Structural and Functional Genomics in Semi-Autonomous Organelles: Composition and Origin of Proteomes of Chloroplasts and Mitochondria and Related Transcriptomics

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

vorgelegt von Erik Richly
aus Gießen

Köln, 2003

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 PD Dr. D. Leister angefertigt.

Berichterstatter:
Prof. Dr. Francesco Salamini
Prof. Dr. Ulf-Ingo Flügge

Tag der letzten mündlichen Prüfung: 03.11.2003

Contents

List of Tables

List of Figures

1. Introduction

All life is organized as cells and based on the most conserved principle of physical compartmentalization from the environment and self-organization of self-contained redox reactions. Both organelles, mitochondria and chloroplasts, are descendants of once free-living eubacteria which were taken up in an endosymbiotic event (Mereschkowsky, 1905; Goksoyr, 1967; Gray, 1992) and are used to protect the cell from redox chemistry involving destructive, mutagenic intermediates. Over evolutionary time the endosymbiosis led to the transfer of organellar DNA to the nucleus of the host cell and subsequent inactivation of the redundant and surplus genes of the organelles which thus became semi-autonomous. For this last step to happen, mechanisms for the re-import of the proteins, now encoded in the nucleus and translated in the cytosol, had to be invented and means of communication between the organelles and the nucleus had to be established. To date, many questions have arisen dealing with topics such as how many proteins make up the proteomes of extant organelles and how many of those proteins can still be traced back to their ancestors. Are proteins always targeted back to the organelle they stem from and how are organelles able to regulate the cytosolic synthesis of proteins they need – i.e. how does the communication between organelles and nucleus function? Powerful tools to tackle questions like these are found in structural and functional genomics and combinations thereof that will increase their strength with the growing number of completely sequenced genomes of various species.

Structural genomics deals with complete genomic DNA sequence(s), the identification of open reading frames (ORFs), and the deduction and structural, as well as, functional classification of proteins. In functional genomics, the underlying concept is based on the monitoring of the complete set of certain compounds in an organism – especially their changes and alterations in response to genetic or environmental stimuli. By this organism-wide surveillance,

overviews of whole biological processes are gained that are not available from single component studies. The compounds monitored can be the complete set of: metabolites expressed (metabolomics; Stitt & Fernie, 2003), proteins synthesized (proteomics; Roberts, 2002), or transcripts produced in a cell (transcriptomics). The ultimate goal of this interesting field will be the combination of all those sub-disciplines leading to *in silico* simulations of whole organisms and a comprehensive understanding of the complex regulatory pathways and their interactions.

1.1. The origin of organelles

Mitochondria are most closely related to extant α-proteobacteria (Andersson *et al.*, 1998; Gray *et al.*, 1999, 2001) and gene sequence data strongly support a monophyletic origin. The same holds true for chloroplasts whose nearest relatives are the cyanobacteria (Douglas, 1998; Delwiche, 1999; Tomitani *et al.*, 1999). The first endosymbiotic event, which ultimately led to the generation of mitochondria, was long thought to have occurred by the endosymbiosis of a proteobacterium by an amitochondriate eukaryote host cell (serial endosymbiosis; Cavalier-Smith, 1987). Recent findings suggest that the mitochondrion arose in a common ancestor of all extant eukaryotes: an uptake of a proteobacterium by another prokaryote – a strictly anaerobic, hydrogen-dependent and autotrophic archaebacterium – serving as the host is hypothesized. This suggests that the first organelle originated at essentially the same time as the nuclear component of the eukaryotic cell – about two billion years ago (Martin & Müller, 1998; Gray *et al.*, 1999; Martin & Russell, 2003) – rather than in a separate, subsequent event. However, this first mitochondriate eukaryote diverged into animals and the progenitor of red and green algae, and plants. For the latter, photosynthetically active species, a second endosymbiotic event involving an ancestor of extant cyanobacteria was necessary, which is assumed to have taken place between 1.2 (Butterfield, 2000) and possibly 1.5 (Javaux *et al.*, 2001) billion years ago. This progenitor also diverged into today's brown algae for which two contrasting views exist of how they evolved: Martin *et al.* (2002) conclude from their analyses that two secondary endosymbiotic events involving a red algae are needed to explain the occurence of plastids in cryptophytes. This view is also supported by the work of De las Rivas *et al.* (2002) – while

Cavalier-Smith (2002) only argues for a single endosymbiotic event. However, this progenitor of brown algae also evolved into the apicomplexans (such as Plasmodium) which have residual plastids, and trypanosomes that have no plastids at all.

Today's organelles have retained many features of their eubacterial heritage, with respect to certain enzymes, transport systems, reproduction, DNA in form of circular molecules not associated with histones or other proteins, transfer RNAs, ribosomes, and other equipment needed to transcribe and translate their DNA into proteins. However, also huge changes are apparent: Directly after the endosymbiotic event plastids and mitochondria must have possessed their own complete genomes and translation machinery. Looking at contemporary organelle genomes though, one realizes that a dramatical change in size has occurred (Martin & Herrmann, 1998) compared to the genomes of their progenitors: only 50–200 genes coding for proteins have been retained in the plastome (Stoebe, 1998), while for mitochondria an even smaller number of 5–60 has been reported (Gray *et al.*, 1998). The proteomes of both organelles, on the other hand, consist of as many, or even more, proteins than that of the organelles' ancestors. How can this discrepancy between genome size and proteome size be explained? For those genes that have not been retained during the process of organelle genome reduction, two possible fates are imaginable: they were either lost or transferred to what is now the nuclear genome and the encoded proteins were later targeted back to the organelle (although there were cases where they were targeted to other locations). Thus, proteins are retained in the organelle but genes are not. This mechanism is often referred to as 'endosymbiotic gene transfer' (EGT).

1.2. Endosymbiotic gene transfer

Endosymbiotic gene transfer is a special kind of lateral gene transfer (Brown, 2003) which appears to be very common in nature: Various DNA fragments of organelles have been found in flowering plant nuclear chromosomes (Ayliffe & Timmis, 1992; The Arabidopsis Genome Initiative, 2000). EGT results in a compartmentalized, but integrated, eukaryotic genetic system under the regulatory dominance of the nucleus (Herrmann, 1997). Consequently, organelles tend to come under the control of nuclear regulatory genes.

Recently, EGT of chloroplast DNA (cpDNA) could even be induced under laboratory con-

ditions in tobacco, estimating that EGT occurs at a surprisingly high frequency of about one successful gene transfer event per 16,000 gametes tested, with 16 independent transfers found (Huang *et al.*, 2003; Stegemann *et al.*, 2003). Furthermore, it was demonstrated that during EGT large, contiguous pieces of bulk DNA are transferred which seem to stem from the occasional lysis of the organelle. In addition, it was established that recombination between this large pieces of cpDNA and nuclear DNA is also taking place (The Arabidopsis Genome Initiative, 2000; Huang *et al.*, 2003; Stegemann *et al.*, 2003).

The reasons for EGT given below are ample and, in most cases, involve the transition from a prokaryotic to a eukaryotic genetic system. However, it should not be assumed that the genetic system of organelles is strictly prokaryotic [e.g. chloroplasts import a 110 kD phage like RNA polymerase encoded by nuclear genes (Hedtke *et al.*, 1997)] but, nonetheless, envisions the advantages such a transition brought about:

(1) The genome size of prokaryotes – and thus the functional diversity – is limited: Prokaryotes possess only one origin of replication per chromosome, leading to a strongly limited size of chromosomes as replication is rate-limiting during exponential growth of prokaryotes and doubling time is the major factor to competition. Subsequently, anything that can be reduced from the genome will eventually be lost. In eukaroytes, genome size is only limited by replication fidelity (Drake, 1999) and this has enabled large numbers of 'selfish' elements (such as transposons) and functional genes to co-exist in eukaryotic genomes (Brosius, 1999; Smit, 1999). Such elements can occasionally be used to build up new functions, as suggested by the 'tinkering concept' in evolution (Jacob, 1977) and thus, EGT helped in the process of functional diversification of the organelles.

(2) Gene duplication of nuclear genes followed by divergence is an effective instrument for the evolution of new functions. This is necessary since the endosymbionts are deprived of their sexual reproduction and furthermore not able to take up DNA from the external milieu. Eventually this would lead to the accumulation of deleterious mutations in their genomes [Muller's ratchet (Muller, 1964)]. Gene duplication and divergence has resulted in major expansions of various gene families, e.g. the homeobox family (Ruddle *et al.*, 1999), and duplication, as in the polyploidy in plants, is considered a driving force of eukaryotic genome evolution (Wolfe & Shields, 1997; Ruddle *et al.*, 1999).

(3) Eukaryotes are capable of extensive transcript processing including mRNA splicing, editing (Sharp, 1994) and snoRNA-mediated cleavage, methylation and pseudouridylation of RNA (Weinstein & Steitz, 1999). Editing and splicing result in multiple RNA or protein products from a single gene, while modifications are employed to fine-tune RNA function (gene silencing, parental imprinting) – but both contribute to a more complex relation between genotype and phenotype. In addition, in prokaryotes transcription is coupled with translation, while in eukaryotes those two processes are physically separated, allowing additional and more sophisticated ways of control.

(4) A gene transferred from the organelle to the nucleus can become subject to altered targeting, i.e. the cell can test whether a protein exhibits equal or an increased functionality in the cytosol or in a different organelle than in its original compartment. Thus, proteins of mitochondrial origin might end up in the chloroplast and vice versa (Martin & Schnarrenberger, 1997).

(5) The chemistry of photosynthetic and respiratory electron transport in plastids and mitochondria leads to the generation of high concentrations of various excessive mutagenic reactive oxygen species. These are responsible for high frequencies of mutation in organellar DNA (Allen & Raven, 1996), thus intensifying the effects of Muller's ratchet (Lynch, 1996; Moran, 1996). Therefore, translocating organellar DNA to the nucleus is advantageous.

However, one has to keep in mind that gene regulation in organelles is not simply a miniature prokaryotic system, but rather part of an integrated eukaryotic gene regulation system (Herrmann, 1997) and it could also be shown that plastid gene expression – at both transcriptional (Hajdukiewicz *et al.*, 1997) and posttranscriptional (Bock & Koop, 1997) levels – is regulated in quite a complex manner (Herrmann, 1997): The genes for σ-factors are a good example of the integrated regulation system, since σ-factors are required by the plastid-encoded RNA core polymerase and are expressed and regulated by the nuclear apparatus (Tanaka *et al.*, 1996, 1997). This is also the case for genes coding for phage-type RNA polymerases which are transported to both mitochondria and plastids (Hedtke *et al.*, 2000) – RpoT;2 even possesses dual targeting properties. This demonstrates that gene regulation in plastids is subordinate to the nucleus. Furthermore, in comparison to animal mitochondria the effects of Muller's ratchet are much less prominent in plant mitochondria and chloroplasts. This could be due to

genetic recombination between organelles [as found in chloroplasts of Chlamydomonas (Fischer *et al.*, 1996)] or by the high polyploidy levels of chloroplasts permitting recombinations between genomes within the same plastid. These assumptions are supported by the fact that plastids actually import a nuclear encoded homologue of RecA which is functionally involved in recombination and repair in chloroplasts (Cerutti *et al.*, 1995).

Obviously, the mere transfer of genetic material from organelle to nucleus and a subsequent recombination are not enough to allow the removal of the organelle copy. The transferred gene copy has to become expressed and, in addition, targeted to the organelle to replace the function of the original. This results in a competition between the single nuclear copy and numerous copies in the many organelles. The acquisition of a transit peptide was found not to be as difficult as expected: For mitochondria one in 30 randomly cloned sequences from *E. coli* DNA was able to act as a mitochondrial transit peptide (mTP) and to direct proteins into the organelle (Baker & Schatz, 1987). For chloroplasts this might not happen as frequently, since chloroplast transit peptides (cTPs) share more distinct features compared to mTPs. However, it is suggested that a stable expression of the newly acquired gene is much more important than the addition of a transit peptide. The adaptation to the eukaryotic genetic system enabling expression includes the change of codon usage, the addition of poly-A signals and intron sequences, and the establishment of a functional promoter. It has already been shown that a stable expression of promoterless constructs in plants occurs very rarely (Herman *et al.*, 1990). In fact, a stable expression as the first step of EGT might also result in another advantage: unless the protein acquires the targeting signal it will be present in the cytosol and the cell is able to test if the transferred gene product is of any value in this compartment. An example for a chloroplast protein targeted to the cytosol is the phosphoglycerate kinase of flowering plants (Martin & Schnarrenberger, 1997).

The fact that EGT was not only a process that took place in the past but is still at work to date is reflected in the presence of large pieces of nuclear DNA still highly homologous (99% identical) to the organellar copy (Stupar *et al.*, 2001), suggesting that this transfer took place as recently as only 2 million years ago (Henze & Martin, 2001). The question arises if at some point in the future all organellar genes will have been transferred to the nucleus. In the case of hydrogenosomes this process of EGT has gone to completion (Martin & Müller, 1998). In

contrast, in mitochondria and chloroplasts it seems that genes coding for key proteins of the electron transport complexes will remain in the organelles as they need to be in direct control of the expression of those genes to react to changes in redox balance (Allen, 2003). For the same reason, genes coding for the translational machinery also tend to remain in the organelles. This explains why in hydrogenosomes all genes were transferred: Since in hydrogenosomes no respiration takes place there is nothing left to be translated and the corresponding genes could be deleted as well. One additional reason why not all genes might be transferred to the nucleus is that gene products can be toxic to the cell if they reach the cytosol. This is the case for the mitochondrial protein cytochrome c, which is exported out of the organelle to induce apoptosis.

1.3. Organellar protein translocons and transit peptides

The synthesis of mitochondrial and chloroplast proteins in the cytosol required the development of protein translocation systems (translocons) ensuring these proteins to be targeted to their respective destinations. Eukaryotic cells target nuclear encoded proteins to several different compartments using N-terminal sequence extensions. Signal peptides direct proteins to the lumen of the ER from where they are trafficked to various cellular destinations or are secreted. Two different classes of so-called transit peptides target proteins into either mitochondria (the mitochondrial signaling peptide should correctly be designated 'presequence') or chloroplasts. Neither mitochondrial nor chloroplast targeting peptides show conserved motifs – but molecular chaperone Hsp70 binding motifs are present in most of them – and both are very variable in their amino acid composition (von Heijne *et al.*, 1989). Nonetheless, they feature a remarkable similarity in their basic amino acid composition: they are rich in hydroxylated, hydrophobic and positively charged amino acid residues and deficient in acidic amino acids to keep the net charge positive. Thus, mTPs and cTPs are very different from the average soluble protein.

The transit peptides of mitochondrial targeted proteins are usually made up of 20–60 amino acid residues and the structure of mTPs could be determined quite well. They seem to be built from two structurally distinct domains: one N-terminal region forming an amphiphilic α-helix, and a second strongly hydrophobic region (10–15 aa) positioned C-terminal to the first. Taken together, mTPs display one hydrophobic and one positively charged face (von Heijne, 1986).

Compared to mitochondrial transit peptides and signal peptides, chloroplast transit peptides are poorly defined. They are usually longer than mTPs and show a wide variation in length (20 to more than 120 aa) (Schnell, 1995; Chen & Schnell, 1999) and sequence. They are especially enriched in serine (19%) and threonine (9%) residues which can be used to distinguish them from mTPs (von Heijne *et al.*, 1989). A semi-conserved motif could be identified around the cleavage site of the stromal processing peptidase (SPP): (I/V)–X–(A/C)↓)A (Gavel & von Heijne, 1990b). In comparison to mTPs, knowledge of the structure of cTPs is limited due to the basic properties of cTPs: in aqueous environment cTPs are largely unstructured (Bruce, 1998; Krimm *et al.*, 1999; Wienk *et al.*, 1999), suggesting that they have evolved to form a random coil (von Heijne & Nishikawa, 1991). As soon as cTPs get in contact with a hydrophobic environment or are inserted into micelles they exhibit significant α-helical structure(s) (Endo *et al.*, 1992; Bruce, 1998; Krimm *et al.*, 1999; Wienk *et al.*, 1999) or amphipathic β-strands (May & Soll, 1999; Wienk *et al.*, 1999). Such membrane-induced secondary structures in cTPs could be used as recognition elements for the import machinery, especially since the outer chloroplast envelope exhibits highly unusual lipids – it is the only cytosolically-exposed, galactolipid-containing membrane in plant cells. Thus, this exceptional hydrophobic environment could induce a common conformation in transit peptides, enabling a single receptor such as Toc159 to facilitate transport of a variety of different precursors (Bölter *et al.*, 1998).

1.3.1. Mitochondrial protein transport and sorting

Protein import of the mitochondrial organelle was the first to be examined and studies have mainly been carried out in yeast. Mitochondria contain two membranes, resulting in four distinct compartments to which proteins can be targeted: the outer membrane (OM), the intermembrane space (IMS), the inner membrane (IM) and the matrix. Two major classes of mitochondrial precursor proteins are known: cleavable proteins featuring the N-terminal extension, or those carrying the targeting signal internally. Three multisubunit translocation complexes have been identified in mitochondria: one translocase of the outer membrane (TOM) and two translocases of the inner membrane (TIM). For an overview see Fig. 1.1, Schnell & Hebert (2003) and Herrmann (2003).

The TOM complex consists of a protein-conducting channel formed by Tom40 subunits,

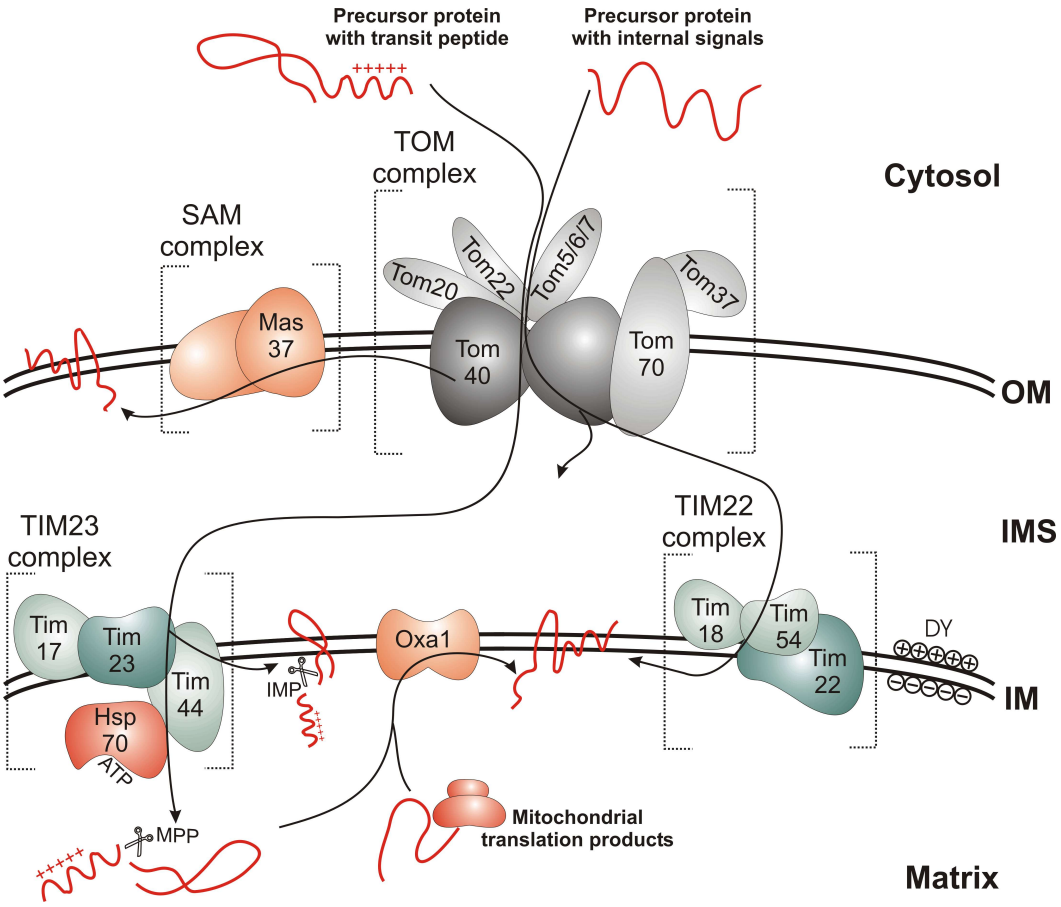

Fig. 1.1.: Overview of the mitochondrial protein translocation and sorting routes. Precursor proteins synthesized in the cytosol bind to the receptors of the TOM complex and are transported into or across the outer membrane (OM). Matrix proteins are further translocated from the intermembrane space (IMS) across the inner membrane by the TIM23 complex and depleted of their targeting signals by the mitochondrial processing peptidase (MPP). Insertion into the inner mitochondrial membrane (IM) can be accomplished by at least three pathways: the TIM22 complex can directly insert proteins not featuring polytopic mTPs; the TIM23 complex can insert monotopic proteins which have to be further processed by the inner membrane protease (IMP); proteins from the matrix (those translocated, as well as mitochondrial translation products) can be inserted via the Oxa1 protein. The driving forces for protein translocation across the inner membrane are ATP hydrolysis and the membrane potential ($\Delta\Psi$).

various targeting signal receptors (Tom5, Tom20, Tom22, Tom37, Tom70) and two accessory factors (Tom6, Tom7). The mitochondrial precursor proteins are kept in an unfolded state by cytosolic factors like MSF (a 14-3-3 protein complex) and targeted to the mitochondria where they are recognized by import receptors (Tom20 being the major one). The subsequent transport through the protein-conducting channel into the IMS does not require the hydrolysis of ATP (Schnell & Hebert, 2003). Proteins to be targeted to the OM are passed from the TOM complex to the separate sorting and assembly complex (SAM) and integrated into the membrane (Wiedemann *et al.*, 2003). After their translocation to the IMS, matrix-targeted proteins are directed by Tim50 (Geissler *et al.*, 2002) and further transported across the IM by the TIM23 complex consisting of Tim23 (targeting signal receptor and channel), Tim17 (channel), and Tim44 (*trans* chaperone binding site). This translocation is driven by membrane potential and ATP hydrolysis involving the matrix chaperone Hsp70 that binds to Tim44 and pulls the protein into the mitochondrial matrix. Proteins to be inserted into the IM are transferred from the IMS by the TIM22 complex consisting of Tim22 (receptor and channel) and the accessory factors Tim54 and Tim18, or first imported into the matrix and then inserted back into the IM by the Oxa1 protein. Oxa1 is also responsible for the insertion of mitochondrial translation products into the IM and many of the proteins inserted by this complex are of bacterial origin. Having arrived in the mitochondrial matrix, proteins are refolded with help of Hsp70 and a 60 kDa molecular chaperone. After this, the signal peptide is cleaved from the protein by the mitochondrial processing protease.

In addition, some proteins can be imported into mitochondria independently from the translocation complexes, most likely because of their ability to spontaneously associate with membrane bilayers or the translocon without the aid of receptors.

1.3.2. Chloroplast protein transport and sorting

Chloroplasts contain three membranes and thus feature six distinct compartments: the outer envelope (OE) membrane, the intermembrane space, the inner envelope (IE), the stroma, the thylakoid membrane (TM) and the thylakoid lumen. One translocon of the OE and one for the IE are known and in accordance to mitochondria designated TOC and TIC, respectively. For an overview see Fig. 1.2, Jarvis & Soll (2001) and Schnell & Hebert (2003). Those two

complexes interact at contact sites and the precursor protein is translocated across both membranes simultaneously (Kouranov *et al.*, 1998) in an energy-dependent manner, involving the hydrolysis of ATP and GTP. Proteins targeted to the thylakoid lumen feature a bi-partite targeting sequence, consisting of a N-terminal cTP followed by a thylakoid transfer domain. The translocation across and sorting into the thylakoid membrane is achieved by at least four different mechanisms, involving the Sec, Tat and signal recognition particle (SRP) pathways originating from cyanobacteria, as well as a spontaneous insertion mechanism for thylakoid membrane proteins of more recent origin.

The TOC complex consists of a trimeric channel-forming core containing Toc75 in association with the two homologous membrane GTPases/receptors Toc34 and Toc159, most likely gating the TOC channel (Bauer *et al.*, 2001), and another receptor, Toc64, interacting with the cytosolic guidance complex. Toc159 exists in both cytoplasmic as well as membrane bound forms and most likely acts as a soluble transit peptide receptor. To import precursor proteins across the OE some are first bound to the cytosolic guidance complex consisting of factors like Hsp70 and a 14-3-3 protein. Successively, they dock to the TOC complex, involving both protein and lipid components of the outer membrane. Transport through the translocon occurs – in accordance to mitochondrial transport – in an unfolded state of the proteins maintained by the chaperone Hsp70 (May & Soll, 2000).

Only a single TIC translocon has been identified to date (Bauer *et al.*, 2001) whose exact composition is not known and discussed controversially – probably due to two or more existing TIC complexes as paralleled by the TIM complexes in mitochondria. Nevertheless, indirect proof connects the two subunits Tic20 (channel, featuring a distant homology to the Tim17/22/23 family) and Tic110 (*trans* chaperone binding site) with protein translocation (van den Wijngaard *et al.*, 2000; Chen *et al.*, 2002). Further Tic40, supposedly a regulator of the chaperones responsible for driving protein import, is closely associated to Tic110. Tic55 is an integral membrane protein featuring a Rieske-type iron-sulfur cluster and possibly involved in response to changes in the redox status of the chloroplast (Jarvis & Soll, 2001). Another component possibly involved in redox signaling is Tic62 which – due to its strong homology to NAD(P)H-dehydrogenases – also might act as a sensor for the ratio of oxidized to reduced NADPH. As a matter of fact, control of protein import into chloroplasts could be linked to

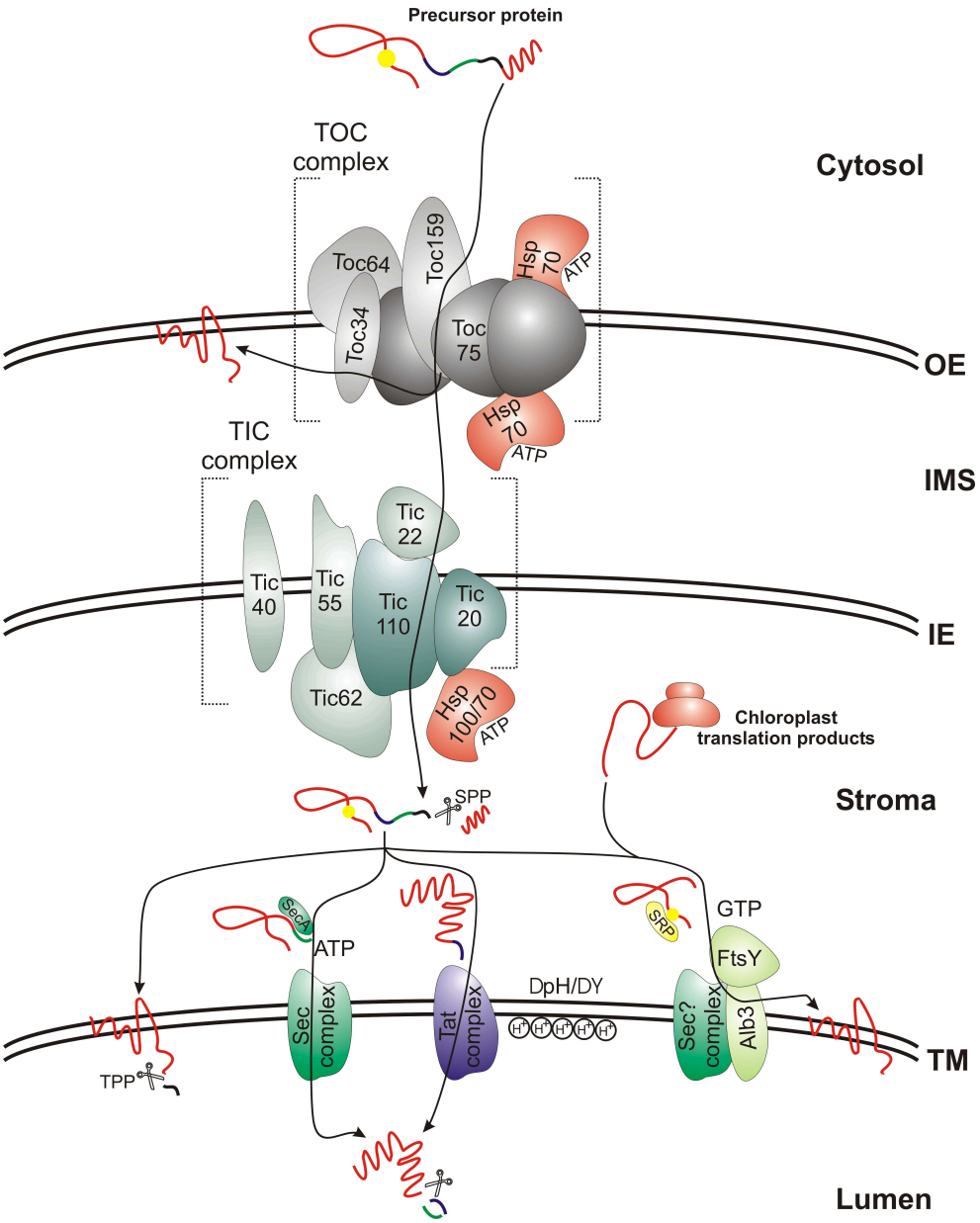

Fig. 1.2.: Overview of the chloroplast protein translocation and sorting routes. Some precursors bind to the cytosolic guidance complex in an ATP consuming manner (not shown) and are brought into the vicinity of the outer envelope (OE) inducing conformational changes in the cTP. After docking to the receptors of the TOC complex they are transported into or across the OE, taken over by Tic22 and Tic20 and threaded through the TIC complex across the inner envelope membrane (IE). Hsp70 and/or Hsp100 pull the precursors into the stroma using up ATP, and the stromal processing peptidase (SPP) cleaves off the transit peptide. For translocation across or into the thylakoid membrane (TM) four different mechanisms are known, relying on additional targeting sequences: 1) the spontaneous insertion (black peptide) possibly involves yet unknown machineries. 2) for the Sec pathway (green peptide) proteins interact with the stromal SecA in an ATP consuming reaction and are translocated unfolded, whereas for 3) the Tat pathway (blue peptide), which is presumably driven by the electric component of the proton gradient (ΔpH) only, no additional factors are needed and translocation occurs in the folded state. 4) the SRP pathway (yellow circle) requires a stromal signal recognition particle (SRP), GTP and the FtsY and Alb3 proteins. This pathway might be involved in translocation of gene products of the plastome. Targeting peptides are cleaved off by the thylakoid processing peptidase (TPP).

the redox state (Hirohashi *et al.*, 2001). The hydrophilic Tic22 is associated with the outer surface of the IE and might act as a receptor for precursor proteins emerging from the TOC complex, or be involved in the association of the TIC and TOC complexes at the contact sites (Kouranov *et al.*, 1998). Tic20 is a hydrophobic protein of the inner membrane also thought to be involved in the recognition and conductance of the precursor proteins, and both Tic22 and Tic20 seem to interact in sequence, passing on the precursors (Kouranov & Schnell, 1997; Kouranov *et al.*, 1998). Translocation is driven mainly by hydrolysis of ATP in the stroma (Olsen & Keegstra, 1992) involving Hsp70 and/or Hsp100 (Nielsen *et al.*, 1997). Once the precursor has entered the stroma it is processed by the stromal processing peptidase cleaving off the transit peptide.

The knowledge about mechanisms of translocation into or across the thylakoid membrane is mostly based on analyses of a few selected proteins and only little is known about the exact structure or composition of the translocons. Proteins which have to be further translocated into the thylakoid feature an additional signaling sequence which is positioned directly C-terminal to the cTP (later cleaved off by the thylakoid processing peptidase) or, in case of the SRP pathway, within the protein. These signaling sequences are recognized by four different mechanisms which are either responsible for a translocation across the TM (Sec and Tat translocons), or into the membrane (SRP and spontaneous insertion). The Sec and Tat translocons have mainly been analyzed in bacteria and differ in one major functional point: while the Sec translocon transports an unfolded protein – which needs to bind to the stromal SecA protein beforehand – across the membrane by use of ATP hydrolysis, the Tat mechanism is able to translocate large (at least up to 132 kD), folded proteins featuring a twin-arginine motif situated upstream of a hydrophobic stretch (Chaddock *et al.*, 1995). This is a remarkable feature, since the TM has to stay sealed in order to prevent large proton efflux. Most likely the Tat translocon is thus used for those lumenal proteins that fold too quickly and/or tightly. The Tat mechanism was long thought to be dependent on the pH gradient across the membrane, but recently Finazzi *et al.* (Finazzi *et al.*, 2003) could demonstrate that thylakoid targeting by the Tat machinery shows no ΔpH dependence *in vivo*. Presumably, the driving force for Tat mechanism is most likely an electric gradient across the membrane ($\Delta\Psi$) that originates from the electric component of the ΔpH. For the insertion of proteins into the thylakoid membrane

the SRP pathway or a spontaneous insertion is employed. Proteins featuring an SRP-type targeting signal are bound by the chloroplast SRP which consists of the two proteins cpSPR54 and cpSPR43 (Schuenemann *et al.*, 1998), and the 7SL RNA molecule (Yukawa *et al.*, 2002). Insertion depends on at least four more factors: GTP hydrolysis, the stromal FtsY protein, an Oxa1 homologue termed Alb3, and a so far unidentified thylakoid-bound translocation machinery. It is hypothesized that the SRP pathway makes use of the Sec translocon (Scotti *et al.*, 2000) and is probably also used for proteins synthesized in the chloroplast (Nilsson *et al.*, 1999). The spontaneous insertion of thylakoid membrane proteins seems to be a very recent invention of the chloroplast, since it is virtually unique to plants (Robinson *et al.*, 2001). Proteins like Psb-W and Psb-X are able to insert into the membrane in absence of SRP, NTPs, a functional Sec machinery, or the proteolysis of thylakoids (Robinson *et al.*, 2001) and thus the term of a spontaneous insertion was coined. However, it can not be completely ruled out that other mechanisms (especially mediated by Alb3) are involved, especially since the targeting signal for the spontaneous import shares similarities to that of the Sec-type signals.

1.3.3. Structural genomics in organelles

Recent advances in large-scale genome sequencing projects of various species have led to the accumulation of a vast number of protein sequences – partly with unknown functions. The automatic identification of transit peptides within these sequences is of highest interest for a number of reasons. For instance, the prediction of the location of unknown proteins may be used to gain indication of its function(s); or it may be used to screen candidate genes for drug discovery by monitoring changes in gene expression in response to drug treatments (Debouck & Goodfellow, 1999). Also, the sizes of the organellar proteomes, which are difficult to be determined experimentally, can be estimated. A computational prediction is feasible, since both mTPs and cTPs share recognizable – yet only indistinct – features. The objective of all attempts made so far is the generalization of information of known transit peptides and transformation into a working algorithm. Such algorithms are in turn used to predict transit peptides in unknown amino acid sequences with a fair rate of success. In general, all employed algorithms are based on either a number of physiochemical parameters (e.g. abundance of certain amino acids, hydrophobicity in certain regions etc.), or the analysis of the

plain residue patterns in (part of) the amino acid sequence. The use of structural information for the prediction – although a very promising concept – has not yet been successful. Most predictors can be grouped into two classes according to their basic algorithms: firstly software like PSORTII (Nakai & Horton, 1999) or iPSORT (Bannai *et al.*, 2002) rely on knowledge-based, multicategory subcellular localization, while in the other class neural networks like in TARGETP (Emanuelsson *et al.*, 2000) or PREDOTAR (Small, 1999) are employed. Other, less prominent, algorithms are based on principal component analysis and linear regression (Schein *et al.*, 2001) or hidden Markov models (Fujiwara *et al.*, 1997).

Structural genomics in *Arabidopsis thaliana* have been conducted before, estimating the proteomes of both mitochondrion and chloroplast, and yielding quite diverging numbers: based on the specificity threshold, for the mitochondrion between 349 and 2,897 and for the chloroplast 2,085 to 3,574 nuclear encoded proteins have been predicted, equaling 1.4 to 11.4% and 8.2 to 14.0% of the total Arabidopsis proteome, respectively (The Arabidopsis Genome Initiative, 2000). Studies based on extrapolation from single chromosomes yielded a proteome size of the chloroplast between 2,900 and 3,500 proteins (Lin *et al.*, 1999; Mayer *et al.*, 1999). In another study, in which also evolutionary aspects by a combination of predictions and comparative genomics were taken into consideration, Pesaresi *et al.* (2001b) predicted a set of 3,100 nuclear encoded *A. thaliana* proteins to be targeted to the chloroplast. Of these 870 stem from the cyanobacterial ancestor and the remaining 2,230 are descendants of the mitochondriate eukaryote. However, these most interesting results are hampered by the still imperfect transit peptide predictions on which they rely, demonstrating that more reliable algorithms and estimates are desperatly needed. This also holds true for functional genomics of organelles (see below) which very heavily depend on correctly identified sets of proteins/genes.

1.4. Organelle-to-nucleus signaling

Since almost all proteins of extant organelles are encoded by nuclear genes which are no longer under immediate control of the mitochondrion and the chloroplast, ways of communication between organelles and nucleus must exist. Organelles have to be able to signal or relay information on their metabolic or photosynthetic state or alterations thereof, ensuring these

changes are countered by a modulation of the expression of the right set of genes. This signaling between the organelles and the nucleus – which is often wrongly referred to as 'cross-talk' – has received wide attention in recent years, but despite the exceptional work of various laboratories only very basic models have been established so far.

For mitochondrion-to-nucleus signaling most work has been done in yeast and especially human, since mitochondrial dysfunctions and mitochondrial stress are associated with severe dysfunctions such as neurodegenerative diseases or various types of cancer, and aging (Delsite *et al.*, 2002; Echtay *et al.*, 2003). Genetic stress is caused, for example, by depletion or damaging of mtDNA. Metabolic stress can result from overproduction of reactive oxygen species often leading to peroxidation of membrane phospholipids and production of reactive toxic aldehydes, uncoupling or disruption of $\Delta\Psi$, or membrane damage (Biswas *et al.*, 2003; Echtay *et al.*, 2003). For mitochondrial signaling in yeast, molecular oxygen and tetrapyrroles, synthesized in the organelle, were shown to regulate the transcription of nuclear genes (Sassa & Nagai, 1996; Forsburg & Guarente, 1989). Furthermore, mitochondria were shown to be capable of stress signaling by Ca^{2+} fluxes and the subsequent activation of the NF-κB/Rel factors (among others) which alter the expression of an array of nuclear genes (Biswas *et al.*, 1999; Amuthan *et al.*, 2002). The NF-κB/Rel family of transcription factors have been studied extensively due to the interesting regulation of various biological processes that they control (overview in Ghosh *et al.*, 1998). NF-κB/Rel is usually complexed with members of the IκB family that constitute an inhibitor of transcription factors restricting their nuclear entry and a subsequent transcription activation or repression of target genes (Beg & Baldwin, 1993; Pahl, 1999). For an activation usually the IκB kinase is employed, which phosphorylates IκB and targets this molecule for ubiquitination and subsequent degradation by the proteasome. In addition, IκB inactivation could recently be linked to calcineurin, a Ca^{2+} and calmodulin-dependent phosphatase, and the Ca^{2+} fluxes mentioned above (Biswas *et al.*, 2003). In its induced state calcineurin dephosphorylates IκB, possibly leading to its degradation or recycling, and activates NFκB/Rel and subsequent gene expressions of several nuclear target genes.

The plastid-to-nucleus signaling has been subject to extensive investigations too (overview in Jarvis, 2001; Rodermel, 2001; Rodermel & Park, 2003) and multiple retrograde signals which coordinate the expression of nuclear genes with the metabolic and/or developmental

state of the plastid could be identified. Most of those plastid signals are components of the photosynthetic apparatus or of the photosynthetic metabolism and include the accumulation of reactive oxygen species, redox poise/state of the thylakoid membrane, carotenoids and tetrapyrroles (Papenbrock & Grimm, 2001; Rodermel, 2001; Surpin *et al.*, 2002). Despite the fact that most retrograde signaling pathways are associated with photosynthesis, it could be demonstrated that not all of them require light and that they do not necessarily interact with the light transduction pathways mediated by phytochromes and cryptochromes. However, in some cases – as with the *laf6* mutant (Møller *et al.*, 2001) – retrograde and light signaling might overlap. Studies of the *immutants* (Wetzel *et al.*, 1994) or the *laf6* mutants furthermore revealed that the plastid is also capable of, so far uncharacterized, 'developmental' signals which control cell differentiation and leaf morphogenesis. Since plastid signaling also influences the expression of non-plastid, as well as, mitochondrial proteins (Barak *et al.*, 2001; Jarvis, 2001), it is thought to exert a major regulatory function on pathways controlling cellular metabolism. It is furthermore clear that interactions among the retrograde signaling pathways and between those and other signal transduction pathways in the cell have to exist (Rodermel, 2001; Oswald *et al.*, 2001).

Of all retrograde signaling pathways, the tetrapyrrole-mediated plastid-to-nucleus signaling is the best understood. At least two different signaling pathways involving tetrapyrroles have been identified (Mochizuki *et al.*, 2001) by characterization of a set of signaling mutants generated in the Chory laboratory (Susek *et al.*, 1993). These mutants are no longer able to repress the *Lhcb* gene of the light-harvesting chlorophyll a/b-binding protein (CAB) of photosystem (PS) II upon treatment with the herbicide norflurazon, causing a complete photodestruction of the chloroplast interior while leaving the rest of the cell intact. This resembles an uncoupling of the nuclear genome from the state of the chloroplast and thus those mutants were designated 'genomes uncoupled' (gun) mutants. The *gun* mutants show a pale green phenotype and are deficient in chlorophyll accumulation. However, the mutations do not affect the light-, tissue-, cell- or circadean-regulated expression of *Lhcb* (Susek *et al.*, 1993). So far, of the five known non-alelic *gun* mutants (*gun1-5*) the four loci *GUN2-5* have been cloned. It was demonstrated that their gene products are on the same retrograde signaling pathway by double mutant analyses and that all four are essential for tetrapyrrole metabolism (Vinti *et al.*, 2000; Mochizuki

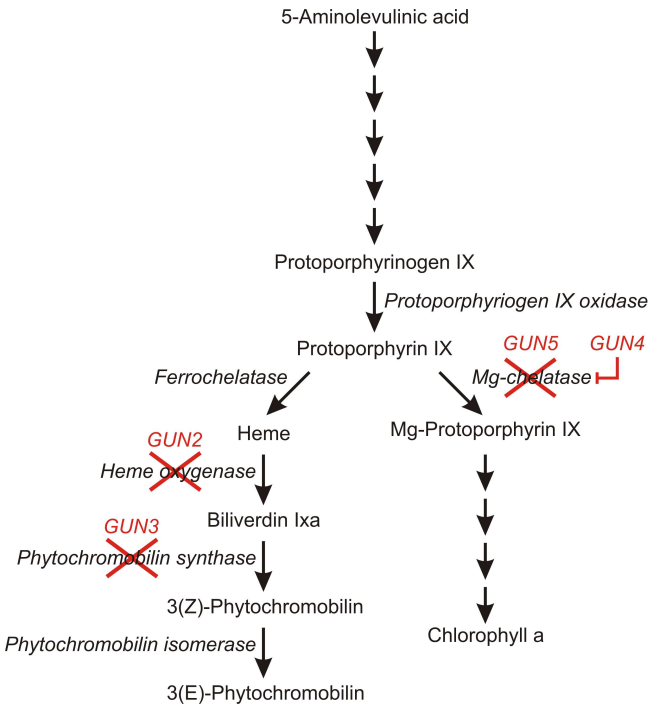

Fig. 1.3.: Overview of gun mutants affecting the tetrapyrrole pathway. Synthesis from 5-aminolevulinic acid to protoporphyrin IX takes place in the stroma, whereas subsequent steps take place on envelope and/or thylakoid membranes (Reinbothe & Reinbothe, 1996). Cloned *GUN* loci and enzymes affected in mutants are depicted/marked in red. Figure adapted from Surpin *et al.* (2002).

et al., 2001; Larkin *et al.*, 2003). While *GUN2, GUN3* and *GUN5* encode enzymes directly involved in the tetrapyrrole pathway (Fig. 1.3), the tetrapyrrole-binding protein GUN4, which is an activator of Mg-chelatase activity, indirectly affects this biochemical pathway (Larkin *et al.*, 2003). In contrast, the *gun1* mutant is not involved in the tetrapyrrole pathway and thus belongs to a different, possibly partially redundant, retrograde pathway (Vinti *et al.*, 2000; Mochizuki *et al.*, 2001). The *laf6* mutant mentioned above is also involved in the tetrapyrrole pathway, even though it is not a *gun* mutant (*Lhcb* expression is not downregulated in photobleached *laf6* mutants). LAF6 is a soluble ATP binding cassette protein that is suggested to be involved in ATP-dependent transport of protoporphyrin IX into the stroma (Møller *et al.*, 2001), so it will be available to the tetrapyrrole pathway. In the *laf6* mutant, the resulting cytoplasmic accumulation of Protoporphyrin IX is proposed to act as a plastid signal affecting light signaling either directly, or indirectly (Møller *et al.*, 2001).

Recent findings have provided further insights into the signaling pathway of the chlorophyll biosynthesis (Larkin *et al.*, 2003; Strand *et al.*, 2003) and Mg-protoporphyrin IX was identified

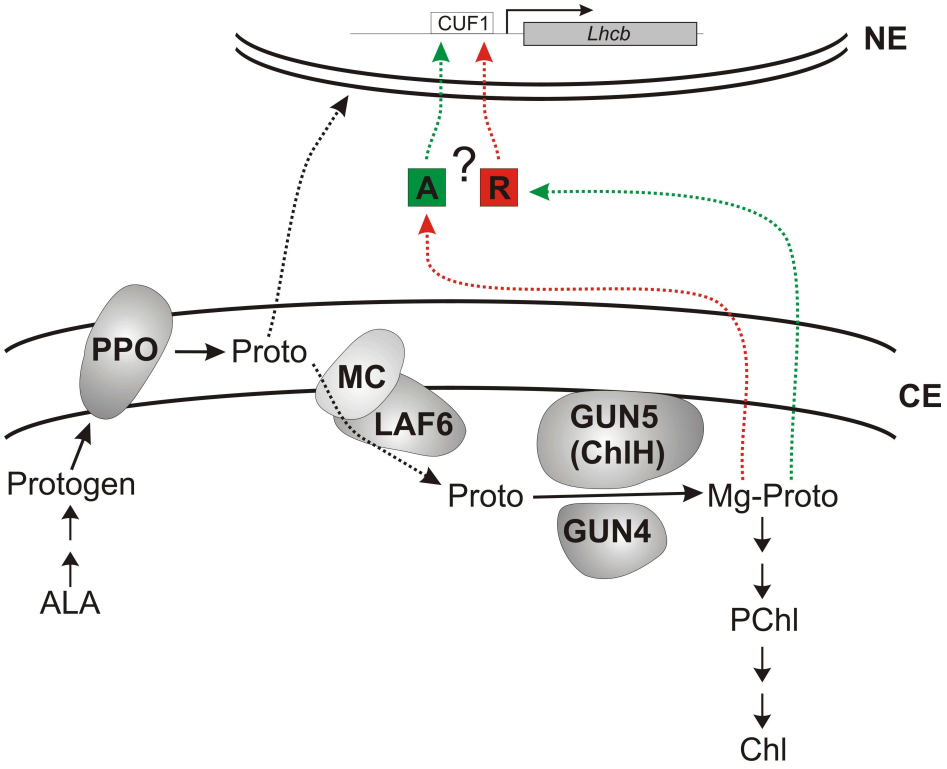

Fig. 1.4.: Models of tetrapyrrole-mediated retrograde signaling. Stromal 5-aminolevulinic acid (ALA) is used to synthesize protoporphyrinogen IX (Protogen), which is oxidized to protoporphyrin IX (Proto) by the protoporphyrinogen oxidase (PPO). A membrane component (MC) and the soluble ATP binding cassette protein LAF6 channel Proto back into the stroma. Mg-protoporphyrin IX (Mg-Proto) is synthesized from Proto by the ChlH subunit of the Mg-chelatase. This process is regulated by the GUN4 protein. Mg-Proto either binds an activator inhibiting its entry to the nucleus, or it promotes the translocation of a repressor to the nucleus. Both, activator and repressor, interact with the CUF1 element of the *Lhcb* gene or other nuclear genes featuring this element. Proto might also directly be used for a different retrograde signaling pathway, as suggested by the findings in the *laf6* mutant. Regular lines depict enzymatic reactions, while dotted lines refer to translocation. Color is used to specify activation (green) or inhibition (red). Abbreviations: chloroplast envelopes (CE), nuclear envelope (NE). Figure adapted from Surpin *et al.* (2002).

to act as the plastid signal (Strand *et al.*, 2003). Even though the exact function of this intermediate in the tetrapyrrole pathways is yet unknown, two different models for the Mg-protoporphyrin IX signal transduction have been established (Fig. 1.4). Both models explain a down-regulation of the *Lhcb* gene in response to an accumulation of the signaling molecule in the chloroplast: firstly, Mg-protoporphyrin IX needs to leave the chloroplast to reach the cytosol in either an active (translocation) or passive (diffusion) manner and in turn (1) binds an activator of *Lhcb* preventing its entry to the nucleus or (2) promotes the translocation of a repressor of *Lhcb* to the nucleus. Both, activator and repressor, are presumed to be capable of binding the CUF1 element of the *Lhcb* gene.

In contrast to the *gun* mutants, another set of mutants, displaying an opposing gene regulation, has been isolated in the Chory lab (Li *et al.*, 1995). These mutants were identified by the aberrant expression of the CAB protein and termed 'CAB-underexpressed' (cue) mutants. Nearly all *cue* mutants are unable to de-repress the *Lhcb* promoter in response to phytochrome activation in the light and thus those mutants are suggested to be light-signaling mutants. However, if the *cue* mutants really are signaling mutants remains to be proven. The signaling phenotype of those mutants might turn out to be a secondary effect in response to metabolic perturbations, which is suggested by the pleiotropic phenotype of the *cue* mutants. The *cue1* mutant is characterized by underexpression of several plastid and nuclear genes encoding plastid-localized proteins, and leaf cell morphology phenotypes, such as, reticulate leaves and fewer palisade cells (Streatfield *et al.*, 1999). Both characteristics are light intensity-dependent. Thus, CUE1 resembles a positive regulator of light-controlled gene expression. *CUE1* encodes the phosphoenolpyruvate/phosphate translocator (PPT) which, in the mutant state, results in a reduced flux through the shikimate pathway. This, in turn, leads to decreased levels of key aromatic metabolites and photosynthetic and light-protective components (Streatfield *et al.*, 1999). Thus, a defective PPT is consistent with the light intensity-dependent phenotypes of the *cue1* mutant.

However, the studies and conclusions summarized here are almost entirely based on a small set of mutants and the expression of only a few genes. Thus, new approaches are needed to improve the understanding of how organelles signal changes in their physiological and developmental state to the nucleus, inducing changes in the nuclear organelle transcriptome. To answer many of these questions a combination of structural and functional genomics will be a valuable tool.

1.5. Organelle-related transcriptomics

The analysis of the transcriptome, which is currently the best developed of the genomic technologies, is a straightforward and well established approach, mostly making use of the micro- or macroarray technique. Both systems are based on standard molecular biology methods at a high miniaturization level. They involve the hybridization of labeled complex cDNA to a

probe immobilized on a solid carrier, and subsequent detection by autoradiography, chemiluminescence or laser-induced fluorescence. Microarrays usually rely on glass slides as a carrier material and detection of fluorescent chemical dyes, allowing for a higher spotting density than macroarrays which make use of nylon membranes and ^{33}P labels. Furthermore, by using two different dyes it is possible to perform one hybridization including two complex cDNA samples at perfectly similar conditions, thus generating a differential expression profile in only one experiment. However, these additional features require very high standards in terms of the working conditions (dust particle filters, constant humidity and temperature etc.) and, most importantly, glass slides can only be reliably used for a single hybridization experiment. Thus, macroarrays are clearly more cost-efficient and require less training time than microarrays.

The prerequisite for the creation of an array is the knowledge of sufficient sequence information for each gene to be monitored. In this work analyses were performed using a gene sequence tag (GST) macroarray. The use of GSTs, as opposed to expressed sequence tags (ESTs; Schenk *et al.*, 2000; Wang *et al.*, 2000; Schaffer *et al.*, 2001), relies on sequence data generated by genome sequencing projects and a subsequent gene prediction and annotation. While GST arrays grant a genome-wide coverage, genes may be selected which are not expressed *in vivo* as they only result from an erroneous automatic prediction.

Following a hybridization experiment, raw data need to be generated, analyzed and evaluated statistically. Firstly, spot detection, background correction and normalization are performed, yielding the raw data. Subsequently, data are evaluated to detect and remove outliers and perform one or more statistical methods, identifying significant changes in gene expression. For such tasks, a two-sample t-test (Devore & Peck, 1997), or a Bonferroni correction (Bland & Altman, 1995) are often employed. However, in order to achieve reliable results, a repetition (at least two to three times) of the hybridization experiment remains mandatory, despite the most sophisticated statistical methods (Lee *et al.*, 2000).

The identified data sets of significantly regulated genes often consist of many thousand entries and need to be refined further. For this purpose, genes can be grouped by their functional classes or the biological pathways they are involved in, to determine if the subsets are over-represented in the data set. In addition, it is now also possible to overlay those subsets with maps of biochemical pathways, resulting in a helpful visualization. Examples

for this kind of software are ARACYC (`http://www.arabidopsis.org/tools/aracyc/`) and GENMAPP (`http://www.genMAPP.org`).

One of the most important methods to reduce the vast amount of expression data – especially from several different treatments/conditions – to a size manageable more easily, is the clustering technique. The most common clustering algorithms are based on hierarchical clustering, self-organizing maps and principal component analysis (Eisen *et al.*, 1998; Tamayo *et al.*, 1999; Sturn *et al.*, 2002). Clustering can be applied in two dimensions, grouping either the genes or the conditions (if more than two are present). By a clustering of genes, insights into biological pathways can become apparent and moreover, functions can be attributed to uncharacterized genes on the basis of co-regulation with known ones (gene function discovery). Clustering of the conditions of a data set visualizes the relatedness of the various 'transcriptome fingerprints', yielding knowledge on the similarities of responses to different mutations or the stimuli applied. Thus, if enough defined 'transcriptome fingerprints' are available, new mutations can be characterized based on a comparison of their transcriptome profile to those of the known mutations.

Another interesting application, which can subsequently be performed to a clustering, is the search for *cis*-acting elements. On the assumption that co-regulated genes should also be regulated by the same promoter elements, members of a cluster can be analyzed for conserved motifs. Such analyses have already been performed successfully (Harmer *et al.*, 2000; Maleck *et al.*, 2000) and will eventually lead to the identification of new transcription factors and DNA-binding proteins. In turn, new insights into gene regulation networks will be obtained.

1.6. Aim of the thesis

This thesis was aimed at the analysis of the semi-autonomous organelles, mitochondrion and chloroplast, by structural and functional genomics. The structural genomics approach concerned questions regarding the sizes, evolution, diversification and functional classification of extant organelles' proteomes and was based on the following strategies:

1) to more reliably estimate the sizes of the organellar proteomes and to lay the grounds for additional analyses, a new method for the *in silico* prediction of precursor proteins had to be

established.

2) to analyze how many organellar proteins can still be traced back to their prokaryotic origin and to determine the extent of diversification of the organellar proteomes between species, comparative and combinatorial BLAST analyses of organellar proteomes of various species were conducted. This approach was also employed to identify mutual subsets within the organellar proteomes which were subsequently subjected to a functional classification.

3) to determine how many genes of the Arabidopsis genome are derived from cyanobacteria, a maximum likelihood phylogenetic inferrence was employed. To make this analysis as reliable as possible, the Arabidopsis genome was compared to 20 other reference genomes of various bacteria and yeast.

The functional genomics approach addressed questions concerning the plastid-to-nucleus signaling in the flowering plant *A. thaliana* and the following approaches were taken:

1) to identify as many different types of transcriptome changes in response to chloroplast signaling as possible, a broadly based survey was conducted. This survey made use of a 3,292 GST array, designed to monitor the nuclear transcriptome of the chloroplast and encompassed 35 different conditions/stimuli of either genetic or environmental nature thought to affect the chloroplast.

2) to draw biological insights from the expression profiles resulting from the survey, a data reduction had to be carried out and was performed by a clustering of both genes and conditions. Cladograms resulting from hierarchical clustering analyses are not suitable to compare large numbers of expression profiles. Thus, a new method for the visualization – a more-dimensional transcriptome response map – had to be developed.

2. Material & Methods

2.1. Protein sequence data

2.1.1. Amino acid sequences used for analyses

Amino acid sequences for the following species were downloaded from the National Center for Biotechnology Information (NCBI, `http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/EG_T.html`): *Agrobacterium tumefaciens* str. C58, *Anopheles gambiae, Aquifex aeolicus* VF5, *Archaeoglobus fulgidus* DSM 4304, *Bacillus cereus* ATCC 14579, *Borrelia burgdorferi, Brucella melitensis, Brucella suis 1330, Caulobacter crescentus* CB15, *Caenorhabditis elegans, Chlamydia muridarum, Drosophila melanogaster, Escherichia coli* CFT073, *Encephalitozoon cuniculi, Haemophilus influenzae* Rd, *Helicobacter hepaticus* ATCC 51449, *Homo sapiens, Methanobacterium ivanovii, Methanococcus vannielii, Mesorhizobium loti, Nostoc* sp. PCC 7120, *Mycobacterium tuberculosis* CDC1551, *Plasmodium falciparum, Pyrococcus horikoshii , Rickettsia conorii, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Rickettsia conorii, Rickettsia prowazekii, Sinorhizobium meliloti, Synechocystis* sp. PCC 6803, *Thermosynechococcus elongatus* BP-1, *Treponema pallidum.*

The proteome sets for *Fugu rubripes* and *Arabidopsis thaliana* were downloaded from the Joint Genome Institute (JGI, `http://genome.jgi-psf.org/fugu6/fugu6.download.ftp.html`) and the Munich Information Center for Protein Sequences (MIPS, `ftp://ftpmips.gsf.de/cress/arabiprot/`), respectively. The protein sequences for *Nostoc punctiforme* and *Prochlorococcus marinus* were retrieved from the U.S. Department of Energy web site (`http://www.jgi.doe.gov/JGI_microbial/html`). Protein sets for *Oryza sativa* were downloaded from The Institute for Genomic Research (TIGR, `ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/`) and from the Rice Genome Automated Annotation

System (RiceGAAS, ftp://ftp.dna.affrc.go.jp/pub/RiceGAAS/current/).

2.1.2. Identities of proteins with experimentally confirmed localization and assembly of test sets

Accession numbers of proteins with experimentally confirmed localization were extracted from the GeneOntology™ Consortium site (http://www.geneontology.org/index.shtml) and had to be identified based on 'traceable author statement' and 'inferred from direct assay' (immunofluorescence, cell fractionating, physical interaction/binding assay).

For the test set used in the analysis of mitochondrial proteomes (section 3.1) the identities of 800 proteins (256 were identified based on 'traceable author statement' and 544 'inferred from direct assay') were downloaded and combined with a set of 106 additional identities downloaded from http://www.gartenbau.uni-hannover.de/genetik/Page3.html, or supplied by Harvey Millar (University of Western Australia). The resulting test set of 906 proteins experimentally confirmed to be targeted to the mitochondrion is enclosed in the appendix (section B.4, p. 108 ff).

For the test set of proteins with known subcellular localization used in the analysis of chloroplast proteomes (section 3.2.3), 2,450 proteins whose subcellular locations are known, were extracted from various publications (Klee *et al.*, 1987; Flügge *et al.*, 1989; Koncz *et al.*, 1990; Block *et al.*, 1991; Keith *et al.*, 1991; Last *et al.*, 1991; Niyogi & Fink, 1992; Schmidt *et al.*, 1994; Sun & Kamiya, 1994; Gaubier *et al.*, 1995; Ohta *et al.*, 1995; Weber *et al.*, 1995; Zhao & Last, 1995; Cunningham *et al.*, 1996; Lindahl *et al.*, 1996; Wu *et al.*, 1997; Essigmann *et al.*, 1998; Kieselbach *et al.*, 1998; Kliebenstein *et al.*, 1998; Ho *et al.*, 1999; Lao *et al.*, 1999; Miège *et al.*, 1999; Nakabayashi *et al.*, 1999; Chen *et al.*, 2000; Moore *et al.*, 2000; Peltier *et al.*, 2000; Weber *et al.*, 2000; Adam *et al.*, 2001; Haussühl *et al.*, 2001; Peltier *et al.*, 2001; Thelen *et al.*, 2001; Block *et al.*, 2002; Knight *et al.*, 2002; Peltier *et al.*, 2002; Schubert *et al.*, 2002; Wilson *et al.*, 2002; Balmer *et al.*, 2003) or from the GeneOntology Consortium™ database. The composition of the test set was designed to reflect the actual protein composition of an angiosperm cell (The Arabidopsis Genome Initiative, 2000), and consisted of 10% with a cTP, 10% with an mTP, and 80% with neither of the two. Thus, 245 of the proteins in this set are targeted to the chloroplast, another 245 to the mitochondrion and the remaining 1,960

proteins were randomly chosen from a set of about 9,000 proteins (GeneOntology Consortium™ database) targeted to neither of these organelles. The resulting test set is enclosed in the appendix (section B.6, p. 115 ff).

2.1.3. Construction of non-redundant sets and protein pools

Since a non-redundant genome annotation for *Oryza sativa* is not yet publicly available the protein sets downloaded from TIGR and RiceGAAS were combined and had to be cleared of redundant entries. For this purpose, proteins of the two sets annotated to be encoded by genes on the same chromosome were compared by BLASTP at an E value threshold of 0. Resulting hits were removed from one of the two protein sets resulting in a non-redundant set of 64,582 ORFs.

The pools of α-proteobacterial and cyanobacterial proteins were assembled by adding the total proteomes of the corresponding species – no verification for redundant entries was performed. For the α-proteobacterial pool the proteomes of *Agrobacterium tumefaciens, Brucella melitensis* and *suis, Caulobacter crescentus, Mesorhizobium loti, Rickettsia conorii* and *prowazekii,* and *Sinorhizobium meliloti* were combined into a set comprising 71,310 protein sequences. In case of the cyanobacterial pool the amino acid sequences encoded in the genomes of *Nostoc* sp. PCC 7120, *Synecchocystis* sp. PCC 6803 and *Thermosynecchocystis elongatus* BP-1 were assembled into a set of 11,771 proteins.

2.2. Software

2.2.1. Sequence comparison software

For sequence comparisons and alignment the programs BLASTP, BLASTN and CLUSTALW were used. BLASTP and BLASTN are part of the NCBI-BLAST (Altschul *et al.*, 1990) software package (v.2.2.6) and obtainable through `ftp://ftp.ncbi.nih.gov/blast/executables/`. All BLAST analyses were performed at an E value threshold of 10^{-10} and standard settings if not stated otherwise.

For the calculation of the distance matrix of the pairwise BLAST analysis (see appendix B.1) the following formula was used: let mTP_x be a mitochondrial protein of species x and $\mathrm{mTP}_{\vec{xy}}$

a mitochondrial protein of species x with homology to a mitochondrial protein of species y then the distance between species a and b is calculated as

$$\text{Distance}_{a,b} = 1 - \frac{\sum \text{mTP}_{\vec{ab}} + \sum \text{mTP}_{\vec{ba}}}{\text{mTP}_a + \text{mTP}_b} \tag{2.1}$$

CLUSTALW (Thompson *et al.*, 1994) v1.82 was downloaded from `ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalW/` and used for sequence alignment. It was set to write in PHYLIP format in non-interleaved mode.

2.2.2. Predictors for targeting peptides

The following freely available or publicly accessible predictors and settings were used in this work and if not noted otherwise run locally with standard settings. If no local versions were available predictions were automated by custom-made PERL scripts (see 2.2.5).

TARGETP (Emanuelsson *et al.*, 2000) v1.01 is available only as an online version at `http://www.cbs.dtu.dk/services/TargetP` and was used with the options: cleavage site prediction turned on and 'winner-takes-all' selected. iPSORT (Bannai *et al.*, 2002) was used as downloaded from `http://biocaml.org/ipsort/ipsort-x86-linux`. PSORTII (Nakai & Horton, 1999) is available from the author (`knakai@ims.u-tokyo.ac.jp`) and was trained with the standard training set supplied. PREDOTAR (Small, 1999) v0.5 (`http://www.inra.fr/predotar/`) was used at a threshold of 0.5 for either cTP or mTP prediction (as suggested by the author) and run under a JAVA environment. PCLR (Schein *et al.*, 2001) v0.9 was run online at `http://apicoplast.cis.upenn.edu/pclr/`

Sensitivity and specificity of a targeting prediction (TP) were assessed as: let tp, fp and fn be the true positives and false positives and negatives of a prediction, respectively, then the sensitivity, which refers to the probability with which any real targeting signal will be identified is calculated as

$$\text{sensitivity}_{\text{TP}} = \frac{\text{tp}}{\text{tp} + \text{fn}} \tag{2.2}$$

while the specificity, which indicates how many of the predicted targeting signals are real is

calculated as

$$\text{specificity}_{\text{TP}} = \frac{\text{tp}}{\text{tp} + \text{fp}} \tag{2.3}$$

To correct for false positive and false negative predictions, actual numbers of proteins containing a cTP or mTP were extrapolated as described previously (Leister, 2003) according to:

$$\text{TP}_{real} = \frac{\text{specificity}_{\text{TP}}}{\text{sensitivity}_{\text{TP}}} \times \text{TP}_{predicted} \tag{2.4}$$

The accuracy of predictors and combinations thereof was assessed using the Matthews correlation coefficient (MCC; Matthews, 1975): let tp, tn, fp, fn be the true positives and negatives and false positives and negatives of a prediction, then

$$\text{MCC} = \frac{\text{tp} \times \text{tn} - \text{fp} \times \text{fn}}{\sqrt{(\text{tp} + \text{fn})(\text{tp} + \text{fp})(\text{tn} + \text{fp})(\text{tn} + \text{fn})}} \tag{2.5}$$

2.2.3. Phylogenetic software

For phylogenetic analyses PROTML of the MOLPHY (Adachi & Hasegawa, 1996) program package (v2.3, `ftp://ftp.ism.ac.jp/pub/ISMLIB/MOLPHY/`), and NEIGHBOR and DRAWTREE of the PHYLIP (Felsenstein, 1989) software package (v.3.5, `ftp://evolution.genetics.washington.edu/pub/phylip/`) were employed.

PROTML was used for protein maximum likelihood (ML) calculations after sequences were aligned using CLUSTALW and gaps removed. Trees were inferred using the JTT-F [Jones, Taylor & Thornton with frequencies (Jones *et al.*, 1992; Cao *et al.*, 1994)] substitution model with the neighbor-joining (NJ; Saitou & Nei, 1987) tree of ML distances as the starting topology and RELL (Adachi & Hasegawa, 1996) bootstrapping (10^4). NEIGHBOR was used for NJ calculations using protein distance matrices, and inferred phylogenetic trees were visualized as unrooted trees using DRAWTREE at standard settings. The phylogenetic ML analyses presented in section 3.2.1 of this work (p. 40 ff.) were performed in a collaboration with W. Martin, T. Rujan, A. Hansen, S. Cornelsen, T. Lins, D. Leister, B. Stoebe, M. Hasegawa and D. Penny (Martin *et al.*, 2002). This work reflects 30% of the performed calculations.

2.2.4. Macroarray software

ARRAYVISION (v6.0; Imaging Research) was used for spot detection, as well as background and artifact removal of array hybridization images. ARRAYSTAT (v1.0; Imaging Research) was employed for a statistical analysis (including outlier detection and removal) of the spot values generated by ARRAYVISION. Part of the evaluations by ARRAYVISION and ARRAYSTAT were performed by A. Biehl, A. Dietzmann and J. Kurth. For hierarchical clustering (Euclidian distance) GENESIS (Sturn *et al.*, 2002) (v1.0.1–v1.2.1, `http://genome.tugraz.at/Software/GenesisCenter.html`) was used. The distance matrix out of the 'all-against-all' comparisons of array results (see appendix B.2) was processed by the PHYLOGRAPHER software (`www.atgc.org/PhyloGrapher/`).

2.2.5. Operating system and additional programs

LINUX (SuSE v7.2–8.0, `http://www.suse.com/us/private/download/index.html`) was used as operating system. Protein sequences, results of calculations and array data were stored in a relational database management system (MYSQL v3.23, `http://www.mysql.com/downloads/mysql-3.23.html`). Custom-made programs and scripts were coded in PERL v5.005 (`http://www.perl.com/CPAN/src/perl5.005_03.tar.gz`) using various extension packages, especially the BIOPERL module (v1.2.1, `http://www.bioperl.org/Core/Latest/index.shtml`).

2.3. Transcriptomics using the 3,292 gene sequence tag array

2.3.1. Design of the GST array

A total of 3,292 GSTs (for identities see: `http://www.mpiz-koeln.mpg.de/~leister/chloro1/GSTs.htm`), consisting of 2,661 proteins most likely to be targeted to the chloroplast and 631 non-chloroplast proteins were designed by use of the GST-PRIME software (Varotto *et al.*, 2001). The primer pairs were synthesized as 35-mers, consisting of a gene-specific 20-mer sequence (chosen by GST-PRIME) and a tailing universal 15-mer sequence at the 5' terminus, suitable for re-amplification with universal tail primers. The amplification of the GSTs, ranging from 150 to 2,050 bp, was performed as decribed (Varotto *et al.*, 2001). All PCR products were analyzed

and quantified by agarose gel electrophoresis and spotted in duplicate onto nylon membrane (Eurogentec) in a geometry depicted in Fig. 3.9, p. 51. Primer synthesis, PCR amplification of GSTs, quality control of PCR products, and spotting were performed by Eurogentec, Seraing, Belgium.

2.3.2. Hybridization experiments

Hybridization experiments were performed by A. Biehl, A. Dietzmann and J. Kurth. Growth of *A. thaliana* plants, RNA preparation, probe synthesis and hybridization were performed according to Kurth *et al.* (2002).

2.3.3. Data analysis

Hybridization images were read using the Storm phosphoimager (Molecular Dynamics), imported into the ARRAYVISION software for background removal and quantification, and statistically evaluated using ARRAYSTAT. Data were normalized with reference to all spots on the array (Kurth *et al.*, 2002) and average expression ratios derived from at least three independent experiments (equaling six data points per GST). This part of the data analysis was performed, in great parts, by A. Biehl, A. Dietzmann and J. Kurth.

For the 'all-against-all' comparisons, ratios of the number of identical genes that were differentially expressed in the same directions versus all genes differentially expressed in each pair of conditions were calculated. The resulting similarity matrix was transformed into a distance matrix by subtracting the values from 1 (see appendix B.2), and subsequently processed by the PHYLOGRAPHER software.

Clustering of expression ratios was performed using the GENESIS software package. Average expression ratios out of the ARRAYSTAT analysis were imported and a base 2 logarithmic transformation was carried out. Hierarchical clustering was performed using Euclidean distances.

3. Results

3.1. Structural genomics in mitochondria

This chapter describes the prediction of the size, diversification and conservation of the mitochondrial proteome for ten species by a combination of computational identification of protein targeting and comparative genomics. Using this approach a core set of conserved proteins present in all non-parasite species under investigation could be identified – many of them associated with inherited human diseases.

3.1.1. Prediction of the size of mitochondrial proteomes of ten different species

The ten eukaryote nuclear genomes considered in this analysis contain between 2,000 and 40,000 open reading frames (Tab. 3.1). For three of them (*S. cerevisia, H. sapiens* and *A. thaliana*), a substantial number of mitochondrial proteins has been experimentally determined but these three sets are far from representing the mitochondrial proteomes completely. Furthermore, protein-encoding genes in the mitochondrial DNA (mtDNA) of nine of the ten species have also been identified and listed in Tab. 3.1.

For a genome-wide identification of nuclear genes coding for proteins that feature a mitochondrial targeting peptide, three different algorithms implemented in the software packages PSORTII (Nakai & Horton, 1999), iPSORT (Bannai *et al.*, 2002) and TARGETP (Emanuelsson *et al.*, 2000) were employed. TARGETP and PSORTII identified an almost equal number of mitochondrial proteins, while iPSORT produced a substantially higher number of predicted mTPs (Fig. 3.1). The specificity and sensitivity of predictors have been evaluated (Emanuelsson *et al.*, 2000; Bannai *et al.*, 2002), enabling a correction of the predicted number of mTPs for each genome (see Eq. 2.4, p. 28). The genomes of the two vertebrates *H. sapiens* and

Organism	ORFs in nuclear genome	mtDNA encoded proteins	Nucleus-encoded proteins experimentally localized to mitochondria	mTPs identified by at least 2 of 3 predictors
H. sapiens	37,891	13	197	4,162
F. rubripes	38,626	13	nd	3,755
A. thaliana	26,620	57	124	2,830
C. elegans	20,206	12	39	1,795
A. gambiae	15,212	13	nd	2,110
D. melanogaster	14,335	13	17	1,291
S. cerevisiae	6,308	28	516	829
S. pombe	5,000	10	7	511
P. falciparum	5,267	3	6	180
E. cuniculi	1,996	Unknown	nd	156

Tab. 3.1.: Number of experimentally confirmed and predicted mitochondrial proteins. Predictions of mTPs were performed by a combination of the three predictors ipsort, psortII and targetp.

F. rubripes, as well as the genome of the flowering plant *A. thaliana* should encode for about 3,000 mTP-featuring proteins. In contrast, the two yeast species should contain less than 1,000 mTPs (Fig. 3.1). An intermediate size of mitochondrial proteomes resulted for the arthropodes *D. melanogaster*, *A. gambiae*, and the nematode *C. elegans*, while for the parasitic species *P. falciparum* and *E. cuniculi* only a small number of nuclear-encoded mitochondrial proteins was predicted. In *P. falciparum*, this parallels an unusually small mitochondrial genome, coding for only three proteins.

When the accuracy of the three mTP predictors was tested on the total of 906 mitochondrial proteins identified experimentally (Tab. 3.1), each 462 (PSORTII), 523 (TARGETP), or 525 (IPSORT), corresponding to 51–58%, were predicted correctly. This value is lower than reported before (Emanuelsson *et al.*, 2000; Schein *et al.*, 2001; Bannai *et al.*, 2002) and, in addition, the specificity of the predictors varied: mTP-featuring proteins predicted by TARGETP were most enriched for experimentally verified mitochondrial proteins in most of the organisms tested, indicating that this program was more specific than the other two. An evaluation of the predictors according to the method described in section 3.2.3 below was not performed, since the test set employed was assembled prior to that one used in section 3.2.3 and did not feature any experimentally confirmed non-mitochondrial proteins. Thus, a calculation of the specificity according to Eq. 2.3 and a subsequent calculation of the Matthews correlation

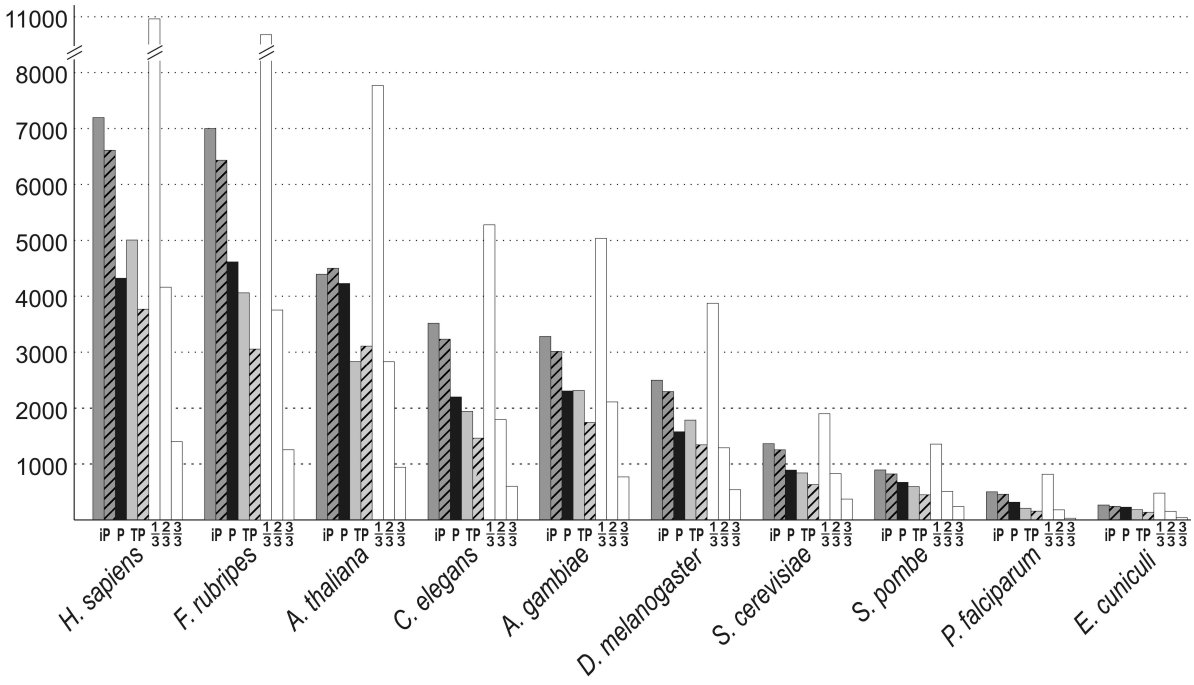

Fig. 3.1.: Prediction of mTP-featuring proteins. The proteomes of the ten eukaryotic species considered in this study were analyzed for mTP-featuring protein sequences. The three predictors ipsort (gray columns, iP), psortII (black columns, P) and targetp (light-gray columns, TP) were employed. For targetp and ipsort predictions the actual number of mTPs (hatched gray and light-gray columns, respectively) was calculated as described (Leister, 2003, see Eq. 2.4, p. 28), considering the specificity/sensitivity of targetp and ipsort (non-plant: 0.67/0.89; plant: 0.90/0.82) and ipsort (non-plant: 0.68/0.74; plant: 0.86/0.84). White columns refer to combinations of the three predictors, indicating the total number of mTPs identified either by at least one ($\frac{1}{3}$), or two ($\frac{2}{3}$), or three ($\frac{3}{3}$) predictors. Maximum sensitivity of mTP prediction was achieved by adding the results of three programs. Using this additive combination, 646 (71%), of experimentally identified mitochondrial proteins were correctly predicted, but at the cost of specificity. Maximum specificity was achieved when an mTP had to be detected by all three predictors. Collections of proteins identified in this way were the most enriched for experimentally verified proteins, but about 60% of true mitochondrial proteins were not detected by the combination of predictors.

coefficient (Eq. 2.5) could not be carried out.

To improve either sensitivity and/or specificity of the mTP identification combinations of the three predictors were employed. Maximum sensitivity of mTP prediction was achieved by adding the results of the three programs. For instance, a sequence was considered to feature an mTP when predicted by at least one of the three programs (Fig. 3.1, column $\frac{1}{3}$). Using this additive combination, 646 (or 71%) of experimentally identified mitochondrial proteins were predicted correctly. However, this increased sensitivity was obtained at the cost of specificity, resulting in a markedly high number of predicted false positive mTPs (compare Fig. 3.1, column $\frac{1}{3}$: more than 2,000 predicted mTPs versus a total proteome size of about 6,000 proteins in *S. cerevisiae*). In contrast, the maximum specificity was achieved when an mTP had to be detected by all three predictors. Although proteins identified in this manner were enriched for experimentally verified proteins, about 60% of true mitochondrial proteins were not detected by this approach (compare Fig. 3.1, column $\frac{3}{3}$: the 373 predicted mitochondrial proteins versus 517 experimentally confirmed in *S. cerevisiae*).

In order to achieve a balance between specificity and sensitivity a third combination of predictors was employed. The results of this combination considering mTPs detected by at least two of the three programs are reported in Fig. 3.1 (column $\frac{2}{3}$). Although this combination predicted a smaller number of mTPs than any of the three programs alone, it resulted in a higher specificity and about the same sensitivity of mTP prediction as with TARGETP or iPSORT (520 of the 906 experimentally verified mitochondrial proteins were detected). The resulting (uncorrected) predictions of the '2 out of 3' combination were closer to the corrected predictions than the original results of the three predictors, and therefore the '2 out of 3' approach was superior to the three original predictors (or any other combination). Thus, as depicted by Fig. 3.1 column $\frac{2}{3}$, functional mitochondria should harbor from a few hundred (budding yeast) to about 4,000 (human) proteins. In all subsequent analyses, mTP predictions were performed using the '2 out of 3' approach.

Organism	Number of proteins predicted or shown to be located in the mitochondrion				
	Total[a]	With α-proteobacterial homologue[b]	Not shared with the mitochondrial proteome of the other species	Not shared with the total proteome of the other species	Conserved[c] among all mitochondrial proteomes
H. sapiens	4,251	553 (13%)	2,524 (59%)	2,000 (47%)	100
F. rubripes	3,768	243 (6%)	2,594 (69%)	1,965 (52%)	75
A. thaliana	2,957	671 (23%)	1,836 (62%)	1,665 (56%)	132
C. elegans	1,820	346 (19%)	967 (53%)	835 (46%)	69
A. gambiae	2,121	447 (21%)	962 (45%)	516 (24%)	65
D. melanogaster	1,483	354 (24%)	562 (38%)	350 (24%)	56
S. cerevisiae	1,082	301 (28%)	535 (49%)	429 (40%)	56
S. pombe	586	191 (33%)	207 (35%)	149 (25%)	46
P. falciparum	187	50 (27%)	115 (61%)	108 (58%)	10
E. cuniculi	156	26 (17%)	94 (60%)	73 (47%)	8

Tab. 3.2.: Evolution of mitochondrial proteomes. [a]This column is derived by the consideration of columns 3 to 5 of Tab. 3.1 [i.e. col. 5 + col. 3 + col. 4 (except those already included in col. 5)]. [b]To identify α-proteobacterial homologues of mitochondrial proteins, amino acid sequences were compared using blastp against a pool of protein sequences from eight α-proteobacteria with fully sequenced genomes. [c]This column refers to the number of proteins with homologues in all mitochondrial proteomes of non-parasite species

3.1.2. BLAST comparison of the predicted mitochondrial proteomes to α-proteobacteria

Because of their endosymbiotic origin, a number of mitochondrial proteins still resemble those of α-proteobacteria. To quantify this prokaryotic heritage, sequences were compared by BLAST analysis (BLASTP at an E value threshold of 10^{-10}). Since the precise prokaryotic lineage giving rise to extant mitochondria is unknown, the mitochondrial proteomes determined for the ten eukaryotic species (Tab. 3.1, columns 3-5; and Tab. 3.2, column 2) were compared to a pool of protein sequences from the genomes of eight α-proteobacterial species. Between 6% (Fugu) and 33% (budding yeast) have an α-proteobacterial homologue (Tab. 3.2, column 3). While the largest fractions of α-proteobacterial homologues were found in unicellular eukaryotes, the smallest were predicted for large mitochondrial proteomes (Fugu and human). This finding indicates that an increase in mitochondrial proteome complexity was accompanied by a post-endosymbiotic relocation of non-prokaryotic gene products into the organelle.

3.1.3. Cross-species BLAST analyses of the mitochondrial proteomes

The comparison of the different mitochondrial proteomes by BLAST analysis (at a threshold E value of 10^{-10}) revealed a variation in the size of the unique fraction – i.e. mitochondrial proteins of one species that are not shared with the mitochondrial proteome of any other species – which ranged from 35% in budding yeast to 69% in Fugu (Tab. 3.2, column 4). This indicates that different degrees of mitochondrial proteome specialization exist. As expected, also non-mitochondrial homologues of species-specific mitochondrial proteins were found by the interspecific BLAST analysis (Tab. 3.2, difference between values in columns 4 and 5). The fraction of these proteins, which feature an mTP in the respective species but do not feature one in any other species, ranged from 3% and 6% (Plasmodium and Arabidopsis, respectively) to 12% and 16% (humans and Fugu, respectively) of the entire mitochondrial proteome.

In parasites, many cellular functions, including those in mitochondria, have been taken over by the host organism. Microsporidia, including *E. cuniculi*, are obligate intracellular parasites that are thought to have lost mitochondria and peroxysomes (Germot *et al.*, 1997). It was recently hypothesized that they might contain a mitochondrion-derived organelle called mitosome (Katinka *et al.*, 2001), and among the 1,996 proteins encoded in the *E. cuniculi* genome, 156 mTP-featuring proteins were predicted. 62 of these tentative mitochondrial proteins had homologues in at least one of the eight non-parasite species (Tab. 3.2), supporting the existence of a mitosome in this organism. Of the 187 predicted mitochondrial proteins of the human malaria parasite *P. falciparum* 72 were homologous to non-parasite mitochondrial proteins.

All sets of mitochondrial proteins were compared with each other set in pairs by BLAST analysis. The two parasite species were excluded due to their extremely specialized metabolism which has lost complexity during the evolution of parasitism. A distance matrix (see appendix, Tab. B.1) based on the fraction of homologous mitochondrion-targeted proteins in each possible pair of species was calculated according to Eq. 2.1, p. 27. In a subsequent step this distance matrix was used to create a similarity tree (Fig. 3.2) using a neighbor-joining algorithm of the NEIGHBOR program, revealing different degrees of relatedness. The mitochondrial proteomes of the two yeast species appeared closely related, sharing about 300 proteins. Similar relationships were found for the two insects *D. melanogaster* and *A. gambiae*, as well as for the

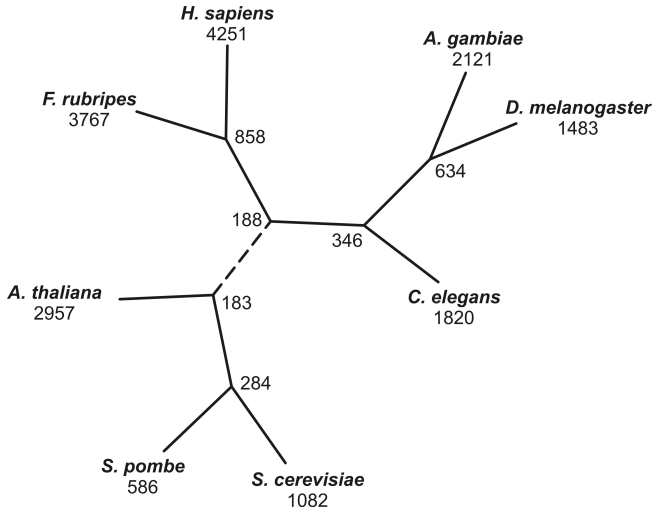

Fig. 3.2.: Similarity tree of the eight non-parasite mitochondrial proteomes. A distance matrix based on the blast analysis of pairs of the mitochondrial proteomes was calculated (see text for details) and subsequently used for the calculation of an unrooted similarity tree applying a neighbor-joining algorithm. Numbers at the branches represent the detected/known mTPs for the corresponding species; numbers at the nodes represent the shared mTPs of the species branching from the according node. The dashed line was introduced to clarify that the 188 and 183 found mTPs are only shared between the five metazoans and the two yeast species and Arabidopsis, respectively. For a set of mTPs shared by all eight species considered see text and Tab. 3.2.

two vertebrates *H. sapiens* and *F. rubripes*, with more than 600 and 800 shared mitochondrial proteins, respectively. About 200 proteins were shared between the five metazoans and, furthermore, proteins with homologues in all of the fully functional mitochondria tested exist. Due to the presence of multiple homologues, the size of this set of conserved mitochondrial proteins varied among species, ranging from 46 in *S. pombe* to 132 in *A. thaliana* (Tab. 3.2).

Since a mitochondrial protein present in all species under analysis should most likely confer essential organellar functions, the corresponding set of 100 orthologues in humans was used and checked against the 'Online Mendelian Inheritance in Man' (OMIM) database. Eighteen of those proteins could directly be associated with human diseases (Table 3.3) and, in addition, of the remaining 82 proteins, 40 (49%) could be mapped to an interval for genetically undefined human disorders (see `http://www.mpiz-koeln.mpg.de/~leister/mito1.html`). In addition, for 61 of the 100 proteins α-proteobacterial homologues were detected by the BLASTP analysis.

A functional classification of the 100 and 132 conserved proteins of *H. sapiens* and *A. thaliana*, respectively, was performed according to the categories described by (Katinka *et al.*, 2001). In humans, 26 proteins with functions in transport, 28 involved in metabolism and

Protein	Disease	OMIM[a]
Glycine dehydrogenase	Glycine encephalopathy	605889
Heat shock 60kDa protein 1	Dominant spastic paraplegia 13	605280
Cytochrome b	Myopathy & excercise intolerance	516020
4-aminobutyrate aminotransferase precursor	GABA transaminase deficiency	137150
Dihydrolipoamide branched chain transacylase	Maple syrup urine disease, type II	248610
α-ketoglutarate dehydrogenase	Oxoglutarate deficiency	203740
Pyruvate dehydrogenase complex, E3-binding protein	Leigh syndrome	266150
Peroxisomal biogenesis factor 7	Rhizomelic chondrodysplasia	215100
ATP-binding cassette, subfamily D, member 3	Zellweger syndrome 2	170995
Paraplegin	Recessive spastic paraplegia 7	602783
ATP-binding cassette, sub-family B, member 7	Sideroblastic anaemia and ataxia	303310
Cystic fibrosis transmembrane conductance regulator, ATP-binding cassette (subfamily C, member 7)	Cystic fibrosis	602421
Dimethylglycine dehydrogenase precursor	Dimethylglycine dehydrogenase deficiency	605580
Solute carrier family 25, member 13	Citrullinemia	603471
Solute carrier family 25 (ornithine transporter) member 15	Ornithine translocase deficiency	238970
Kinesin family member 1B	Charcot Marie Tooth disease 2A	118210
Cytochrome c oxidase I	Mitochondrial disorder	516030
Cytochrome c oxidase II	Mitochondrial disorder	516040

Tab. 3.3.: Conserved proteins directly associated with inherited human diseases. The 100 conserved mitochondrial proteins found in the human species were checked against the OMIM database and yielded 18 directly associated disorders. [a]OMIM: Online Mendelian Inheritance in Man database of human genes and genetic disorders (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM).

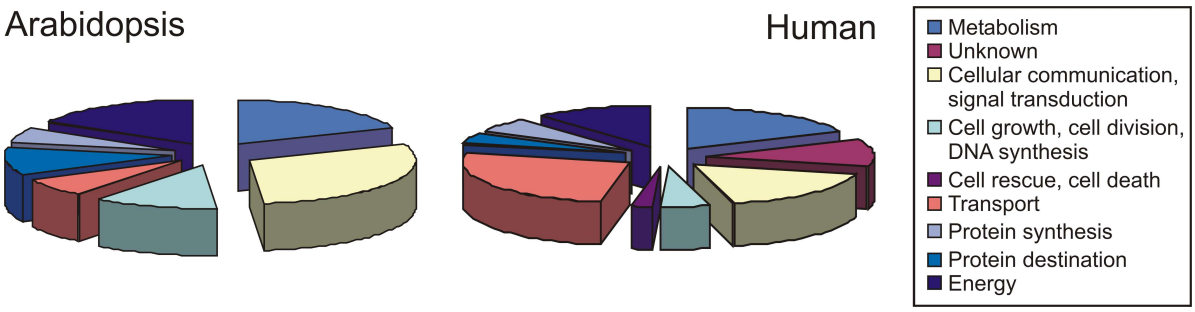

Fig. 3.3.: Distribution of the conserved mitochondrial proteins of humans and Arabidopsis among functional groups. The categories introduced by (Katinka *et al.*, 2001) were used to classify the 132 and 100 core set proteins from Arabidopsis and humans, respectively.

energy production, 17 in cellular communication and signal transduction, and 13 related to protein synthesis, folding and targeting were assigned (Fig. 3.3). In Arabidopsis, 39 proteins with functions in cellular communication and signal transduction, 47 in metabolism and energy production, and 13 related to cell growth, cell division and DNA synthesis were found. These results suggest that the functions of mitochondria in plant and non-plant species have diverged – possible due to a second organelle present in plants: the chloroplast.

3.2. Structural genomics in chloroplasts

Chloroplasts are the descendants of cyanobacteria (or their progenitors) that were taken up during an endosymbiotic event. In this section, an estimate on the number of proteins present in the proteome of *Arabidopsis thaliana* that stem from cyanobacteria is given. Additional estimates have been made regarding the size, composition and diversification of the chloroplast proteome of Arabidopsis in comparison to that of rice. By this approach conclusions about the evolution and function of the chloroplast proteomes of flowering plants could be drawn.

The approximation of the number of proteins in Arabidopsis that descend from cyanobacteria involved both a BLAST analysis, and a more sophisticated phylogenetic approach using protein maximum likelihood inferrence. The BLAST analysis was performed for two reasons: (1) to directly compare the Arabidopsis proteins to cyanobacterial proteins on a genome-wide scale and detect homologues. (2) to define a subset of Arabidopsis proteins that can be used for a phylogenetic inferrence, as such an analysis requires a set of proteins that share sufficient sequence conservation (Nei, 1996; Nei & Kumar, 2000).

The analyses of the chloroplast proteomes of rice and Arabidopsis were carried out in analogy to those for the prediction of the mitochondrial proteome. An improved method for the detection of cTPs was developed and – in combination with BLAST analyses – applied to predict the size, composition and diversification of the chloroplast proteome of the two species.

3.2.1. Comparative BLAST and phylogenetic analyses of Arabidopsis and 20 reference genomes

The analyses presented in this section resulted from a collaboration with W. Martin, T. Rujan, A. Hansen, S. Cornelsen, T. Lins, D. Leister, B. Stoebe, M. Hasegawa and D. Penny (Martin *et al.*, 2002). The share of this thesis reflects 30% of the performed work.

To identify the subset of Arabidopsis proteins satisfying the prerequisite for a phylogenetic analysis 24,990 protein sequences from *A. thaliana* were individually compared to 51,361 protein sequences of the 20 reference proteomes depicted in Fig. 3.4. By this approach 9,368 Arabidopsis proteins were found that – at an E value of 10^{-10} or lower – share a homologue in one of the other species. All subsequent analyses were performed on this subset.

This BLAST analysis also revealed that the largest fraction of the Arabidopsis proteins was most similar to yeast (Fig. 3.4). These genes were probably present in the host cell that acquired plastids (Douglas, 1998; Cavalier-Smith, 2000; Delwiche, 1999; Rujan & Martin, 2001) and have been preserved in both yeast and Arabidopsis. The second largest fraction of Arabidopsis proteins are cyanobacterial acquisitions. BLASTP detected a set of 866 proteins that are shared by *A. thaliana* and cyanobacteria only: 677 proteins are detected in one cyanobacterium but no other genome; 133 proteins were found in two, and additional 56 Arabidopsis proteins in all three cyanobacteria only. These 866 proteins – shared only between Arabidopsis and cyanobacteria among the genomes under study – are most likely to be of cyanobacterial origin.

Within the subset of 9,368 proteins 7,304 protein sequences showed a cyanobacterial homologue at an E value lower than 10^{-4}, and among these in 2,363 cases a cyanobacterial protein was the closest homologue. In an additional 1,265 cases, the cyanobacterial homologue was among the best matches. For those 3,628 $(2{,}363 + 1{,}265)$ Arabidopsis proteins the identified homologues of all species were collected and aligned using CLUSTALW. In addition, gapped positions were removed and the alignments subsequently used for a phylogenetic inferrence analysis using neighbor-joining and the PROTML software for ML calculations. The evaluation of the phylogenetic trees yielded an additional 834 proteins – to the 866 identified by the BLAST analysis – that branch specifically with cyanobacterial homologues in the phylogenetic analysis: 513 Arabidopsis proteins shared a common branch with one cyanobacterium, 179 branched basal to two cyanobacteria, and 142 branched basal to all three cyanobacteria sampled. Fig. 3.5 depicts an exemplary tree resulting from this analysis which clearly shows the Arabidopsis protein branching with the cyanobacteria.

In consequence, based on their similarity these 1,700 $(834 + 866$, or 18% of the 9,368 that could be used for phylogenetic inferrence) Arabidopsis proteins are encoded by genes that were transferred from plastids to the nucleus. Additional 354 ambiguous Arabidopsis proteins were not considered: for 300 of these BLASTP detected either only two homologues (one from cyanobacteria and one from another genome), or, as in the remaining 54 cases, detected only three homologues, two from cyanobacteria and one from another genome. Obviously, all of those combinations would result in a branch between Arabidopsis and cyanobacteria and are

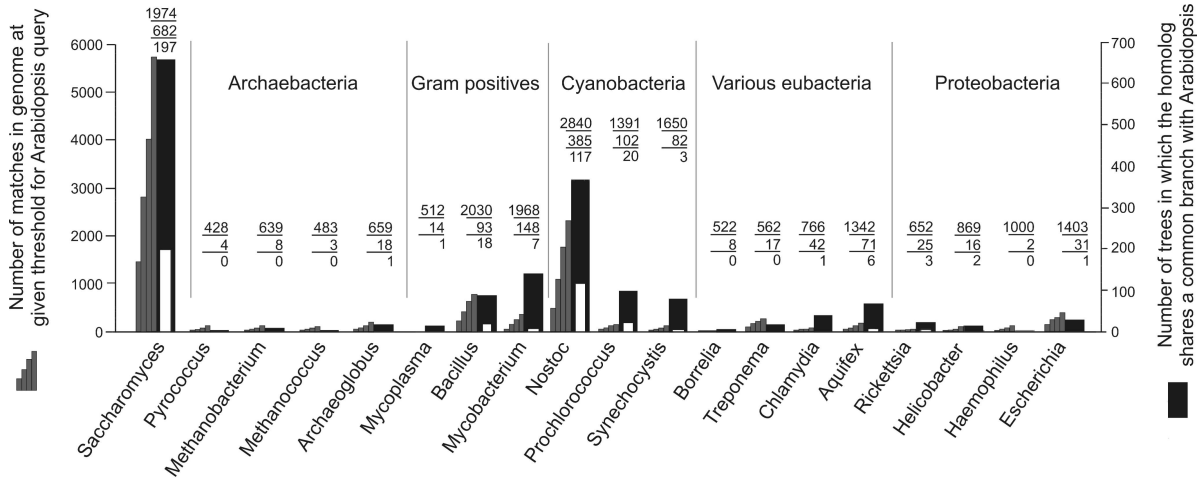

Fig. 3.4.: Comparison of Arabidopsis to 20 reference proteomes. 24,990 Arabidopsis proteins were compared to 51,361 proteins from 20 reference species (two Mycoplasma genome sequences are treated as one species) by both blast and phylogenetic analysis. Gray columns indicate the number of times that the genome gave the best match against Arabidopsis when blast was used at the four *E* value thresholds: 10^{-40}, 10^{-20}, 10^{-10}, and 10^{-4} (from left to right). The number of times that a homologue from the genome occurred in any tree is indicated (top number above columns). Black columns refer to the number of times that proteins from the genome indicated gave a common branch with the Arabidopsis homologue in protml analysis using the JTT-F matrix (middle number). White columns indicate the number of trees in which the branch was supported at a bootstrap probability ≥ 0.95 (bottom number)

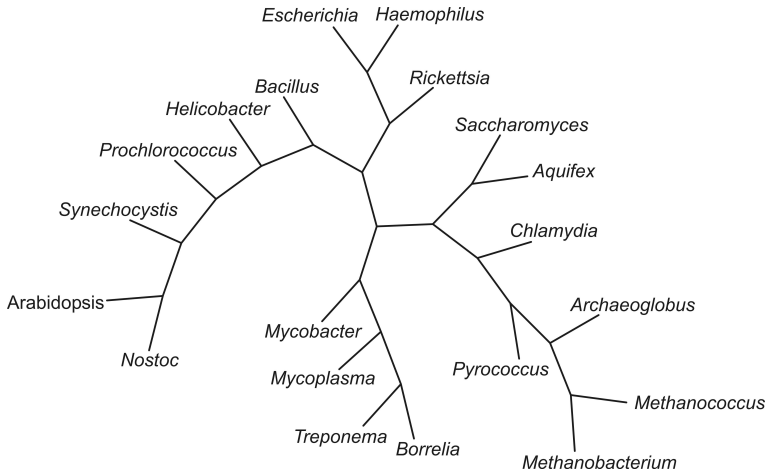

Fig. 3.5.: Phylogenetic tree by ML inferrence. Exemplary unrooted tree of an Arabidopsis translation elongation factor (GI: 7268831; At4g20360) and homologues found in all 20 reference species. The homologues were aligned using clustalw and cleared of gaps, resulting in a 364-site dataset which was used for ML calculations by protml. For the visualization by drawtree the display of branch lengths was turned off.

also most likely to be of cyanobacterial origin. Nonetheless, they were treated as false negatives and not considered further. In contrast, also false positives were detected as Arabidopsis on average branched 32 times with each non-cyanobacterial prokaryote sampled (Fig. 3.4, sum of middle numbers divided by the prokaryote proteomes under analysis). Thus, if one allows the 32 false negatives to be outweighed by the 354 false positives, still about 1,700 (or 18%) of the 9,368 investigated proteins stem from cyanobacteria and allow an extrapolation for the whole genome of Arabidopsis. In this way, about 4,800 ($\frac{26{,}620}{9{,}368} \times 1{,}700$) proteins in *Arabidopsis thaliana* should be descendants of cyanobacteria.

3.2.2. Analysis of protein compartmentalization and functional categories

To estimate how many of the nuclear genes donated by the chloroplast are targeted back to the organelle, the 3,628 Arabidopsis proteins were subjected to a TARGETP analysis (Fig. 3.6). The targeting predictions were evaluated for the five different reliability classes that TARGETP uses to indicate how probable the given (best) prediction is compared to the second best. The analysis revealed that more than half of the cyanobacterial proteins are not targeted to the plastid, whereas many non-cyanobacterial proteins are. In addition, many proteins of cyanobacterial origin appear to enter the secretory pathway. These findings strongly support the idea of re-routing of former chloroplast proteins (altered targeting).

The functional classification of the 1,700 proteins of cyanobacterial origin was performed according to the categories given by The Arabidopsis Genome Initiative (2000) and is displayed in Tab. 3.4: Most proteins analyzed showed to be involved in metabolism (562), signal transduction (189) and cellular response (137). Furthermore, a great number of proteins could not, or not clearly, be categorized (342) because of a lack of functional annotation.

3.2.3. Evaluation and combination of cTP predictors

The identification of the chloroplast proteomes of rice and Arabidopsis was carried out according to the prediction of the mitochondrial proteome (see section 3.1). The four predictors TARGETP (Emanuelsson *et al.*, 2000), iPSORT (Bannai *et al.*, 2002), PCLR (Schein *et al.*, 2001), and PREDOTAR (Small, 1999) were employed and tested on a test set of 2,450 proteins with experimentally confirmed localization. The test set used was assembled to reflect the actual

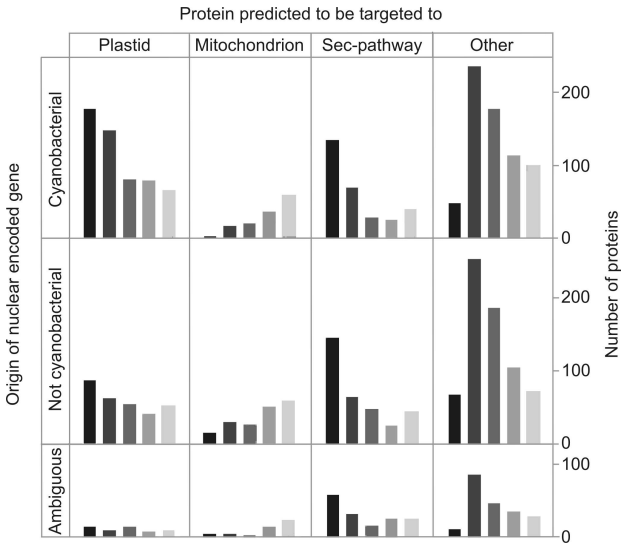

Fig. 3.6.: Targeting predictions for 3,628 Arabidopsis proteins examined. targetp predictions were conducted for the 3,628 Arabidopsis proteins under study. Columns indicate the number of genes predicted to be targeted to the compartment designated in the top row. The shading reflects the five reliability classes of the targetp prediction. Dark bars (left) indicate the most reliable class, light bars (right) indicate the least reliable class.

Functional category	Number of proteins
Biosynthesis and metabolism	562
Signal transduction	189
Cellular response	137
Energy generation	93
Cell organization	71
Protein synthesis	68
Protein destination	63
Cell growth and division	31
Transcription	54
Biogenesis	38
Transport facilitators	35
Intracellular transport	12
Homeostasis	5
Unclassified	303
Classification not clear-cut	39

Tab. 3.4.: Functional classification of Arabidopsis proteins of cyanobacterial origin. The functional categories according to The Arabidopsis Genome Initiative (2000) and the number of proteins per category among the 1,700 gene products originating from cyanobacteria are given.

Predictor	Sensitivity	Specificity	MCC
iPSORT	0.62 (0.86)	0.42 (0.71)	0.45 (0.64)
PCLR	0.83 (0.83)	0.32 (0.30)	0.49
PREDOTAR	0.67 (0.82)	0.45 (0.77)	0.43
TARGETP	0.83 (0.85)	0.52 (0.69)	0.61 (0.72)

Tab. 3.5.: Accuracy of the cTP predictors. Predictors were tested on a test set of 2,450 proteins with experimentally confirmed localization. Values in parentheses are taken from the literature: ipsort (Bannai *et al.*, 2002), pclr (Schein *et al.*, 2001), predotar (Emanuelsson & von Heijne, 2001), targetp (Emanuelsson *et al.*, 2000). MCC, Matthews correlation coefficient (Matthews, 1975). The sensitivity value refers to the probability/frequency with which any real targeting signal is identified, while the specificity value indicates how many of the predicted targeting signals are real.

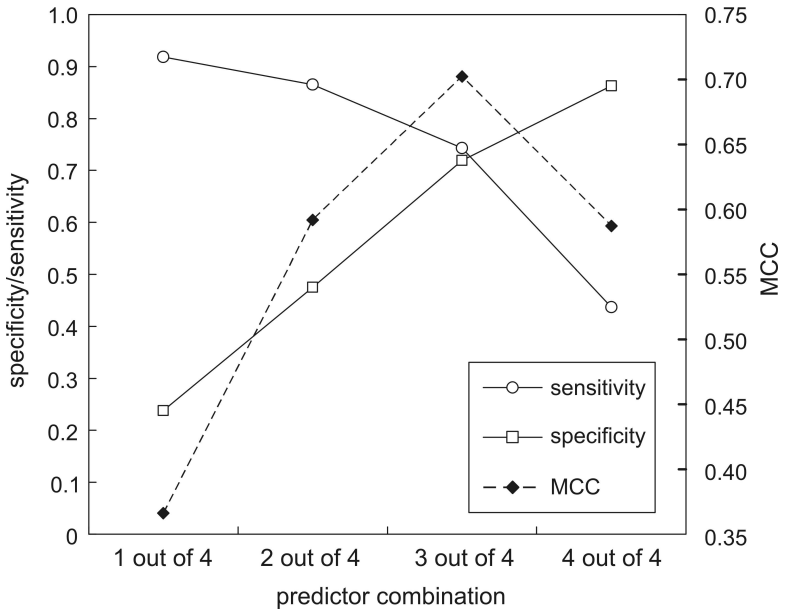

Fig. 3.7.: Evaluation of the combinatorial prediction of cTPs. Sensitivity, specificity and the Matthews correlation coefficient were calculated for the four employed combinations of predictors used on the 2,450 protein test set.

protein composition of an angiosperm cell (The Arabidopsis Genome Initiative, 2000), i.e. one tenth (245) of the proteins was proven to be localized in the chloroplasts, another tenth was proven to be targeted to the mitochondria and the remaining 1,960 proteins were verified not to be targeted to either of the two types of organelles.

Sensitivity and specificity of cTP prediction were assessed for all four predictors (Tab. 3.5). Sensitivity ranged from 0.62 (iPSORT) to 0.83 (PCLR/TARGETP), while the specificity was computed ranging from 0.32 (PCLR) to 0.52 (TARGETP). Since the accuracy of cTP predictors determined here is substantially lower than originally reported (Small, 1999; Emanuelsson *et al.*, 2000; Schein *et al.*, 2001; Bannai *et al.*, 2002) combinations of the predictors were

applied to either increase sensitivity and/or specificity (Fig. 3.7). As with the prediction of mTPs, the specificity of combinatorial cTP prediction increased with the number of predictors used. This was also accompanied by a decrease in sensitivity.

In terms of protein composition, the test set used in this analysis was superior to the one which was employed in section 3.1. Using this extended test set, also sensitivity and specificity of mTP prediction for TARGETP were calculated. Compared to the cTP prediction, both values showed to be even more unsatisfactory than previously reported: 0.62 and 0.30 versus 0.82 and 0.90 (Emanuelsson *et al.*, 2000) for sensitivity and specificity, respectively. These values were used to correct the number of predicted Arabidopsis proteins that feature an mTP (Eq. 2.4) resulting in only about 1,400 proteins which are targeted to the mitochondrion.

The Matthews correlation coefficient (Matthews, 1975, Eq. 2.5, p. 28) was used as a measure for the overall prediction accuracy and revealed that the '3 out of 4' approach – i.e. at least three out of the four employed predictors had to detect a cTP within a given amino acid sequence – was superior to all other combinations and single predictors (Fig. 3.7). The specificity (see Eq. 2.3, p. 28) of the '3 out of 4' prediction was markedly increased by 38% (0.72) compared to TARGETP which was the most specific of the four individual predictors. The sensitivity of 0.74 (see Eq. 2.2, p. 27) decreased only by 11% compared to the most sensitive predictors TARGETP and PCLR. Therefore, the '3 out of 4' approach was used in all further analyses.

3.2.4. Prediction of the size of the chloroplast proteomes of Arabidopsis and Oryza

The '3 out of 4' approach was employed to determine the number of nuclear encoded proteins in *Arabidopsis thaliana* and *Oryza sativa* that feature a cTP (Tab. 3.6). For Arabidopsis a set of 2,090 cTPs (or 7.9% of the total proteome) was found, while in rice 4,853 cTPs could be detected (7.5% of the total proteome). Furthermore, a correction for false positive and false negative predictions (see Eq. 2.4, p. 28) according to Leister (2003) was carried out, based on sensitivity and specificity of the '3 out of 4' approach. As a result of this correction the numbers of chloroplast-targeted nuclear proteins in Arabidopsis and rice are reduced to 2,023 (7.6%) and 4,698 (7.3%), respectively. This finding suggests, that the chloroplast proteome size linearly increases with the size of the total proteome ($\frac{64{,}582}{26{,}535} \approx \frac{4{,}698}{2{,}023} \approx 2.4$).

Species	ORFs in nuclear genome	cTP identified by '3 out of 4' approach	cTPs after accuracy correction
A. thaliana	26,535	2,090 (7.9%)	2,023 (7.6%)
O. sativa	64,582	4,853 (7.5%)	4,698 (7.3%)

Tab. 3.6.: Number of predicted chloroplast proteins. The '3 out of 4' combination of predictors was used to identify those proteins within the Arabidopsis and rice proteomes that are targeted to the chloroplast. In addition, the resulting numbers were corrected for false positives and false negatives according to Leister (2003). Also see Eq. 2.4, p. 28).

3.2.5. Comparative BLAST analyses of the predicted chloroplast proteomes

In analogy to the BLAST analyses of the mitochondrial proteome (section 3.1.2, p. 35) the prokaryotic heritage of the chloroplast proteome was also surveyed. At an E value threshold of 10^{-10} all Arabidopsis and rice protein sequences with a predicted or known cTP were compared to a pool of cyanobacterial sequences assembled from the fully sequenced genomes of *Nostoc* sp. PCC 7120, *Synechocystis* sp. PCC 6803 and *Thermosynechococcus elongatus* BP-1. In Arabidopsis 38.6% (807) of the known or predicted chloroplast protein homologues were detected in cyanobacteria, in rice only 16.8% (817) of homologous proteins could be found (Tab. 3.7). Comparative analysis by BLASTP showed that 434 (54%) out of the predicted 807 chloroplast proteins of Arabidopsis which have a cyanobacterial counterpart also possess a homologue with a predicted cTP in rice.

Organism	Number of proteins predicted or experimentally shown to be chloroplast-located			
	Total	With cyanobacterial homologue	Shared with the chloroplast proteome of the other species	Shared with the total proteome of the other species
A. thaliana	2,134	807 (38.6%)	844 (39.6%)	1,876 (87.9%)
O. sativa	4,854	817 (16.8%)	1,004 (20.7%)	2,683 (53.3%)

Tab. 3.7.: Size and evolution of chloroplast proteomes. Column 2 considers the chloroplast proteins experimentally confirmed and present in the 2,450 protein test set and columns 3 of Tab. 3.6. To identify cyanobacterial homologues of chloroplast proteins, amino acid sequences were compared to a pool of protein sequences from three cyanobacteria with fully sequenced genomes (*Nostoc* sp. PCC 7120, *Synechocystis* sp. PCC 6803 and *Thermosynechococcus elongatus* BP-1; column 3). Diversification of the chloroplast proteomes was analyzed by a comparison of the chloroplast proteins against chloroplast and total proteome of the second species (columns 4 and 5).

In section 3.2.1, a more accurate method of identifying proteins in Arabidopsis that originated from the cyanobacterial endosymbiont has been provided, and a subset of 1,700 proteins was detected. Among these 1,700 cyanobacterium-derived Arabidopsis proteins 385 (23%)

feature an actual or predicted cTP. A subset of 228 (59%) of these also share a rice homologue featuring a cTP. This indicates that a core set of at least 228 chloroplast proteins can be traced back to the cyanobacterial endosymbiont and has persisted in the two species *Arabidopsis thaliana* and *Oryza sativa*. A functional analysis of these 228 evolutionarily 'ancient' proteins revealed that they are mostly related to metabolism, energy and protein fate, while the majority of 37% is still unclassified (Fig. 3.8). In comparison to the functional distribution in the whole Arabidopsis proteome, the set of 228 proteins clearly shows an expected over-representation of functions in energy, metabolism and protein synthesis, while proteins involved in transcription or signal transduction are less frequent.

Since the 1,700 identified cyanobacterial proteins were only based on a subset of 9,368 proteins, the number of cyanobacterial proteins in the two flowering plant species featuring a cTP has to be extrapolated to the whole genome. This is performed by multiplying 228 by the fraction of the total ORFs in the nuclear genome (26,535) and the number of proteins used in the phylogenetic analysis (9,368) – equaling a factor of 2.83 – thus extrapolating from 228 to 645.

Correction of this number for the accuracy of the '3 out of 4' prediction (Eq. 2.4) yields an estimated total of 628 actual cTP-featuring proteins of cyanobacterial origin. The same calculations were applied to the set of the 385 of 1,700 cyanobacterial Arabidopsis proteins with a cTP, resulting in a number of 1,061. This equals 52% of the 2,023 proteins (Tab. 3.6) predicted or known to be targeted to the chloroplast.

The diversification of the two chloroplast proteomes was investigated by a BLASTP analysis comparing the proteomes of Oryza and Arabidopsis. The relative size of the species-specific fraction was found to be 60% in Arabidopsis and 79% in rice (Tab. 3.7, column 4). This indicates that the degree of chloroplast specialization differs between the two species. Paralleling the findings for mitochondria, also non-chloroplast homologues of species-specific chloroplast proteins were found by this interspecific analysis (Tab. 3.7, difference between values in columns 4 and 5). These proteins represent a substantial fraction of the chloroplast proteome of the two species (48% and 35% for Arabidopsis and rice, respectively), indicating that chloroplast proteome diversity is generated by both gene evolution leading to novel proteins located in the chloroplast, and by relocation of conserved gene products due to altered targeting.

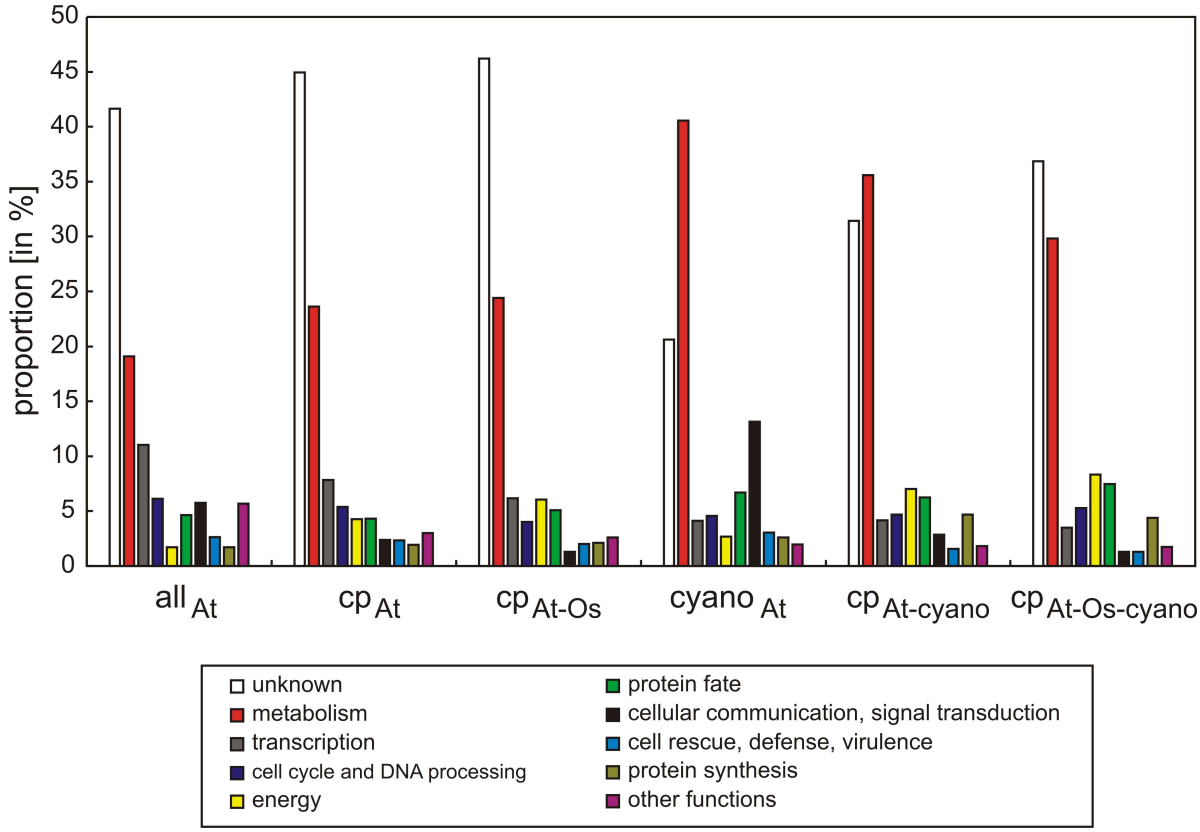

Fig. 3.8.: Comparative functional classification in Arabidopsis. The distribution of different functional categories of several sets of chloroplast proteins of *A. thaliana* is depicted and compared to the proportion of all predicted Arabidopsis proteins ('all_{At}', shown on the left). 'cp_{At}' refers to the set of 2,134 Arabidopsis proteins with a predicted or actual cTP (Tab. 3.7); '$cp_{At\text{-}Os}$' indicates those 844 Arabidopsis proteins of 'cp_{At}' which possess a rice homologue with a cTP; '$cyano_{At}$' refers to the set of 1,700 cyanobacteria-derived Arabidopsis proteins identified in section 3.2.1; '$cp_{At\text{-}cyano}$' indicates the 385 proteins of '$cyano_{At}$' that have a cTP; and '$cp_{At\text{-}Os\text{-}cyano}$' includes the 228 proteins of '$cp_{At\text{-}Os}$' which were shown to be derived from cyanobacteria in section 3.2.1

3.3. Analysis of the nuclear chloroplast transcriptome

In this section the creation and subsequent use of a chloroplast GST array for 3,292 genes is described. A broadly based transcriptome survey of 35 different environmental and genetic conditions was carried out (hybridizations and most of the ARRAYVISION and ARRAYSTAT evaluations were performed by A. Biehl, A. Dietzmann and J. Kurth), and data of the expression profiles were subjected to a clustering procedure, resulting in the creation of a transcriptome response map. Only three main classes of response from the nuclear chloroplast transcriptome could be detected, suggesting a regulatory master-switch.

3.3.1. A chloroplast GST array for 3,292 genes

Based on the set of 1,827 GSTs described by Kurth *et al.* (2002) an additional 1,465 GSTs from genes, whose gene products feature a cTP predicted by a TARGETP analysis, were automatically designed as described by Varotto *et al.* (2001). The resulting set comprises only of GSTs deduced from proteins for which a high probability of chloroplast targeting was predicted, i.e. cTPs with low reliability levels of the TARGETP prediction were excluded. Together, both sets include 2,661 GSTs of proteins that are most likely to be targeted to the chloroplast. The remaining 631 GSTs were selected non-chloroplast (e.g. mitochondrial) proteins. To test the possibility of cross-hybridization between the 3,292 GSTs and with the rest of the Arabidopsis genome a BLASTN analysis was performed. Only 14 GSTs contained stretches of more than 100 bp matching another GST. Only 90 GSTs showed a sequence homology to any other Arabidopsis ORF. Thus, possible cross-hybridizations should be of minor concern. The 3,292 GSTs were spotted onto nylon membranes in duplicates as depicted in Fig. 3.9.

For about 75% of the genes corresponding to the GSTs, expressed sequence tags could be found in public databases. To test whether the remaining ORFs are also transcribed, their expression was studied under 35 different conditions given in Tab. 3.8, p. 52. In at least one of these 35 conditions for each of the 3,292 GSTs spotted a signal was detected, reflecting a significant change in expression of the corresponding gene. This indicates that those ORFs are transcribed and most likely annotated correctly and, furthermore, that a constitutive expression of the genes sampled is unusual under the conditions tested. In turn,

Fig. 3.9.: Organization of the 3,292 GST array. Exemplary picture of a hybridization with ^{33}P-labeled oligonucleotides complementary to the tailing universal 15-mer sequence of every GST (see section 2.3.1, p. 29). By this method the quality of spotted GSTs was verified. The array is organized in 48×16 modules with up to 6 different GSTs or control spots each. The geometry of duplicates spotted in the modules is depicted in the lower left corner. By this arrangement each duplicate will exhibit a different orientation and thus be most easily distinguishable from any other duplicate within a module.

also the validity of the GST approach is demonstrated as the spotted gene sequence tags detect expressed sequences.

3.3.2. Responses of the nuclear chloroplast transcriptome under 35 conditions

Using the GST array, the chloroplast transcriptome of *Arabidopsis thaliana* under 35 different environmental and genetic conditions given in Tab. 3.8 was analyzed. The results revealed a substantial variation in the levels of transcriptional response: in photosynthetic tissue, the significant differential expression levels of the 3,292 genes ranged from 15 to 76% (Fig. 3.10). With respect to the ratio of up-regulated to down-regulated genes, three main types of transcriptome response were identified: two types of response were predominantly associated with either up-regulation or with down-regulation. A third mixed type of response involved approximately equal numbers of up-regulated and down-regulated genes. Most of the 35 conditions analyzed resulted mainly either in the first two types of response (up- or down-regulation), whereas only eight led to a mixed response (Fig. 3.10). Interestingly, the regulation of the 631 non-chloroplast-protein genes seems to be linked with that of the chloroplast-protein genes (gray shadings within columns in Fig. 3.10 are oriented in the same direction as columns), suggesting that the regulation of chloroplast function is integrated into larger regulatory networks.

Abbreviation	Genotype or treatment	Primarily affected chloroplast function
T r e a t m e n t s		
PSII/I	Growth under PSII-specific light *vs* PSI-specific light[a]	Photosynthesis, plastid redox state
L_30'	30 min light *vs* darkness	Various functions
HL_15'	15 min highlight stress[b] *vs* normal light[c]	Photosynthesis and others
HL_1h	1 h highlight stress[b] *vs* normal light[c]	Photosynthesis and others
HLrec_2h	2 h recovery[c] after 1 h highlight stress[b] *vs* before stress[c]	Various functions
HLrec_48h	48 h recovery[c] after 1 h highlight stress[b] *vs* before stress[c]	Various functions
ML_4C	24 h medium light[d] at 4 °C *vs* 20 °C	Photoinhibition of photosystems
+Par	Treatment with benzoquinone herbicide paraquat *vs* no treatment	Reactive oxygen species, photosynthesis, oxidative stress
+Bro	Treatment with nitrile herbicide bromoxynil *vs* untreated	Photosynthesis and others
–CO_2_2d	Low-CO_2 stress: 2 d of 0.003% (v/v) CO_2 *vs* normal CO_2 level	Calvin cycle, respiration
–CO_2_4d	Low-CO_2 stress: 4 d of 0.003% (v/v) CO_2 *vs* normal CO_2 level	Calvin cycle, respiration
+CO_2_1d	High-CO_2 stress: 1 d of 1% (v/v) CO_2 *vs* normal CO_2 levels	Calvin cycle, respiration
+CO_2_10d	High-CO_2 stress: 10 d of 1% (v/v) CO_2 *vs* normal CO_2 levels	Calvin cycle, respiration
+Fe	High-iron stress (spraying with iron solution[e]) *vs* no treatment	Oxidative stress
+Pro	48 h 100 mM proline *vs* no treatment[f]	Increased levels of cellular proline
+CK	Cytokinin-treated (2 h) cell culture *vs* untreated cell culture[g]	Various functions
M u t a n t s		
prpl11	*prpl11-1* (Pesaresi *et al.*, 2001a) *vs* WT	Lacks plastid ribosomal protein PRPL11; plastid protein synthesis
psad	*psad1-1*[h] *vs* WT	Lacks PS I protein D1; photosynthesis
psae	*psae1-1* (Varotto *et al.*, 2000) *vs* WT	Lacks PS I protein E1; photosynthesis
psan	*psan-1*[h] *vs* WT	Lacks PS I protein N; photosynthesis
psao	*psao-1*[h] *vs* WT	Lacks PS I protein O; photosynthesis
atpc	*atpc1-1*[i] *vs* WT	Lacks ATP synthase subunit; photophosphorylation
atpd	*atpd-1*[h] *vs* WT	Lacks ATP synthase subunit; photophosphorylation
hcf145	*hcf145*[i] *vs* WT	Assembly of PS I perturbed; chloroplast biogenesis
gun1	*gun1-1* (Susek *et al.*, 1993) *vs* WT	Unknown, plastid signaling
gun5	*gun5* (Susek *et al.*, 1993; Mochizuki *et al.*, 2001) *vs* WT	ChlH subunit of Mg-chelatase; plastid signaling
cue1	*cue1-1* (Streatfield *et al.*, 1999) *vs* WT	Lacks phosphoenol pyruvate translocator, metabolite exchange, plastid signaling
flu_D	*flu* (Meskauskiene *et al.*, 2001) (dark) *vs* WT (dark)	Lacks FLU, a regulator of chlorophyll biosynthesis
flu_L	*flu* (Meskauskiene *et al.*, 2001) (light) *vs* WT (dark)	Lacks FLU, a regulator of chlorophyll biosynthesis
ppi1	*ppi1* (Jarvis *et al.*, 1998) *vs* WT	Defective in import of chloroplast precursor proteins
kn09	*kno9*[f] *vs* WT	Blocked in mitochondrial proline catabolism
sut2	*sut2*[f] *vs* WT	Lacks plasma-membrane sucrose sensor/transporter: carbohydrate partitioning
mak3	*atmak3-1*[h] *vs* WT	Defective in cytoplasmic N-acetyltransferase: various functions
pam48	*pam48*[h] *vs* WT	Defective in unknown protein: altered photosynthesis
pam46	*pam46*[h] *vs* WT	Defective in unknown protein: altered photosynthesis

Tab. 3.8.: Overview of conditions used for expression profiling. [a] RNA provided by T. Pfannschmidt; [b] 2,000 µmol photons $m^{-2}s^{-1}$; [c] 80 µmol photons $m^{-2}s^{-1}$; [d] 100 µmol photons $m^{-2}s^{-1}$, RNA provided by A. Haldrup; [e] leaves were sprayed with a 0.06% Fe^{2+}-chelate solution; [f] RNA provided by W. Frommer and co-workers; [g] provided by T. Schmülling; [h] unpublished mutant isolated in our laboratory; [i] provided by J. Meurer.

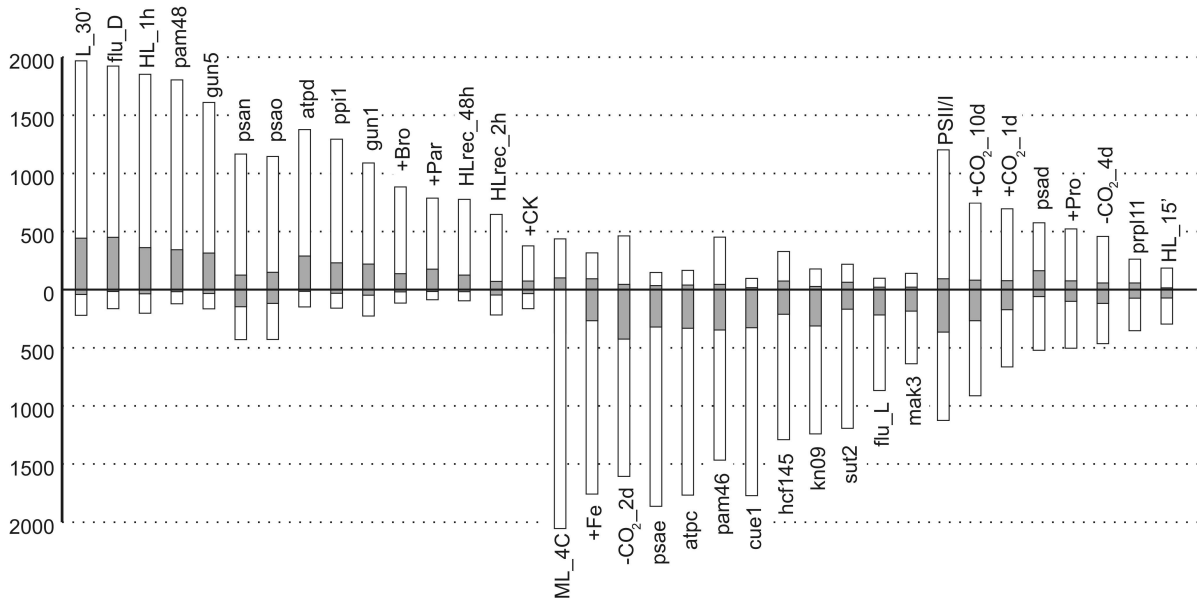

Fig. 3.10.: Effects of 35 environmental or genetic states on the nuclear chloroplast transcriptome. Three groups of differential transcriptome changes, including preferential up-regulation, down-regulation, or a mixed type of gene expression, are shown. Columns represent differential expression of nuclear chloroplast genes, while gray shading that of genes encoding non-chloroplast proteins. For abbreviations of conditions see Tab. 3.8.

3.3.3. Classification of transcriptomes by clustering and 'all-against-all' comparison

For the analysis described above only the total number of genes up- or down-regulated was taken into consideration, i.e. it was not discriminated whether similar or completely different sets of genes were co-regulated under the different conditions tested. Therefore, a hierarchical clustering of the expression profiles was performed on a subset of 1,972 genes (including 1,542 encoding for chloroplast proteins). These 1,972 genes were selected because they respond transcriptionally under at least 33 of the 35 conditions tested. Due to this filtering a set with sufficient data points for a reliable hierarchical clustering was generated. From the result of the hierarchical clustering (Fig. 3.11) it is evident that sets consisting mostly of the same genes were coordinately up-regulated or down-regulated.

Due to their opposing signaling phenotype the two classes of plastid signaling mutants (*gun1* and *gun5*, and *cue1*, respectively) should also exhibit contrasting transcriptome responses and were analyzed in more detail. The hierarchical clustering analysis and the cladogram show

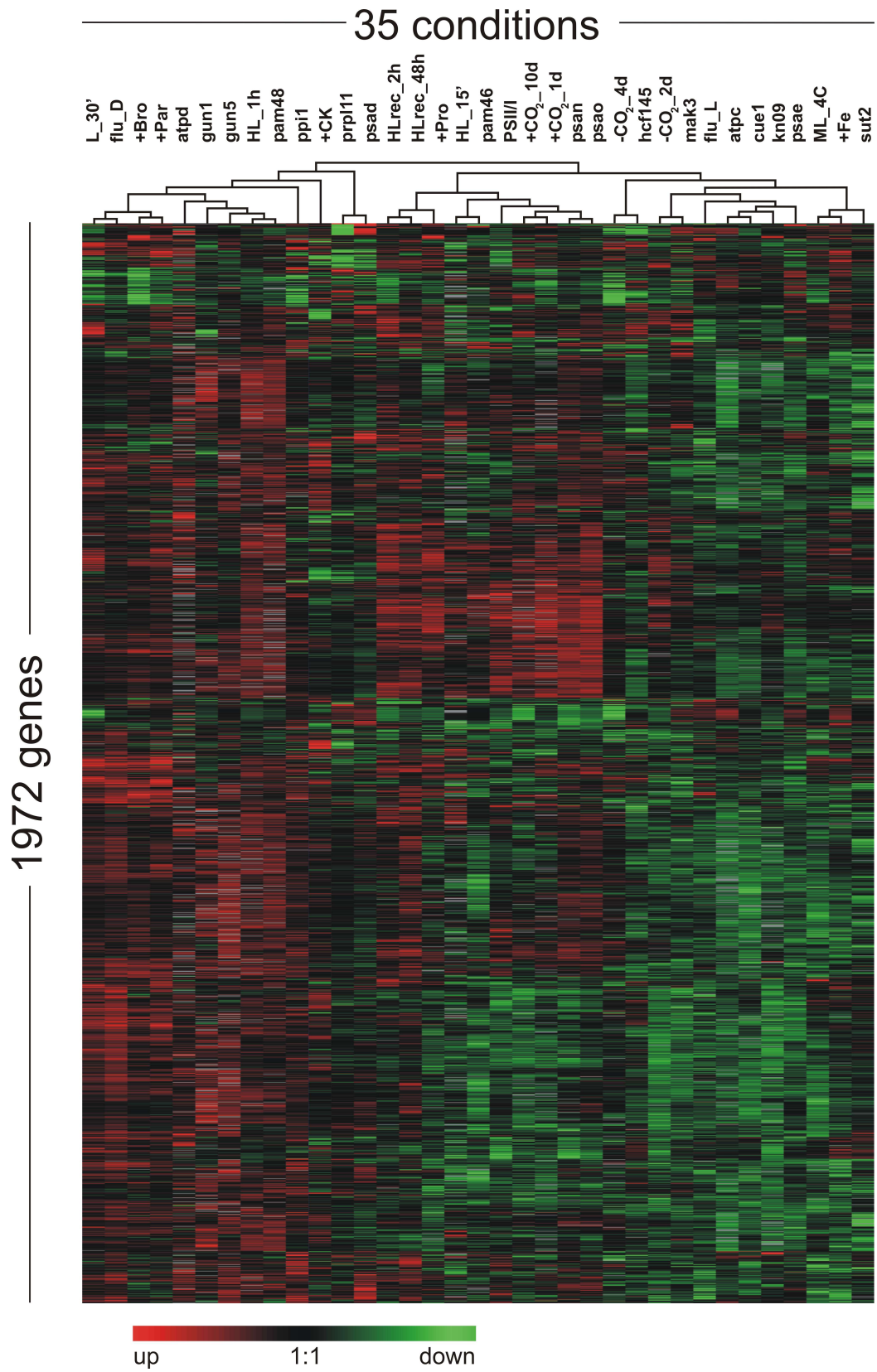

Fig. 3.11.: Hierarchical clustering of transcriptome responses for 35 conditions. Clustering of the differential expression of 1,972 genes from at least 33 out of 35 conditions using Euclidean distances. The cladogram at the top summarizes the relatedness of transcriptome responses. Each gene is represented by a single horizontal line whose colors indicate up-regulation (red) or down-regulation (green). For abbreviations of conditions see Tab. 3.8.

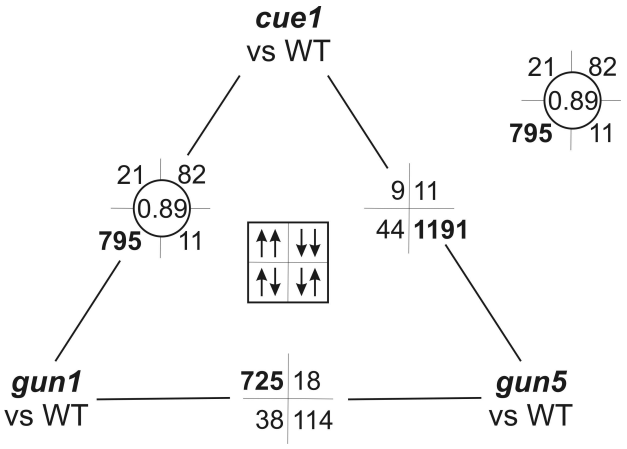

Fig. 3.12.: Comparison of significantly differentially regulated genes. All significantly differentially regulated genes of the three mutants *gun1, gun5* and *cue1* were compared in pairs. The results of this comparison are given as the numbers positioned between the corresponding two mutants, i.e. the same (up/up or down/down) or opposite (up/down or down/up) trends in expression of identical genes are stated. The distance between two mutants was calculated as explained in section 2.3.3. For example, for the two mutants *cue1* and *gun5* a value of: $1 - \frac{9+11}{9+11+44+1,191} = 0.98$ results. For the 'all-against-all' comparison this calculation was repeated for all mutants and conditions, yielding the distance matrix given in appendix B.2.

that this is the case. Genes up-regulated in the *gun1* and *gun5* mutants clearly show a down-regulation in the *cue1* mutant and vice versa. An additional, and more thorough, analysis explained in detail below ('all-against-all') was performed. The result of the comparison is depicted in Fig. 3.12 and again shows that almost all genes up-regulated in *gun1* and *gun5* were down-regulated in *cue1*. These findings are also reflected in the calculated distance values: 0.17 for the *gun1* and *gun5* mutants, and 0.89 and 0.98 for a comparison between *gun1* and *cue1* and *gun5* and *cue1*, respectively.

For a more detailed characterization of the similarities between the 35 transcriptome responses, hierarchical clustering was combined with an analysis comparing many gene–transcription interrelationships. For this purpose, an 'all-against-all' comparison in pairs (see section 2.3.3) of transcriptome responses was performed and an exemplary comparison involving only three mutants instead of all 35 conditions, is depicted in Fig. 3.12. For this 'all-against-all' analysis, a comparison of each possible pair of the 35 conditions was carried out, and every significantly regulated gene present in both conditions was concerned. Genes either up-regulated or down-regulated, or regulated in opposite directions in every pair of conditions were counted and a distance value was calculated. An example of such a calculation is given in Fig. 3.12

and the resulting distance matrix for all 35 conditions is appended in section B.2.

The results of the complete combined 'all-against-all' analysis are depicted in the transcriptome response map in Fig. 3.13. The transcriptome responses for the 35 conditions are arranged in the order resulting from the hierarchical clustering (Fig. 3.11) and the many relationships between the transcriptome responses are indicated by connecting lines. A well organized structure emerged, in which three main types of transcriptome response are apparent. Two of those correspond well to the classes already identified as involving predominantly up-regulation (class 1) or down-regulation (class 3), whereas class 2 is characterized by expression profiles showing some relationship to class 1 as well as to class 3.

When only the 23 environmental or genetic perturbations that resulted in transcriptome responses belonging to classes 1 and 3 were considered, a set of 1,349 genes (including 1,022 encoding chloroplast proteins) was identified, which were expressed in all 23 conditions. In this set 641 genes encode proteins with still unknown function, while among those with known function 116 are involved in transcription or translation, 135 in metabolism, 71 in photosynthesis, 46 in transport, and 53 encode for protein kinases or phosphatases. A hierarchical clustering of the expression profiles for the 1,349 genes showed that almost all genes belonging to class 3 were affected in opposite ways compared to those of class 1: class-1 genes were up-regulated and class-3 genes were down-regulated (Fig. 3.14). Furthermore, detailed analysis of the expression profiles revealed that, within a particular class, additional levels of variation were present – a few cases even showed reversed behavior for a particular gene tested under specific conditions. This implies the existence of additional regulatory mechanisms (induced by class-1 and class-3 conditions) that differentiate and modify the general responses of the nuclear chloroplast transcriptome. Indeed, some class-2 transcriptome responses [e.g. in the *pam46* mutant; or with 1 h of high-light stress, followed by recovery for 48 h (HLrec_48h); Fig. 3.13], might result from such combined mechanisms.

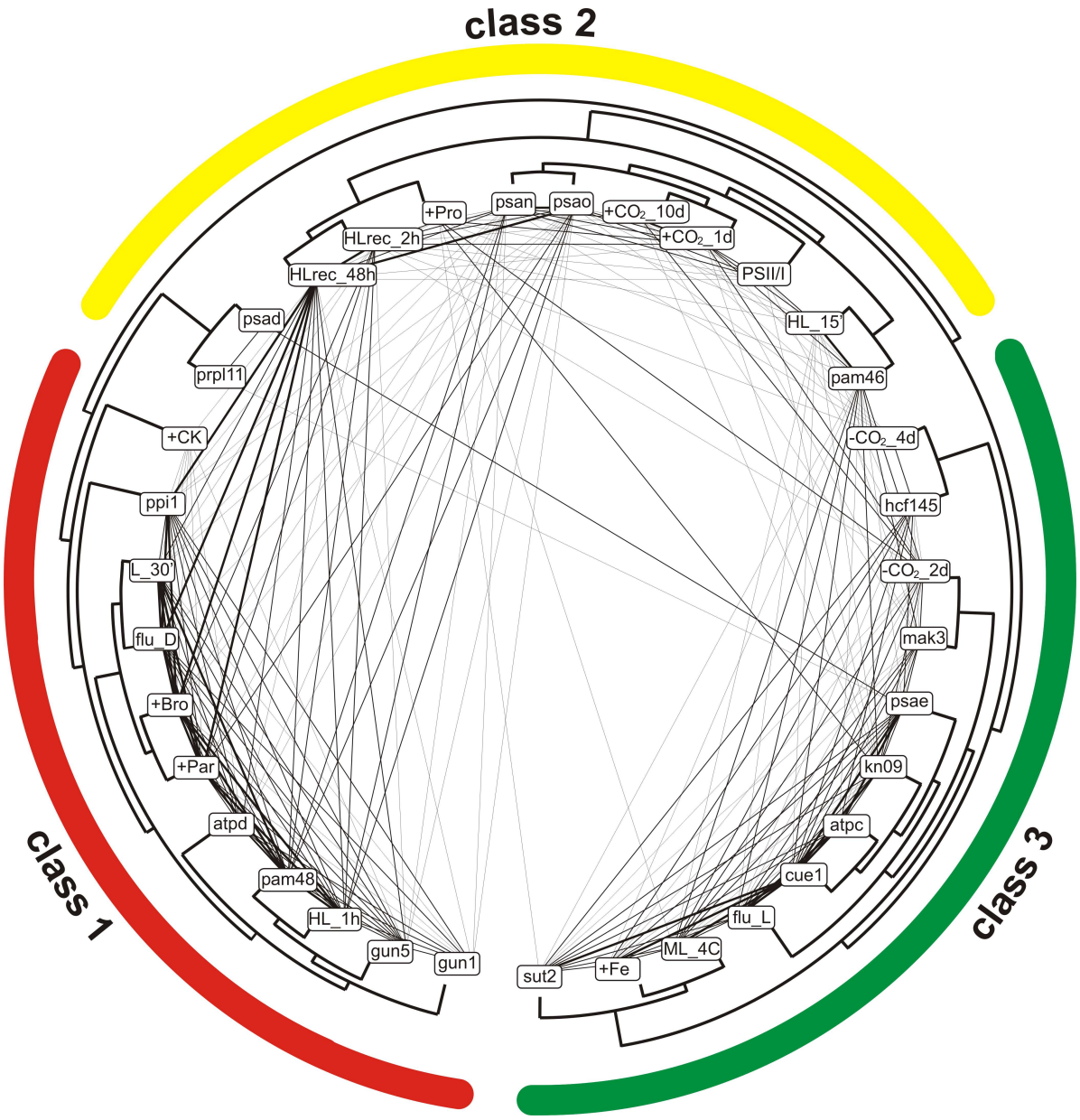

Fig. 3.13.: Transcriptome response map of 35 different conditions. The topographical arrangement of environmental and genetic conditions is derived from Fig. 3.11, and the cladogram in the same figure is shown at the periphery. Connecting lines inside the circle reflect the relationships between the expression profiles – the thickness of a line being a measure of similarity (hairlines: 65–80%, medium-thick lines: 80–90%, and thick lines: 90–100% of genes showing the same trend in expression in each pair of transcriptomes). In some cases, cladogram branches were rotated, without changing overall topography, to minimize the length of the connecting lines. Main classes of transcriptome change are indicated in red (class 1), yellow (class 2) and green (class 3).

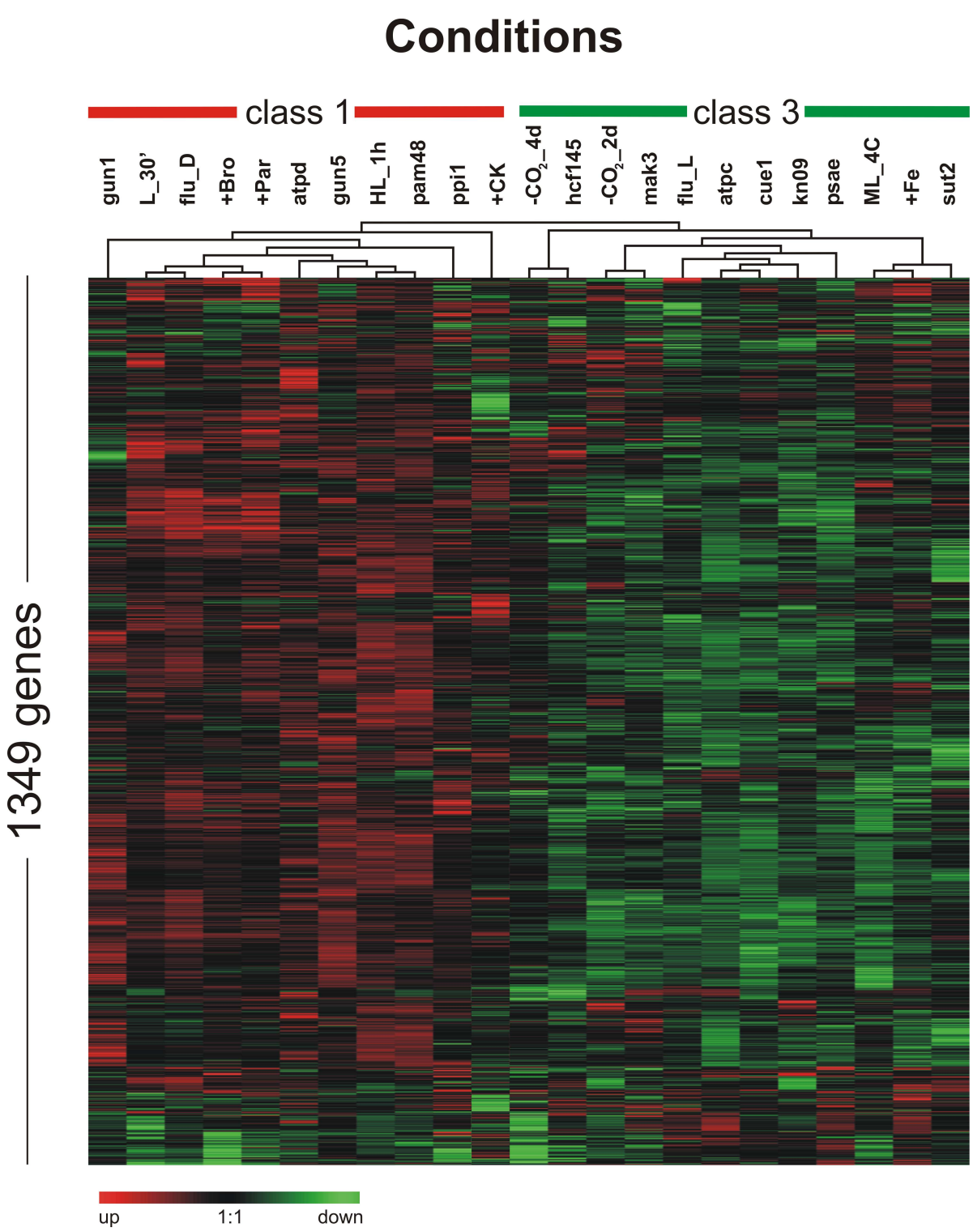

Fig. 3.14.: Class-1-type and class-3-type responses of the nuclear chloroplast transcriptome. Both types of responses generally involve the same set of genes. 1,349 genes differently expressed in all of the 23 conditions that elicited up-regulation (class-1 response) or down-regulation (class-3 response), were clustered hierarchically. Red and green colors correspond to up-regulation and down-regulation, respectively. Each gene is represented by a horizontal line, which, in most cases, shifts its color from red (left; class-1 condition) to green (right; class-3 condition).

4. Discussion

4.1. Comparative structural genomics of mitochondria: size, origin and composition of mitochondrial proteomes

4.1.1. An improved approach for the prediction of mitochondrial transit peptides

To determine the size and composition of the mitochondrial proteomes, with both high specificity and sensitivity, is currently a demanding task. Here, a novel approach for the identification of N-terminal mitochondrial targeting sequences has been established by combining the three predictors TARGETP, PSORTII and iPSORT. When used on a test set containing 906 proteins with experimentally confirmed localization this new method proved to be superior to any of the employed predictors alone. A '2 out of 3' combination – i.e. at least two out of the three employed predictors identified an mTP within a given amino acid sequence – resulting in an increased specificity at about the same sensitivity in comparison to TARGETP or iPSORT. As a result less false positives were called, leading to an optimized overall performance of the prediction. This improvement can best be explained by the fact that (a) the predictors employed are based on dissimilar algorithms, each having certain advantages and drawbacks, and (b) the protein sets used to train those predictors were assembled sub-optimally.

TARGETP relies on the use of a neural network architecture to detect the cleavage site of a transit peptide. If such a site is found, the transit peptide is subsequently categorized by an additional layer of neurons. Neural networks are self-teaching networks of computational units that learn by a process of trial and error and are known to yield the best prediction results. Although neural networks are often successful in practice because they have the advantage that a strict input model is not required, it is very difficult to determine exactly what a certain

decision of the network was based on. Thus improving and optimizing such a system can be a perplexing task. PSORTII – an expert system, i.e. a type of artificial intelligence – does not only investigate the N-terminus of a given protein sequence. Instead, it makes use of a machine-learning technique and employs a basic but robust pattern-recognition algorithm [k-nearest-neighbor method (Gaasterland *et al.*, 1997)] to compare the complete sequence under analysis to a training set. However, the resulting performance of PSORTII is unsatisfactory compared to that of TARGETP. iPSORT, a successor to PSORTII, makes use of a set of derived and interpretable rules (e.g. biochemical features of the amino acid residues of N-terminal sorting signals). The overall idea of this approach is the accumulation of significant amounts of very simple attributes and their subsequent combination yielding in a prediction.

One major general disadvantage of all presented approaches is that – in the case of mTP prediction – the knowledge about unambiguous features, for example, conserved, clear-cut motifs around the cleavage site of the transit peptide are lacking, since it is unlikely that such distinct features exist for mTPs (Gavel & von Heijne, 1990a,b; Emanuelsson *et al.*, 1999). Only very basic facts on mTP composition (e.g. often arginine, alanine and serine are incorporated, while negatively charged amino acid residues like aspartic acid or glutamic acid are rare) are available (Emanuelsson & von Heijne, 2001). In consequence, all presented predictors very heavily depend on the composition and quality of the training set. In this respect, all three predictors rely on, or have been tested on data that have only partially been confirmed experimentally (Nakai & Horton, 1999; Emanuelsson *et al.*, 2000; Bannai *et al.*, 2002): to increase the number of available sequences – which is a critical factor for the training of neural networks – also sequences annotated as 'potential', 'probable' or those only based on homology to known mTPs have been incorporated. This causes the predictions to be prone to error and is reflected in the still unsatisfactory results of the analysis of the 906 proteins with experimentally proven localization used in this work.

Some other general difficulties in mTP prediction arise from the fact that some matrix proteins are cleaved a second time by the mitochondrial intermediate peptidase removing an additional eight to nine residues from the protein (Kalousek *et al.*, 1988; Isaya & Kalousek, 1994). Thus, training of predictors is hampered. Further, some mitochondrial proteins are first imported into the matrix where the mTP is cleaved off, and in a second step re-targeted back to

the intermembrane space by a second signal that shares similarities with the targeting sequence for the signaling pathway (Gasser *et al.*, 1982; van Loon & Schatz, 1987). Finally, information on targeting can also be encoded internally or C-terminal of the protein, and furthermore it could be shown that few proteins can directly be inserted into the outer membrane or intermembrane space without passing the matrix first (Diekert *et al.*, 1999; Lee *et al.*, 1999).

This study shows that a combination (the '2 out of 3' approach) integrating the benefits of three differently working predictors outweighs the combination of their disadvantages and yields more reliable results. Still, the *in silico* identification of mTPs is far from perfect and much less trustworthy than the prediction of targeting of proteins to the chloroplast or the secretory pathway (Millar *et al.*, 2001). Thus, ever new approaches are needed. One promising proposal could be the combination of a neural network with a support vector machine (SVM) algorithm that has already proven to be a reliable tool for the prediction of subcellular localizations (Hua & Sun, 2001). An additional promising approach could be the consideration of structural information such as secondary structure of the targeting peptides, but this has not yet been successful for mitochondria nor other subcellular compartments that require a transit peptide for their import. This is mostly due to the lack of sufficient and reliable structural information on transit peptides. Another prerequisite for improved predictions are, undoubtedly, larger training sets containing both, true positive and true negative protein sequences which have been confirmed experimentally. Owing to 2D-gel-based proteomics, proteome-scale epitope tagging and immunolocalization, the number of reliably identified mitochondrial proteins is increasing. However, only for Arabidopsis, the human species and baker's yeast more than 100 experimentally identified mitochondrial proteins are known and it could be potentially difficult to identify low-abundance mitochondrial proteins by protein-gel based approaches.

4.1.2. Size and α-proteobacterial heritage of mitochondrial proteomes

The novel approach for the prediction of mTPs described in this work was used to estimate the sizes of the mitochondrial proteomes for ten different species. For further improvement of the approximation, sequences of proteins experimentally localized in the mitochondrion were used for complementation. To determine the fraction of the proteins that trace back to their endosymbiotic origin these organelle proteomes were compared to a pool of α-proteobacterial

proteins as the precise lineage of the predecessor of the mitochondria is unknown.

The approximation of the mitochondrial proteome sizes showed that – in the case of non-parasite species – functional mitochondria harbor from a few hundred to about 4,000 proteins (Tab. 3.2, p. 35 and Fig. 3.1, p. 33). Furthermore, this study also includes the microsporidian *E. cuniculi* (a protist once believed never to have possessed mitochondria), for which a number of 156 proteins featuring an mTP was predicted. These findings are of exceptional interest to cell biologists and evolutionary biologists alike.

The number of predicted mitochondrial proteins increased with the total number of nuclear genes (see Tab. 3.1, p.32). This is either an indication of a substantial systematic artifact in mTP prediction or, more likely, implies that an expanding genome complexity is accompanied by a corresponding increase in complexity of the proteome of mitochondria and other organelles or compartments.

The BLAST analyses of the identified mitochondrial proteomes against a pool of α-proteobacterial protein sequences (Tab. 3.2) revealed that between 6% and 33% (Fugu and budding yeast, respectively) of the proteins are of prokaryotic descent. This is in conflict with former studies stating that 50–60% of the mitochondrial proteins of *S. cerevisiae* still resemble those of the former endosymbiont (Karlberg *et al.*, 2000; Marcotte *et al.*, 2000). Interestingly, the relationship of the mitochondrial proteomes to the endosymbiotic ancestor decreases over evolutionary time, i.e. unicellular eukaryotes show the largest, while higher eukaryotes feature the lowest fraction of α-proteobacterial homologues. This either suggests that a relocation of proteins of non-prokaryote origin into the mitochondria took place after the endosymbiotic event, or that many mitochondrial proteins in Fugu and humans are poorly conserved or annotated, making them more difficult to detect by sequence comparison methods. In multicellular organisms the number of increased nuclear mitochondrial genes can also be interpreted in the light of tissue differentiation: different cell types may harbor isoforms of mitochondria with different repertoires of proteins.

4.1.3. Diversification and core sets of mitochondrial proteomes

The cross-species BLAST comparison showed a variation in the size of the fraction of mitochondrial proteins unique to a given species. The values ranged from 35% (budding yeast)

to 69% (Fugu; Tab. 3.2, column 4), indicating that different degrees of mitochondrial proteome specialization exist. In addition, the magnitude of the values constitute a markedly increased percentage compared to the fractions of 15.5–22% given in the literature for baker's yeast (Karlberg *et al.*, 2000; Marcotte *et al.*, 2000). However, the existence of a large fraction of species-specific mitochondrial proteins strongly suggests that species evolution has been accompanied by a diversification of mitochondrial proteins and, in turn, of species-related organelle functions.

The species-specific diversification of mitochondrial functions is also reflected in the composition of the mitochondrial proteomes of parasites and plants as mitochondria of parasites are miniaturized. For parasites this can be explained by the fact that the original functions of the mitochondria are no longer needed since they are covered by the host organism. The human parasite *E. cuniculi* was thought to have completely lost the mitochondrial organelle over evolutionary time, possibly retaining a mitochondrion-derived organelle, the mitosome (Katinka *et al.*, 2001). In this work, 156 most likely mitochondrial proteins were detected in *E. cuniculi*, eight of which are even shared by all ten species under study (compare Tab. 3.2, p. 35), suggesting the existence of a mitochondrion in the parasite. In plants an opposing scenario emerges: mitochondria have acquired a substantial number of proteins originating from the progenitor of extant chloroplasts, since in Arabidopsis about 6% of the proteins tracing back to cyanobacteria were shown to be re-directed to the mitochondrion (see Fig. 3.6, p. 44). This considerable relocation of proteins from the protochloroplast to the mitochondrion supports the view that in the course of eukaryote evolution any endogenous or introduced gene may be tested and selected for its usefulness by altered targeting to the different compartments of the cell. Already in 1977 Francois Jacob introduced his model of 'evolution as a tinkerer' (Jacob, 1977) that explains such a continuous subcellular relocation of proteins in eukaryotic cells with the aim to build up new functions by using components of existing ones.

The detected non-mitochondrial homologues of species-specific mitochondrial proteins (Tab. 3.2; difference between values in columns 4 and 5) suggest that mitochondrial proteome diversity is on one hand based on gene evolution leading to novel proteins located in the mitochondrion, and on the other hand generated by relocation of conserved genes due to altered targeting. The fraction of these non-mitochondrial homologues of species-specific mitochon-

drial proteins, which feature an mTP in the respective species but do not feature one in any other species, is undersized in Arabidopsis (6% of the total mitochondrial proteome) and even smaller in Plasmodium (3%), which used to be a photosynthetical active species before it turned into a parasite with highly degenerate plastid genome (Wilson & Williamson, 1997). An explanation for these observations might be that both species feature (or used to feature in case of Plasmodium) one additional organelle – the chloroplast – to which former mitochondrial proteins may have been relocated.

The cross-species comparison not only detected differences among the mitochondrial proteomes of the eight non-parasite species but also revealed similarities. This is depicted by means of a similarity tree based on a distance matrix that was calculated from the relatedness of the eight mitochondrial proteomes (Fig. 3.2, p. 37). The similarity tree reflects the commonly accepted phylogenetic relations of the analyzed species. As expected, the amount of shared (i.e. conserved) proteins between a set of species decreased with the number of compared species. For example, the two insects Anopheles and Drosophila share about 600 common mitochondrial proteins, the intersection of Anopheles, Drosophila and Caenorhabditis comprises of about 350 gene products, while all five metazoans share only close to 200 proteins. If all species with fully functional mitochondria were considered, the number of conserved proteins – or core sets – among the mitochondrial proteomes ranged from 46 to 132 (see Tab. 3.2; budding yeast and Arabidopsis, respectively). This variation in numbers can be explained by the presence of multiple homologues in the various species: one protein sequence in budding yeast shares, for example, four homologues in Arabidopsis and vice versa.

The members of the mitochondrial core set are highly conserved among species and evolutionary time and should in turn be accountable for indispensable organellar functions. To test this hypothesis, the 100 proteins of the identified human mitochondrial core set were used to query the OMIM database. Eighteen of those proteins could directly be associated with human diseases (Tab. 3.3, p. 38), which is considerably more than one would expect by chance: the 'morbid map', which is a compilation of all known and suspected disease genes, currently comprises of 3,206 entries (`ftp://ftp.ncbi.nih.gov/repository/OMIM/morbidmap`). Compared to the total human genome (currently 37,891 genes) this only equals 8.5%. Furthermore, of the 82 proteins left, 40 (or 49%) could be mapped to an interval for genetically undefined

human disorders and thus might be attractive and valuable candidates for future studies of gene-disease relationships in humans and other species. These findings strongly support the validity of the hypothesis. For 61 of the 100 identified proteins α-proteobacterial homologues were detected, implying that essential mitochondrial functions trace back to the endosymbiont.

The functional classification of the 100 human and 132 Arabidopsis core set proteins (Fig. 3.3, p. 39) according to Katinka *et al.* (2001) showed that these proteins mainly maintain functions involved in (a) metabolism and energy, (b) cellular communication and signal transduction, and (c) transport. It is noteworthy, that in Arabidopsis categories (a) and (b) are more highly represented than in the human species (35% and 30% in Arabidopsis versus 28% and 26% in humans, respectively) but that other categories are undersized (e.g. transport: 7% in Arabidopsis versus 26% in humans) or do not even exist (cell rescue and cell death). This suggests that in Arabidopsis former functions of the chloroplast – which is also strongly involved in metabolism – have been retargeted to the mitochondrion (Martin *et al.*, 2002; Leister, 2003) and that additional ways for communication between the organelles must exist. An opposed retargeting from mitochondrion to chloroplast seems feasible (as for example in the case of the cell rescue and cell death category) but remains to be analyzed further.

4.2. Chloroplasts – the heirs of cyanobacteria

4.2.1. The cyanobacterial heritage of the Arabidopsis genome

Previous work has shown that many gene transfers from the organelle to the nucleus have occurred during plastid evolution (Martin *et al.*, 1998; Millen *et al.*, 2001), but estimates for the total number of transferred genes have been elusive. To determine the number of Arabidopsis proteins of cyanobacterial origin a phylogenetic approach using ML calculations was employed. Many methods for inferring phylogenetic trees from DNA and protein sequence data have been developed (Felsenstein, 1988, 1989). Among these methods, the maximum likelihood method (Felsenstein, 1981) is based on an explicit model for the substitution process of nucleotides or amino acids. The method has a sound statistical ground (Reeves, 1992) and has proved to be powerful in recovering correct tree topologies by computer simulation studies (Hasegawa *et al.*, 1991; Hasegawa & Fujiwara, 1993; Yang, 1996). Thus, although heavily dependent on computer power, ML analysis – due to its probabilistic basis – is the method of choice for the inferrence of reliable phylogenies.

The ML analysis provided in this work resulted from a collaboration with W. Martin, T. Rujan, A. Hansen, S. Cornelsen, T. Lins, D. Leister, B. Stoebe, M. Hasegawa and D. Penny (Martin *et al.*, 2002) and this thesis reflects 30% of the performed work in all aspects of the cited paper. The analysis revealed that about 18% (or close to 4,800 proteins) of the Arabidopsis proteome trace back to the prokaryotic endosymbiont. This represents a significantly higher number compared to the 800–2,000 proteins previously estimated by BLAST analyses (Cavalier-Smith, 2000; The Arabidopsis Genome Initiative, 2000). Since these 18% only rely on a subset of 9,368 protein sequences (equalling 35% of the proteome) that sufficed the prerequisite for a phylogenetic inferrence the question arises, whether an extrapolation to the whole genome is justified. Two reasons indicate that this is the case:

(1) There is no *a priori* reason to suspect that this subset of Arabidopsis proteins should preferentially be populated by those proteins stemming from cyanobacteria as opposed to, for example, those stemming from the Arabidopsis host lineage.

(2) The remaining 65% of the Arabidopsis genome must have come from somewhere. Either they arose *de novo* from non-coding DNA – which is a very improbable assumption – or they

were generated by mechanisms such as recombination and duplication involving pre-existing coding sequences. The latter is more possible and demonstrated by the cyanobacterial component in the conserved fraction of proteins analyzed here – the second largest fraction of Arabidopsis proteins was shown to be most similar to cyanobacteria by the BLAST (section 3.2.1, p. 40).

Another question arising is, whether the estimated 18% are an underestimate or an overestimate. Since the only non-photosynthetic eukaryote species among the reference proteomes is yeast, one concern might be that including other eukaryotes could increase the number of homologues detected at an E value of 10^{-10} and thereby the subset of 9,368 analyzed proteins. Salzberg *et al.* (2001) showed in a comparable study that by increasing the reference set by five whole lineages the resulting homologues only increased by a few hundred. This indicates that the concern is probably not too severe.

One additional circumstance that might lead to an overestimate is based on the relationship between cyanobacteria and the chloroplast: if the cyanobacteria used in this work diverged from the endosymbiotic ancestor of plastids more recently than yeast diverged from Arabidopsis, then the cyanobacterial genes in Arabidopsis would share a more recent common ancestor with their homologues than the 'host' genes in Arabidopsis would share with their yeast homologues. This would bias the BLAST analysis, making cyanobacterial homologues easier to detect and, in turn, easier to use in the phylogenetic analysis, resulting in an overestimate. Obviously, the converse is feasible, too: if the cyanobacteria used in this work diverged from the cyanobacterium giving rise to plastids before yeast and Arabidopsis, the resulting bias would lead to an underestimate. This second possibility is even more likely since there is no evidence to suggest that the cyanobacteria and the plastid's ancestor have diverged after yeast and Arabidopsis. Furthermore, the commonly accepted view exists that the cyanobacterial lineage is a very ancient one. Two further observations suggest the estimated 18% to be an underestimate:

(1) the efficiency of phylogenetic inferrence heavily depends on the similarity of analyzed sequences and decreases with sequence divergence (Nei, 1996; Nei & Kumar, 2000). Therefore, many Arabidopsis–cyanobacterium branches may not have been detected due to the poor ability of phylogenetic methods to cope with weakly conserved proteins.

(2) a large fraction of Arabidopsis proteins were found branching with homologues from Gram-positive bacteria (see Fig. 3.4, p. 42). For example, more Arabidopsis proteins branched with their homologues from Mycobacterium (148 proteins) than with either Prochlorococcus (102) or Synechocystis (82). This may suggest that Arabidopsis acquired genes especially from Gram-positive after diverging from yeast. However, another conclusion would then also have to be that Arabidopsis acquired genes from all organisms in this study, because branches between Arabidopsis and all those organisms where detected by phylogenetic inferrence. Another assumption would have to be that Arabidopsis acquired genes from all three cyanobacteria in this analysis and not from just a single origin (and this even at a bootstrap probability ≥ 0.95). Such interpretations are unlikely and, in the case of Arabidopsis, sampling from all three cyanobacteria has already been contradicted suggesting that plastids trace back to a single ancestor (Stoebe & Kowallik, 1999; Moreira *et al.*, 2000). Thus, the detected similarity between the Gram-positive bacteria and Arabidopsis most probably reflects an overall similarity between Gram-positive and cyanobacteria. This could be shown for various cases, e.g. rRNA (Woese, 1987), protein trees (Hansmann & Martin, 2000; Schütz *et al.*, 2000; Xiong *et al.*, 2000), operon organization (Wächtershäuser, 1998), and lipoprotein components (Maeda & Omata, 1997). Therefore, the detected Gram-positive signal in the analysis presented is most likely due to genes that entered the plant lineage through the ancestors of plastids (Rujan & Martin, 2001). This Gram-positive signal – though substantial and probably of cyanobacterial origin – did not add to the 18% estimated which might therefore well be an underestimate.

To conclude, the traces of the cyanobacterial ancestors are found in a considerable portion of the genome of the flowering plant *A. thaliana* and those detected 4,800 proteins or 18% of the total proteome are more likely to be an underestimate.

4.2.2. Compartmentalization and functions of cyanobacteria-derived proteins

To analyze whether proteins of the nuclear genes that were derived from the chloroplast are targeted back to this organelle, a targeting prediction using TARGETP was performed on the set of 3,628 protein sequences. More than 50% of the cyanobacterial proteins were found to be not targeted to the plastid whereas many non-cyanobacterial proteins are (see Fig. 3.6, p. 44; also see section 4.2.4). In addition, proteins stemming from cyanobacteria often seem

to enter the secretory pathway. Previous studies have suggested that protein compartmentalization and gene origin do not necessarily correspond. However, this is still a commonly accepted view (Horiike *et al.*, 2001). Those previous findings have also indicated that proteins encoded by genes of cyanobacterial origin are targeted to different compartments, and furthermore, that proteins which do not stem from cyanobacteria might eventually be directed to the chloroplast by altered targeting (Abdallah *et al.*, 2000; Martin & Herrmann, 1998; Martin & Schnarrenberger, 1997).

Though the number of chloroplast targeted proteins stemming from cyanobacteria does not fully comply with the more accurate findings presented and explained in section 4.2.4, this work clearly shows that gene origin and protein compartmentalization do not strictly correspond.

The functional classification of the 1,700 proteins of cyanobacterial origin in *A. thaliana* (see Tab. 3.4, p. 44) revealed that they are involved in (a) functions anticipated, but also (b) a number of different functions not typically brought into connection with cyanobacteria. As an example for the first category, the proteins functionally involved in biosynthesis and metabolism (562 proteins or 33%) or energy generation (93 proteins, equalling 0.5%) are given. Those functions were already present in the cyanobacterial ancestor. In contrast, functions such as intracellular protein routing or disease resistance suggest that genes acquired from the ancestor of plastids were a rich source of genetic raw material for the evolution of new functions. This suggestion is also supported by the finding of the large number of 342 (or 20%) unclassified proteins. The translocation of genes to the nucleus helped the process of diversification, since those newly acquired nuclear genes can undergo duplication and diversification like any pre-existing gene. Many Arabidopsis genes are indeed recent duplicates (The Arabidopsis Genome Initiative, 2000).

4.2.3. An improved approach for the prediction of chloroplast transit peptides

This work presents an evaluation of four predictors for the detection of chloroplast transit peptides. It was shown that using a set of 2,450 proteins with experimentally confirmed location – their accuracy is far lower than originally reported (Emanuelsson *et al.*, 2000; Emanuelsson & von Heijne, 2001; Schein *et al.*, 2001; Bannai *et al.*, 2002). A more reliable method for the prediction of chloroplast protein targeting was introduced by the combination of the four pre-

dictors TARGETP, PCLR, iPSORT and PREDOTAR performing better than any of the predictors alone. A '3 out of 4' combination (at least tree out of the four used predictors identified a cTP) resulted in the best performance (sensitivity: 0.74, specificity: 0.72). This was confirmed by the Matthews correlation coefficient calculated for the four tested combinations (see Fig. 3.7, p. 45). In analogy to the prediction of mTPs, this improvement can be explained by the use of predictors that function differently and thus disadvantages of one algorithm can be compensated by the superiority of another, and secondly, by too small and/or suboptimally assembled protein sets used in the training of the predictors.

The function of TARGETP and iPSORT has been discussed in section 4.1.1. The additional predictors PREDOTAR and PCLR have been especially designed for the use in plants (Small, 1999; Emanuelsson & von Heijne, 2001; Schein *et al.*, 2001), the latter only being able to discriminate between cTP and non-cTP. PREDOTAR – like TARGETP – relies on a neural network architecture and thus is a very robust and accurate predictor. Since it has been trained on nearly the same protein set that was used for TARGETP (personal communication with I. Small) comparable accuracies in predictions are resulting. Nevertheless, PREDOTAR only distinguishes between cTP and mTP, while TARGETP integrates several neural network programs to predict cleavage sites and to discriminate between secretory, mitochondrial and chloroplast signals. Thus the two predictors – although relying on the same basic type of algorithm – are very dissimilar in their function. PCLR combines a principle component analysis with a logistic regression. The principal component analysis is a method of factoring co-linearity out of data and reducing dimensionality for a machine learning algorithm, (Rawlings *et al.*, 1998) while the logistic regression makes predictions or decisions between 0 (false) and 1 (true) if a certain regression coefficient has been reached. In the case of PREDOTAR a stepwise logistic regression was performed on the first 12 components of the principal component matrix using a regression coefficient of 0.42 (Schein *et al.*, 2001). PCLR has also been tested on the TARGETP training set and exhibits a comparable sensitivity at a lower specificity.

Another question is, why the prediction of cTPs is much more reliable compared to prediction of mTPs. This is probably due to the fact that cTPs differ more from random sequences than mTPs: they are longer and much more biased in amino acid composition. As an example, putative mTPs are at least five times more frequent in the Arabidopsis genome (all frames,

genes and intergenic regions) than putative cTPs. Further, cleavage site prediction for mTPs is not as reliable as for cTPs, and a portion of proteins targeted to the mitochondrion carries the targeting signal inside its sequence instead of an N-terminal transit peptide, which makes a reliable prediction a lot more difficult.

Summarizing, the same conclusions as with the prediction of mTPs are reached: although cTP prediction is more reliable than mTP prediction, it still needs further improvement. This can either be achieved by combinations of differently working algorithms, and/or by the optimization of the protein sets used to train those predictors. For this purpose, more protein sequences with experimentally confirmed localization of additional species are needed. Prediction algorithms for protein targeting will improve considerably when new proteins whose subcellular location is known, such as those in the test set comprising 2,450 proteins, are routinely incorporated into the training sets for predictors of subcellular targeting. Proteins predicted to be targeted to the chloroplast on the basis of improved algorithms can then be tested by proteomics for their actual location, giving rise to a re-iterative process of prediction, testing, and improvement of the prediction algorithm.

4.2.4. Size and diversification of chloroplast proteomes in flowering plants

The '3 out of 4' approach identified about 2,000 cTP-featuring proteins in the genome of Arabidopsis, while in rice a total of about 4,700 chloroplast targeted proteins were estimated for the first time (see Tab. 3.6, p. 47). In comparison to previous analyses in Arabidopsis, (Abdallah *et al.*, 2000; Emanuelsson *et al.*, 2000; The Arabidopsis Genome Initiative, 2000; Peltier *et al.*, 2002; Leister, 2003) predicting up to 4,000 cTPs, these estimates are strikingly lower and provide a new view of subcellular protein sorting in the model plant *A. thaliana*. At the same time, the importance of using exclusively experimentally validated protein test sets to ensure the accuracy of *in silico* studies is clearly demonstrated.

The chloroplast proteome of the monocot species rice showed to be more than twice the size of that of the dicot species Arabidopsis. This indicates that the size of the chloroplast proteome is surprisingly variable over evolutionary time. The existence of a large fraction of species-specific chloroplast proteins in the two species (60% and 79% in Arabidopsis and rice, respectively; see Tab. 3.7, p. 47) strongly suggests that the evolution of flowering plants has been accompanied

by a diversification of the chloroplast proteins, leading to species-related differences in organelle function. These findings are also reflected in the results of the presented analysis of the mitochondrial proteomes of ten different eukaryotes, which support a similar interpretation of organellar evolution. However, in some cases the difference noted in chloroplast proteomes might be due to interspecific variation in the number of specific isoforms of some chloroplast proteins (Osteryoung & McAndrew, 2001; Hedtke *et al.*, 2002).

A mutual set of 844 homologous chloroplast targeted proteins between the two species was identified (see Tab. 3.7). These 40% of the total chloroplast proteome of Arabidopsis suggest that a core set in the chloroplasts of flowering plants exists, which again parallels the findings in mitochondria. Based on an extrapolation for the whole Arabidopsis genome (compare section 3.2.5, p. 47 ff.), a number of around 650, or about three quarters, of these conserved proteins are descendants of the cyanobacterial endosymbiont. The functional classification of the core set of 228 proteins (see Fig. 3.8, p. 49) revealed that these conserved gene products are mainly related to metabolism, energy and protein fate and that this should hold true for most flowering plants. In comparison to the functional classification of the mitochondrial core set of Arabidopsis (see Fig. 3.3, p. 39), it is interesting to note that for both organelles the largest fraction of the proteome is represented by the same functional group – even at nearly identical values (metabolism plus energy ∼35%). In contrast, mitochondria of flowering plants seem to be involved in cellular communication a lot more than chloroplasts, since this functional class is overrepresented in mitochondria (30%) while it is very low in chloroplasts (2%), which instead feature a huge fraction of unclassified proteins (37%). This hints at a major role of mitochondria in the concerted interaction of the two organelles.

For Arabidopsis about half of the cTP proteins are predicted to derive from the endosymbiont. This is significantly more than extrapolated in previous estimates (Leister, 2003) or in section 3.2.2. This discrepancy can be explained by the fact that those estimates only made use of one cTP predictor (TARGETP) rather than four leading to more false positives, and furthermore estimates were only corrected by means of the sensitivity and specificity values given in the literature which were shown to have been overestimated in this work. As such, the necessity of a thorough evaluation of the reliability of computational predictions of subcellular targeting is demonstrated.

This work provides an example of how improved cTP prediction, in combination with a comparative genome analysis, can support the compositional and evolutionary examination of the proteome of chloroplasts. Of course, this type of analysis will develop its full potential only when the genome sequences of additional plant species become available.

4.3. Origin and intracellular targeting of the organellar proteomes in Arabidopsis thaliana – a summary

This chapter combines the findings from sections 3.1 and 3.2 to give a final overview of the genetic origin of mitochondrion and chloroplast in *A. thaliana*.

In this work, about 3,000 nuclear encoded proteins were predicted to most likely be targeted to the mitochondrion of which about 700 can still be traced back to the α-proteobacterial ancestor. However, a most recent analysis employing an extended test set (section 3.2.3, p. 43 ff.) suggests that these 3,000 are still an overestimate due to the extremely poor accuracy of the mTP predictors. This improved analysis also suggests a value of 1,400 mTP featuring proteins in the flowering plant *A. thaliana*. For cTP-featuring gene products a number close to 2,000 (with about 1,100 of cyanobacterial origin) was estimated, and about 4,800 genes within the genome of *A. thaliana* were shown to trace back to the cyanobacterial ancestor.

In addition, the number of proteins of the complete Arabidopsis genome that, at an E value of 10^{-10} or lower, feature a homologue in the α-proteobacterial pool was determined (data not shown). Of these 4,351 detected proteins, 735 were shown to be among the set of 1,700 cyanobacterial proteins identified in section 3.2.1. In these 735 cases, homology to both α-proteobacteria and cyanobacteria was detected by BLASTP analysis, but a cyanobacterial origin could unequivocally be inferred by the use of a more precise phylogenetic approach. Since those 735 detected proteins are based on a subset of the Arabidopsis genome, an extrapolation according to section 3.2.1 has to be performed ($735 \times \frac{26{,}620}{9{,}368}$), resulting in an estimated 2,100 proteins that have to be subtracted from the 4,351 found α-proteobacterial homologues. In this way, an estimate of about 2,300 genes in the nuclear genome of Arabidopsis that are derived from the α-proteobacterial endosymbiont is reached. These conclusions are summarized in Fig. 4.3 and demonstrate that for both organelles about half of the proteome still traces back

to the prokaryote ancestor.

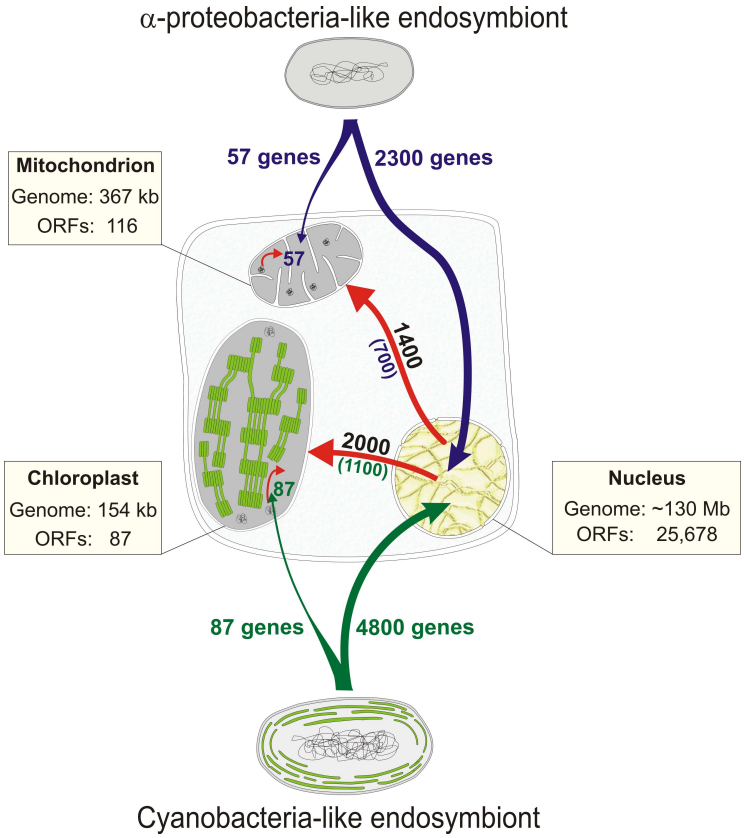

Fig. 4.1.: Overview of the intracellular targeting of prokaryote-derived proteins in A. thaliana. The figure summarizes the prokaryotic heritage of the Arabidopsis genome and the intracellular targeting of the inherited proteins. Values were taken from sections 3.1 and 3.2, respectively, or calculated as described above. Blue lines and numbers indicate genes and their respective gene products (red arrows) of α-proteobacterial origin. Green lines and numbers refers to genes and proteins derived from cyanobacteria. Black numbers indicate the total number of proteins targeted to the respective organelle. The values in parentheses refers to the subsets of mTP- and cTP-featuring proteins of α-proteobacterial and cyanobacterial origin, respectively). Figure adapted from Leister (2003).

The 1,100 cTP-featuring proteins of cyanobacterial origin are the result of a phylogenetic inferrence and, even though an extrapolation for the whole genome had to be performed, are quite reliable. However, the 2,300 genes of α-proteobacterial origin are only based on this extrapolation and thus have only been inferred indirectly. A phylogenetic maximum likelihood analysis including both the cyanobacterial and the α-proteobacterial protein pools would yield a more credible estimate. Additionally, this type of analysis could be used to determine the number of proteins that were relocated from one organelle to the other by altered targeting.

4.4. Nuclear chloroplast transcriptome responses in A. thaliana – communication between organelle and nucleus

4.4.1. A 3,292 GST array to monitor the nuclear chloroplast transcriptome

The expression of 3,292 genes, mainly coding for proteins of the chloroplast proteome, has been monitored using a GST-based array. An automated sequence-retrieval and primer design system was used to derive primers for the PCR amplification of corresponding ORFs from *Arabidopsis thaliana* genomic DNA (Varotto *et al.*, 2001). The use of GSTs instead of ESTs (Schenk *et al.*, 2000; Wang *et al.*, 2000; Schaffer *et al.*, 2001) offers several advantages – an improved representation of sequences being the most important one, since only about 75% of the corresponding genes of the spotted GSTs are represented by ESTs and publicly available through databases. This advantage is also reflected in the fact that, among the 35 conditions applied, for every of the remaining 25% of GSTs spotted on the array at least one significant change in expression of the corresponding gene could be detected. Thus, the transcription – and in turn – exsistence of those 25% of ORFs, which are only based on an automated annotation, was proven.

One possible drawback of GSTs in hybridization experiments is the presence of intron sequences that might interfere with an efficient annealing. However, the average fraction of intron sequences relative to the total length of sequences in this study did not exceed 35% and furthermore, primers were designed to amplify preferentially more than 500 bp of exon sequence (Varotto *et al.*, 2001). Thus, the efficient formation of DNA–DNA hybrids should be granted in all cases.

The reliability of the GST array has been verified in previous work employing quantitative reverse transcription PCR (Kurth *et al.*, 2002) and could even be improved in this study, making use of the statistical evaluation supplied by the ARRAYSTAT software.

4.4.2. Responses of the nuclear chloroplast transcriptome

In plants, expression profiles have been analyzed for various conditions and stimuli such as: responses to light (Harmer *et al.*, 2000; Ma *et al.*, 2001; Schaffer *et al.*, 2001; Kurth *et al.*, 2002),

analysis of mutants or transgenic plants (Kurth *et al.*, 2002; Wang *et al.*, 2002; Glazebrook *et al.*, 2003; Zik & Irish, 2003), seed development (Girke *et al.*, 2000; Ruuska *et al.*, 2002), hormone treatments (Goda *et al.*, 2002; Sawa *et al.*, 2002), changes in nutriments sources (Wang *et al.*, 2000; Negishi *et al.*, 2002), and abiotic and biotic stresses (Maleck *et al.*, 2000; Reymond *et al.*, 2000; Seki *et al.*, 2001; Vranová *et al.*, 2002). In this work, for the first time a survey of the nuclear chloroplast transcriptome in response to 35 different environmental and genetic conditions mostly affecting the chloroplast is presented. It revealed that the transcriptome response is subject to a broad variation in the number of significantly regulated genes. Depending on the condition, between 15% and 75% of the 3,292 genes were shown to be altered significantly in their expression levels (see Fig. 3.10, p. 53). The observed changes in expression levels of the chloroplast transcriptome seem to group into three basic types: they either consisted of (1) mainly upregulated or (2) mainly downregulated genes, or (3) in few cases exhibited a mixed response with a balanced ratio of upregulated to downregulated genes. This points to the conclusion that despite the 35 differing conditions (or stimuli) applied, the organelle reacts in only a few elementary ways, leading to a concerted regulation of huge sets of nuclear encoded genes. It is noteworthy that also genes of non-plastid and mitochondrial proteins seem to be affected by the three basic types: Clearly, most of the 631 non-chloroplast-protein genes followed the prevailing trend of expression, suggesting that chloroplast signaling affects larger regulatory networks like the control of cellular metabolism or development.

A hierarchical clustering was performed on the 35 expression profiles (see Fig. 3.11, p. 54), because the clustering technique not only distinguishes between up- and down-regulation but additionally takes into consideration the level of the change in expression. This way, a more detailed picture of the relations between various conditions emerged, which is also represented by the cladogram on top of Fig. 3.11. As expected many nearest neighbors were detected at the outermost branches of the cladogram: mutants like *gun1-1* and *gun5*, or *psan-1* and *psao-1*, or the high-CO_2 stress conditions ($+CO_2$_1d and $+CO_2$_10d) were found to exhibit highly similar expression profiles, demonstrating the validity and robustness of the clustering approach. Also unexpected pairs of conditions, as in the case of the *hcf145* mutant, which exhibits a perturbed assembly of PS I, and the low-CO_2 stress condition ($-CO_2$_4d) showed highly similar expression profiles, although they strongly differ in their phenotypes. Further,

conditions that initially were expected to share a common node in the cladogram did not do so, as with the *psad1-1* and *psae1-1* mutants which neither cluster with each other nor with the *psan-1* and *psao-1* mutants. These observations support the conclusion introduced above and developed further in the next section: the chloroplast seems to integrate the various stimuli, resulting in one of three basic transcriptome response types.

In cladograms, relationships between members – especially those more distant – are very tedious to compare, and it is not possible to unequivocally determine the relationship between the members of two pairs of nearest neighbors branching from different nodes. Therefore, the information stemming from the clustering was combined with an 'all-against-all' comparison, relating each transcriptome response to every other on the basis of mutually versus totally regulated genes. The results of this analysis yielded the transcriptome response map which is depicted in Fig. 3.13, p. 57, clearly showing the three types of transcriptome responses and a well-organized overall structure. This transcriptome response map provides a framework for the classification of novel chloroplast-function mutants and/or additional treatments based on their expression profiles. Furthermore, the identification of potential *cis*-acting regulatory elements that should be overrepresented in the promoters of co-expressed genes is feasible, and detection specificity will increase with every additional condition added to the transcriptome response map. Identified promoters will, in turn, provide further insights into the complex network of gene regulation, since they function as transducers, integrating information about the internal state of a cell.

4.4.3. Evidence for a binary master switch in chloroplast signaling

Organellar functions are regulated at several levels such as: transcriptional regulation (reviewed by Maliga, 1998), regulation of RNA processing and stability (reviewed by Schuster *et al.*, 1999; Bock, 2000; Hoffmann *et al.*, 2001), translational regulation (reviewed by Zerges, 2000, 2002), and the regulation of protein stability and assembly (reviewed by Choquet & Vallon, 2000; Choquet *et al.*, 2001).

Chloroplast and leaf development are regulated by light signal transduction involving phytochromes and cryptochromes, and it is known that the photosynthetic process itself contributes signals to its own regulation, as well as that of other plastid and non-plastid processes

(Allen *et al.*, 1995; Jarvis, 2001; Rodermel, 2001; Mullineaux & Karpinski, 2002). Intra-organellar signaling, sensing the redox state of the photosynthetic apparatus, modulates the expression of certain plastome genes (Pfannschmidt *et al.*, 1999) and possibly the activity of plastid protein kinases (Forsberg *et al.*, 2001). As shown in this work and elsewhere (Martin & Herrmann, 1998), a massive transfer of chloroplast genes into the nucleus occurred during evolution of the plant lineage by EGT. This, in turn, required the invention of an inter-organellar signaling, so the chloroplast remained able to affect the coordinate expression of whole sets of – now nuclear – genes (Jarvis, 2001).

In previous studies the emitter side of these plastid-to-nucleus signal relays was subject to extensive research providing the first working models. It could be demonstrated that plastid signals are mediated by porphyrins, reactive oxygen intermediates or carotenoids (overview in Rodermel, 2001). The *gun* mutants, which express *Lhcb* in the absence of chloroplast development and despite photobleaching, were used to dissect the plastid-to-nucleus signaling in Arabidopsis, and the tetrapyrrole Mg-protoporphyrin IX could be identified as the plastid signal which is sent out to the nucleus (Surpin *et al.*, 2002; Strand *et al.*, 2003). Two working models for the signaling by Mg-protoporphyrin IX have been proposed and either involve the inhibition of a cytoplasmic activator, or the activation of a cytoplasmic inhibitor – both capable of binding the CUF-1 element of the *Lhcb* gene (compare Fig. 1.4, p. 19). In contrast, the *cue* signaling mutants exhibit an opposing gene regulation: they underexpress light-regulated nuclear genes encoding chloroplast genes (Streatfield *et al.*, 1999; Li *et al.*, 1995). Since the *cue1* mutant is defective in the phosphoenolpyruvate/phosphate translocator and the *cue1* is light intensity-dependent, it was suggested that the plastid signal is modulated through metabolites involved in the shikimate pathway and thus dependent on a plastid supply of phosphoenolpyruvate (Streatfield *et al.*, 1999).

In this work, the two classes of signaling mutants (*gun1*, *gun5*, and *cue1*, respectively) were present among the 35 conditions tested. As expected, highly similar expression profiles between the two *gun* mutants, and opposing transcriptome responses between the *gun* mutants and the *cue1* mutant were detected (see Fig. 3.14, p. 58 and Fig. 3.12, p. 55). The *gun* mutants showed a similarity of more than 80%, supporting the findings by Strand *et al.* (2003), while almost all genes up-regulated in *gun1* and *gun5* were down-regulated in *cue1*. These

findings are also reflected in the combined analysis (Fig. 3.13, p. 57), clearly indicating different classes for both types of signaling mutants. The detection of two major and diametrically opposed responses, implies the existence of a two-state switch. This binary switch seems to integrate the information on perturbations and subsequent homeostatic adjustments due to altered/environmental states, regulating an array of nuclear genes in concert. Once activated, this switch either promotes class-1- or class-3-type expression changes in vast parts of the nuclear chloroplast transcriptome. This is unequivocally depicted in Fig. 3.14, in which the transcriptome response for nearly all genes monitored 'switches' from red (up-regulated) to green (down-regulated).

Also other lines of evidence suggest the existence of this binary master switch: For the *flu* mutant (Meskauskiene *et al.*, 2001) a very similar scenario emerges: FLU acts as a negative regulator of tetrapyrrole biosynthesis that, when mutated, leads to an accumulation of protochlorophyllide – even if the plants are kept growing in the dark. Such dark-adapted *flu* mutants showed a clear class-1-type expression, which after exposure to light for 30 min was completely shifted to the class-3-type profile (see Fig. 3.13 and Fig. 3.14: flu_D and flu_L, respectively). This can be explained by the fact that the intermediate protochlorophyllide is extremely destructive when illuminated (Matringe *et al.*, 1989) and thus its accumulation leads to a stress response in the plant, and ultimately to a class-3-type response. Since tetrapyrroles play an important role in GUN-type signaling, (Strand *et al.*, 2003) FLU, as a regulator of their biosynthesis, may further be involved in controlling the signal stability.

Not all conditions that – based on present knowledge – should produce similar physiological perturbations, were paralleled by similar transcriptome responses. Instead, opposite responses were observed: the mutants *psae1-1*, *psad1-1*, *psan-1* and *psao-1*, all affecting PS I, clearly differed in both their photosynthetic phenotype (data not shown) and also their expression profile, since they either elicited a class-2-type or class-3-type response. Further, the *psae1-1* mutation, as well as high-iron treatment (+Fe) and irradiation in the cold (ML_4C), are thought to affect the cellular redox state (Varotto *et al.*, 2000) or to increase the level of oxidative stress (Halliwell & Gutteridge, 1984), and all conditions lead to a class-3-type response. But conversely, herbicides (+Bro, +Par), also supposed to increase the level of oxidative stress (Böger & Sandmann, 1998; Tjus *et al.*, 2001), induce the opposed class-1-type responses.

Various perturbations in signaling (*gun5, cue1, flu*), in ATP/ADP ratio (*atpd-1, atpc1-1*), or in redox state (*psae1-1*, herbicide treatments, photoinhibitory conditions) all lead to one of the two possible modes of the binary master switch – even though all should represent differing physiological states. This up- or down-regulation of largely the same set of genes strongly supports the existence of a master switch, clearly demonstrating its integrative function. Furthermore, since the two main responses, induced by most of the conditions tested and including diverse and apparently unrelated conditions, are characterized by expression profiles typical of, or opposite to, those of GUN-type mutants of plastid signaling (Surpin *et al.*, 2002; Strand *et al.*, 2003), it appears that the GUN-type signaling is embedded in the larger type of reactions controlled by the binary master switch.

A coordinate expression of whole sets of nuclear genes in response to different treatments is not uncommon to both prokaryotes and eukaryotes. Examples are the at least 30 SOS response genes in *E. coli* exhibiting a coordinate expression upon treatments leading to DNA damage (Sutton *et al.*, 2000; Khil & Camerini-Otero, 2002) and the ESR response in yeast that leads to an activation of 900 genes upon exposure to multiple stresses (Gasch *et al.*, 2000). This work shows that specific treatments or genetic defects result in three discrete types of coordinate transcriptional responses of nuclear genes coding for chloroplast proteins.

However, the results presented in this work can not finally prove whether *cue1* is a signaling mutant and whether Mg-protoporphyrin IX, identified by characterization of the *gun* mutants, is a plastid signal. Especially the latter statement is currently under debate in the plastid-to-nucleus signaling field for a number of reasons:

(1) the norflurazon-based screen that led to the identification of the *gun* mutants is quite artificial and unspecific. It results in various perturbations, such as an impairment of accumulation of carotenoids, chlorophyll-binding proteins and photosystems, or production of reactive oxygen species in the bleached chloroplasts. One effect on transcriptional level, which was used in the screen for mutants, is the down-regulation of the *Lhcb* gene. However, those mutants identified, that do not show a decrease in *Lhcb* mRNA, must not necessarily be affected in plastid-to-nucleus signaling, but can just be less sensitive to norflurazon treatment (a biochemical effect instead of a signaling effect). Furthermore, it is also imaginable that downregulation of the chlorophyll biosynthesis due to a *gun* mutation leads to a desensitization to norflurazon

treatment and, only secondarily, to the observed transcriptional effects.

(2) in the green algae Chlamydomonas, the role of chlorophyll synthesis intermediates, particularly Mg-protoporphyrin, as regulators of nuclear gene expression has been subject to extensive research for years (Kropat *et al.*, 2000; Xu *et al.*, 2001). In this species, tetrapyrroles were shown to induce transcript accumulation of nuclear genes, as opposed to the repression postulated for Arabidopsis. Even though chlorophyll biosynthesis in Chlamydomonas differs from that in plants (e.g. they can become green in the dark), these findings indicate that the models for Mg-protoporphyrin IX signaling might have to be revised in the future.

(3) it still remains to be proven that Mg-protoporphyrin can leave the chloroplast to act as an activator or repressor in the cytoplasma. So far, the proposed 'factor X' responsible for the translocation of Mg-protoporphyrin (Surpin *et al.*, 2002) could not be identified. Furthermore, the characterizations that led to the model of Mg-protoporphyrin as a signaling molecule are mostly of genetic nature. Satisfactory biochemical analyses such as measurements of the accumulation of intermediates of tetrapyrrole, or the activity of the involved enzymes are still lacking and might result in diverging conclusions.

4.4.4. Reasons for a class-2-type response

In contrast to the clearly distinguishable transcriptome responses of classes 1 and 3, the expression profiles of class-2-type transcriptome changes fall into several subclasses, which seem more related to a class-1-type response than to the class-3-type (see Fig. 3.13). Conditions that induce a class-2-type response involve shift in light quality (PSII/I), high CO_2 concentrations ($+CO_2$_1d, $+CO_2$_10d), or mutations of the photosynthetic machinery (*psad-1, psan-1, psao-1*). The commonalities of all these genetic/environmental situations are a change in the redox state and/or in the concentration of reactive oxygen species in the chloroplast (Pfannschmidt *et al.*, 2001; Mullineaux & Karpinski, 2002; Noctor *et al.*, 2002; Pfannschmidt, 2003).

The redox state – or more precisely the redox state of plastoquinone (the ratio of PQH_2 to PQ) – is known to be responsible for a direct regulation of organelle gene transcription, especially for genes coding for proteins in the reaction centers of both photosystems (Pfannschmidt *et al.*, 1999). This allows for a rapid response of the chloroplast to changes in the physical environment via a short, simple signaling pathway to avoid, for example, the generation of reactive

oxygen species. By this mechanism, the plant is able to efficiently convert energy and to protect itself from damage caused by its own electron transport chain. Furthermore, chloroplast redox signals were also shown to influence to expression of nuclear genes and thus represent an additional class of plastid signals (Pfannschmidt, 2003). This utilization of changes in redox potential for the regulation of expression of chloroplast (and mitochondrial) key proteins in the electron transport chains is most likely a very ancient mechanism that was already present in the ancestors of today's organelles (Allen, 1993). The observed responses, in terms of the regulation of the nuclear chloroplast transcriptome, indicate that information on the redox state or the concentration of reactive oxygen species feeds into a more complex signaling pathway inducing some genes while repressing others. Thus the response of the nuclear transcriptome is either directly influenced by the redox state signaling, or it is modulated by the antique signaling still present within the chloroplast.

However, by increasing the number of conditions, additional mechanisms of transcriptional responses, acting in parallel or down-stream of the three responses described here, could be unraveled, and might enable a further clarification of the class-2-type response.

5. Summary/Zusammenfassung

5.1. Summary

Mitochondria and chloroplasts are semi-autonomous organelles that have arisen through an endosymbiotic event and, over evolutionary time, have donated most of their genome to the nucleus of the host cell. Due to this transfer of genetic material, the expression of many proteins of the organellar proteomes, now synthesized in the cytosol and re-targeted to the organelles, came under control of the nucleus. Subsequently, means of communication between organelles and nucleus must exist, enabling the organelles to take influence on the nuclear gene expression. This thesis focused on structural and functional genomics in mitochondria and chloroplasts, addressing questions related to the composition, origin and evolution of the organelles, as well as, chloroplast-to-nucleus signaling.

The accuracy of five different predictors for the detection of N-terminal targeting peptides was evaluated employing test sets consisting of proteins with experimentally proven subcellular localization, and found to be substantially lower than reported before. Combinations of the predictors showed to be more accurate than any of the predictors alone and were subsequently used to estimate the size and composition of the organellar proteomes.

A prediction of the mitochondrial proteomes for ten species was performed and revealed that functional mitochondria harbor from a few hundred to more than 3,000 gene products. A core set of conserved mitochondrial proteins could be identified whose functions are mostly related to transport and metabolism, and – if mutated – are frequently associated with disease in humans.

In collaboration with W. Martin (Universität Düsseldorf) and co-workers (Martin *et al.*, 2002), the cyanobacterial heritage of the Arabidopsis genome was estimated by phylogenetic

inferrence and about 4,800 genes (or 18% of the genome) were shown to have been acquired from the prokaryotic ancestor. In both flowering plants, *A. thaliana* and *O. sativa*, about 7% of the whole proteome were predicted to be targeted to the chloroplast, with close to 600 of those proteins shared by both species and most likely to be derived from cyanobacteria. The functions of this subset are mainly related to metabolism and energy.

In both organelles, species-specific proteins were detected indicating a functional diversification. Even though in *A. thaliana* the cTP-featuring proteins are predominantly of prokaryotic origin (more than 50%), this indicates that post-endosymbiotic relocations of proteins from/to the chloroplast occurred by altered targeting. These findings were also confirmed for mitochondria.

A differential-expression analysis of the nuclear chloroplast transcriptome under 35 environmental and genetic conditions was performed. It revealed, that most of those conditions elicit only three main classes of transcriptome response. Two of these classes, probably involving GUN-type plastid signaling, are characterized by alterations, in opposite directions, in the expression of largely overlapping sets of genes. Thus these findings, suggest the existence of a regulatory, binary master-switch.

5.2. Zusammenfassung

Mitochondrien und Chloroplasten sind semi-autonome Organelle, die durch Endosymbiose entstanden sind und deren nahezu gesamtes genetisches Erbgut im Laufe der Zeit an den Kern der Wirtszelle übertragen wurde. Die Transkription der transferierten Gene und die Synthese der zugehörigen Proteine, die nun im Cytoplasma stattfinden, unterliegen folglich nicht mehr der direkten Kontrolle der Organelle. Deswegen müssen Mitochondrien und Chloroplasten eine Möglichkeit zur Kommunikation mit dem Kern besitzen um, je nach Bedarf, Einfluss auf die Genexpression nehmen zu können. Diese Arbeit behandelt Fragestellungen zu Zusammensetzung, Ursprung und Evolution von Chloroplast und Mitochondrion, welche mit Hilfe der Genomstrukturanalyse angegangen wurden. Weiterhin wurde die Kommunikation zwischen Chloroplast und Kern unter Zuhilfenahme der Genomfunktionsanalyse untersucht.

Die Vorhersagegenauigkeit fünf verschiedener Programme zur Detektion von N-terminalen Transitpeptiden wurde an Proteinen mit experimentell bestätigter subzellulärer Lokalisation bestimmt. Die ermittelten Genauigkeiten waren deutlich geringer als zuvor veröffentlicht, weswegen Kombinationen der Vorhersageprogramme verwandt wurden, die eine verbesserte Vorhersagegenauigkeit ergaben. Dieser kombinatorische Ansatz wurde deshalb zur Bestimmung von Grösse und Zusammensetzung der Organellproteome eingesetzt.

Eine Vorhersage der Grössen von mitochondrialen Proteomen in zehn Spezies zeigte, dass funktionelle Proteome einige hundert bis zu mehr als 3.000 Proteine beinhalten. Weiterhin wurde eine in allen untersuchten Spezies konservierte Menge von Proteinen identifiziert, die sich zum grössten Teil den Funktionsklassen 'Transport', oder 'Metabolismus' zuordnen liessen. Mutationen der korrespondierenden Gene konnten zudem in vielen Fällen mit menschlichen Erbkrankheiten in Verbindung gebracht werden.

In Zusammenarbeit mit W. Martin (Universität Düsseldorf) und Mitarbeitern (Martin *et al.*, 2002) wurde das cyanobakterielle Erbe im Arabidopsisgenom mit Hilfe von phylogenetischen Untersuchungen bestimmt: circa 4.800 Gene (entsprechend 18% des Genoms) leiten sich vom prokaryotischen Vorfahren ab. In den höheren Pflanzen *A. thaliana* und *O. sativa* konnten jeweils 7% des Gesamtgenoms dem chloroplastidären Proteom zugeordnet werden. Beide Spezies teilen hieraus eine Untergruppe von 600 Proteinen, welche sehrwahrscheinlich cyanobakterieller

Herkunft sind und überwiegend zu den funktionellen Klassen ‘Metabolismus’ und ‘Energie’ gehören.

In beiden Organellen konnten spezies-spezifische Proteine nachgewiesen werden, die auf eine funktionelle Diversifizierung hindeuten. Obwohl mehr als die Hälfte der detektierten chloroplastidären Proteine in *A. thaliana* cyanobakteriellen Ursprungs sind, weist dies auf eine post-endosymbiotische Relokalisierung von Proteinen zum/vom Chloroplasten hin. Gleiches konnte für die Mitochondrien bestätigt werden.

Eine differentielle Expressionsanalyse der nukleären Chloroplastengene unter 35 verschiedenen Konditionen (Mutationen oder Umweltfaktoren) ergab, dass die meisten der getesteten Bedingungen nur eine der drei identifizierten Hauptantworten des Transkriptoms hervorrufen. Zwei dieser Klassen zeichnen sich wahrscheinlich durch eine GUN-ähnliche Transkriptomantwort aus und zeigen eine gegenläufige Expression einer grösstenteils identischen Genmenge. Diese Resultate weisen auf einen binär arbeitenden, regulatorischen Hauptschalter hin.

References

Abdallah F, Salamini F & Leister D. A prediction of the size and evolutionary origin of the proteome of chloroplasts of Arabidopsis. *Trends Plant Sci*, 5:141–2, **2000**.

Adachi J & Hasegawa M. MOLPHY. Institue of Statistical Mathematics, Tokyo, **1996**.

Adam Z, Adamska I, Nakabayashi K, Ostersetzer O, Haussuhl K, Manuell A, Zheng B, Vallon O, Rodermel S, Shinozaki K & Clarke A. Chloroplast and mitochondrial proteases in Arabidopsis. A proposed nomenclature. *Plant Physiol*, 125:1912–8, **2001**.

Allen J. Control of gene expression by redox potential and the requirement for chloroplast and mitochondrial genomes. *J Theor Biol*, 165:609–31, **1993**.

Allen J. The function of genomes in bioenergetic organelles. *Philos Trans R Soc Lond B Biol Sci*, 358:19–37; discussion 37–8, **2003**.

Allen J, Alexciev K & Håkansson G. Photosynthesis. regulation by redox signalling. *Curr Biol*, 5:869–72, **1995**.

Allen J & Raven J. Free-radical-induced mutation vs redox regulation: costs and benefits of genes in organelles. *J Mol Evol*, 42:482–92, **1996**.

Altschul S, Gish W, Miller W, Myers E & Lipman D. Basic local alignment search tool. *J Mol Biol*, 215:403–10, **1990**.

Amuthan G, Biswas G, Ananadatheerthavarada H, Vijayasarathy C, Shephard H & Avadhani N. Mitochondrial stress-induced calcium signaling, phenotypic changes and invasive behavior in human lung carcinoma A549 cells. *Oncogene*, 21:7839–49, **2002**.

Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, Podowski RM, Naslund AK, Eriksson AS, Winkler HH & Kurland CG. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature*, 396:133–40, **1998**.

Ayliffe M & Timmis J. Plastid DNA sequence homologies in the tobacco nuclear genome. *Mol Gen Genet*, 236:105–12, **1992**.

Baker A & Schatz G. Sequences from a prokaryotic genome or the mouse dihydrofolate reductase gene can restore the import of a truncated precursor protein into yeast mitochondria. *Proc Natl Acad Sci U S A*, 84:3117–21, **1987**.

Balmer Y, Koller A, del Val G, Manieri W, Schürmann P & Buchanan B. Proteomics gives insight into the regulatory function of chloroplast thioredoxins. *Proc Natl Acad Sci U S A*, 100:370–5, **2003**.

Bannai H, Tamada Y, Maruyama O, Nakai K & Miyano S. Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics*, 18:298–305, **2002**.

Barak S, Nejidat A, Heimer Y & Volokita M. Transcriptional and posttranscriptional regulation of the glycolate oxidase gene in tobacco seedlings. *Plant Mol Biol*, 45:399–407, **2001**.

Bauer J, Hiltbrunner A & Kessler F. Molecular biology of chloroplast biogenesis: gene expression, protein import and intraorganellar sorting. *Cell Mol Life Sci*, 58:420–33, **2001**.

Beg A & Baldwin A. The I kappa B proteins: multifunctional regulators of Rel/NF-kappa B transcription factors. *Genes Dev*, 7:2064–70, **1993**.

Böger P & Sandmann G. *Photosynthesis: A Comprehensive Treatise*, pp. 337–351. University Press, Cambridge, UK, **1998**.

Biswas G, Adebanjo O, Freedman B, Anandatheerthavarada H, Vijayasarathy C, Zaidi M, Kotlikoff M & Avadhani N. Retrograde Ca2+ signaling in C2C12 skeletal myocytes in response to mitochondrial genetic and metabolic stress: a novel mode of inter-organelle crosstalk. *EMBO J*, 18:522–33, **1999**.

Biswas G, Anandatheerthavarada H, Zaidi M & Avadhani N. Mitochondria to nucleus stress signaling: a distinctive mechanism of NFkappaB/Rel activation through calcineurin-mediated inactivation of IkappaBbeta. *J Cell Biol*, 161:507–19, **2003**.

Bland J & Altman D. Multiple significance tests: the Bonferroni method. *BMJ*, 310:170, **1995**.

Block M, Joyard J & Douce R. Purification and characterization of E37, a major chloroplast envelope protein. *FEBS Lett*, 287:167–70, **1991**.

Block M, Tewari A, Albrieux C, Maréchal E & Joyard J. The plant S-adenosyl-L-methionine:Mg-protoporphyrin IX methyltransferase is located in both envelope and thylakoid chloroplast membranes. *Eur J Biochem*, 269:240–8, **2002**.

Bölter B, May T & Soll J. A protein import receptor in pea chloroplasts, Toc86, is only a proteolytic fragment of a larger polypeptide. *FEBS Lett*, 441:59–62, **1998**.

Bock R. Sense from nonsense: how the genetic information of chloroplasts is altered by RNA editing. *Biochimie*, 82:549–57, **2000**.

Bock R & Koop H. Extraplastidic site-specific factors mediate RNA editing in chloroplasts. *EMBO J*, 16:3282–8, **1997**.

Brosius J. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene*, 238:115–34, **1999**.

Brown J. Ancient horizontal gene transfer. *Nat Rev Genet*, 4:121–32, **2003**.

Bruce B. The role of lipids in plastid protein transport. *Plant Mol Biol*, 38:223–46, **1998**.

Butterfield NJ. *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the mesoproterozoicneoproterozoic radiation of eukaryotes. *Paleobiology*, 26:386–404, **2000**.

Cao Y, Adachi J, Janke A, Pääbo S & Hasegawa M. Phylogenetic relationships among eutherian orders estimated from inferred sequences of mitochondrial proteins: instability of a tree based on a single gene. *J Mol Evol*, 39:519–27, **1994**.

Cavalier-Smith T. Eukaryotes with no mitochondria. *Nature*, 326:332–3, **1987**.

Cavalier-Smith T. Membrane heredity and early chloroplast evolution. *Trends Plant Sci*, 5:174–82, **2000**.

Cavalier-Smith T. The phagotrophic origin of eukaryotes and phylogenetic classification of protozoa. *Int J Syst Evol Microbiol*, 52:297–354, **2002**.

Cerutti H, Johnson A, Boynton J & Gillham N. Inhibition of chloroplast DNA recombination and repair by dominant negative mutants of *Escherichia coli RecA*. *Mol Cell Biol*, 15:3003–11, **1995**.

Chaddock A, Mant A, Karnauchov I, Brink S, Herrmann R, Klösgen R & Robinson C. A new type of signal peptide: central role of a twin-arginine motif in transfer signals for the delta pH-dependent thylakoidal protein translocase. *EMBO J*, 14:2715–22, **1995**.

Chen M, Choi Y, Voytas D & Rodermel S. Mutations in the arabidopsis *VAR2* locus cause leaf variegation due to the loss of a chloroplast FtsH protease. *Plant J*, 22:303–13, **2000**.

Chen X & Schnell D. Protein import into chloroplasts. *Trends Cell Biol*, 9:222–7, **1999**.

Chen X, Smith M, Fitzpatrick L & Schnell D. In vivo analysis of the role of atTic20 in protein import into chloroplasts. *Plant Cell*, 14:641–54, **2002**.

Choquet Y & Vallon O. Synthesis, assembly and degradation of thylakoid membrane proteins. *Biochimie*, 82:615–34, **2000**.

Choquet Y, Wostrikoff K, Rimbault B, Zito F, Girard-Bascou J, Drapier D & Wollman F. Assembly-controlled regulation of chloroplast gene translation. *Biochem Soc Trans*, 29:421–6, **2001**.

Cunningham F, Pogson B, Sun Z, McDonald K, DellaPenna D & Gantt E. Functional analysis of the beta and epsilon lycopene cyclase enzymes of Arabidopsis reveals a mechanism for control of cyclic carotenoid formation. *Plant Cell*, 8:1613–26, **1996**.

Debouck C & Goodfellow P. DNA microarrays in drug discovery and development. *Nat Genet*, 21:48–50, **1999**.

Delsite R, Kachhap S, Anbazhagan R, Gabrielson E & Singh K. Nuclear genes involved in mitochondria-to-nucleus communication in breast cancer cells. *Mol Cancer*, 1:6, **2002**.

Delwiche CF. Tracing the thread of plastid diversity through the tapestry of life. *Am Nat*, 154:S164–S177, **1999**.

Devore J & Peck R. *Statistics: the Exploration and Analysis of Data.* Duxbury Press, Pacific Grove, CA, **1997**.

Diekert K, Kispal G, Guiard B & Lill R. An internal targeting signal directing proteins into the mitochondrial intermembrane space. *Proc Natl Acad Sci U S A*, 96:11752–7, **1999**.

Douglas SE. Plastid evolution: origins, diversity, trends. *Curr Opin Genet Dev*, 8:655–61, **1998**.

Drake J. The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes. *Ann N Y Acad Sci*, 870:100–7, **1999**.

Echtay K, Esteves T, Pakay J, Jekabsons M, Lambert A, Portero-Otín M, Pamplona R, Vidal-Puig A, Wang S, Roebuck S & Brand M. A signalling role for 4-hydroxy-2-nonenal in regulation of mitochondrial uncoupling. *EMBO J*, 22:4103–4110, **2003**.

Eisen M, Spellman P, Brown P & Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A*, 95:14863–8, **1998**.

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

Emanuelsson O, Nielsen H & von Heijne G. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci*, 8:978–84, **1999**.

Emanuelsson O & von Heijne G. Prediction of organellar targeting signals. *Biochim Biophys Acta*, 1541:114–9, **2001**.

Endo T, Kawamura K & Nakai M. The chloroplast-targeting domain of plastocyanin transit peptide can form a helical structure but does not have a high affinity for lipid bilayers. *Eur J Biochem*, 207:671–5, **1992**.

Essigmann B, Güler S, Narang R, Linke D & Benning C. Phosphate availability affects the thylakoid lipid composition and the expression of *SQD1*, a gene required for sulfolipid biosynthesis in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A*, 95:1950–5, **1998**.

Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol*, 17:368–76, **1981**.

Felsenstein J. Phylogenies from molecular sequences: inference and reliability. *Annu Rev Genet*, 22:521–65, **1988**.

Felsenstein J. PHYLIP – phylogeny inference package (version 3.2). *Cladistics*, 5:164–166, **1989**.

Finazzi G, Chasen C, Wollman F & de Vitry C. Thylakoid targeting of Tat passenger proteins shows no delta pH dependence in vivo. *EMBO J*, 22:807–15, **2003**.

Fischer N, Stampacchia O, Redding K & Rochaix J. Selectable marker recycling in the chloroplast. *Mol Gen Genet*, 251:373–80, **1996**.

Flügge U, Fischer K, Gross A, Sebald W, Lottspeich F & Eckerskorn C. The triose phosphate-3-phosphoglycerate-phosphate translocator from spinach chloroplasts: nucleotide sequence of a full-length cDNA clone and import of the in vitro synthesized precursor protein into chloroplasts. *EMBO J*, 8:39–46, **1989**.

Forsberg J, Rosenquist M, Fraysse L & Allen J. Redox signalling in chloroplasts and mitochondria: genomic and biochemical evidence for two-component regulatory systems in bioenergetic organelles. *Biochem Soc Trans*, 29:403–7, **2001**.

Forsburg S & Guarente L. Communication between mitochondria and the nucleus in regulation of cytochrome genes in the yeast *Saccharomyces cerevisiae*. *Annu Rev Cell Biol*, 5:153–80, **1989**.

Fujiwara Y, Asogawa M & Nakai K. Prediction of mitochondrial targeting signals using hidden markov model. *Genome Inform Ser Workshop Genome Inform*, 8:53–60, **1997**.

Gaasterland T, Karp PD, Karplus K, Ouzounis CA, Sander C & Valencia A, eds. *Proceedings of the 5th International Conference on Intelligent Systems for Molecular Biology, Halkidiki, Greece, June 21-26, 1997*. AAAI, **1997**.

Gasch A, Spellman P, Kao C, Carmel-Harel O, Eisen M, Storz G, Botstein D & Brown P. Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell*, 11:4241–57, **2000**.

Gasser SM, Ohashi A, Daum G, Bohni PC, Gibson J, Reid GA, Yonetani T & Schatz G. Imported mitochondrial proteins cytochrome b2 and cytochrome c1 are processed in two steps. *Proc Natl Acad Sci U S A*, 79:267–71, **1982**.

Gaubier P, Wu H, Laudié M, Delseny M & Grellet F. A chlorophyll synthetase gene from *Arabidopsis thaliana*. *Mol Gen Genet*, 249:58–64, **1995**.

Gavel Y & von Heijne G. Cleavage-site motifs in mitochondrial targeting peptides. *Protein Eng*, 4:33–7, **1990**a.

Gavel Y & von Heijne G. A conserved cleavage-site motif in chloroplast transit peptides. *FEBS Lett*, 261:455–8, **1990**b.

Geissler A, Chacinska A, Truscott K, Wiedemann N, Brandner K, Sickmann A, Meyer H, Meisinger C, Pfanner N & Rehling P. The mitochondrial presequence translocase: an essential role of Tim50 in directing preproteins to the import channel. *Cell*, 111:507–18, **2002**.

Germot A, Philippe H & Le Guyader H. Evidence for loss of mitochondria in microsporidia from a mitochondrial-type HSP70 in *Nosema locustae*. *Mol Biochem Parasitol*, 87:159–68, **1997**.

Ghosh S, May M & Kopp E. NF-kappa B and Rel proteins: evolutionarily conserved mediators of immune responses. *Annu Rev Immunol*, 16:225–60, **1998**.

Girke T, Todd J, Ruuska S, White J, Benning C & Ohlrogge J. Microarray analysis of developing Arabidopsis seeds. *Plant Physiol*, 124:1570–81, **2000**.

Glazebrook J, Chen W, Estes B, Chang H, Nawrath C, Métraux J, Zhu T & Katagiri F. Topology of the network integrating salicylate and jasmonate signal transduction derived from global expression phenotyping. *Plant J*, 34:217–28, **2003**.

Goda H, Shimada Y, Asami T, Fujioka S & Yoshida S. Microarray analysis of brassinosteroid-regulated genes in Arabidopsis. *Plant Physiol*, 130:1319–34, **2002**.

Goksoyr J. Evolution of eucaryotic cells. *Nature*, 214:1161, **1967**.

Gray M. The endosymbiont hypothesis revisited. *Int Rev Cytol*, 141:233–357, **1992**.

Gray M, Burger G & Lang B. Mitochondrial evolution. *Science*, 283:1476–81, **1999**.

Gray M, Lang B, Cedergren R, Golding G, Lemieux C, Sankoff D, Turmel M, Brossard N, Delage E, Littlejohn T, Plante I, Rioux P, Saint-Louis D, Zhu Y & Burger G. Genome structure and gene content in protist mitochondrial DNAs. *Nucleic Acids Res*, 26:865–78, **1998**.

Gray MW, Burger G & Lang BF. The origin and early evolution of mitochondria. *Genome Biol*, 2:REVIEWS1018, **2001**.

Hajdukiewicz P, Allison L & Maliga P. The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. *EMBO J*, 16:4041–8, **1997**.

Halliwell B & Gutteridge J. Role of iron in oxygen radical reactions. *Methods Enzymol*, 105:47–56, **1984**.

Hansmann S & Martin W. Phylogeny of 33 ribosomal and six other proteins encoded in an ancient gene cluster that is conserved across prokaryotic genomes: influence of excluding poorly alignable sites from analysis. *Int J Syst Evol Microbiol*, 50 Pt 4:1655–63, **2000**.

Harmer S, Hogenesch J, Straume M, Chang H, Han B, Zhu T, Wang X, Kreps J & Kay S. Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. *Science*, 290:2110–3, **2000**.

Hasegawa M & Fujiwara M. Relative efficiencies of the maximum likelihood, maximum parsimony, and neighbor-joining methods for estimating protein phylogeny. *Mol Phylogenet Evol*, 2:1–5, **1993**.

Hasegawa M, Kishino H & Saitou N. On the maximum likelihood method in molecular phylogenetics. *J Mol Evol*, 32:443–5, **1991**.

Haussühl K, Andersson B & Adamska I. A chloroplast DegP2 protease performs the primary cleavage of the photodamaged D1 protein in plant photosystem II. *EMBO J*, 20:3–22, **2001**.

Hedtke B, Börner T & Weihe A. Mitochondrial and chloroplast phage-type RNA polymerases in Arabidopsis. *Science*, 277:809–11, **1997**.

Hedtke B, Börner T & Weihe A. One RNA polymerase serving two genomes. *EMBO Rep*, 1:435–40, **2000**.

Hedtke B, Legen J, Weihe A, Herrmann R & Börner T. Six active phage-type RNA polymerase genes in *Nicotiana tabacum*. *Plant J*, 30:625–37, **2002**.

Henze K & Martin W. How do mitochondrial genes get into the nucleus? *Trends Genet*, 17:383–7, **2001**.

Herman L, Jacobs A, Montagu MV & Depicker A. Plant chromosome/marker gene fusion assay for study of normal and truncated T-DNA integration events. *Mol Gen Genet*, 224:248–56, **1990**.

Herrmann J. Converting bacteria to organelles: evolution of mitochondrial protein sorting. *Trends Microbiol*, 11:74–9, **2003**.

Herrmann RG. *Eukaryotism, towards a new interpretation.*, pp. 73–118. Springer Verlag, Heidelberg, **1997**.

Hirohashi T, Hase T & Nakai M. Maize non-photosynthetic ferredoxin precursor is mis-sorted to the intermembrane space of chloroplasts in the presence of light. *Plant Physiol*, 125:2154–63, **2001**.

Ho C, Noji M, Saito M & Saito K. Regulation of serine biosynthesis in Arabidopsis. Crucial role of plastidic 3-phosphoglycerate dehydrogenase in non-photosynthetic tissues. *J Biol Chem*, 274:397–402, **1999**.

Hoffmann M, Kuhn J, Däschner K & Binder S. The RNA world of plant mitochondria. *Prog Nucleic Acid Res Mol Biol*, 70:119–54, **2001**.

Horiike T, Hamada K, Kanaya S & Shinozawa T. Origin of eukaryotic cell nuclei by symbiosis of archaea in bacteria is revealed by homology-hit analysis. *Nat Cell Biol*, 3:210–4, **2001**.

Hua S & Sun Z. Support vector machine approach for protein subcellular localization prediction. *Bioinformatics*, 17:721–8, **2001**.

Huang C, Ayliffe M & Timmis J. Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature*, 422:72–6, **2003**.

Isaya G & Kalousek F. Mitochondrial intermediate peptidase. In G von Heijne, ed., *Signal Peptidases*, pp. 87–103. R.G. Landes Publishing Co., Austin, **1994**.

Jacob F. Evolution and tinkering. *Science*, 196:1161–6, **1977**.

Jarvis P. Intracellular signalling: the chloroplast talks! *Curr Biol*, 11:R307–10, **2001**.

Jarvis P, Chen L, Li H, Peto C, Fankhauser C & Chory J. An *Arabidopsis* mutant defective in the plastid general protein import apparatus. *Science*, 282:100–3, **1998**.

Jarvis P & Soll J. Toc, tic, and chloroplast protein import. *Biochim Biophys Acta*, 1590:177–89, **2001**.

Javaux E, Knoll A & Walter M. Morphological and ecological complexity in early eukaryotic ecosystems. *Nature*, 412:66–9, **2001**.

Jones D, Taylor W & Thornton J. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci*, 8:275–82, **1992**.

Kalousek F, Hendrick JP & Rosenberg LE. Two mitochondrial matrix proteases act sequentially in the processing of mammalian matrix enzymes. *Proc Natl Acad Sci U S A*, 85:7536–40, **1988**.

Karlberg O, Canback B, Kurland CG & Andersson SG. The dual origin of the yeast mitochondrial proteome. *Yeast*, 17:170–87, **2000**.

Katinka MD, Duprat S, Cornillot E, Metenier G, Thomarat F, Prensier G, Barbe V, Peyretaillade E, Brottier P, Wincker P, Delbac F, El Alaoui H, Peyret P, Saurin W, Gouy M, Weissenbach J & Vivares CP. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature*, 414:450–3, **2001**.

Keith B, Dong X, Ausubel F & Fink G. Differential induction of 3-deoxy-D-arabinoheptulosonate 7-phosphate synthase genes in *Arabidopsis thaliana* by wounding and pathogenic attack. *Proc Natl Acad Sci U S A*, 88:8821–5, **1991**.

Khil P & Camerini-Otero R. Over 1000 genes are involved in the DNA damage response of *Escherichia coli*. *Mol Microbiol*, 44:89–105, **2002**.

Kieselbach T, Hagman T, Andersson B & Schröder W. The thylakoid lumen of chloroplasts. Isolation and characterization. *J Biol Chem*, 273:6710–6, **1998**.

Klee H, Muskopf Y & Gasser C. Cloning of an *Arabidopsis thaliana* gene encoding 5-enolpyruvylshikimate-3-phosphate synthase: sequence analysis and manipulation to obtain glyphosate-tolerant plants. *Mol Gen Genet*, 210:437–42, **1987**.

Kliebenstein D, Monde R & Last R. Superoxide dismutase in Arabidopsis: an eclectic enzyme family with disparate regulation and protein localization. *Plant Physiol*, 118:637–50, **1998**.

Knight J, Duckett C, Sullivan J, Walker A & Gray J. Tissue-specific, light-regulated and plastid-regulated expression of the single-copy nuclear gene encoding the chloroplast Rieske FeS protein of *Arabidopsis thaliana*. *Plant Cell Physiol*, 43:522–31, **2002**.

Koncz C, Mayerhofer R, Koncz-Kalman Z, Nawrath C, Reiss B, Redei G & Schell J. Isolation of a gene encoding a novel chloroplast protein by T-DNA tagging in *Arabidopsis thaliana*. *EMBO J*, 9:1337–46, **1990**.

Kouranov A, Chen X, Fuks B & Schnell D. Tic20 and Tic22 are new components of the protein import apparatus at the chloroplast inner envelope membrane. *J Cell Biol*, 143:991–1002, **1998**.

Kouranov A & Schnell D. Analysis of the interactions of preproteins with the import machinery over the course of protein import into chloroplasts. *J Cell Biol*, 139:1677–85, **1997**.

Krimm I, Gans P, Hernandez J, Arlaud G & Lancelin J. A coil-helix instead of a helix-coil motif can be induced in a chloroplast transit peptide from *Chlamydomonas reinhardtii*. *Eur J Biochem*, 265:171–80, **1999**.

Kropat J, Oster U, Rüdiger W & Beck C. Chloroplast signalling in the light induction of nuclear HSP70 genes requires the accumulation of chlorophyll precursors and their accessibility to cytoplasm/nucleus. *Plant J*, 24:523–31, **2000**.

Kurth J, Varotto C, Pesaresi P, Biehl A, Richly E, Salamini F & Leister D. Gene-sequence-tag expression analyses of 1,800 genes related to chloroplast functions. *Planta*, 215:101–9, **2002**.

Lao N, Schoneveld O, Mould R, Hibberd J, Gray J & Kavanagh T. An Arabidopsis gene encoding a chloroplast-targeted beta-amylase. *Plant J*, 20:519–27, **1999**.

Larkin R, Alonso J, Ecker J & Chory J. GUN4, a regulator of chlorophyll synthesis and intracellular signaling. *Science*, 299:902–6, **2003**.

Last R, Bissinger P, Mahoney D, Radwanski E & Fink G. Tryptophan mutants in Arabidopsis: the consequences of duplicated tryptophan synthase beta genes. *Plant Cell*, 3:345–58, **1991**.

Lee CM, Sedman J, Neupert W & Stuart RA. The DNA helicase, Hmi1p, is transported into mitochondria by a C-terminal cleavable targeting signal. *J Biol Chem*, 274:20937–42, **1999**.

Lee M, Kuo F, Whitmore G & Sklar J. Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc Natl Acad Sci U S A*, 97:9834–9, **2000**.

Leister D. Chloroplast research in the genomic age. *Trends Genet*, 19:47–56, **2003**.

Li H, Culligan K, Dixon R & Chory J. CUE1: A mesophyll cell-specific positive regulator of light-controlled gene expression in Arabidopsis. *Plant Cell*, 7:1599–1610, **1995**.

Lin X, Kaul S, Rounsley S, Shea T, Benito M, Town C, Fujii C, Mason T, Bowman C, Barnstead M, Feldblyum T, Buell C, Ketchum K, Lee J, Ronning C, Koo H, Moffat K, Cronin L, Shen M, Pai G, Aken SV, Umayam L, Tallon L, Gill J & Venter J. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature*, 402:761–8, **1999**.

Lindahl M, Tabak S, Cseke L, Pichersky E, Andersson B & Adam Z. Identification, characterization, and molecular cloning of a homologue of the bacterial FtsH protease in chloroplasts of higher plants. *J Biol Chem*, 271:29329–34, **1996**.

Lynch M. Mutation accumulation in transfer RNAs: molecular evidence for Muller's ratchet in mitochondrial genomes. *Mol Biol Evol*, 13:209–20, **1996**.

Ma L, Li J, Qu L, Hager J, Chen Z, Zhao H & Deng X. Light control of Arabidopsis development entails coordinated regulation of genome expression and cellular pathways. *Plant Cell*, 13:2589–607, **2001**.

Maeda S & Omata T. Substrate-binding lipoprotein of the cyanobacterium *Synechococcus* sp. strain pcc 7942 involved in the transport of nitrate and nitrite. *J Biol Chem*, 272:3036–41, **1997**.

Maleck K, Levine A, Eulgem T, Morgan A, Schmid J, Lawton K, Dangl J & Dietrich R. The transcriptome of *Arabidopsis thaliana* during systemic acquired resistance. *Nat Genet*, 26:403–10, **2000**.

Maliga P. Two plastid RNA polymerases of higher plants: an evolving story. *Trends Plant Sci*, 3:4–6, **1998**.

Marcotte EM, Xenarios I, van Der Bliek AM & Eisenberg D. Localizing proteins in the cell from their phylogenetic profiles. *Proc Natl Acad Sci U S A*, 97:12115–20, **2000**.

Martin W & Herrmann R. Gene transfer from organelles to the nucleus: how much, what happens, and why? *Plant Physiol*, 118:9–17, **1998**.

Martin W & Müller M. The hydrogen hypothesis for the first eukaryote. *Nature*, 392:37–41, **1998**.

Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M & Penny D. 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*, 99:12246–51, **2002**.

Martin W & Russell M. 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*, 358:59–83; discussion 83–5, **2003**.

Martin W & Schnarrenberger C. The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. *Curr Genet*, 32:1–18, **1997**.

Martin W, Stoebe B, Goremykin V, Hapsmann S, Hasegawa M & Kowallik KV. Gene transfer to the nucleus and the evolution of chloroplasts. *Nature*, 393:162–5, **1998**.

Matringe M, Camadro J, Labbe P & Scalla R. Protoporphyrinogen oxidase as a molecular target for diphenyl ether herbicides. *Biochem J*, 260:231–5, **1989**.

Matthews B. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim Biophys Acta*, 405:442–51, **1975**.

May T & Soll J. Chloroplast precursor protein translocon. *FEBS Lett*, 452:52–6, **1999**.

May T & Soll J. 14-3-3 proteins form a guidance complex with chloroplast precursor proteins in plants. *Plant Cell*, 12:53–64, **2000**.

Mayer K, Schüller C, Wambutt R, Murphy G, Volckaert G, Pohl T, Düsterhöft A, Stiekema W, Entian K, Terryn N, Harris B, Ansorge W, Brandt P, Grivell L, Rieger M, Weichselgartner M, de Simone V, Obermaier B, Mache R, Müller M, Kreis M, Delseny M, Puigdomenech P, Watson M & McCombie W. Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. *Nature*, 402:769–77, **1999**.

Mereschkowsky C. Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. *Biol Centralbl*, 25:593–604, **1905**.

Meskauskiene R, Nater M, Goslings D, Kessler F, op den Camp R & Apel K. FLU: a negative regulator of chlorophyll biosynthesis in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A*, 98:12826–31, **2001**.

Miège C, Maréchal E, Shimojima M, Awai K, Block M, Ohta H, Takamiya K, Douce R & Joyard J. Biochemical and topological properties of type A MGDG synthase, a spinach chloroplast envelope enzyme catalyzing the synthesis of both prokaryotic and eukaryotic MGDG. *Eur J Biochem*, 265:990–1001, **1999**.

Millar AH, Sweetlove LJ, Giege P & Leaver CJ. Analysis of the Arabidopsis mitochondrial proteome. *Plant Physiol*, 127:1711–27, **2001**.

Millen R, Olmstead R, Adams K, Palmer J, Lao N, Heggie L, Kavanagh T, Hibberd J, Gray J, Morden C, Calie P, Jermiin L & Wolfe K. Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell*, 13:645–58, **2001**.

Møller S, Kunkel T & Chua N. A plastidic ABC protein involved in intercompartmental communication of light signaling. *Genes Dev*, 15:90–103, **2001**.

Mochizuki N, Brusslan J, Larkin R, Nagatani A & Chory J. 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*, 98:2053–8, **2001**.

Moore M, Harrison M, Peterson E & Henry R. Chloroplast oxa1p homolog albino3 is required for post-translational integration of the light harvesting chlorophyll-binding protein into thylakoid membranes. *J Biol Chem*, 275:1529–32, **2000**.

Moran N. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc Natl Acad Sci U S A*, 93:2873–8, **1996**.

Moreira D, Guyader HL & Philippe H. The origin of red algae and the evolution of chloroplasts. *Nature*, 405:69–72, **2000**.

Muller HJ. The relation of recombination to mutual advance. *Mut Res*, 1:2–9, **1964**.

Mullineaux P & Karpinski S. Signal transduction in response to excess light: getting out of the chloroplast. *Curr Opin Plant Biol*, 5:43–8, **2002**.

Nakabayashi K, Ito M, Kiyosue T, Shinozaki K & Watanabe A. Identification of *clp* genes expressed in senescing Arabidopsis leaves. *Plant Cell Physiol*, 40:504–14, **1999**.

Nakai K & Horton P. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem Sci*, 24:34–6, **1999**.

Negishi T, Nakanishi H, Yazaki J, Kishimoto N, Fujii F, Shimbo K, Yamamoto K, Sakata K, Sasaki T, Kikuchi S, Mori S & Nishizawa N. cDNA microarray analysis of gene expression during Fe-deficiency stress in barley suggests that polar transport of vesicles is implicated in phytosiderophore secretion in Fe-deficient barley roots. *Plant J*, 30:83–94, **2002**.

Nei M. Phylogenetic analysis in molecular evolutionary genetics. *Annu Rev Genet*, 30:371–403, **1996**.

Nei M & Kumar S. *Molecular Evolution and Phylogenetics*. Oxford University Press, Oxford, **2000**.

Nielsen E, Akita M, Davila-Aponte J & Keegstra K. Stable association of chloroplastic precursors with protein translocation complexes that contain proteins from both envelope membranes and a stromal Hsp100 molecular chaperone. *EMBO J*, 16:935–46, **1997**.

Nilsson R, Brunner J, Hoffman N & van Wijk K. Interactions of ribosome nascent chain complexes of the chloroplast-encoded D1 thylakoid membrane protein with cpSRP54. *EMBO J*, 18:733–42, **1999**.

Niyogi K & Fink G. Two anthranilate synthase genes in Arabidopsis: defense-related regulation of the tryptophan pathway. *Plant Cell*, 4:721–33, **1992**.

Noctor G, Veljovic-Jovanovic S, Driscoll S, Novitskaya L & Foyer C. Drought and oxidative load in the leaves of C3 plants: a predominant role for photorespiration? *Ann Bot (Lond)*, 89 Spec No:841–50, **2002**.

Ohta M, Sugita M & Sugiura M. Three types of nuclear genes encoding chloroplast RNA-binding proteins (cp29, cp31 and cp33) are present in *Arabidopsis thaliana*: presence of cp31 in chloroplasts and its homologue in nuclei/cytoplasms. *Plant Mol Biol*, 27:529–39, **1995**.

Olsen L & Keegstra K. The binding of precursor proteins to chloroplasts requires nucleoside triphosphates in the intermembrane space. *J Biol Chem*, 267:433–9, **1992**.

Osteryoung K & McAndrew R. The plastid division machine. *Annu Rev Plant Physiol Plant Mol Biol*, 52:315–333, **2001**.

Oswald O, Martin T, Dominy P & Graham I. Plastid redox state and sugars: interactive regulators of nuclear-encoded photosynthetic gene expression. *Proc Natl Acad Sci U S A*, 98:2047–52, **2001**.

Pahl H. Activators and target genes of Rel/NF-kappaB transcription factors. *Oncogene*, 18:6853–66, **1999**.

Papenbrock J & Grimm B. Regulatory network of tetrapyrrole biosynthesis–studies of intracellular signalling involved in metabolic and developmental control of plastids. *Planta*, 213:667–81, **2001**.

Peltier J, Emanuelsson O, Kalume D, Ytterberg J, Friso G, Rudella A, Liberles D, Söderberg L, Roepstorff P, von Heijne G & van Wijk K. Central functions of the lumenal and peripheral thylakoid proteome of Arabidopsis determined by experimentation and genome-wide prediction. *Plant Cell*, 14:211–36, **2002**.

Peltier J, Friso G, Kalume D, Roepstorff P, Nilsson F, Adamska I & van Wijk K. Proteomics of the chloroplast: systematic identification and targeting analysis of lumenal and peripheral thylakoid proteins. *Plant Cell*, 12:319–41, **2000**.

Peltier J, Ytterberg J, Liberles D, Roepstorff P & van Wijk K. Identification of a 350-kDa ClpP protease complex with 10 different Clp isoforms in chloroplasts of *Arabidopsis thaliana*. *J Biol Chem*, 276:16318–27, **2001**.

Pesaresi P, Varotto C, Meurer J, Jahns P, Salamini F & Leister D. Knock-out of the plastid ribosomal protein L11 in Arabidopsis: effects on mRNA translation and photosynthesis. *Plant J*, 27:179–89, **2001**a.

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

Pfannschmidt T. Chloroplast redox signals: how photosynthesis controls its own genes. *Trends Plant Sci*, 8:33–41, **2003**.

Pfannschmidt T, Allen JF & Oelmüller R. Principles of redoc control in photosynthesis gene expression. *Physiol Planta*, 112, **2001**.

Pfannschmidt T, Nilsson A & Allen JF. Photosynthetic control of chloroplast gene expression. *Nature*, 397:625–628, **1999**.

Rawlings J, Pantula S & Dickey D. *Applied Regression Analysis, a Research Tool*. Springer-Verlag, New York, NY, **1998**.

Reeves J. Heterogeneity in the substitution process of amino acid sites of proteins coded for by mitochondrial DNA. *J Mol Evol*, 35:17–31, **1992**.

Reinbothe S & Reinbothe C. Regulation of chlorophyll biosynthesis in angiosperms. *Plant Physiol*, 111:1–7, **1996**.

Reymond P, Weber H, Damond M & Farmer E. Differential gene expression in response to mechanical wounding and insect feeding in Arabidopsis. *Plant Cell*, 12:707–20, **2000**.

Rivas JDL, Lozano J & Ortiz A. Comparative analysis of chloroplast genomes: functional annotation, genome-based phylogeny, and deduced evolutionary patterns. *Genome Res*, 12:567–83, **2002**.

Roberts J. Proteomics and a future generation of plant molecular biologists. *Plant Mol Biol*, 48:143–54, **2002**.

Robinson C, Thompson S & Woolhead C. Multiple pathways used for the targeting of thylakoid proteins in chloroplasts. *Traffic*, 2:245–51, **2001**.

Rodermel S. Pathways of plastid-to-nucleus signaling. *Trends Plant Sci*, 6:471–8, **2001**.

Rodermel S & Park S. Pathways of intracellular communication: tetrapyrroles and plastid-to-nucleus signaling. *Bioessays*, 25:631–6, **2003**.

Ruddle F, Amemiya C, Carr J, Kim C, Ledje C, Shashikant C & Wagner G. Evolution of chordate *hox* gene clusters. *Ann N Y Acad Sci*, 870:238–48, **1999**.

Rujan T & Martin W. How many genes in Arabidopsis come from cyanobacteria? An estimate from 386 protein phylogenies. *Trends Genet*, 17:113–20, **2001**.

Ruuska S, Girke T, Benning C & Ohlrogge J. Contrapuntal networks of gene expression during Arabidopsis seed filling. *Plant Cell*, 14:1191–206, **2002**.

Saitou N & Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*, 4:406–25, **1987**.

Salzberg SL, White O, Peterson J & Eisen JA. Microbial genes in the human genome: lateral transfer or gene loss? *Science*, 292:1903–6, **2001**.

Sassa S & Nagai T. The role of heme in gene expression. *Int J Hematol*, 63:167–78, **1996**.

Sawa S, Ohgishi M, Goda H, Higuchi K, Shimada Y, Yoshida S & Koshiba T. The *HAT2* gene, a member of the HD-Zip gene family, isolated as an auxin inducible gene by DNA microarray screening, affects auxin response in Arabidopsis. *Plant J*, 32:1011–22, **2002**.

Schaffer R, Landgraf J, Accerbi M, Simon V, Larson M & Wisman E. Microarray analysis of diurnal and circadian-regulated genes in Arabidopsis. *Plant Cell*, 13:113–23, **2001**.

Schein AI, Kissinger JC & Ungar LH. Chloroplast transit peptide prediction: a peek inside the black box. *Nucleic Acids Res*, 29:E82, **2001**.

Schenk P, Kazan K, Wilson I, Anderson J, Richmond T, Somerville S & Manners J. Coordinated plant defense responses in Arabidopsis revealed by microarray analysis. *Proc Natl Acad Sci U S A*, 97:11655–60, **2000**.

Schmidt H, Dresselhaus T, Buck F & Heinz E. Purification and PCR-based cDNA cloning of a plastidial n-6 desaturase. *Plant Mol Biol*, 26:631–42, **1994**.

Schnell D. Shedding light on the chloroplast protein import machinery. *Cell*, 83:521–4, **1995**.

Schnell D & Hebert D. Protein translocons: multifunctional mediators of protein translocation across membranes. *Cell*, 112:491–505, **2003**.

Schütz M, Brugna M, Lebrun E, Baymann F, Huber R, Stetter K, Hauska G, Toci R, Lemesle-Meunier D, Tron P, Schmidt C & Nitschke W. Early evolution of cytochrome bc complexes. *J Mol Biol*, 300:663–75, **2000**.

Schubert M, Petersson U, Haas B, Funk C, Schröder W & Kieselbach T. Proteome map of the chloroplast lumen of *Arabidopsis thaliana*. *J Biol Chem*, 277:8354–65, **2002**.

Schuenemann D, Gupta S, Persello-Cartieaux F, Klimyuk V, Jones J, Nussaume L & Hoffman N. A novel signal recognition particle targets light-harvesting proteins to the thylakoid membranes. *Proc Natl Acad Sci U S A*, 95:10312–6, **1998**.

Schuster G, Lisitsky I & Klaff P. Polyadenylation and degradation of mRNA in the chloroplast. *Plant Physiol*, 120:937–44, **1999**.

Scotti P, Urbanus M, Brunner J, de Gier J, von Heijne G, van der Does C, Driessen A, Oudega B & Luirink J. YidC, the *Escherichia coli* homologue of mitochondrial Oxa1p, is a component of the Sec translocase. *EMBO J*, 19:542–9, **2000**.

Seki M, Narusaka M, Abe H, Kasuga M, Yamaguchi-Shinozaki K, Carninci P, Hayashizaki Y & Shinozaki K. Monitoring the expression pattern of 1300 Arabidopsis genes under drought and cold stresses by using a full-length cDNA microarray. *Plant Cell*, 13:61–72, **2001**.

Sharp P. Split genes and RNA splicing. *Cell*, 77:805–15, **1994**.

Small I. Predotar v0.5. **1999**. `http://www.inra.fr/predotar/`.

Smit A. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr Opin Genet Dev*, 9:657–63, **1999**.

Stegemann S, Hartmann S, Ruf S & Bock R. High-frequency gene transfer from the chloroplast genome to the nucleus. *Proc Natl Acad Sci U S A*, 100:8828–33, **2003**.

Stitt M & Fernie A. From measurements of metabolites to metabolomics: an 'on the fly' perspective illustrated by recent studies of carbon-nitrogen interactions. *Curr Opin Biotechnol*, 14:136–44, **2003**.

Stoebe B. Distribution and nomenclature of protein-coding genes in 12 sequenced chloroplast genomes. *Plant Mol Biol Rep*, 16:243–255, **1998**.

Stoebe B & Kowallik K. Gene-cluster analysis in chloroplast genomics. *Trends Genet*, 15:344–7, **1999**.

Strand A, Asami T, Alonso J, Ecker J & Chory J. Chloroplast to nucleus communication triggered by accumulation of Mg-protoporphyrin IX. *Nature*, 421:79–83, **2003**.

Streatfield S, Weber A, Kinsman E, Häusler R, Li J, Post-Beittenmiller D, Kaiser W, Pyke K, Flügge U & Chory J. The phosphoenolpyruvate/phosphate translocator is required for phenolic metabolism, palisade cell development, and plastid-dependent nuclear gene expression. *Plant Cell*, 11:1609–22, **1999**.

Stupar R, Lilly J, Town C, Cheng Z, Kaul S, Buell C & Jiang J. Complex mtDNA constitutes an approximate 620-kb insertion on *Arabidopsis thaliana* chromosome 2: implication of potential sequencing errors caused by large-unit repeats. *Proc Natl Acad Sci U S A*, 98:5099–103, **2001**.

Sturn A, Quackenbush J & Trajanoski Z. Genesis: cluster analysis of microarray data. *Bioinformatics*, 18:207–8, **2002**.

Sun T & Kamiya Y. The Arabidopsis *GA1* locus encodes the cyclase ent-kaurene synthetase A of gibberellin biosynthesis. *Plant Cell*, 6:1509–18, **1994**.

Surpin M, Larkin R & Chory J. Signal transduction between the chloroplast and the nucleus. *Plant Cell*, 14 Suppl:S327–38, **2002**.

Susek R, Ausubel F & Chory J. Signal transduction mutants of Arabidopsis uncouple nuclear *CAB* and *RBCS* gene expression from chloroplast development. *Cell*, 74:787–99, **1993**.

Sutton M, Smith B, Godoy V & Walker G. The SOS response: recent insights into umuDC-dependent mutagenesis and DNA damage tolerance. *Annu Rev Genet*, 34:479–497, **2000**.

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

Tanaka K, Oikawa K, Ohta N, Kuroiwa H, Kuroiwa T & Takahashi H. Nuclear encoding of a chloroplast RNA polymerase sigma subunit in a red alga. *Science*, 272:1932–5, **1996**.

Tanaka K, Tozawa Y, Mochizuki N, Shinozaki K, Nagatani A, Wakasa K & Takahashi H. Characterization of three cDNA species encoding plastid RNA polymerase sigma factors in *Arabidopsis thaliana*: evidence for the sigma factor heterogeneity in higher plant plastids. *FEBS Lett*, 413:309–13, **1997**.

The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 408:796–815, **2000**.

Thelen J, Mekhedov S & Ohlrogge J. Brassicaceae express multiple isoforms of biotin carboxyl carrier protein in a tissue-specific manner. *Plant Physiol*, 125:2016–28, **2001**.

Thompson J, Higgins D & Gibson T. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, 22:4673–80, **1994**.

Tjus S, Scheller H, Andersson B & Møller B. Active oxygen produced during selective excitation of photosystem I is damaging not only to photosystem I, but also to photosystem II. *Plant Physiol*, 125:2007–15, **2001**.

Tomitani A, Okada K, Miyashita H, Matthijs H, Ohno T & Tanaka A. Chlorophyll b and phycobilins in the common ancestor of cyanobacteria and chloroplasts. *Nature*, 400:159–62, **1999**.

van den Wijngaard P, Demmers J, Thompson S, Wienk H, de Kruijff B & Vredenberg W. Further analysis of the involvement of the envelope anion channel PIRAC in chloroplast protein import. *Eur J Biochem*, 267:3812–7, **2000**.

van Loon AP & Schatz G. Transport of proteins to the mitochondrial intermembrane space: the 'sorting' domain of the cytochrome c1 presequence is a stop-transfer sequence specific for the mitochondrial inner membrane. *Embo J*, 6:2441–8, **1987**.

Varotto C, Pesaresi P, Meurer J, Oelmüller R, Steiner-Lange S, Salamini F & Leister D. Disruption of the Arabidopsis photosystem I gene *psaE1* affects photosynthesis and impairs growth. *Plant J*, 22:115–24, **2000**.

Varotto C, Richly E, Salamini F & Leister D. GST-PRIME: a genome-wide primer design software for the generation of gene sequence tags. *Nucleic Acids Res*, 29:4373–7, **2001**.

Vinti G, Hills A, Campbell S, Bowyer J, Mochizuki N, Chory J & López-Juez E. Interactions between *hy1* and *gun* mutants of Arabidopsis, and their implications for plastid/nuclear signalling. *Plant J*, 24:883–94, **2000**.

von Heijne G. Mitochondrial targeting sequences may form amphiphilic helices. *EMBO J*, 5:1335–42, **1986**.

von Heijne G & Nishikawa K. Chloroplast transit peptides. The perfect random coil? *FEBS Lett*, 278:1–3, **1991**.

von Heijne G, Steppuhn J & Herrmann R. Domain structure of mitochondrial and chloroplast targeting peptides. *Eur J Biochem*, 180:535–45, **1989**.

Vranová E, Atichartpongkul S, Villarroel R, Montagu MV, Inzé D & Camp WV. Comprehensive analysis of gene expression in *Nicotiana tabacum* leaves acclimated to oxidative stress. *Proc Natl Acad Sci U S A*, 99:10870–5, **2002**.

Wang H, Ma L, Habashi J, Li J, Zhao H & Deng X. Analysis of far-red light-regulated genome expression profiles of phytochrome A pathway mutants in Arabidopsis. *Plant J*, 32:723–33, **2002**.

Wang R, Guegler K, LaBrie S & Crawford N. Genomic analysis of a nutrient response in Arabidopsis reveals diverse expression patterns and novel metabolic and potential regulatory genes induced by nitrate. *Plant Cell*, 12:1491–509, **2000**.

Wächtershäuser G. Towards a reconstruction of ancestral genomes by gene cluster alignment. *Syst Appl Microbiol*, 21:473–477, **1998**.

Weber A, Menzlaff E, Arbinger B, Gutensohn M, Eckerskorn C & Flügge U. The 2-oxoglutarate/malate translocator of chloroplast envelope membranes: molecular cloning of a transporter containing a 12-helix motif and expression of the functional protein in yeast cells. *Biochemistry*, 34:2621–7, **1995**.

Weber A, Servaites J, Geiger D, Kofler H, Hille D, Gröner F, Hebbeker U & Flügge U. Identification, purification, and molecular cloning of a putative plastidic glucose translocator. *Plant Cell*, 12:787–802, **2000**.

Weinstein L & Steitz J. Guided tours: from precursor snoRNA to functional snoRNP. *Curr Opin Cell Biol*, 11:378–84, **1999**.

Wetzel C, Jiang C, Meehan L, Voytas D & Rodermel S. Nuclear-organelle interactions: the *immutans* variegation mutant of Arabidopsis is plastid autonomous and impaired in carotenoid biosynthesis. *Plant J*, 6:161–75, **1994**.

Wiedemann N, Kozjak V, Chacinska A, Schönfisch B, Rospert S, Ryan M, Pfanner N & Meisinger C. Machinery for protein sorting and assembly in the mitochondrial outer membrane. *Nature*, 424:565–71, **2003**.

Wienk H, Czisch M & de Kruijff B. The structural flexibility of the preferredoxin transit peptide. *FEBS Lett*, 453:318–26, **1999**.

Wilson K, McManus M, Gordon M & Jordan T. The proteomics of senescence in leaves of white clover, *Trifolium repens* (l.). *Proteomics*, 2:1114–22, **2002**.

Wilson R & Williamson D. Extrachromosomal DNA in the apicomplexa. *Microbiol Mol Biol Rev*, 61:1–16, **1997**.

Woese C. Bacterial evolution. *Microbiol Rev*, 51:221–71, **1987**.

Wolfe K & Shields D. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*, 387:708–13, **1997**.

Wu K, Rooney M & Ferl R. The Arabidopsis 14-3-3 multigene family. *Plant Physiol*, 114:1421–31, **1997**.

Xiong J, Fischer W, Inoue K, Nakahara M & Bauer C. Molecular evidence for the early evolution of photosynthesis. *Science*, 289:1724–30, **2000**.

Xu H, Vavilin D & Vermaas W. Chlorophyll b can serve as the major pigment in functional photosystem II complexes of cyanobacteria. *Proc Natl Acad Sci U S A*, 98:14168–73, **2001**.

Yang Z. Phylogenetic analysis using parsimony and likelihood methods. *J Mol Evol*, 42:294–307, **1996**.

Yukawa Y, Matousek J, Grimm M, Vrba L, Steger G, Sugiura M & Beier H. Plant *7SL RNA* and *tRNA(Tyr)* genes with inserted antisense sequences are efficiently expressed in an in vitro transcription system from *Nicotiana tabacum* cells. *Plant Mol Biol*, 50:713–23, **2002**.

Zerges W. Translation in chloroplasts. *Biochimie*, 82:583–601, **2000**.

Zerges W. Does complexity constrain organelle evolution? *Trends Plant Sci*, 7:175–82, **2002**.

Zhao J & Last R. Immunological characterization and chloroplast localization of the tryptophan biosynthetic enzymes of the flowering plant *Arabidopsis thaliana*. *J Biol Chem*, 270:6081–7, **1995**.

Zik M & Irish V. Global identification of target genes regulated by APETALA3 and PISTILLATA floral homeotic gene action. *Plant Cell*, 15:207–22, **2003**.

A. Abbreviations

aa	amino acid (residues)
ALA	5-aminolevulinic acid
bp	base pair(s)
CAB	chlorophyll a/b binding protein
cDNA	complementary DNA
Col-0	Columbia 0
cTP	chloroplast transit peptide
cue	chlorophyll a/b-binding-protein-underexpressed
dCTP	desoxycytidine-triphosphate
DNA	desoxiribonucleic acid
dNTP	desoxynucleotide-triphosphate
DTT	dithiothreitol
EDTA	ethylenediaminetetra-acetate
EGT	endosymbiotic gene transfer
EST	expressed sequence tag
Fig.	figure
GST	gene sequence tag
gun	genomes uncoupled
h	hour(s)
IE	inner envelope
IM	inner membrane
IMS	inner membrane space
JGI	Joint Genome Institute
laf	long after far red
LHC	light harvesting complex
Mg-proto	Mg-protoporphyrin IX
min	minute(s)
MIPS	Munich Information Center for Protein Sequences
ML	maximum likelihood
mRNA	messenger ribonucleic acid
mtDNA	mitochondrial DNA
mTP	mitochondrial transit peptide
NCBI	National Center for Biotechnology Information
nd	not determined
NJ	neighbor-joining
OE	outer envelope
OM	outer membrane
OMIM	Online Mendelian Inheritance in Man
ORF	open reading frame
PCR	polymerase chain reaction
PPT	phosphoenolpyruvate/phosphate translocator

Protogen	Protoporphyrinogen IX
PS	photosystem
RDBMS	relational database management system
RiceGAAS	Rice Genome Automated Annotation System
RNA	ribonucleic acid
SDS	sodium dodecyl sulfate
snoRNA	small nucleolar RNA
SPP	stromal processing peptidase
SRP	signal recognition particle
Tab.	table
TIC	translocase of the inner membrane of chloroplasts
TIM	translocase of the inner membrane
TIGR	The Institute for Genomic Research
TM	thylakoid membrane
TOC	translocase of the inner membrane of chloroplasts
TOM	translocase of the outer membrane
TRIS	tris(hydroxymethyl)methanamine
WT	wild-type

B. Tables

Distance matrices of the mitochondrial proteomes of ten species

The following tables depicts the distance matrices calculated from the BLAST analysis of the proteins of identified mitochondrial proteomes from 10 species (see section 3.1.3, p. 36 ff) and used for the generation of a similarity tree.

		1	2	3	4	5	6	7	8	9	10
1	*A. gambiae*	0									
2	*A. thaliana*	0.72	0								
3	*E. cuniculi*	0.93	0.92	0							
4	*D. melanogaster*	0.48	0.72	0.90	0						
5	*C. elegans*	0.62	0.76	0.92	0.61	0					
6	*S. pombe*	0.77	0.76	0.85	0.72	0.79	0				
7	*F. rubripes*	0.70	0.83	0.96	0.74	0.76	0.89	0			
8	*H. sapiens*	0.65	0.76	0.94	0.65	0.70	0.82	0.65	0		
9	*P. falciparum*	0.90	0.88	0.89	0.87	0.89	0.79	0.96	0.92	0	
10	*S. cerevisiae*	0.76	0.76	0.86	0.73	0.78	0.51	0.88	0.81	0.83	0

Tab. B.1.: Distance matrix for the mitochondrial proteome of 10 species. The distance matrix based on the fraction of homologous mitochondrion-targeted proteins in each possible pair of species was calculated according to Eq. 2.1, p. 27. Please note that values are given rounded.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
1 atpc	0																																		
2 atpd	0.88	0																																	
3 –CO_2_2d	0.30	0.71	0																																
4 –CO_2_4d	0.61	0.47	0.39	0																															
5 +CO_2_10d	0.48	0.55	0.17	0.26	0																														
6 +CO_2_1d	0.57	0.48	0.18	0.29	0.04	0																													
7 cue1	0.05	0.89	0.19	0.59	0.45	0.49	0																												
8 fluD	0.88	0.12	0.89	0.50	0.68	0.59	0.92	0																											
9 fluL	0.06	0.89	0.20	0.57	0.36	0.40	0.07	0.88	0																										
10 gun1	0.87	0.21	0.75	0.57	0.61	0.52	0.89	0.17	0.78	0																									
11 gun5	0.91	0.13	0.80	0.41	0.53	0.51	0.98	0.11	0.91	0.17	0																								
12 hcf145	0.19	0.72	0.30	0.34	0.45	0.61	0.17	0.77	0.30	0.89	0.80	0																							
13 HL_15	0.45	0.71	0.25	0.49	0.35	0.15	0.30	0.67	0.28	0.47	0.67	0.76	0																						
14 HL_1h	0.95	0.12	0.67	0.44	0.49	0.36	0.92	0.12	0.92	0.14	0.13	0.84	0.42	0																					
15 HLrec_2h	0.85	0.29	0.27	0.31	0.20	0.12	0.73	0.36	0.74	0.40	0.33	0.70	0.21	0.13	0																				
16 HLrec_48h	0.85	0.17	0.60	0.40	0.38	0.30	0.88	0.10	0.88	0.27	0.14	0.83	0.43	0.11	0.03	0																			
17 kn09	0.11	0.83	0.15	0.54	0.25	0.24	0.09	0.88	0.11	0.81	0.87	0.35	0.26	0.80	0.53	0.6	0																		
18 L_30	0.92	0.14	0.80	0.37	0.59	0.53	0.92	0.01	0.91	0.26	0.10	0.67	0.66	0.11	0.27	0.11	0.86	0																	
19 mak3	0.21	0.81	0.13	0.47	0.27	0.34	0.14	0.97	0.17	0.80	0.84	0.23	0.30	0.77	0.50	0.90	0.17	0.88	0																
20 ML_4C	0.21	0.76	0.18	0.48	0.39	0.37	0.08	0.86	0.17	0.84	0.92	0.20	0.30	0.75	0.51	0.81	0.20	0.81	0.11	0															
21 pam46	0.26	0.76	0.14	0.55	0.26	0.20	0.16	0.81	0.16	0.71	0.82	0.43	0.09	0.67	0.41	0.67	0.15	0.80	0.14	0.17	0														
22 pam48	0.99	0.09	0.68	0.38	0.48	0.36	0.95	0.11	0.95	0.14	0.10	0.79	0.55	0.02	0.14	0.12	0.87	0.09	0.79	0.78	0.73	0													
23 +Bro	0.91	0.12	0.74	0.35	0.46	0.37	0.93	0.04	0.65	0.22	0.12	0.76	0.58	0.09	0.17	0.10	0.80	0.05	0.86	0.80	0.75	0.07	0												
24 +CK	0.63	0.34	0.76	0.39	0.59	0.59	0.72	0.22	0.74	0.42	0.29	0.61	0.55	0.38	0.55	0.32	0.70	0.24	0.79	0.70	0.72	0.36	0.34	0											
25 +Fe	0.15	0.81	0.24	0.62	0.47	0.45	0.08	0.84	0.10	0.88	0.92	0.20	0.42	0.84	0.63	0.79	0.23	0.83	0.24	0.10	0.22	0.86	0.90	0.67	0										
26 +Par	0.92	0.15	0.70	0.39	0.56	0.43	0.89	0.03	0.72	0.29	0.19	0.77	0.50	0.09	0.17	0.07	0.79	0.05	0.88	0.71	0.67	0.08	0.03	0.30	0.69	0									
27 ppi1	0.85	0.16	0.79	0.44	0.61	0.52	0.92	0.07	0.87	0.19	0.11	0.76	0.74	0.13	0.25	0.08	0.79	0.11	0.89	0.87	0.79	0.10	0.09	0.31	0.90	0.13	0								
28 +Pro	0.54	0.49	0.20	0.38	0.21	0.15	0.42	0.57	0.51	0.66	0.58	0.39	0.36	0.43	0.16	0.26	0.14	0.45	0.36	0.33	0.31	0.44	0.41	0.64	0.38	0.37	0.50	0							
29 prpl11	0.40	0.53	0.52	0.53	0.62	0.67	0.45	0.61	0.62	0.57	0.45	0.40	0.67	0.69	0.62	0.49	0.55	0.57	0.62	0.67	0.65	0.54	0.60	0.41	0.56	0.74	0.41	0.59	0						
30 psad	0.47	0.39	0.64	0.53	0.68	0.63	0.60	0.47	0.74	0.53	0.37	0.43	0.71	0.51	0.55	0.44	0.57	0.45	0.49	0.56	0.68	0.48	0.68	0.48	0.50	0.62	0.38	0.46	0.20	0					
31 psae	0.04	0.89	0.25	0.60	0.49	0.55	0.06	0.93	0.11	0.89	0.89	0.19	0.46	0.94	0.77	0.87	0.13	0.92	0.11	0.18	0.26	0.96	0.94	0.73	0.10	0.94	0.89	0.46	0.26	0.14	0				
32 psan	0.76	0.28	0.38	0.24	0.15	0.14	0.77	0.37	0.76	0.30	0.25	0.75	0.31	0.19	0.12	0.11	0.47	0.29	0.56	0.67	0.47	0.17	0.25	0.43	0.79	0.27	0.29	0.27	0.65	0.56	0.77	0			
33 psao	0.80	0.26	0.38	0.25	0.21	0.12	0.79	0.32	0.77	0.30	0.25	0.74	0.32	0.15	0.10	0.09	0.53	0.25	0.62	0.64	0.47	0.15	0.25	0.45	0.69	0.18	0.29	0.22	0.58	0.50	0.77	0.02	0		
34 PSII_I	0.54	0.47	0.27	0.46	0.20	0.16	0.55	0.56	0.47	0.42	0.48	0.70	0.18	0.42	0.22	0.26	0.31	0.53	0.36	0.56	0.26	0.43	0.42	0.56	0.65	0.57	0.46	0.31	0.44	0.59	0.54	0.17	0.18	0	
35 sut2	0.12	0.78	0.30	0.66	0.56	0.56	0.09	0.76	0.16	0.83	0.89	0.18	0.56	0.84	0.65	0.71	0.20	0.74	0.23	0.18	0.31	0.89	0.80	0.69	0.11	0.71	0.81	0.29	0.53	0.46	0.14	0.83	0.75	0.62	0

Tab. B.2.: Distance matrix of 'all-against-all' comparisons for 35 conditions. Values were calculated as described in section 2.3.3, p. 30. Please note that values are given rounded.

Distance matrices of transcriptome responses resulting from 35 different conditions

Test set used for the evaluation of mTP predictors

The following table holds the 906 gene identifiers of the proteins that were experimentally confirmed to be targeted to the mitochondrion. This test set was used to analyze the accuracy of the mTP predictors and the combinations thereof.

GI/MIPS	Species	GI/MIPS	Species	GI/MIPS	Species
At1g03860	*A. thaliana*	6912518	*H. sapiens*	6324716	*S. cerevisiae*
At1g06530	*A. thaliana*	6912712	*H. sapiens*	6324724	*S. cerevisiae*
At1g07900	*A. thaliana*	6912714	*H. sapiens*	6324732	*S. cerevisiae*
At1g09210	*A. thaliana*	7019365	*H. sapiens*	6324750	*S. cerevisiae*
At1g13440	*A. thaliana*	7242140	*H. sapiens*	6324761	*S. cerevisiae*
At1g14980	*A. thaliana*	7262378	*H. sapiens*	6324770	*S. cerevisiae*
At1g17350	*A. thaliana*	7657369	*H. sapiens*	6324775	*S. cerevisiae*
At1g24180	*A. thaliana*	7657581	*H. sapiens*	6324789	*S. cerevisiae*
At1g27390	*A. thaliana*	7657585	*H. sapiens*	6324790	*S. cerevisiae*
At1g32580	*A. thaliana*	7661730	*H. sapiens*	6324796	*S. cerevisiae*
At1g47260	*A. thaliana*	7705626	*H. sapiens*	6324800	*S. cerevisiae*
At1g47420	*A. thaliana*	7705630	*H. sapiens*	6324806	*S. cerevisiae*
At1g48030	*A. thaliana*	7705927	*H. sapiens*	6324815	*S. cerevisiae*
At1g50400	*A. thaliana*	7706391	*H. sapiens*	6324820	*S. cerevisiae*
At1g50940	*A. thaliana*	8922511	*H. sapiens*	6324822	*S. cerevisiae*
At1g51980	*A. thaliana*	9845267	*H. sapiens*	6324825	*S. cerevisiae*
At1g53240	*A. thaliana*	9845297	*H. sapiens*	6324840	*S. cerevisiae*
At1g56340	*A. thaliana*	9910372	*H. sapiens*	6324845	*S. cerevisiae*
At1g59900	*A. thaliana*	10835025	*H. sapiens*	6324848	*S. cerevisiae*
At1g79230	*A. thaliana*	10863931	*H. sapiens*	6324871	*S. cerevisiae*
At1g79010	*A. thaliana*	10947137	*H. sapiens*	6324909	*S. cerevisiae*
At1g79440	*A. thaliana*	11024700	*H. sapiens*	6324932	*S. cerevisiae*
At1g80230	*A. thaliana*	11128019	*H. sapiens*	6324950	*S. cerevisiae*
At2g04030	*A. thaliana*	11321581	*H. sapiens*	6324984	*S. cerevisiae*
At2g05710	*A. thaliana*	11321583	*H. sapiens*	6324985	*S. cerevisiae*
At2g14200	*A. thaliana*	11545827	*H. sapiens*	6324993	*S. cerevisiae*
At2g20420	*A. thaliana*	13435350	*H. sapiens*	6325004	*S. cerevisiae*
At2g20530	*A. thaliana*	13435356	*H. sapiens*	6325020	*S. cerevisiae*
At2g21640	*A. thaliana*	13631037	*H. sapiens*	6325032	*S. cerevisiae*
At2g21870	*A. thaliana*	13639114	*H. sapiens*	6325038	*S. cerevisiae*
At2g27730	*A. thaliana*	13641644	*H. sapiens*	6325041	*S. cerevisiae*
At2g28000	*A. thaliana*	13647558	*H. sapiens*	6325058	*S. cerevisiae*
At2g33040	*A. thaliana*	13699868	*H. sapiens*	6325068	*S. cerevisiae*
At2g33150	*A. thaliana*	14725900	*H. sapiens*	6325084	*S. cerevisiae*
At2g33210	*A. thaliana*	14763423	*H. sapiens*	6325085	*S. cerevisiae*
At2g35240	*A. thaliana*	14765959	*H. sapiens*	6325089	*S. cerevisiae*
At2g41380	*A. thaliana*	14786823	*H. sapiens*	6325092	*S. cerevisiae*
At2g44350	*A. thaliana*	14790138	*H. sapiens*	6325098	*S. cerevisiae*
At2g47510	*A. thaliana*	15321446	*H. sapiens*	6325109	*S. cerevisiae*
At3g01280	*A. thaliana*	15718706	*H. sapiens*	6325122	*S. cerevisiae*

continued from last page

GI/MIPS	Species	GI/MIPS	Species	GI/MIPS	Species
At3g02090	*A. thaliana*	16157047	*H. sapiens*	6325123	*S. cerevisiae*
At3g03100	*A. thaliana*	16554611	*H. sapiens*	6325125	*S. cerevisiae*
At3g04120	*A. thaliana*	16950591	*H. sapiens*	6325139	*S. cerevisiae*
At3g06050	*A. thaliana*	16950593	*H. sapiens*	6325149	*S. cerevisiae*
At3g07480	*A. thaliana*	17437217	*H. sapiens*	6325154	*S. cerevisiae*
At3g07770	*A. thaliana*	17441315	*H. sapiens*	6325158	*S. cerevisiae*
At3g08580	*A. thaliana*	18426967	*H. sapiens*	6325160	*S. cerevisiae*
At3g10920	*A. thaliana*	18559169	*H. sapiens*	6325179	*S. cerevisiae*
At3g12260	*A. thaliana*	18644883	*H. sapiens*	6325184	*S. cerevisiae*
At3g15000	*A. thaliana*	19743875	*H. sapiens*	6325187	*S. cerevisiae*
At3g15020	*A. thaliana*	19923315	*H. sapiens*	6325197	*S. cerevisiae*
At3g16480	*A. thaliana*	20127406	*H. sapiens*	6325198	*S. cerevisiae*
At3g17240	*A. thaliana*	20336214	*H. sapiens*	6325228	*S. cerevisiae*
At3g20000	*A. thaliana*	20336335	*H. sapiens*	6325253	*S. cerevisiae*
At3g22200	*A. thaliana*	20482186	*H. sapiens*	6325257	*S. cerevisiae*
At3g22330	*A. thaliana*	20554413	*H. sapiens*	6325258	*S. cerevisiae*
At3g22370	*A. thaliana*	20554449	*H. sapiens*	6325263	*S. cerevisiae*
At3g23990	*A. thaliana*	20560255	*H. sapiens*	6325277	*S. cerevisiae*
At3g27070	*A. thaliana*	20562312	*H. sapiens*	6325281	*S. cerevisiae*
At3g27080	*A. thaliana*	21314627	*H. sapiens*	6325290	*S. cerevisiae*
At3g27240	*A. thaliana*	21359837	*H. sapiens*	6325304	*S. cerevisiae*
At3g27280	*A. thaliana*	21359867	*H. sapiens*	6325315	*S. cerevisiae*
At3g45300	*A. thaliana*	21361103	*H. sapiens*	6325318	*S. cerevisiae*
At3g48000	*A. thaliana*	21361114	*H. sapiens*	6325324	*S. cerevisiae*
At3g48680	*A. thaliana*	21361280	*H. sapiens*	6325357	*S. cerevisiae*
At3g52200	*A. thaliana*	21361324	*H. sapiens*	6325373	*S. cerevisiae*
At3g52300	*A. thaliana*	21361331	*H. sapiens*	6325377	*S. cerevisiae*
At3g54110	*A. thaliana*	21361565	*H. sapiens*	6325382	*S. cerevisiae*
At3g59760	*A. thaliana*	21389315	*H. sapiens*	6325392	*S. cerevisiae*
At3g61440	*A. thaliana*	21396489	*H. sapiens*	7839207	*S. cerevisiae*
At4g01100	*A. thaliana*	21536274	*H. sapiens*	6325398	*S. cerevisiae*
At4g02580	*A. thaliana*	22024385	*H. sapiens*	6325424	*S. cerevisiae*
At4g02930	*A. thaliana*	22041648	*H. sapiens*	6325449	*S. cerevisiae*
At4g11010	*A. thaliana*	22057417	*H. sapiens*	6319370	*S. cerevisiae*
At4g11120	*A. thaliana*	22058461	*H. sapiens*	6319374	*S. cerevisiae*
At4g13850	*A. thaliana*	23507957	*P. falciparum*	6319381	*S. cerevisiae*
At4g15940	*A. thaliana*	23508137	*P. falciparum*	6319407	*S. cerevisiae*
At4g23900	*A. thaliana*	23508719	*P. falciparum*	6319425	*S. cerevisiae*
At4g26970	*A. thaliana*	23508823	*P. falciparum*	6319426	*S. cerevisiae*
At4g27585	*A. thaliana*	23509134	*P. falciparum*	6319433	*S. cerevisiae*
At4g28060	*A. thaliana*	23612115	*P. falciparum*	6319441	*S. cerevisiae*
At4g28510	*A. thaliana*	19112859	*S. pombe*	6319449	*S. cerevisiae*
At4g31500	*A. thaliana*	19113244	*S. pombe*	6319456	*S. cerevisiae*
At4g31810	*A. thaliana*	19114063	*S. pombe*	6319458	*S. cerevisiae*
At4g32470	*A. thaliana*	19114131	*S. pombe*	6319471	*S. cerevisiae*

continued from last page

GI/MIPS	Species	GI/MIPS	Species	GI/MIPS	Species
At4g35260	*A. thaliana*	19115272	*S. pombe*	6319475	*S. cerevisiae*
At4g35830	*A. thaliana*	19115434	*S. pombe*	6319481	*S. cerevisiae*
At4g35850	*A. thaliana*	19115831	*S. pombe*	6319488	*S. cerevisiae*
At4g37910	*A. thaliana*	6319272	*S. cerevisiae*	6319498	*S. cerevisiae*
At5g07440	*A. thaliana*	6319278	*S. cerevisiae*	6319500	*S. cerevisiae*
At5g08300	*A. thaliana*	6319304	*S. cerevisiae*	6319503	*S. cerevisiae*
At5g08670	*A. thaliana*	6319337	*S. cerevisiae*	6319511	*S. cerevisiae*
At5g08690	*A. thaliana*	6322252	*S. cerevisiae*	6319513	*S. cerevisiae*
At5g09590	*A. thaliana*	6322253	*S. cerevisiae*	6319518	*S. cerevisiae*
At5g10860	*A. thaliana*	6322270	*S. cerevisiae*	6319558	*S. cerevisiae*
At5g13430	*A. thaliana*	6322281	*S. cerevisiae*	6319560	*S. cerevisiae*
At5g13440	*A. thaliana*	6322295	*S. cerevisiae*	6319567	*S. cerevisiae*
At5g13450	*A. thaliana*	6322314	*S. cerevisiae*	6319596	*S. cerevisiae*
At5g13490	*A. thaliana*	6322318	*S. cerevisiae*	6319598	*S. cerevisiae*
At5g14040	*A. thaliana*	6322328	*S. cerevisiae*	6319599	*S. cerevisiae*
At5g15090	*A. thaliana*	6322330	*S. cerevisiae*	6319604	*S. cerevisiae*
At5g18170	*A. thaliana*	6322349	*S. cerevisiae*	6319622	*S. cerevisiae*
At5g19760	*A. thaliana*	6322357	*S. cerevisiae*	6319628	*S. cerevisiae*
At5g20080	*A. thaliana*	6322365	*S. cerevisiae*	6319662	*S. cerevisiae*
At5g23140	*A. thaliana*	6322390	*S. cerevisiae*	6319669	*S. cerevisiae*
At5g23250	*A. thaliana*	6322395	*S. cerevisiae*	6319698	*S. cerevisiae*
At5g37510	*A. thaliana*	6322398	*S. cerevisiae*	6319704	*S. cerevisiae*
At5g40770	*A. thaliana*	6322406	*S. cerevisiae*	6319707	*S. cerevisiae*
At5g40810	*A. thaliana*	6322407	*S. cerevisiae*	6319728	*S. cerevisiae*
At5g40930	*A. thaliana*	6322409	*S. cerevisiae*	6319740	*S. cerevisiae*
At5g47030	*A. thaliana*	6322476	*S. cerevisiae*	6319739	*S. cerevisiae*
At5g50850	*A. thaliana*	6322480	*S. cerevisiae*	6319745	*S. cerevisiae*
At5g52840	*A. thaliana*	6322494	*S. cerevisiae*	6319759	*S. cerevisiae*
At5g54100	*A. thaliana*	6322505	*S. cerevisiae*	10383751	*S. cerevisiae*
At5g55070	*A. thaliana*	6322508	*S. cerevisiae*	10383753	*S. cerevisiae*
At5g57300	*A. thaliana*	6322537	*S. cerevisiae*	6319793	*S. cerevisiae*
At5g59880	*A. thaliana*	6322555	*S. cerevisiae*	6319805	*S. cerevisiae*
At5g60640	*A. thaliana*	6322560	*S. cerevisiae*	10383773	*S. cerevisiae*
At5g62530	*A. thaliana*	6322561	*S. cerevisiae*	6319837	*S. cerevisiae*
At5g63400	*A. thaliana*	6322564	*S. cerevisiae*	6319848	*S. cerevisiae*
At5g63510	*A. thaliana*	6322574	*S. cerevisiae*	6319872	*S. cerevisiae*
At5g66510	*A. thaliana*	6322581	*S. cerevisiae*	10383791	*S. cerevisiae*
At5g66760	*A. thaliana*	7839183	*S. cerevisiae*	6319894	*S. cerevisiae*
At5g67500	*A. thaliana*	6322604	*S. cerevisiae*	6319925	*S. cerevisiae*
10727067	*D. melanogaster*	6322653	*S. cerevisiae*	6319984	*S. cerevisiae*
7291797	*D. melanogaster*	6322654	*S. cerevisiae*	6319999	*S. cerevisiae*
7292404	*D. melanogaster*	6322656	*S. cerevisiae*	6320003	*S. cerevisiae*
7292599	*D. melanogaster*	6322678	*S. cerevisiae*	6320020	*S. cerevisiae*
7294849	*D. melanogaster*	6322681	*S. cerevisiae*	6320023	*S. cerevisiae*
7293902	*D. melanogaster*	6322694	*S. cerevisiae*	6320027	*S. cerevisiae*

continued from last page

GI/MIPS	Species	GI/MIPS	Species	GI/MIPS	Species
7296255	*D. melanogaster*	6322699	*S. cerevisiae*	6320044	*S. cerevisiae*
7297280	*D. melanogaster*	6322701	*S. cerevisiae*	6320059	*S. cerevisiae*
7298099	*D. melanogaster*	6322708	*S. cerevisiae*	6320072	*S. cerevisiae*
7298102	*D. melanogaster*	6322711	*S. cerevisiae*	6320083	*S. cerevisiae*
7298325	*D. melanogaster*	6322715	*S. cerevisiae*	6320096	*S. cerevisiae*
7298326	*D. melanogaster*	6322729	*S. cerevisiae*	6320099	*S. cerevisiae*
7298546	*D. melanogaster*	6322743	*S. cerevisiae*	6320118	*S. cerevisiae*
7300119	*D. melanogaster*	6322765	*S. cerevisiae*	6320135	*S. cerevisiae*
7300630	*D. melanogaster*	6322795	*S. cerevisiae*	6320136	*S. cerevisiae*
10726716	*D. melanogaster*	6322811	*S. cerevisiae*	6320137	*S. cerevisiae*
7302274	*D. melanogaster*	6322823	*S. cerevisiae*	6320158	*S. cerevisiae*
17506561	*C. elegans*	6322836	*S. cerevisiae*	6320160	*S. cerevisiae*
17506737	*C. elegans*	6322841	*S. cerevisiae*	6320200	*S. cerevisiae*
17507307	*C. elegans*	6322847	*S. cerevisiae*	6320211	*S. cerevisiae*
17508139	*C. elegans*	6322850	*S. cerevisiae*	6320222	*S. cerevisiae*
17509685	*C. elegans*	6322858	*S. cerevisiae*	6320235	*S. cerevisiae*
17533037	*C. elegans*	6322868	*S. cerevisiae*	6320246	*S. cerevisiae*
17533915	*C. elegans*	6322905	*S. cerevisiae*	6320269	*S. cerevisiae*
17534033	*C. elegans*	6322918	*S. cerevisiae*	6320284	*S. cerevisiae*
17534525	*C. elegans*	6322919	*S. cerevisiae*	6320325	*S. cerevisiae*
17535649	*C. elegans*	6322923	*S. cerevisiae*	6320352	*S. cerevisiae*
17535897	*C. elegans*	6322938	*S. cerevisiae*	6320354	*S. cerevisiae*
17536047	*C. elegans*	6322987	*S. cerevisiae*	6320380	*S. cerevisiae*
17536071	*C. elegans*	6323001	*S. cerevisiae*	6320383	*S. cerevisiae*
17552212	*C. elegans*	6323020	*S. cerevisiae*	6320400	*S. cerevisiae*
17552308	*C. elegans*	6323028	*S. cerevisiae*	6320403	*S. cerevisiae*
17552470	*C. elegans*	6323067	*S. cerevisiae*	6320410	*S. cerevisiae*
17553560	*C. elegans*	6323083	*S. cerevisiae*	6320432	*S. cerevisiae*
17553574	*C. elegans*	6323088	*S. cerevisiae*	6320437	*S. cerevisiae*
17553678	*C. elegans*	6323096	*S. cerevisiae*	6320438	*S. cerevisiae*
17554166	*C. elegans*	6323106	*S. cerevisiae*	6320442	*S. cerevisiae*
17554318	*C. elegans*	6323111	*S. cerevisiae*	6320443	*S. cerevisiae*
17554946	*C. elegans*	6323118	*S. cerevisiae*	6320464	*S. cerevisiae*
17555174	*C. elegans*	6323168	*S. cerevisiae*	6320469	*S. cerevisiae*
17555558	*C. elegans*	6323171	*S. cerevisiae*	6320474	*S. cerevisiae*
17556823	*C. elegans*	6323192	*S. cerevisiae*	6320502	*S. cerevisiae*
17538940	*C. elegans*	6323193	*S. cerevisiae*	6320504	*S. cerevisiae*
17539378	*C. elegans*	6323217	*S. cerevisiae*	6320528	*S. cerevisiae*
17539560	*C. elegans*	6323230	*S. cerevisiae*	6320529	*S. cerevisiae*
17539604	*C. elegans*	6323232	*S. cerevisiae*	6320544	*S. cerevisiae*
17539650	*C. elegans*	6323247	*S. cerevisiae*	6320554	*S. cerevisiae*
17541224	*C. elegans*	6323288	*S. cerevisiae*	6320583	*S. cerevisiae*
17562024	*C. elegans*	6323312	*S. cerevisiae*	6320584	*S. cerevisiae*
17549919	*C. elegans*	6323319	*S. cerevisiae*	6320585	*S. cerevisiae*
17550100	*C. elegans*	6323321	*S. cerevisiae*	6320599	*S. cerevisiae*

continued from last page

GI/MIPS	Species	GI/MIPS	Species	GI/MIPS	Species
17550194	*C. elegans*	6323326	*S. cerevisiae*	6320613	*S. cerevisiae*
17551134	*C. elegans*	6323335	*S. cerevisiae*	6320670	*S. cerevisiae*
17567829	*C. elegans*	6323344	*S. cerevisiae*	6320701	*S. cerevisiae*
17569045	*C. elegans*	6323387	*S. cerevisiae*	6320719	*S. cerevisiae*
17569479	*C. elegans*	6323401	*S. cerevisiae*	6320738	*S. cerevisiae*
4501859	*H. sapiens*	6323414	*S. cerevisiae*	7839166	*S. cerevisiae*
4501867	*H. sapiens*	6323425	*S. cerevisiae*	6320785	*S. cerevisiae*
4502037	*H. sapiens*	6323427	*S. cerevisiae*	6320796	*S. cerevisiae*
4502083	*H. sapiens*	6323458	*S. cerevisiae*	6320805	*S. cerevisiae*
4502215	*H. sapiens*	6323472	*S. cerevisiae*	6320811	*S. cerevisiae*
4502295	*H. sapiens*	6323483	*S. cerevisiae*	6320815	*S. cerevisiae*
4502297	*H. sapiens*	6323506	*S. cerevisiae*	6320840	*S. cerevisiae*
4502301	*H. sapiens*	6323507	*S. cerevisiae*	6320851	*S. cerevisiae*
4502303	*H. sapiens*	6323509	*S. cerevisiae*	6320854	*S. cerevisiae*
4502375	*H. sapiens*	6323515	*S. cerevisiae*	6320889	*S. cerevisiae*
4502855	*H. sapiens*	6323526	*S. cerevisiae*	6320891	*S. cerevisiae*
4502987	*H. sapiens*	6323548	*S. cerevisiae*	6320894	*S. cerevisiae*
4503021	*H. sapiens*	6323553	*S. cerevisiae*	6320900	*S. cerevisiae*
4503265	*H. sapiens*	6323558	*S. cerevisiae*	6320904	*S. cerevisiae*
4503301	*H. sapiens*	6323562	*S. cerevisiae*	6320913	*S. cerevisiae*
4503423	*H. sapiens*	6323565	*S. cerevisiae*	6320924	*S. cerevisiae*
4503607	*H. sapiens*	6323579	*S. cerevisiae*	6320989	*S. cerevisiae*
4503609	*H. sapiens*	6323580	*S. cerevisiae*	6321001	*S. cerevisiae*
4503785	*H. sapiens*	6323587	*S. cerevisiae*	6321002	*S. cerevisiae*
4503943	*H. sapiens*	6323599	*S. cerevisiae*	6321018	*S. cerevisiae*
4504069	*H. sapiens*	6323611	*S. cerevisiae*	6321026	*S. cerevisiae*
4504107	*H. sapiens*	6323616	*S. cerevisiae*	6321030	*S. cerevisiae*
4504133	*H. sapiens*	6323620	*S. cerevisiae*	14318482	*S. cerevisiae*
4504327	*H. sapiens*	6323634	*S. cerevisiae*	14318489	*S. cerevisiae*
4504575	*H. sapiens*	6323659	*S. cerevisiae*	16740526	*S. cerevisiae*
4505277	*H. sapiens*	6323665	*S. cerevisiae*	14318501	*S. cerevisiae*
4505355	*H. sapiens*	6323666	*S. cerevisiae*	14318504	*S. cerevisiae*
4505357	*H. sapiens*	6323678	*S. cerevisiae*	14318533	*S. cerevisiae*
4505359	*H. sapiens*	6323684	*S. cerevisiae*	14318558	*S. cerevisiae*
4505361	*H. sapiens*	6323701	*S. cerevisiae*	14318570	*S. cerevisiae*
4505363	*H. sapiens*	6323705	*S. cerevisiae*	14318574	*S. cerevisiae*
4505365	*H. sapiens*	6323711	*S. cerevisiae*	6321179	*S. cerevisiae*
4505367	*H. sapiens*	6323717	*S. cerevisiae*	6321190	*S. cerevisiae*
4505369	*H. sapiens*	6323729	*S. cerevisiae*	6321202	*S. cerevisiae*
4505371	*H. sapiens*	6323736	*S. cerevisiae*	6321212	*S. cerevisiae*
4505685	*H. sapiens*	6323745	*S. cerevisiae*	6321247	*S. cerevisiae*
4505689	*H. sapiens*	6323755	*S. cerevisiae*	6321251	*S. cerevisiae*
4505693	*H. sapiens*	6323757	*S. cerevisiae*	6321295	*S. cerevisiae*
4505699	*H. sapiens*	6323778	*S. cerevisiae*	6321302	*S. cerevisiae*
4505937	*H. sapiens*	6323794	*S. cerevisiae*	6321309	*S. cerevisiae*

continued from last page

GI/MIPS	Species	GI/MIPS	Species	GI/MIPS	Species
4506001	*H. sapiens*	6323800	*S. cerevisiae*	6321313	*S. cerevisiae*
4506673	*H. sapiens*	6323801	*S. cerevisiae*	6321319	*S. cerevisiae*
4506865	*H. sapiens*	6323807	*S. cerevisiae*	6321362	*S. cerevisiae*
4507009	*H. sapiens*	6323808	*S. cerevisiae*	6321374	*S. cerevisiae*
4507173	*H. sapiens*	6323810	*S. cerevisiae*	6321420	*S. cerevisiae*
4507315	*H. sapiens*	6323817	*S. cerevisiae*	6321440	*S. cerevisiae*
4507319	*H. sapiens*	6323831	*S. cerevisiae*	6321445	*S. cerevisiae*
4507401	*H. sapiens*	6323843	*S. cerevisiae*	6321452	*S. cerevisiae*
4507807	*H. sapiens*	6323847	*S. cerevisiae*	7839171	*S. cerevisiae*
4507841	*H. sapiens*	6323859	*S. cerevisiae*	7839170	*S. cerevisiae*
4507879	*H. sapiens*	6323881	*S. cerevisiae*	6681846	*S. cerevisiae*
4557044	*H. sapiens*	6323884	*S. cerevisiae*	6321499	*S. cerevisiae*
4557231	*H. sapiens*	6323897	*S. cerevisiae*	6321513	*S. cerevisiae*
4557235	*H. sapiens*	6323898	*S. cerevisiae*	6321521	*S. cerevisiae*
4557237	*H. sapiens*	6323902	*S. cerevisiae*	6321523	*S. cerevisiae*
4557353	*H. sapiens*	6323908	*S. cerevisiae*	6321524	*S. cerevisiae*
4557355	*H. sapiens*	6323912	*S. cerevisiae*	6321538	*S. cerevisiae*
4557361	*H. sapiens*	6323913	*S. cerevisiae*	6321550	*S. cerevisiae*
4557525	*H. sapiens*	6323923	*S. cerevisiae*	6321571	*S. cerevisiae*
4557767	*H. sapiens*	6323938	*S. cerevisiae*	6321608	*S. cerevisiae*
4757732	*H. sapiens*	6323942	*S. cerevisiae*	6321610	*S. cerevisiae*
4757852	*H. sapiens*	6323943	*S. cerevisiae*	6321613	*S. cerevisiae*
4757860	*H. sapiens*	6323950	*S. cerevisiae*	6321620	*S. cerevisiae*
4758034	*H. sapiens*	6323959	*S. cerevisiae*	6321622	*S. cerevisiae*
4758050	*H. sapiens*	6323960	*S. cerevisiae*	6321631	*S. cerevisiae*
4758072	*H. sapiens*	6324001	*S. cerevisiae*	6321632	*S. cerevisiae*
4758076	*H. sapiens*	6324014	*S. cerevisiae*	6321654	*S. cerevisiae*
4758424	*H. sapiens*	6324023	*S. cerevisiae*	6321659	*S. cerevisiae*
4758744	*H. sapiens*	6324045	*S. cerevisiae*	6321661	*S. cerevisiae*
4758768	*H. sapiens*	6324054	*S. cerevisiae*	6321670	*S. cerevisiae*
4758770	*H. sapiens*	6324069	*S. cerevisiae*	6321673	*S. cerevisiae*
4758772	*H. sapiens*	6324073	*S. cerevisiae*	6321674	*S. cerevisiae*
4758774	*H. sapiens*	6324077	*S. cerevisiae*	6321683	*S. cerevisiae*
4758776	*H. sapiens*	6324116	*S. cerevisiae*	6321696	*S. cerevisiae*
4758778	*H. sapiens*	6324118	*S. cerevisiae*	6321716	*S. cerevisiae*
4758786	*H. sapiens*	6324144	*S. cerevisiae*	6321725	*S. cerevisiae*
4758788	*H. sapiens*	6324152	*S. cerevisiae*	6321749	*S. cerevisiae*
4758790	*H. sapiens*	6324160	*S. cerevisiae*	6321753	*S. cerevisiae*
4758792	*H. sapiens*	6324192	*S. cerevisiae*	6321766	*S. cerevisiae*
4758886	*H. sapiens*	6324196	*S. cerevisiae*	6321782	*S. cerevisiae*
4758894	*H. sapiens*	6324198	*S. cerevisiae*	6321787	*S. cerevisiae*
4758956	*H. sapiens*	6324204	*S. cerevisiae*	6321788	*S. cerevisiae*
4759068	*H. sapiens*	6324225	*S. cerevisiae*	6321789	*S. cerevisiae*
4759286	*H. sapiens*	6324229	*S. cerevisiae*	6321790	*S. cerevisiae*
4826848	*H. sapiens*	6324246	*S. cerevisiae*	6321793	*S. cerevisiae*

continued from last page

GI/MIPS	Species	GI/MIPS	Species	GI/MIPS	Species
4826850	*H. sapiens*	6324258	*S. cerevisiae*	6321796	*S. cerevisiae*
4826852	*H. sapiens*	6324259	*S. cerevisiae*	6321798	*S. cerevisiae*
4826854	*H. sapiens*	6324260	*S. cerevisiae*	6321805	*S. cerevisiae*
4826856	*H. sapiens*	6324263	*S. cerevisiae*	6321813	*S. cerevisiae*
4826862	*H. sapiens*	6324266	*S. cerevisiae*	6321826	*S. cerevisiae*
4826880	*H. sapiens*	6324273	*S. cerevisiae*	6321827	*S. cerevisiae*
4826992	*H. sapiens*	7839195	*S. cerevisiae*	6321842	*S. cerevisiae*
4885079	*H. sapiens*	6324276	*S. cerevisiae*	6321858	*S. cerevisiae*
4885081	*H. sapiens*	6324291	*S. cerevisiae*	6321867	*S. cerevisiae*
4885083	*H. sapiens*	6324302	*S. cerevisiae*	6321875	*S. cerevisiae*
4885387	*H. sapiens*	6324323	*S. cerevisiae*	6321892	*S. cerevisiae*
4885401	*H. sapiens*	6324328	*S. cerevisiae*	6321898	*S. cerevisiae*
4885545	*H. sapiens*	6324344	*S. cerevisiae*	6321912	*S. cerevisiae*
5031709	*H. sapiens*	6324365	*S. cerevisiae*	6321941	*S. cerevisiae*
5031751	*H. sapiens*	6324370	*S. cerevisiae*	6321993	*S. cerevisiae*
5031787	*H. sapiens*	6324373	*S. cerevisiae*	6322002	*S. cerevisiae*
5032181	*H. sapiens*	6324394	*S. cerevisiae*	6322034	*S. cerevisiae*
5174567	*H. sapiens*	6324452	*S. cerevisiae*	6322036	*S. cerevisiae*
5174615	*H. sapiens*	6324458	*S. cerevisiae*	6322050	*S. cerevisiae*
5453902	*H. sapiens*	6324476	*S. cerevisiae*	6322055	*S. cerevisiae*
5454070	*H. sapiens*	6324477	*S. cerevisiae*	6322059	*S. cerevisiae*
5454120	*H. sapiens*	6324495	*S. cerevisiae*	6322066	*S. cerevisiae*
5454122	*H. sapiens*	6324525	*S. cerevisiae*	6322077	*S. cerevisiae*
5454124	*H. sapiens*	6324540	*S. cerevisiae*	6322080	*S. cerevisiae*
5454152	*H. sapiens*	6324546	*S. cerevisiae*	6322093	*S. cerevisiae*
5729935	*H. sapiens*	6324550	*S. cerevisiae*	6322097	*S. cerevisiae*
5802970	*H. sapiens*	6324591	*S. cerevisiae*	6322098	*S. cerevisiae*
5803217	*H. sapiens*	6324594	*S. cerevisiae*	6322104	*S. cerevisiae*
5901896	*H. sapiens*	6324611	*S. cerevisiae*	6322105	*S. cerevisiae*
5901982	*H. sapiens*	6324614	*S. cerevisiae*	6322113	*S. cerevisiae*
5902010	*H. sapiens*	6324639	*S. cerevisiae*	6322119	*S. cerevisiae*
6005723	*H. sapiens*	6324674	*S. cerevisiae*	6322124	*S. cerevisiae*
6005804	*H. sapiens*	6324688	*S. cerevisiae*	6322138	*S. cerevisiae*
6031192	*H. sapiens*	6324695	*S. cerevisiae*	6322167	*S. cerevisiae*
6041669	*H. sapiens*	6324699	*S. cerevisiae*	6322229	*S. cerevisiae*
6274550	*H. sapiens*	6324704	*S. cerevisiae*	6322231	*S. cerevisiae*
6912238	*H. sapiens*	6324709	*S. cerevisiae*	6322233	*S. cerevisiae*

Tab. B.4.: Test set for the evaluation of mTP predictors.

Test set used for the evaluation of cTP predictors

The following table holds the 2 450 gene identifiers of the proteins whose localization was experimentally confirmed. This test set was used to analyze the accuracy of the cTP predictors and the combinations thereof. **Proof** refers to the evidence for the localization of a given protein: TAS = tracable author statement, IDA = inferred from direct assay. **PMID** refers to the PubmedID under which a given publication is available

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
132161	Chloroplast	*T. repens*	Chloroplast	IDA	Wilson *et al.* (2002)	12362330
1806283	Chloroplast	*S. rostrata*	Lumen/Peripheral	IDA	Peltier *et al.* (2000)	10715320
2290402	Chloroplast	*H. annuus*	Lumen/Peripheral	IDA	Peltier *et al.* (2000)	10715320
2330647	Chloroplast	*P. sativum*	Lumen/Peripheral	IDA	Peltier *et al.* (2000)	10715320
4090533	Chloroplast	*C. papaya*	Lumen/Peripheral	IDA	Peltier *et al.* (2000)	10715320
4105180	Chloroplast	*P. sativum*	Lumen/Peripheral	IDA	Peltier *et al.* (2000)	10715320
4731322	Chloroplast	*M. sativa*	Chloroplast	IDA	Wilson *et al.* (2002)	12362330
At1g02560	Chloroplast	*A. thaliana*	Stroma/Thylakoid	IDA	Peltier *et al.* (2001)	11278690
At1g03130	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At1g03600	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Peltier *et al.* (2002)	11826309
At1g03680	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Peltier *et al.* (2002)	11826309
At1g06680	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Peltier *et al.* (2002)	11826309
At1g07320	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At1g07780	Chloroplast	*A. thaliana*	Chloroplast	IDA	Zhao & Last (1995)	7890741
At1g08540	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	
At1g08550	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At1g09130	Chloroplast	*A. thaliana*	Stroma	IDA	Peltier *et al.* (2001)	11278690
At1g11750	Chloroplast	*A. thaliana*	Stroma	IDA	Peltier *et al.* (2001)	11278690
At1g12410	Chloroplast	*A. thaliana*	Stroma/Thylakoid	IDA	Peltier *et al.* (2001)	11278690
At1g14150	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At1g15700	Chloroplast	*A. thaliana*	ATP-synthase complex	TAS	GeneOntology	
At1g17745	Chloroplast	*A. thaliana*	Chloroplast	IDA	Ho *et al.* (1999)	9867856
At1g18640	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	
At1g20340	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At1g20810	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At1g22300	Chloroplast	*A. thaliana*	Stroma	IDA	Wu *et al.* (1997)	9276953
At1g22940	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	
At1g27770	Chloroplast	*A. thaliana*	Chlp inner membrane	IDA	GeneOntology	
At1g29410	Chloroplast	*A. thaliana*	Chloroplast	IDA	Zhao & Last (1995)	7890741
At1g48860	Chloroplast	*A. thaliana*	Chloroplast	TAS	Klee *et al.* (1987)	3481024
At1g49970	Chloroplast	*A. thaliana*	Stroma/Thylakoid	IDA	Peltier *et al.* (2001)	11278690
At1g50250	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Chen *et al.* (2000)	10849347
At1g54780	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At1g64860	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	
At1g66670	Chloroplast	*A. thaliana*	Stroma	IDA	Peltier *et al.* (2001)	11278690
At1g67090	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At1g67280	Chloroplast	*A. thaliana*	Chloroplast	IDA	Schubert *et al.* (2002)	11719511
At1g67740	Chloroplast	*A. thaliana*	Thylakoid	IDA	GeneOntology	
At1g73990	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	GeneOntology	
At1g76100	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At1g76450	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At1g77090	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At1g77750	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	
At1g78630	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At2g05990	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At2g21330	Chloroplast	*A. thaliana*	Chloroplast	IDA	Schubert *et al.* (2002)	11719511
At2g21970	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	GeneOntology	
At2g22330	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	
At2g26670	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	
At2g28000	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Peltier *et al.* (2002)	11826309
At2g28190	Chloroplast	*A. thaliana*	Chloroplast	IDA	Kliebenstein *et al.* (1998)	9765550
At2g28800	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Moore *et al.* (2000)	10636840
At2g29630	Chloroplast	*A. thaliana*	Stroma	IDA	Balmer *et al.* (2003)	12509500
At2g29690	Chloroplast	*A. thaliana*	Chloroplast	TAS	Niyogi & Fink (1992)	1392592
At2g30790	Chloroplast	*A. thaliana*	Thylakoid lumen	TAS	Schubert *et al.* (2002)	11719511
At2g30950	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Lindahl *et al.* (1996)	8910594
At2g33800	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At2g35410	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At2g35490	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At2g36250	Chloroplast	*A. thaliana*	Stroma	IDA	GeneOntology	
At2g36990	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	
At2g37220	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At2g39470	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At2g39730	Chloroplast	*A. thaliana*	Chloroplast	IDA	Wilson *et al.* (2002)	12362330
At2g42130	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At2g42540	Chloroplast	*A. thaliana*	Stroma	IDA	GeneOntology	
At2g42590	Chloroplast	*A. thaliana*	Stroma	IDA	Wu *et al.* (1997)	9276953
At2g43560	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At2g44650	Chloroplast	*A. thaliana*	Stroma	IDA	GeneOntology	
At2g44810	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	
At2g44920	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At2g45300	Chloroplast	*A. thaliana*	Chloroplast	TAS	Klee *et al.* (1987)	3481024
At2g47450	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	
At2g47940	Chloroplast	*A. thaliana*	Thylakoid	IDA	Haussühl *et al.* (2001)	11179216
At2g48120	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
At3g01440	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At3g01480	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At3g01500	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Peltier *et al.* (2002)	11826309
At3g02510	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	
At3g02520	Chloroplast	*A. thaliana*	Stroma	IDA	Wu *et al.* (1997)	9276953
At3g09150	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	
At3g10060	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Peltier *et al.* (2002)	11826309
At3g10230	Chloroplast	*A. thaliana*	Chloroplast	TAS	Cunningham *et al.* (1996)	8837512
At3g11630	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Peltier *et al.* (2002)	11826309
At3g15360	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Peltier *et al.* (2002)	11826309
At3g15520	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At3g17820	Chloroplast	*A. thaliana*	Stroma	IDA	Balmer *et al.* (2003)	12509500
At3g22890	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	
At3g23400	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At3g27850	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At3g45140	Chloroplast	*A. thaliana*	Chloroplast	TAS	GeneOntology	
At3g47520	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	
At3g48110	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	
At3g48870	Chloroplast	*A. thaliana*	Stroma	IDA	Nakabayashi *et al.* (1999)	10427773
At3g50820	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At3g51820	Chloroplast	*A. thaliana*	Thylakoid membrane	TAS	Gaubier *et al.* (1995)	8552034
At3g52150	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At3g52380	Chloroplast	*A. thaliana*	Chloroplast	IDA	Ohta *et al.* (1995)	7894017
At3g52750	Chloroplast	*A. thaliana*	Stroma	IDA	GeneOntology	
At3g52960	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Peltier *et al.* (2002)	11826309
At3g53460	Chloroplast	*A. thaliana*	Chloroplast	IDA	Ohta *et al.* (1995)	7894017
At3g53920	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	
At3g54640	Chloroplast	*A. thaliana*	Chloroplast	IDA	Zhao & Last (1995)	7890741
At3g55330	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At3g55870	Chloroplast	*A. thaliana*	Chloroplast	TAS	Niyogi & Fink (1992)	1392592
At3g56650	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At3g58010	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At3g60370	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At3g62030	Chloroplast	*A. thaliana*	Stroma	IDA	GeneOntology	
At3g63190	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At4g02530	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At4g02610	Chloroplast	*A. thaliana*	Chloroplast	IDA	Zhao & Last (1995)	7890741
At4g02780	Chloroplast	*A. thaliana*	Chloroplast	IDA	Sun & Kamiya (1994)	7994182
At4g03280	Chloroplast	*A. thaliana*	Thylakoid membrane	TAS	Knight *et al.* (2002)	12040099
At4g03520	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Peltier *et al.* (2002)	11826309
At4g04020	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At4g04640	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At4g05180	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At4g09010	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At4g09650	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At4g10340	Chloroplast	*A. thaliana*	Thylakoid	IDA	GeneOntology	
At4g14210	Chloroplast	*A. thaliana*	Chloroplast	TAS	GeneOntology	
At4g15510	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At4g17040	Chloroplast	*A. thaliana*	Stroma	IDA	Peltier *et al.* (2001)	11278690
At4g17090	Chloroplast	*A. thaliana*	Stroma	IDA	Lao *et al.* (1999)	10652124
At4g17740	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At4g18370	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At4g18480	Chloroplast	*A. thaliana*	Stroma	IDA	Koncz *et al.* (1990)	2158442
At4g21280	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At4g22240	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At4g24280	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Peltier *et al.* (2002)	11826309
At4g24770	Chloroplast	*A. thaliana*	Chloroplast	IDA	Ohta *et al.* (1995)	7894017
At4g24930	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At4g25080	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Block *et al.* (2002)	11784318
At4g25100	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At4g25370	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At4g27070	Chloroplast	*A. thaliana*	Chloroplast	IDA	Last *et al.* (1991)	1840915
At4g27670	Chloroplast	*A. thaliana*	Chloroplast	TAS	GeneOntology	
At4g28750	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At4g31780	Chloroplast	*A. thaliana*	Chlp inner membrane	IDA	GeneOntology	
At4g31990	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	
At4g32770	Chloroplast	*A. thaliana*	Chlp inner membrane	TAS	GeneOntology	
At4g33030	Chloroplast	*A. thaliana*	Chloroplast	IDA	Essigmann *et al.* (1998)	9465123
At4g33510	Chloroplast	*A. thaliana*	Chloroplast	TAS	Keith *et al.* (1991)	1681544
At4g33650	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	
At4g34190	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	GeneOntology	
At4g37000	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	
At4g39710	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Peltier *et al.* (2002)	11826309
At4g39950	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	
At4g39980	Chloroplast	*A. thaliana*	Chloroplast	TAS	Keith *et al.* (1991)	1681544
At5g03940	Chloroplast	*A. thaliana*	Stroma	IDA	GeneOntology	
At5g05590	Chloroplast	*A. thaliana*	Chloroplast	IDA	Zhao & Last (1995)	7890741
At5g05730	Chloroplast	*A. thaliana*	Chloroplast	TAS	Niyogi & Fink (1992)	1392592
At5g11450	Chloroplast	*A. thaliana*	Lumen	IDA	Schubert *et al.* (2002)	11719511
At5g11710	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Peltier *et al.* (2002)	11826309
At5g13120	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At5g13410	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At5g15530	Chloroplast	*A. thaliana*	Chloroplast	IDA	Thelen *et al.* (2001)	11299381
At5g16050	Chloroplast	*A. thaliana*	Stroma	IDA	Wu *et al.* (1997)	9276953
At5g16390	Chloroplast	*A. thaliana*	Stroma	IDA	Thelen *et al.* (2001)	11299381
At5g20720	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Peltier *et al.* (2002)	11826309
At5g23120	Chloroplast	*A. thaliana*	Lumen	IDA	Schubert *et al.* (2002)	11719511
At5g23120	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At5g23140	Chloroplast	*A. thaliana*	Stroma	IDA	Peltier *et al.* (2001)	11278690

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
At5g23310	Chloroplast	*A. thaliana*	Chloroplast	IDA	Kliebenstein *et al.* (1998)	9765550
At5g24020	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	
At5g35100	Chloroplast	*A. thaliana*	Chloroplast	IDA	Schubert *et al.* (2002)	11719511
At5g35360	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	
At5g39830	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At5g40950	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At5g42080	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	GeneOntology	
At5g43780	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	
At5g45390	Chloroplast	*A. thaliana*	Stroma	IDA	Peltier *et al.* (2001)	11278690
At5g45680	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At5g45930	Chloroplast	*A. thaliana*	Stroma	IDA	GeneOntology	
At5g46390	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At5g50920	Chloroplast	*A. thaliana*	Stroma	IDA	GeneOntology	
At5g51070	Chloroplast	*A. thaliana*	Stroma	TAS	Adam *et al.* (2001)	11299370
At5g51820	Chloroplast	*A. thaliana*	Stroma	IDA	GeneOntology	
At5g52970	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At5g53490	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At5g54770	Chloroplast	*A. thaliana*	Stroma	IDA	Balmer *et al.* (2003)	12509500
At5g55280	Chloroplast	*A. thaliana*	Stroma	IDA	GeneOntology	
At5g57030	Chloroplast	*A. thaliana*	Chloroplast	TAS	Cunningham *et al.* (1996)	8837512
At5g58250	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Peltier *et al.* (2002)	11826309
At5g60600	Chloroplast	*A. thaliana*	Stroma	IDA	Balmer *et al.* (2003)	12509500
At5g62790	Chloroplast	*A. thaliana*	Chloroplast	IDA	GeneOntology	
At5g64040	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
At5g66190	Chloroplast	*A. thaliana*	Thylakoid membrane	IDA	Peltier *et al.* (2002)	11826309
At5g66570	Chloroplast	*A. thaliana*	Thylakoid lumen	IDA	Schubert *et al.* (2002)	11719511
O20250	Chloroplast	*S. oleracea*	Stroma	IDA	Balmer *et al.* (2003)	12509500
O20252	Chloroplast	*S. oleracea*	Stroma	IDA	Balmer *et al.* (2003)	12509500
O22591	Chloroplast	*O. viciifolia*	Lumen/Peripheral	IDA	Peltier *et al.* (2000)	10715320
O23960	Chloroplast	*G. max*	Stroma	IDA	Balmer *et al.* (2003)	12509500
O49939	Chloroplast	*S. oleracea*	Lumen	IDA	Kieselbach *et al.* (1998)	9506969
O50036	Chloroplast	*S. oleracea*	Stroma	IDA	Balmer *et al.* (2003)	12509500
P00289	Chloroplast	*S. oleracea*	Lumen	IDA	Kieselbach *et al.* (1998)	9506969
P00455	Chloroplast	*S. oleracea*	Lumen	IDA	Kieselbach *et al.* (1998)	9506969
P00869	Chloroplast	*P. sativum*	Lumen/Peripheral	IDA	Peltier *et al.* (2000)	10715320
P04717	Chloroplast	*P. sativum*	Lumen/Peripheral	IDA	Peltier *et al.* (2000)	10715320
P08926	Chloroplast	*P. sativum*	Stroma	IDA	Balmer *et al.* (2003)	12509500
P08926	Chloroplast	*P. sativum*	Lumen/Peripheral	IDA	Peltier *et al.* (2000)	10715320
P08927	Chloroplast	*P. sativum*	Stroma	IDA	Balmer *et al.* (2003)	12509500
P09559	Chloroplast	*S. oleracea*	Stroma	IDA	Balmer *et al.* (2003)	12509500
P10871	Chloroplast	*S. oleracea*	Stroma	IDA	Balmer *et al.* (2003)	12509500
P11869	Chloroplast	*S. oleracea*	IE membrane	IDA	Flügge *et al.* (1989)	2714257
P12301	Chloroplast	*S. oleracea*	Lumen	IDA	Kieselbach *et al.* (1998)	9506969
P12302	Chloroplast	*S. oleracea*	Lumen	IDA	Kieselbach *et al.* (1998)	9506969
P12359	Chloroplast	*S. oleracea*	Lumen	IDA	Kieselbach *et al.* (1998)	9506969
P12860	Chloroplast	*S. oleracea*	Stroma	IDA	Balmer *et al.* (2003)	12509500
P14226	Chloroplast	*P. sativum*	Lumen/Peripheral	IDA	Peltier *et al.* (2000)	10715320
P16002	Chloroplast	*P. sativum*	Lumen/Peripheral	IDA	Peltier *et al.* (2000)	10715320
P16016	Chloroplast	*S. oleracea*	Stroma	IDA	Balmer *et al.* (2003)	12509500
P16059	Chloroplast	*P. sativum*	Lumen/Peripheral	IDA	Peltier *et al.* (2000)	10715320
P16098	Chloroplast	*H. vulgare*	Stroma	IDA	Balmer *et al.* (2003)	12509500
P23525	Chloroplast	*S. oleracea*	IE membrane	IDA	Block *et al.* (1991)	1879527
P27789	Chloroplast	*Z. mays*	Lumen/Peripheral	IDA	Peltier *et al.* (2000)	10715320
P28552	Chloroplast	*P. sativum*	Lumen/Peripheral	IDA	Peltier *et al.* (2000)	10715320
P29344	Chloroplast	*S. oleracea*	Stroma	IDA	Balmer *et al.* (2003)	12509500
P29518	Chloroplast	*Z. mays*	Lumen/Peripheral	IDA	Peltier *et al.* (2000)	10715320
P31093	Chloroplast	*H. vulgare*	Lumen/Peripheral	IDA	Peltier *et al.* (2000)	10715320
P31541	Chloroplast	*L. esculentum*	Stroma	IDA	Balmer *et al.* (2003)	12509500
P31593	Chloroplast	*N. tabacum*	Stroma	IDA	Balmer *et al.* (2003)	12509500
P32260	Chloroplast	*S. oleracea*	Stroma	IDA	Balmer *et al.* (2003)	12509500
P34811	Chloroplast	*G. max*	Stroma	IDA	Balmer *et al.* (2003)	12509500
P41346	Chloroplast	*V. faba*	Lumen/Peripheral	IDA	Peltier *et al.* (2000)	10715320
P48496	Chloroplast	*S. oleracea*	Stroma	IDA	Balmer *et al.* (2003)	12509500
P48629	Chloroplast	*S. oleracea*	chloroplast envelope	IDA	Schmidt *et al.* (1994)	7948918
P52426	Chloroplast	*S. oleracea*	Stroma	IDA	Balmer *et al.* (2003)	12509500
Q02028	Chloroplast	*P. sativum*	Lumen/Peripheral	IDA	Peltier *et al.* (2000)	10715320
Q02758	Chloroplast	*P. sativum*	Lumen/Peripheral	IDA	Peltier *et al.* (2000)	10715320
Q40739	Chloroplast	*O. sativa*		IDA	GeneOntology	
Q41229	Chloroplast	*N. sylvestris*	Lumen/Peripheral	IDA	Peltier *et al.* (2000)	10715320
Q41364	Chloroplast	*S. oleracea*	IE membrane	IDA	Weber *et al.* (1995)	7873543
Q42967	Chloroplast	*N. tabacum*	Stroma	IDA	Balmer *et al.* (2003)	12509500
Q43157	Chloroplast	*S. oleracea*	Stroma	IDA	Balmer *et al.* (2003)	12509500
Q43467	Chloroplast	*G. max*	Stroma	IDA	Balmer *et al.* (2003)	12509500
Q43832	Chloroplast	*S. oleracea*	Stroma	IDA	Balmer *et al.* (2003)	12509500
Q9FSS6	Chloroplast	*O. sativa*	Stroma	IDA	Balmer *et al.* (2003)	12509500
Q9LEJ0	Chloroplast	*H. brasiliensis*	Stroma	IDA	Balmer *et al.* (2003)	12509500
Q9LLE2	Chloroplast	*S. oleracea*	IE membrane	IDA	Weber *et al.* (1995)	10810150
Q9LRC5	Chloroplast	*L. longiflorum*	Stroma	IDA	Balmer *et al.* (2003)	12509500
Q9SM43	Chloroplast	*S. oleracea*	Lumen	IDA	Kieselbach *et al.* (1998)	9506969
Q9SM44	Chloroplast	*S. oleracea*	IE membrane	IDA	Miège *et al.* (1999)	10518794
Q9SP64	Chloroplast	*A. annua*	Stroma	IDA	Balmer *et al.* (2003)	12509500
6319426	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6319458	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6319475	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6319498	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6319511	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6319518	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6319567	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6319598	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
6319793	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6319837	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6319872	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6319894	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6319925	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6320003	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6320027	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6320044	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6320099	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6320158	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6320160	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6320200	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6320235	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6320246	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6320383	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6320464	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6320502	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6320583	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6320585	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6320785	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6320851	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6320889	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6321030	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6321179	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6321202	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6321251	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6321452	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6321513	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6321521	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6321550	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6321613	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6321631	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6321670	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6321796	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6321826	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6321858	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6321867	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6321875	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6321912	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6322002	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6322034	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6322055	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6322138	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6322328	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6322365	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6322398	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6322407	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6322409	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6322476	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6322505	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6322564	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6322654	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6322708	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6322765	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6322811	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6322850	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6323001	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6323028	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6323118	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6323193	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6323217	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6323232	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6323312	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6323319	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6323387	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6323506	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6323526	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6323808	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6323817	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6323831	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6323881	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6323897	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6323908	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6323959	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6323960	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6324014	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6324054	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6324069	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6324118	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6324160	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6324192	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6324196	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6324198	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6324258	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6324260	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6324263	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6324302	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6324365	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6324546	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
6324591	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6324594	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6324704	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6324716	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6324750	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6324775	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6324789	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6324825	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6324909	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6324932	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6324993	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6325020	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6325085	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6325123	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6325158	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6325160	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6325197	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6325198	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6325228	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6325281	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
6325304	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
14318482	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
14318558	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
14318570	Mitochondrion	*S. cerevisiae*		TAS/IDA	GeneOntology	
At1g03860	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At1g06530	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At1g07900	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At1g09210	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At1g13440	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At1g14980	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At1g17350	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At1g24180	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At1g27390	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At1g32580	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At1g47260	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At1g47420	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At1g48030	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At1g50400	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At1g50940	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At1g51980	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At1g53240	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At1g56340	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At1g59900	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At1g79010	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At1g79230	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At1g79440	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At1g80230	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At2g04030	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At2g05710	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At2g14200	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At2g20420	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At2g20530	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At2g21640	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At2g21870	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At2g27730	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At2g28000	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At2g33040	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At2g33150	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At2g33210	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At2g35240	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At2g41380	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At2g44350	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At2g47510	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g01280	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g02090	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g03100	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g04120	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g06050	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g07480	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g07770	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g08580	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g10920	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g12260	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g15000	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g15020	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g16480	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g17240	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g20000	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g22200	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g22330	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g22370	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g23990	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g27070	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g27080	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g27240	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g27280	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g45300	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g48000	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g48680	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
At3g52200	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g52300	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g54110	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g59760	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At3g61440	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At4g01100	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At4g02580	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At4g02930	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At4g11010	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At4g11120	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At4g13850	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At4g15940	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At4g23900	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At4g26970	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At4g27585	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At4g28060	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At4g28510	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At4g31500	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At4g31810	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At4g32470	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At4g35260	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At4g35830	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At4g35850	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At4g37910	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g07440	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g08300	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g08670	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g08690	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g09590	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g10860	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g13430	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g13440	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g13450	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g13490	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g14040	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g15090	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g18170	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g19760	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g20080	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g23140	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g23250	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g37510	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g40770	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g40810	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g40930	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g47030	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g50850	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g52840	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g54100	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g55070	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g57300	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g59880	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g60640	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g62530	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g63400	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g63510	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g66510	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g66760	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
At5g67500	Mitochondrion	*A. thaliana*		TAS/IDA	GeneOntology	
FB|FBgn00000	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00001	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00001	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00001	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00001	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00002	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00002	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00002	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00003	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00003	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00004	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00004	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00004	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00005	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00006	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00008	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00009	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00009	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00009	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00010	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00011	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00011	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00012	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00012	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00012	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00020	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00021	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00024	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00025	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00027	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
FB|FBgn00027	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00027	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00029	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00029	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00029	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00029	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00029	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00030	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00030	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00031	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00032	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00033	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00034	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00034	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00035	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00035	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00036	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00037	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00037	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00041	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00042	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00043	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00043	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00044	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00046	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00049	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00049	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00054	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00055	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00056	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00086	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00101	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00101	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00102	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00103	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00103	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00103	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00103	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00112	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00115	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00115	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00116	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00116	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00116	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00116	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00117	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00117	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00117	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00135	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00137	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00140	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00141	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00152	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00155	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00156	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00158	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00166	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00175	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00175	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00203	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00205	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00205	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00242	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00248	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00256	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00258	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00261	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00262	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00264	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00264	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00270	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00285	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00286	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00286	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00299	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00354	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00372	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00391	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00411	NON	*D. melanogaster*	nucleus	TAS/IDA	GeneOntology	
FB|FBgn00430	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
FB|FBgn00430	NON	*D. melanogaster*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1019	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1019	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1030	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1030	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:1030	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1032	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:1047	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:1047	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
MGI|MGI:1050	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1073	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1073	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1073	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1074	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1077	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1079	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1079	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:1080	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:1083	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1092	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1097	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:1098	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:1100	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1194	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1201	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:1202	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1270	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:1276	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1298	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:1306	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1313	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:1329	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1330	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1333	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1333	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:1333	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1333	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:1334	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:1334	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1337	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1338	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:1338	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:1338	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1341	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1343	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1344	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1345	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1345	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1345	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1347	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:1349	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1349	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1351	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:1858	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1858	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1858	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:1858	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1858	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:1860	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1889	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1891	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:1921	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1927	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1927	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1927	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1927	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1927	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1928	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1929	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:1929	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:2135	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:2135	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:2136	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:2137	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:2137	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:2137	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:2137	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:2137	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:8798	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:8807	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:8807	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:8825	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:9490	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:9566	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:9572	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:9641	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:9652	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:9717	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:9735	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:9743	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:9747	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:9751	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:9755	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:9790	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:9831	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:9831	NON	*M. musculus*	nucleus	TAS/IDA	GeneOntology	
MGI|MGI:9948	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	
MGI|MGI:9967	NON	*M. musculus*	cytoplasma	TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
SGD|S0000008	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000021	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000028	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000037	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000040	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000040	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000041	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0000066	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0000109	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000110	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0000119	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0000120	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0000121	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0000130	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000159	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000188	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000193	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0000264	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0000283	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000283	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000284	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000309	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000319	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000323	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0000324	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000326	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000330	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000334	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000335	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000358	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0000364	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000369	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0000386	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0000386	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000387	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000389	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000399	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0000406	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000408	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000416	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000426	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000451	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0000461	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0000472	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000486	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000515	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000537	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000540	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000557	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000596	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000598	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000630	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000630	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000631	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0000631	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000640	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000689	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000752	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0000753	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000755	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000757	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000758	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000758	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0000768	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000773	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000782	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0000782	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000797	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000808	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0000808	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0000810	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000811	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0000814	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000814	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0000823	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000831	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0000846	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000872	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000933	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000956	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000972	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000977	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000993	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0000996	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001032	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001039	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001050	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001060	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001065	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
SGD|S0001069	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001070	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001107	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001119	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001121	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001130	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001131	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001140	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001151	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001160	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001190	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001195	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001203	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001209	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001212	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001215	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001218	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001236	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001238	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001283	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001288	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001325	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001363	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001375	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001376	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001384	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001386	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001387	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001411	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001421	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001441	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001443	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001444	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001451	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001454	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001470	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001473	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001477	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001485	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001490	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001496	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001505	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001507	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001512	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001528	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001531	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001561	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001563	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001563	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001568	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001570	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001572	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001589	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001592	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001603	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001613	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001622	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001624	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001627	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001639	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001655	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001668	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001675	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001693	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001695	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001717	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001742	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001754	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001760	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001762	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001762	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001774	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001780	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001783	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001846	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001857	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001863	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001865	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001890	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001891	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001898	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001901	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0001915	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0001944	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002135	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002178	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002180	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002209	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002218	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
SGD|S0002227	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002228	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002244	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002245	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002251	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002255	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002264	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002265	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002283	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002287	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002293	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002295	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002319	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002324	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002354	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002367	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002386	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002414	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002450	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002459	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002499	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002520	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002520	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002528	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002536	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002552	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002553	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002573	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002577	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002577	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002578	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002580	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002596	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002600	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002609	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002637	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002642	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002643	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002648	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002653	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002678	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002681	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002701	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002709	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002718	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002723	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002739	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002743	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002766	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002772	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002793	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002802	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002835	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002851	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002854	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002868	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002868	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002870	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002876	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002886	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002888	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002893	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002906	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002906	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002918	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0002919	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002969	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002969	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0002999	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003003	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003017	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003026	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003043	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003054	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003059	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003060	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003063	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003065	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003073	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003083	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003092	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003103	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003111	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003113	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003115	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003122	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003124	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003140	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003147	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
SGD|S0003173	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003174	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003175	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003176	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003193	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003205	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003206	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003220	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003237	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003242	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003245	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003260	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003262	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003286	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003304	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003304	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003310	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003323	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003324	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003330	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003330	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003335	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003336	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003337	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003337	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003351	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003361	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003362	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003364	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003368	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003380	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003387	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003389	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003390	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003391	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003395	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003401	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003402	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003407	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003410	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003412	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003425	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003427	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003443	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003447	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003452	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003454	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003454	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003461	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003464	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003471	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003472	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003478	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003487	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003517	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003540	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003566	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003586	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003595	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003609	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003610	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003636	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003648	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003677	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003684	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003708	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003716	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003739	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003744	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003766	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003811	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003813	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003818	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003821	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003829	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003838	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003850	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003852	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003853	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003873	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003874	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003886	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003924	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003927	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003931	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0003965	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003969	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003973	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0003992	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
SGD|S0003995	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004007	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004015	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004015	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004049	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004065	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004068	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004069	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004070	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004083	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004092	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004093	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004093	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004095	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004095	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004096	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004123	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004129	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004131	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004132	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004138	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004140	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004141	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004143	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004145	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004156	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004206	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004223	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004249	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004255	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004262	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004264	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004267	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004282	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004289	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004295	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004322	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004332	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004334	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004342	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004370	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004390	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004413	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004421	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004425	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004434	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004442	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004460	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004470	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004521	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004526	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004540	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004540	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004556	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004564	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004571	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004572	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004593	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004596	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004598	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004624	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004627	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004628	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004638	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004640	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004658	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004663	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004674	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004695	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004698	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004710	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004735	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004762	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004795	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004801	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004810	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004829	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004831	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004842	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004843	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004848	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004860	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004873	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004881	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004895	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004900	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004903	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0004907	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004912	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
SGD|S0004916	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004921	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004942	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004950	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004954	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0004970	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005000	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005004	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005010	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005023	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005026	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005046	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005048	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005048	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005057	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005061	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005061	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005068	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005092	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005096	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005105	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005111	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005111	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005113	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005116	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005119	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005133	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005138	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005165	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005175	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005176	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005182	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005188	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005196	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005216	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005221	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005226	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005228	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005256	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005261	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005286	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005318	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005321	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005324	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005335	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005358	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005369	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005373	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005398	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005404	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005442	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005451	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005482	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005483	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005499	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005504	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005518	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005554	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005563	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005566	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005574	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005575	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005591	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005600	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005617	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005641	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005643	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005656	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005659	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005662	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005668	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005676	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005684	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005686	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005697	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005700	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005706	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005724	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005735	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005743	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005750	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005756	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005770	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005783	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005795	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005798	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005800	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005800	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005821	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
SGD|S0005823	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005839	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005853	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005853	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005859	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005859	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005861	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005874	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005902	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005902	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005915	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0005928	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005941	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005952	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0005970	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0006000	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0006006	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0006008	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0006021	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0006022	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0006037	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0006046	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0006050	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0006055	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0006056	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0006076	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0006095	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0006130	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0006138	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0006145	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0006169	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0006183	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0006189	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0006209	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0006220	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0006222	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0006223	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0006234	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0006238	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0006276	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0006311	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0006312	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0006314	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0006341	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0006370	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0006372	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0006377	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0006386	NON	*S. cerevisiae*	nucleus	TAS/IDA	GeneOntology	
SGD|S0006387	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0006393	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0007222	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SGD|S0007348	NON	*S. cerevisiae*	cytoplasma	TAS/IDA	GeneOntology	
SPTR|O00154	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O00165	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O00203	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O00208	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O00221	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O00291	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O00291	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O00330	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O00410	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O00461	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O00539	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O00623	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O08547	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O08919	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O14519	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O14521	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O14645	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O14655	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O14727	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O14773	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O14795	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O14818	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O14833	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O14893	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O14925	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O14960	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O14965	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O14980	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O15065	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O15111	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O15117	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O15118	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O15143	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O15144	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O15144	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O15195	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O15239	NON		cytoplasma	TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
SPTR|O15265	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O15297	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O15298	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O15317	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O15318	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O15344	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O15360	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O15360	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O15381	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O15392	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O15446	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O15446	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O15504	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O15511	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O15527	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O16011	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O18388	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O22870	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O35153	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O35231	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O35381	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O35499	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O35625	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O35625	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O35980	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O43143	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O43289	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O43290	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O43390	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O43402	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O43427	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O43427	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O43446	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O43447	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O43516	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O43598	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O43602	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O43615	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O43630	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O43663	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O43681	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O43683	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O43707	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O43707	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O43852	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O43903	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O55070	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O55135	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O60437	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O60499	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O60592	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O60620	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O60682	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O60725	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O60763	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O74700	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O75027	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O75299	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O75312	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O75319	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O75341	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O75433	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O75436	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O75438	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O75521	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O75766	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O75766	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O75781	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O75791	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O75818	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O75822	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O75832	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O75880	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O75912	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O75912	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O76022	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O88695	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O88720	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O88814	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O88848	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O88942	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O88973	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O88981	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O88992	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O89087	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O89110	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O94742	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O94763	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O94805	NON		nucleus	TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
SPTR|O95218	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O95239	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O95243	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O95251	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O95453	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O95644	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O95707	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|O95793	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O95793	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O95822	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|O95834	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P00175	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P00522	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P00546	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P00572	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P00966	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P01106	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P01123	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P01123	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P02545	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P02786	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P02836	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P04424	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P04821	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P04843	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P04843	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P05030	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P05387	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P05552	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P05661	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P05771	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P06002	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P06105	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P06105	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P06182	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P06400	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P06576	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P06785	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P06843	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P07244	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P07256	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P07260	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P07269	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P07283	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P07884	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P07954	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P08044	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P08118	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P08525	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P08559	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P08727	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P09103	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P09234	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P09493	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P09622	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P09661	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P09938	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P09950	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P09959	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P10071	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P10080	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P10085	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P10180	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P10355	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P10834	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P10962	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P11163	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P11309	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P11325	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P11416	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P11612	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P11655	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P11745	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P11745	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P12272	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P12272	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P12398	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P12683	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P12688	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P12754	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P12868	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P12883	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P13096	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P13298	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P13433	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P13667	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P13674	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P13804	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P14020	NON		cytoplasma	TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
SPTR|P14142	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P14209	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P14416	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P14680	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P14736	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P14785	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P14785	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P14866	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P14904	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P15021	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P15151	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P15315	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P15348	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P15348	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P15367	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P15425	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P15442	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P15646	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P15884	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P16523	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P16861	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P17157	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P17225	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P17558	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P17980	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P17980	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P17987	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P18283	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P18289	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P18414	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P18825	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P18850	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P18850	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P18850	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P19075	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P19658	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P19783	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P19880	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P20339	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P20438	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P20459	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P21190	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P21268	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P21359	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P21462	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P21519	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P21549	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P21619	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P22057	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P22213	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P22213	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P22307	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P22361	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P22532	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P22674	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P22695	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P22745	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P22745	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P23201	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P23250	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P23284	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P23500	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P23511	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P23528	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P23644	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P23833	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P24279	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P24279	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P24539	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P25028	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P25116	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P25344	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P25367	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P25372	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P25373	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P25375	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P25385	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P25454	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P25491	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P25628	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P25635	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P25719	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P25846	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P26019	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P26364	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P26599	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P26637	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P26651	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P27476	NON		nucleus	TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
SPTR|P27577	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P27694	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P27801	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P28062	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P28288	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P28329	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P28742	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P29374	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P29461	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P29469	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P30518	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P30518	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P30622	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P30624	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P30626	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P30665	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P31040	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P31115	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P31244	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P31378	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P31378	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P31384	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P31539	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P31949	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P32067	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P32074	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P32336	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P32337	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P32368	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P32381	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P32447	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P32474	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P32485	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P32497	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P32521	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P32522	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P32523	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P32559	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P32562	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P32578	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P32599	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P32605	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P32606	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P32629	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P32639	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P32644	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P32795	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P32803	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P32835	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P32855	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P32864	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P32875	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P32910	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P32912	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P32915	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P32916	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P32969	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P33176	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P33244	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P33311	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P33334	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P33411	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P33748	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P33755	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P33759	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P33892	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P33894	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P34021	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P34896	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P34897	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P35182	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P35182	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P35184	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P35191	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P35192	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P35222	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P35237	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P35269	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P35554	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P35600	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P35601	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P35610	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P35680	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P35869	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P35908	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P36048	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P36051	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P36056	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P36060	NON		cytoplasma	TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
SPTR|P36068	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P36093	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P36107	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P36161	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P36421	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P36542	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P36776	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P37108	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P37238	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P37238	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P38009	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38072	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38117	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38120	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38127	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38132	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P38137	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38165	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38203	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P38205	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P38208	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P38238	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38283	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38290	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P38291	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P38300	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38307	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38336	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P38336	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P38339	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38342	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38342	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38353	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38398	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P38532	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P38533	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P38571	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38630	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P38719	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P38720	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38731	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38741	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38779	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38792	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38792	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P38805	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P38825	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38855	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38855	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38874	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P38875	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38888	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38910	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P38985	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P39076	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P39104	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P39109	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P39112	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P39112	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P39656	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P39683	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P39685	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P39704	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P39723	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P39925	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P39943	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P39968	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P39984	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P39990	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P40005	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P40007	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P40010	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P40047	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P40048	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P40064	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P40068	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P40068	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P40069	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P40075	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P40202	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P40227	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P40306	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P40316	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P40317	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P40339	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P40341	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P40343	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P40356	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P40368	NON		nucleus	TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
SPTR|P40395	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P40466	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P40495	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P40518	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P40548	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P40557	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P40571	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P40582	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P40764	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P40961	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P40971	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P41222	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P41227	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P41252	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P41544	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P41697	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P41806	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P41835	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P41910	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P41920	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P41940	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P42223	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P42228	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P42685	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P42768	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P42773	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P42834	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P42838	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P42951	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P43120	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P43243	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P43535	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P43638	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P43897	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P45437	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P45877	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P46063	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P46087	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P46108	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P46199	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P46673	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P46783	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P46948	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P46956	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P47001	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P47025	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P47043	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P47083	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P47104	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P47142	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P47160	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P47818	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P47912	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P47939	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P48382	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P48412	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P48432	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P48556	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P48562	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P48570	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P48593	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P48725	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P48810	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P48810	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P49023	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P49028	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P49137	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P49354	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P49367	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P49411	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P49418	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P49588	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P49589	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P49615	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P49736	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P49748	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P49790	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P49792	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P49815	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P49848	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P49917	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P49959	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P49959	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P50086	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P50091	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P50224	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P50289	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P50607	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P50607	NON		nucleus	TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
SPTR|P50613	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P51003	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P51003	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P51114	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P51141	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P51150	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P51532	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P51592	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P51601	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P51858	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P51955	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P52304	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P52434	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P53011	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P53062	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P53067	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P53164	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P53197	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P53204	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P53218	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P53218	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P53250	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P53281	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P53337	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P53378	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P53397	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P53550	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P53551	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P53618	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P53618	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P53762	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P53801	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P53845	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P53847	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P53848	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P53858	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P53859	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P53893	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P53900	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P53946	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P54005	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P54113	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P54136	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P54199	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P54253	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P54278	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P54358	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P54362	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P54577	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P54867	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P54999	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P55010	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P55209	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P55265	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P55916	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P56181	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P56371	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P56945	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P57740	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P70201	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P78317	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P78346	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P78382	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P78406	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P78406	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P80428	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P81117	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P81117	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P89102	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P91927	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P97443	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P97474	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P97801	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P97801	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P97822	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P97929	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P97931	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|P98164	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|P98177	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q00597	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q00597	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q00684	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q00723	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q00772	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q00816	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q00873	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q00947	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q00955	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q00963	NON		cytoplasma	TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
SPTR|Q01097	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q01163	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q01649	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q01780	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q01820	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q01826	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q02046	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q02067	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q02100	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q02100	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q02199	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q02410	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q02455	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q02486	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q02521	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q02645	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q02724	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q02775	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q02884	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q03017	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q03081	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q03181	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q03201	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q03220	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q03330	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q03455	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q03530	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q03660	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q03750	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q03750	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q03769	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q03898	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q03923	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q03935	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q04110	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q04207	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q04307	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q04401	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q04410	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q04651	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q04725	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q04739	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q04951	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q05123	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q05209	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q05513	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q05636	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q05652	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q05682	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q06106	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q06141	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q06142	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q06219	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q06265	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q06287	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q06323	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q06340	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q06497	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q06512	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q06609	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q06630	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q06704	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q06787	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q06787	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q07468	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q07468	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q07540	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q07938	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q08032	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q08096	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q08170	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q08268	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q08369	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q08473	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q08484	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q08500	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q08558	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q08601	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q08605	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q08645	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q08650	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q08746	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q08826	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q08887	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q08965	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q09028	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q10713	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q12039	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q12045	NON		cytoplasma	TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
SPTR|Q12045	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q12056	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q12067	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q12143	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q12207	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q12207	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q12236	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q12250	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q12289	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q12306	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q12320	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q12341	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q12341	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q12363	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q12375	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q12382	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q12417	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q12420	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q12425	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q12433	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q12445	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q12446	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q12460	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q12509	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q12674	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q12749	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q12798	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q12852	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q12888	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q12934	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q12996	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q13006	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q13023	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q13077	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q13087	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q13115	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q13185	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q13310	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q13342	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q13342	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q13352	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q13404	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q13435	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q13439	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q13515	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q13569	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q13588	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q13608	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q13625	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q13823	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q13838	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q13907	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q14032	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q14203	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q14211	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q14221	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q14746	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q14764	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q14790	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q14805	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q14848	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q14872	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q14978	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q14981	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q15029	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q15046	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q15118	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q15154	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q15172	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q15287	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q15299	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q15327	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q15393	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q15427	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q15435	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q15542	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q15916	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q16181	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q16384	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q16514	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q16549	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q16586	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q16595	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q16621	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q16644	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q16698	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q16799	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q16822	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q24342	NON		cytoplasma	TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
SPTR|Q24572	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q27487	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q30201	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q60698	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q60698	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q60873	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q60954	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q60954	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q61048	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q61152	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q61315	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q61501	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q61769	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q61804	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q61884	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q61937	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q62377	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q64028	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q64317	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q8K1M4	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q8K443	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q8K443	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q8K4J6	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q8R0N8	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q8R4E4	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q8R4L0	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q8R4L0	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q8R507	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q8R5L3	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q8VBU8	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q8VCY4	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q8VHC3	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q8VIE6	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q91V37	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q91VF6	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q91W92	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q91X44	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q91XU9	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q91YE4	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q91YJ0	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q921C5	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q921F3	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q922A5	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q922E3	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q922E3	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q922I7	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q92498	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q924C5	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q924L5	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q92535	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q92581	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q92636	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q92731	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q92747	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q92791	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q92793	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q92797	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q92797	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q92922	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q92966	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q92973	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q93009	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q93063	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q99207	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q99257	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q99436	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q99442	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q99460	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q99541	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q99551	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q99567	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q99575	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q99575	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q99589	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q99613	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q99627	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q99633	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q99708	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q99738	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q99933	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q99956	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q99972	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q99996	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q99J03	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q99JF1	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q99JG7	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q99JX7	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q99L06	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q99L58	NON		nucleus	TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
SPTR|Q99LH8	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q99M04	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q99M55	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q99M87	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q99MQ2	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q99N95	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q99N96	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q99NB8	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q99P71	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q99P71	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q99P72	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q99PJ2	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q99PQ5	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9C0J8	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9CR11	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9CR42	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9CUX1	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9CXT1	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9CXZ1	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9CY82	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9CZD2	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9CZD6	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9CZP7	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9D0J0	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9D1Z9	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9D335	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9D5V6	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9D6P6	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9D704	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9D7Z6	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9D870	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9D9U4	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9DBH1	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9DC44	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9DC44	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9DC50	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9DC61	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9DC77	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9DCD6	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9DCD6	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9DCI9	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9EPK1	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9EPM5	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9EQ61	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9EQF7	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9EQF8	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9EQG3	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9EQY0	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9ERK2	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9ESL3	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9ESL4	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9ET42	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9ET47	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9H1K1	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9H254	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9H2X6	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9HC62	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9HCN0	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9JHG6	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9JHR9	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9JI11	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9JIF7	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9JII1	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9JIQ4	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9JJ11	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9JJA4	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9JJI6	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9JJT0	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9JJT9	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9JK11	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9JK37	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9JKM7	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9JKW0	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9JKX4	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9JKY7	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9JL19	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9JL72	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9JLA1	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9JLD3	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9JLI3	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9JLV1	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9JM95	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9JMB0	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9JMB6	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9JMC4	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9NPB1	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9NQT5	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9NR62	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9NS25	NON		nucleus	TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
SPTR|Q9NS64	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9NS87	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9NUT2	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9NWB1	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9P0I1	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9QWN8	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9QX60	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9QX60	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9QXB7	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9QXE4	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9QXI1	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9QXN0	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9QXN5	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9QXV0	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9QY31	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9QYB2	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9QYC3	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9QZA0	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9QZD8	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9QZE7	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9QZF3	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9QZK7	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9QZS3	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9QZT6	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9QZT7	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9QZU4	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9QZV9	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9R0B7	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9R0Q1	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9R0Y5	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9R1K7	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9R1M5	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9R1R2	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9UBE8	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9UBS8	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9UGK9	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9UH59	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9UH92	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9UH99	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9UHB9	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9UHC1	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9UHI6	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9UHK0	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9UHL0	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9UI32	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9UII2	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9UJ83	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9UJX5	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9UJZ1	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9UKD1	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9UKX3	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9UL40	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9ULV4	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9UM54	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9UMW8	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9UMZ2	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9UNK0	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9UPM8	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9UQ80	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9UQ90	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9UQL6	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9W4X9	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9WTJ2	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9WTZ7	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9WU12	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9WUB1	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9WUL8	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9WUL8	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9WV00	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9WV55	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9WV86	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9WVA6	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9WVI9	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9WVJ2	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9WVP6	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9XYP8	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9Y223	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9Y237	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9Y265	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9Y2B4	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9Y2D0	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9Y2W1	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9Y2X3	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9Y3D5	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9Y3E1	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9Y3P9	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9Y474	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9Y4Z0	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9Y573	NON		cytoplasma	TAS/IDA	GeneOntology	

continued from last page

GI/MIPS	Organelle	Species	Localization	Proof	Reference	PMID
SPTR|Q9Y5L4	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9Y5P8	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9Y5X6	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9Y5Z7	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9Y664	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9Y697	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9Y6A9	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9Y6K8	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9Y6M0	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9Y6M1	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9Y6N4	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9Y6X1	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9Y6Y0	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9Z0E9	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9Z0R0	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9Z0Y1	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9Z1J3	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9Z248	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9Z2V5	NON		cytoplasma	TAS/IDA	GeneOntology	
SPTR|Q9Z2V6	NON		nucleus	TAS/IDA	GeneOntology	
SPTR|Q9Z326	NON		nucleus	TAS/IDA	GeneOntology	
TAIR|gene:22	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:33	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:33	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:33	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:33	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:33	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:33	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:33	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:33	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:33	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:33	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:33	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:33	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:33	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:33	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:33	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:33	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:33	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:33	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:36	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:37	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:37	NON	*A. thaliana*	nucleus	TAS/IDA	GeneOntology	
TAIR|gene:37	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:37	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:37	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:37	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:50	NON	*A. thaliana*	cytoplasma	TAS/IDA	GeneOntology	
TAIR|gene:50	NON	*A. thaliana*	nucleus	TAS/IDA	GeneOntology	
TAIR|gene:50	NON	*A. thaliana*	nucleus	TAS/IDA	GeneOntology	
TAIR|gene:50	NON	*A. thaliana*	nucleus	TAS/IDA	GeneOntology	
WB|CE02166	NON	*C. elegans*	nucleus	TAS/IDA	GeneOntology	
WB|CE03573	NON	*C. elegans*	nucleus	TAS/IDA	GeneOntology	
WB|CE04629	NON	*C. elegans*	cytoplasma	TAS/IDA	GeneOntology	
WB|CE09012	NON	*C. elegans*	cytoplasma	TAS/IDA	GeneOntology	
WB|CE11778	NON	*C. elegans*	nucleus	TAS/IDA	GeneOntology	
WB|CE18542	NON	*C. elegans*	nucleus	TAS/IDA	GeneOntology	
WB|CE18884	NON	*C. elegans*	cytoplasma	TAS/IDA	GeneOntology	
WB|CE25146	NON	*C. elegans*	cytoplasma	TAS/IDA	GeneOntology	
WB|CE25147	NON	*C. elegans*	cytoplasma	TAS/IDA	GeneOntology	
WB|CE28017	NON	*C. elegans*	cytoplasma	TAS/IDA	GeneOntology	

Tab. B.6.: Test set for the evaluation of cTP predictors.

Erklärung

Ich versichere, daß ich die von mir vorgelegte Dissertation selbständg 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 Prof. Dr. Francesco Salamini betreut worden.

Köln, den 15. September 2003

Erik Richly

Teilpublikationen:

Richly E, Leister D. **2003** An improved prediction of chloroplast proteins reveals diversities and commonalities in the chloroplast proteomes of Arabidopsis and rice. Submitted.

Richly E, Dietzmann A, Biehl A, Kurth J, Laloi C, Apel K, Salamini F, Leister D. **2003** Covariations in the nuclear chloroplast transcriptome reveal a regulatory master-switch. *EMBO Rep* 4:491-498

Richly E, Chinnery PF, Leister D. **2003** Evolutionary diversification of mitochondrial proteomes: implications for human disease. *Trends Genet* 19:356-362.

Richly E, Dietzmann A, Biehl A, Kurth J, Laloi C, Apel K, Salamini F, Leister D. **2003** Covariations in the nuclear chloroplast transcriptome reveal a regulatory master-switch. *EMBO Rep* 4:491-498

Martin, W. and Rujan, T. and Richly, E. and Hansen, A. and Cornelsen, S. and Lins, T. and Leister, D. and Stoebe, B. and Hasegawa, M. and Penny, D. **2002** Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *PNAS* 19:12246-51

Kurth J, Varotto C, Pesaresi P, Biehl A, Richly E, Salamini F, Leister D. **2002** Gene-sequence-tag expression analyses of 1800 genes related to chloroplast functions. *Planta* 215:101-109

Varotto C, Richly E, Salamini F, and Leister D. **2001** GST-PRIME - a genome-wide primer design software for the generation of gene sequence tags. *Nucl Acids Res* 29, 4373-4377

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

. . . Elvis has left the building!