A. Biehl, E. Richly, F. Salamini und D. Leister (2002). Gene-sequence-tag expression analyses of 1,800 genes related to chloroplast functions. Planta **215**(1): 101-9.