# Analysis of targets and functions of the chloroplast intron maturase MatK

Dissertation

zur Erlangung des akademischen Grades
d o c t o r  r e r u m  n a t u r a l i u m
(Dr. rer. nat.)

im Fach Biologie

eingereicht an der

Lebenswissenschaftliche Fakultät
der Humboldt-Universität zu Berlin

Von

Yujiao Qu

Präsident der Humboldt-Universität zu Berlin
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Lebenswissenschaftliche Fakultät
Prof. Dr. Richard Lucius

Gutachter/in: 1. Prof. Dr. Christian Schmitz-Linneweber

2. Prof. Dr. Bernhard Grimm

3. Associate Prof. Oren Ostersetzer-Biran

Datum der Einreichung: 17.12.2014

Datum der Promotion: 18.06.2015

# Abstract

In chloroplasts, primary transcripts are subjected to a number of processing events. These events play important roles in the regulation of gene expression and are extensively controlled by protein factors, especially by RNA-binding proteins.

Chloroplast splicing factor MatK is related to prokaryotic group II intron maturases. *Nicotiana tabacum* MatK interacts with its home intron *trnK* and six additional group IIA introns. In this study, binding sites of MatK were narrowed down to varying regions of its group II targets by RIP-seq in *Nicotiana tabacum*. The results obtained demonstrate that MatK has gained versatility in RNA recognition relative to its bacterial ancestors. MatK thus exemplifies how a maturase could have gained the ability to act *in trans* on multiple introns during the dispersion of the group II introns through the eukaryotic genome early in the eukaryote evolution.

Quantitative investigation and mathematical modeling of the expression of MatK and its targets revealed a complex pattern of possible feedback regulatory interactions. In this study, one possible feedback regulation mechanism was ruled out by the analysis of polysome associated transcripts.

Stable binding of proteins to specific RNA sites and subsequent degradation of the unprotected RNA regions can result in small RNA, footprint of the RNA binding protein. Such footprints were identified by examining small RNA datasets of *Chlamydomonas reinhardtii*. Two of the sRNAs correspond to the 5’ ends of mature *psbB* and *psbH* mRNAs. Both sRNAs are dependent on Mbb1, a nuclear-encoded TPR (Tetratrico-peptide repeat) protein. The two sRNAs have high similarity in primary sequence, and both are absent in the *mbb1* mutant. This suggests that sRNAs at the 5’ ends of chloroplast mRNAs identified here generally represent the binding sites of proteins, which function in RNA processing and RNA stabilization in *Chlamydomonas* chloroplast.

Key words: Chloroplast, Group II intron maturase, RNA binding protein, small RNA

# Zusammenfassung

In Chloroplasten durchlaufen primäre Transkripte eine großen Anzahl von bzw. Reifungsprozesse. Diese Ereignisse spielen eine wichtige Rolle bei der Regulation der Genexpression und sind im Wesentlichen durch Proteinfaktoren, insbesondere RNA-Bindeproteine, reguliert.

Der plastidäre Spleißfaktor MatK zählt zu den prokaryotischen Gruppe-II-Intron. MatK aus *Nicotiana tabacum* interagiert mit seinem Heimatintron *trnK* und sechs weiteren Gruppe IIA Introns. In dieser Untersuchung, MatK-Bindestellen konnten unterschiedlichen Regionen der Gruppe-II-Introns zugewiesen werden mit RIP-seq in *Nicotiana tabacum*. Die vorliegenden Ergebnisse zeigen, dass MatK im Vergleich zu seinen bakteriellen Vorfahren an Vielseitigkeit in der RNA-Erkennung gewonnen hat. MatK zeigt somit beispielhaft, wie eine Maturase die Fähigkeit erworben haben könnte, in trans auf mehrere Introns zu wirken.

Quantitative Untersuchung und mathematische Modellierung der Expression von MatK und dessen Zielen offenbart ein komplexes Muster möglicher regulatorischer Feedback-Mechanismen. In dieser Studie konnte ein möglicher Feedback- Mechanismus durch Analyse von polysomal gebundenen Transkripten ausgeschlossen werden.

Stabile Bindung von Proteinen an spezifische RNA-Bindestellen und anschließender Abbau der ungeschützten RNA kann zu Akkumulation von kleinen RNAs (sRNAs) führen. Solche Footprints von RNA-Bindeproteinen wurden durch die Untersuchung von Datensätzen kleiner RNAs in Chlamydomonas reinhardtii identifiziert. Zwei der sRNAs entsprechen den 5' Enden der reifen *psbB* und *psbH* mRNAs. Beide sRNAs sind abhängig von Mbb1, einem TPR (Tetratrico-peptide repeat) Protein. Die beiden sRNAs besitzen eine hohe Ähnlichkeit in ihrer Primärsequenz und fehlen in der *mbb1* Mutante. Dies legt nahe, dass auch andere der hier identifizierten sRNAs an 5' Enden plastidärer mRNAs Protein-Bindestellen repräsentieren, die für die korrekte RNA-Prozessierung und RNA-Stabilisierung in *Chlamydomonas* Chloroplasten erforderlich sind.

Schlagwörter: Chloroplasten, Gruppe II Intron Maturase, RNA-bindende Protein, kleine RNA

**Table of contents**

# 1. Introduction

## 1.1 Post-transcriptional regulation of gene expression in chloroplast

RNA molecules in plastids undergo a series of multiple-level post-transcriptional processing steps including splicing, 5' and 3' maturation, editing and finally degradation. Together with transcription and translation, these post-transcriptional steps regulate the gene expression in different layers rapidly and precisely (Deshpande et al. 1997; Baginsky and Gruissem 2002; Garcia-Andrade et al. 2013). RNA-binding proteins (RBPs) play diverse roles in these levels of post- transcriptional processes. With the exception of one chloroplast encoded factor-MatK, all RBPs are encoded in the nuclear genome and are posttranslationally imported into the chloroplasts.

## 1.2 Intergenic processing of plastid RNAs

The RNA population in plastids encompasses both primary and processed transcripts. Many genes are primarily transcribed in the form of polycistronic transcripts, a property of plastids' endosymbiont ancestor, by a bacterial-type polymerase (PEP for plastid-encoded plastid RNA polymerase) and two phage-type polymerases (NEP for nuclear-encoded plastid RNA polymerase) (Weihe 2012). These polycistronic transcripts are further processed into smaller units and monocistronic transcripts. However, the exact mechanism by which these processing sites are determined has not been known for a long time. Nonetheless, a number of RNA-binding proteins have been implicated in this process (Nickelsen 2003). A large number of proteins implicated in intergenic processing contain degenerate repeats of 34 to 38 amino acids. According to the number of amino acids in one repeat, these proteins are named TPR (tetratricopeptide repeats), PPR (pentatricopeptide repeat) and OPR (octatricopeptide repeat) proteins (Small and Peeters 2000; D'Andrea and Regan 2003; Schmitz-Linneweber and Small 2008).

In the chloroplast of *Chlamydomonas reinhardtii*, nuclear-encoded stabilization factors have been shown to bind to the 5'UTR region of the transcripts and protect them from degradation. Nuclear proteins Nac2 and MCD1 determine the stability of the *psbD* and *petD* transcripts by acting on their respective 5'UTR (Drager et al. 1998; Nickelsen et al. 1999; Vaistij et al. 2000; Bollenbach and Stern 2003; Murakami et al. 2005; Loiselay et al.

2008). Many of such factors belong to the aforementioned classes of TPR, PPR or OPR proteins (Auchincloss et al. 2002; Rahire et al. 2012).

Mbb1 is a tetratricopeptide-like protein with 10 tandemly arranged repeats in *Chlamydomonas*. It is encoded in the nucleus and acts as an RNA binding protein in the chloroplast. In the *mbb1* mutant 222E, the *psbB/T/H* transcription unit fails to accumulate (Monod et al. 1992). Mbb1 acts directly on the 5'UTR of *psbB* (Vaistij et al. 2000). Although orthologs of *Chlamydomonas* proteins can rarely be found in higher plants, HCF107 (high chlorophyll fluorescence 107), the ortholog of Mbb1 in *Arabidopsis thaliana*, is required for *psbB* and *psbH* expression (Felder et al. 2001; Sane et al. 2005). The 3'and 5' maturation is a result of RNA protection from digestion of ribonuclease, including endoribonucleases and exoribonucleases (5' to 3' or 3' to 5') (Stoppel and Meurer 2011). By binding to its 5' end *in vitro*, recombinant HCF107 can protect *psbH* from the 5' to 3' exonucleases degradation. Similarly, the 3' to 5' exonucleases could also be blocked by HCF107. *In vivo*, degradation by these two types of exonucleases would, in theory, leave a small RNA (sRNA) protected by HCF107. Indeed, in maize, a sRNA corresponding to the *psbH* 5' end can be detected (Hammani et al. 2012).

While the OPRs are expanded in *Chlamydomonas*, PPRs are prominent in terrestrial plants; for example, 450 PPRs are found in *Arabidopsis thaliana*. Mutants of PPR proteins are defective in splicing, editing, stabilization, or translation of subsets of organellar RNAs, as indicated in the review (Schmitz-Linneweber and Small 2008). An amino acid code for PPR proteins binding to single strand RNA (ssRNA) was suggested and confirmed *in vitro*. Two amino acids in each PPR repeat were shown to determine the binding specificity (Barkan et al. 2012; Ke et al. 2013). PPR10 is an extensively studied protein. It binds to *atpI-atpH* and *psaJ-rpl33* intergenic regions and defines the 5' and 3' terminal positions of processed transcripts (Pfalz et al. 2009). Pfalz *et al*. described a sRNA in maize, which represents the footprint of PPR10 and overlaps with the ends of the processed transcripts. The binding of an RBP or a presence of stable RNA secondary structure could lead to sRNA in the corresponding region, and sRNAs were found to coincide with the processed 5' or 3' transcript ends (Ruwe and Schmitz-Linneweber 2012; Zhelyazkova et al. 2012b). As a result, the presence of a sRNA is an indicator of the existence of an RBP binding site (Ruwe and Schmitz-Linneweber 2012; Zhelyazkova et al. 2012b).

Based on these observations, a model for processing in plastids was suggested, whereby endoribonucleases first cleave RNA sequences that are unstructured or unprotected by

proteins. The cleavd products serve as substrates for exonucleases, and nucleotides at either the 5' or 3' end are progressively removed until the point is reached where the sequences are blocked by the structure or an RBP (Pfalz et al. 2009; Prikryl et al. 2011). According to this model, there is a very fine boundary between intergenic processing and RNA degradation.

## 1.3 Group II intron splicing

### 1.3.1 Group II intron

Group II introns are large catalytic RNAs, which are often referred to as ribozymes. They are usually found in bacteria, lower eukaryotes and organellar genomes of plants. There are about 20 group II introns in the chloroplast genome of higher plants. Despite the diversity of the primary sequences, group II introns are characterized by their self-splicing ability and formation of a secondary structure of six double helical domains (DI-DVI) radiating out from a central core (Michel et al. 1982; Michel and Dujon 1983). Group II introns and nuclear spliceosome pre-mRNAs are similar in their splicing mechanism and the formation of lariat intronic RNA (Butcher and Brow 2005; Marcia and Pyle 2012).

The six domains of group II introns have different functions: DI is the largest domain and potentially builds the scaffold for folding; DII is not conserved phylogenetically and its deletion scarcely affects splicing efficiency; DIII could enhance the catalyzation step; DIV is variable in length and sequence. This domain usually encodes a protein (intron encoded protein, IEP, also called maturase), which exhibits both splicing and retrohoming activities in bacteria and some organelles; DV is the most conserved domain and is critical for the catalytic reaction; and DVI contains the branch adenosine (branch "A") which is necessary for lariat RNA (branched circular RNA) formation during splicing (Qin and Pyle 1998). Along with these common features, several exceptions are noteworthy. For example, *trnV* in plant chloroplasts lacks the branch adenosine and is spliced in a hydrolytic way (Vogel and Borner 2002). Based on their structural features, group II introns can be further classified into group IIA, IIB (Michel et al. 1989), IIC and IID (Toor et al. 2001). Novel lineages IIE and IIF were also suggested in bacteria based on the ongoing identification of new group II introns and bioinformatic analyses (Nagy et al. 2013). The major structural forms of group II intron RNAs (IIA and IIB) in bacteria

and organelles developed through coevolution with their respective maturase rather than as independent catalytic RNAs (Toor et al. 2001).

*In vitro* self-splicing of group II introns under a high $Mg^{2+}$ concentration condition was first reported in 1986 (Peebles et al. 1986; van der Veen et al. 1986). These initial studies characterized the formation of a circular product-lariat. A self-splicing model with two transesterification steps was also proposed based on the results. More specifically, in the first step, the 2' OH of a bulged adenosine in DVI attacks the 5' splice site, resulting in the cleavage of 5' exon and the formation of a lariat RNA intermediate. In the second step, 3' OH of the upstream exon nucleophilically attacks the 3' splice site, resulting in ligated exons and a release of the intron as a lariat (Sharp 1987).

Group II introns have to fold into a correct conformation before performing the catalytic reaction. $Mg^{2+}$ was shown to be required in intron folding. All intron regions fold uniformly at the same $Mg^{2+}$ concentration (Swisher et al. 2002). A further study showed that folding of D1 is the rate limiting step of the whole RNA folding, eventhough this folding step requires the same salt concentration as the whole RNA folding (Su et al. 2005).

### 1.3.2 Yeast and bacterial maturases

The most extensively studied group II introns are intron of *Ll.LtrB* in *Lactococcus lactis* (Mills et al. 1996) as well as *al1* and *al2* in yeast mitochondria (mt) (van der Veen et al. 1986). Although group II introns can self-splice in high $Mg^{2+}$ concentrations *in vitro*, the addition of protein factors could promote the splicing in low $Mg^{2+}$ concentration conditions (Saldanha et al. 1999). Moreover, proteins are required for catalyzing the splicing reaction *in vivo* (Lambowitz and Belfort 1993; Wank et al. 1999). In yeast and bacteria, each group II intron encodes a maturase mostly in DIV. The maturase contains four domains with different functions: reverse transcriptase (RT) domain for intron duplication, DNA binding domain and endonuclease domain for site-specific cleavage. These three domains are necessary for retrohoming, while the fourth X domain is needed for promoting splicing (Figure 1) (Mohr et al. 1993).

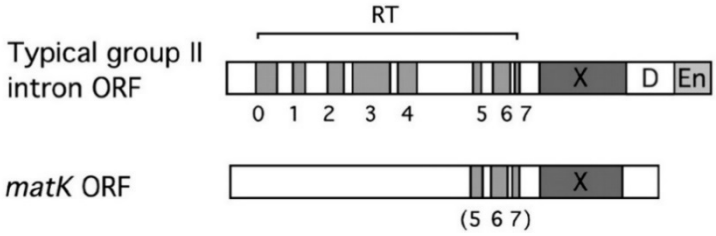



**Figure 1: Structures of a typical group II intron ORF and *matK* ORF.** A typical group II intron ORF contains four domains, reverse transcriptase domain for intron duplication, DNA binding domain (D) and endonuclease domain (En domain) for site-specific cleavage. These three domains are necessary for retrohoming while the fourth X domain is needed to promote RNA binding and splicing; *matK* ORF has only remnants of RT domain and a well-conserved domain X. The schematics are derived from the *Lactococcus lactis Ll.ltrB* intron (U50902) and *Arabidopsis trnK1*I1 (NC_000932), from (Hausner et al. 2006).

Yeast mtDNA contains two group II introns in some alleles of *cox1* gene: *aI1* and *aI2*. They have a sequence identity of 50%. The open reading frames of their maturases extend to the upstream exons and the maturases are generated after proteolytic processing (Carignani et al. 1983; Moran et al. 1994). Group II intron in yeast or bacteria could be excised as a donor sequence and inserted into a specific reception. This process is known as retrohoming (Cousineau et al. 1998; Yang et al. 1998; Watanabe and Lambowitz 2004). Owing to their high mobility, group II introns can be used in other applications, such as site-specific deletion and insertion (Enyeart et al. 2014).

LtrA is the maturase of *Lactococcus lactis LtrB*. It is encoded entirely in DIV which also shows the highest affinity of binding to LtrA. This binding is important for splicing, retrohoming and regulation (Watanabe and Lambowitz 2004). The deletion of DIV decreases the binding affinity of LtrA to RNA by five orders of magnitude, while also dramatically decreasing the splicing efficiency (Wank et al. 1999). LtrA was found to bind primarily to DIV, after which it makes further contact with upstream DI and downstream DVI. In this way, this maturase promotes the folding of group II introns to the splicing structure (Matsuura et al. 2001). Further *in vitro* mutagenesis studies determined a stem-loop at the distal end of DIV. As the LtrA key binding site, it is a 25 nt minimal binding region containing the Shine-Dalgarno (Keene et al.) sequence and the start codon of LtrA (Singh et al. 2002; Keene et al. 2006). Both the sequence and the RNA structure of this binding site are important for LtrA recognition (Matsuura et al. 2001). The binding of LtrA to a number of bases in the SD sequences implies autoregulation of its expression (Matsuura et al. 2001). An autoregulation model was

suggested based on the fact that binding of LtrA to this region down-regulates LtrA translation. According to this premise, the binding of LtrA to the SD region competes with the ribosome loading and/or the binding stabilizes the RNA structure and thus inhibits the translation. This model was tested with a lacZ report vector in *E. coli*. Indeed, the presence of LtrA was found to down regulate lacZ expression (Singh et al. 2002).
At least three conformation statuses of LtrA were found based on the limited proteolysis and fluorescence quenching experiments. The domains of LtrA undergo conformation change in the presence of DIV of RNA (Rambo and Doudna 2004). The first interaction between the LtrA protein and group II introns is encountered as a result of the attraction of positively charged residues in the protein to the negatively charged RNA. Further specific binding of LtrA was approached via the recognition of the DI and DIV. This study further revealed that the formation of maturase dimmer offers the possibility of making contact with two domains (Rambo and Doudna 2004). In contrast, findings of another study that applied the same method under a different experimental condition suggested that the formation of protein conformation is not dependent on the presence of RNA (Blocker et al. 2005).

### 1.3.3 Plant maturases in chloroplast

During coevolution between the host and the endosymbiont, the endosymbionts degenerated into organelles, the plastid and mitochondrion. Most of the endosymbiont genes were transferred to the host nuclear genome, as were the maturases (Ayliffe et al. 1998; Rousseau-Gueutin et al. 2012; Yoshida et al. 2014). In recent years, significant number of nuclear maturases have been identified. However, maturase can still be found in organelles of plant and mitochondria of fungi (de Longevialle et al. 2010). The only retained maturase in the chloroplast is MatK, which is encoded in *trnK* intron, while in mitochondrion, the only maturase, MatR, is encoded in *nad1* intron 4 (Mohr et al. 1993). Splicing of organellar RNA is catalyzed by organellar maturases and nuclear splicing factors that with organelle localization (de Longevialle et al. 2010; Brown et al. 2014).

#### 1.3.3.1 Group II intron and MatK

Thus far, 20 group II introns have been found for *Arabidopsis thaliana* and *Nicotiana tabacum* (Shinozaki et al. 1986), while 17 group II introns have been identified for *Zea mays* and *Oryza sativa* (Michel et al. 1989). To date, many nuclear-encoded plastid

splicing factors have been characterized. Their roles in splicing plastid group II introns are summarized in Figure 2. They belong to different protein families such as PPR protein and CRM (Chloroplast RNA splicing and ribosome maturation) domain protein (de Longevialle et al. 2010; Asakura et al. 2012; Hammani and Barkan 2014). MatK is likely to serve the splicing role for several group II introns (Zoschke et al. 2010). Different from the bacterial maturase, MatK only retains the conserved X domain and the partial reverse transcriptase domain (Figure 1 and (Mohr, 1993)).

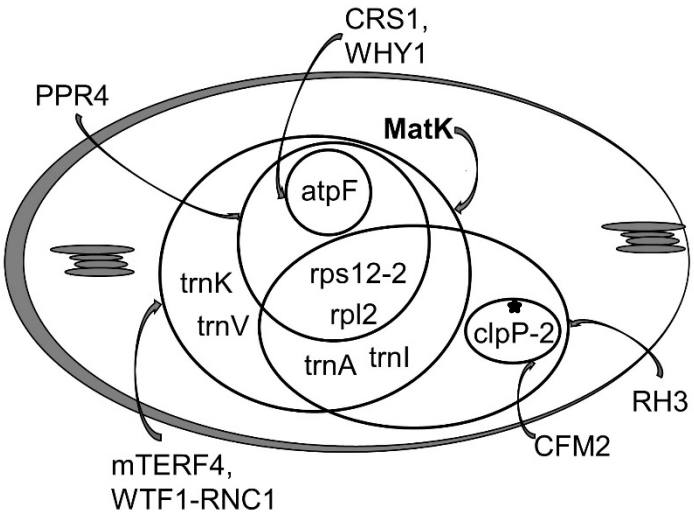



**Figure 2: Group IIA introns in chloroplast and their splicing factors.** The group IIA and IIB introns are classified according to Michel (Michel et al. 1989). The intron that is found in *Arabidopsis thaliana* but not in maize and rice is marked with asterisks. PPR4 is a PPR protein, CRS1 and CFM2 are CRM domain proteins, WTF1 is a PORR (Plant Organelle RNA Recognition) domain protein, mTERF4 belongs to the mTERF (mitochondrial transcription termination factor) protein family, RH3 is a member of DEAD box RNA helicase family and WHY1 is a 'Whirly' protein. The only chloroplast maturase-MatK, is depicted in bold, modified from de Longevialle (de Longevialle et al. 2010).

### 1.3.4 Spliceosomal splicing is believed to descend from group II splicing

Spliceosome is a uniform RNA-protein composition that catalyzes the splicing of most nuclear pre-mRNA introns. Same as group II introns, the splicing reaction consists of two consecutive transesterification steps. The assembly of spliceosome is based on snRNPs (small nuclear ribonucleoprotein particles) U1, U2, U4/U6 and U5, and numerous protein factors (reviewed by Matera and Wang (Matera and Wang 2014)). This assembly process commences with the recognition of U1 snRNP to the 5' splicing site (5' SS), after which the intron sequence flanking branch point is recognized by U2 snRNP. Subsequently, a

number of RNA and protein factors are recruited, including a snRNP complex consists U5, extensively base-paired U4 and U6, as well as their corresponding protein factors. Finally, conformational rearrangement occurs to facilitate the splicing. U6 snRNA has been shown to play the same role as DVI of the group II intron (Peebles et al. 1995), whereby both could form stem-loop and interact with metal ion (Yean et al. 2000; Gordon and Piccirilli 2001). Functional similarity between U6 and DVI was proven by the fact that the replacement of U6atac snRNA with DVI can still promote the *in vivo* splicing (Shukla and Padgett 2002). Prp8 is one of the U5 snRNP proteins that with a size of 280kDa in yeast. The sequence of its central domain resembles the reverse transcriptases domain of the group II intron maturases. The Prp8 palm-fingers region (designated Th/X) was also speculated to be functionally equivalent to the X/maturase domain in group II maturases (Dlakic and Mushegian 2011). The crystal structure of a large piece of a Prp8 fragment is similar to the bacterial group II intron reverse transcriptase, revealing that the spliceosome and bacterial group II intron splicing pathways share the same evolutional origin. Moreover, Prp8 forms a large cavity based on its reverse transcriptase thumb, the endonuclease-like and RNaseH-like domains. The space in this cavity is sufficient for accommodating the catalytic core of group II intron RNA (Galej et al. 2013).

### 1.3.5 Distribution of MatK and group II introns suggests the splicing role of MatK

MatK is present in all the chloroplast genomes of autotrophic land plants that contain group II introns, and is also found in charophycean green algae (Turmel et al. 2006). The location of *matK* in the intron of *trnK* suggests its function in splicing its home intron, in the manner akin to that of its bacterial counterparts. However, *matK* is present as a stand-alone reading frame, in the species where the *trnK* gene has been lost, such as streptophyte algae *Zygnema*, the fern *Adiantum capillus-veneris* and also the parasitic land plants *Epifagus virginiana* and four *Cuscuta* species (*C. exaltata, C. reflexa, C. campestris* and *C. obtusiflora*) (Wolfe et al. 1992; Turmel et al. 2005; Funk et al. 2007; McNeal et al. 2007; Braukmann et al. 2013). As the loss of *trnK* is not in line with that of *matK*, the latter is likely to have other functions and, it most likely targets other introns. To date, the loss of *matK* has only been observed in *Rhizanthella gardneri* of mycoheterotrophic orchid and some members of *Grammica* of Cuscuta subgenus from the parasitic angiosperm. These species also lost several group IIA introns that may require activity of MatK (McNeal et al. 2009; Delannoy et al. 2011; Braukmann et al.

2013). The evidence presented above supports the hypothesis that MatK splices other introns, rather than only its home intron.

### 1.3.6 MatK and its targets, association and splicing

MatK is significantly similar to bacterial maturase. The direct binding of MatK to introns was demonstrated *in vitro* (Liere and Link 1995). Via RIP- Chip analysis, seven out of eight group II introns were identified (*trnK, trnV, trnI, trnA, rps12-in2, rpl2* and *atpF*) to be associated with MatK in chloroplast, while no group IIB intron was found to be associated. This association was verified by slot-blot analysis. Further microarray analysis narrowed down the binding site for MatK to the intron region of *trnK* excluding the entire *matK* ORF, DV and parts of DVI (Zoschke et al. 2010).
Microarray analysis was performed with other targets of MatK, and an association between MatK and DIV was found for *trnA, trnI* and *atpF*. However, no enrichment was found by microarray for *rpl2, trnV* and *rps12-in2* (Neumann 2011). In contrast to the bacterial group II introns, which are able to disperse to different genome locations, chloroplast group II introns of embryophytes are no longer mobile. In accordance with this feature of the chloroplast group II introns, the putative splicing factor, MatK, lost the protein domains required for intron mobility (Mohr et al. 1993)(Mohr, Perlman, and Lambowitz 1993; Barthet and Hilu 2007). The role of MatK in splicing was studied with the barley albostrian that lack the functional ribosomes for translation in plastids. It was found that neither the *trnK* precursor nor any other member of the subgroup IIA is spliced in this strian (Hess et al. 1994; Vogel et al. 1997; Vogel et al. 1999).

### 1.3.7 Autoregulation of maturases

The aforementioned bacterial maturase LtrA autoregulates its own expression via feedback loop circuits mediated by a stem-loop structure at its start codon region. Multiple negative feedback loops were also found in the chloroplast of *Chlamydomonas* (Ramundo et al. 2013). In higher plants, the large subunit of Rubisco was shown to undergo autoregulation of translation (Wostrikoff and Stern 2007). The maintenance of *matK* in chloroplast might be a prerequisite for a similar autoregulation mechanism. A stem-loop structure was found in the *matK* start codon region in *Zea mays*. To examine the autoregulation *in vivo*, a reporter gene, driven by different lengths of 5'UTRs surrounding the *matK* start codon, was coexpressed with MatK in a heterologous *E. coli*

expression system. No difference in reporter gene expression was observed either before or after the induction of MatK expression, indicating that, in this heterologous system, chloroplast MatK does not influence its own gene expression via its 5'UTR region, which might also be the case in higher plants (Zoschke et al. 2009).

The regulatory role of protein can be examined by mutagenesis. However, *matK* knock out by insertion mutagenesis only results in heteroplastomic lines (Drescher 2003). An attempt to produce MatK with lower activity by introducing point mutants was also unsuccessful (Zoschke et al. 2010). As these results suggest the importance of MatK, subsequently studies were performed without disrupting the *matK* expression (Zoschke et al. 2010). Quantitative analyses of protein and RNA levels of MatK demonstrated a strong discrepancy between them through tobacco development, especially in the early development stages. The most pronounced discrepancy was found in 7-day-old tissue, as the MatK protein reached the highest level while the RNA was almost undetectable (Hertel et al. 2013). This discrepancy indicates that *matK* expression is not simply determined by the transcript amount, but that a regulation in translation or protein stability may also play a role. With the additional information of the dynamic interaction between MatK and its targets during development, a mathematical model with feedback regulation was generated. This model reflects the characteristics of the MatK gene expression network. Accordingly, MatK protein autoregulates its own translation via the formation of pre-*trnK*/MatK repression complexes (Hertel et al. 2013).

## 1.4 RNA-Protein association analysis

The control of post-transcriptional level is largely relies on RBPs. Some RBPs modulate a single RNA. For example, the PPR protein HCF152 binds to the *psbH–petB* intergenic region in Arabidopsis (Meierhoff et al. 2003), yet some RBPs, such as hnRNP-like proteins, interact with multiple RNA targets (Yeap et al. 2014). In order to understand the manner in which RBPs control RNAs, the first step is to identify the RNA targets of RBPs. A variety of methods have been applied to achieve this goal. RNAs bound to RBP can be identified by *in vitro* methods such as RNA electromobility shift assays (REMSAs) (Hellman and Fried 2007). The yeast three hybrids system is another solution, with the aim of analyzing RNA-protein interaction *in vivo* (SenGupta et al. 1996). More recently, several newly developed high-throughput methods were utilized in large-scale analyses. These studies included systematic evolution of ligands by exponential enrichment

(SELEX) (Ogawa and Biggin 2012), RIP-chip (RNA immunoprecipitation and Chip-hybridization) (Keene et al. 2006), RIP-seq (Zhao et al. 2010) and PAR-Clip (Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation) (Hafner et al. 2010), or related techniques (HITS-Clip, etc.) (Licatalosi et al. 2008; Konig et al. 2011). Although each method has its advantages and disadvantages, combining genetic and biochemical approaches with computational analysis is typically required when applying these methods.

## 1.5 The aim of this work

The identification of protein-bound RNA targets has been widely carried out using various methods. MatK, the only chloroplast maturase, has been shown to be associated with seven group II introns. This leads to the need to understand how a maturase can recognize introns with diverse sequences. Determining the binding preference of MatK to different domains of group II introns is thus the main goal of this work. In this study, the newly developed RIP-seq method will be used to analyze the binding sites for MatK within HA-tagged and non-tagged MatK of *Nicotiana tabacum*; using this method, a quantitative analysis of the binding will be performed simultaneously.

Thus far, the function of MatK in chloroplast intron splicing has only been implied based on the binding data, phylogenetic data and pharmacological considerations. There is functional similarity between MatK and bacterial group II maturase. Moreover, bacterial maturase LtrA was proved to be autoregulated, and the available data indicates that MatK may possess this ability. Consequently, further investigation of MatK regulation is an important aspect of this work. As the knockout of MatK has failed in previous studies, in this work, an attempt to overexpress MatK has been made, both in chloroplasts and in the nucleus. Overexpression might help obtain more direct evidence for the function of splicing, and the nuclear overexpression of MatK may be able to disrupt its potential autoregulation circuits.

Two autoregulation models have been proposed based on the characteristics of MatK: (1) Binding of MatK to the stem-loop in its own start codon region inhibits the translation; and (2) The *trnK* precursor serves as the translational template of MatK, whereby splicing of this template decreases the MatK translation. The first model has been tested in an *E.coli* system and the results indicated that MatK is insufficient to repress its expression

in this system (Zoschke et al. 2009). Here, the second model will be studied by identifying the translational-active transcripts.

In the second part of this thesis, attempts are made to analyze the binding sites of RBPs via a different approach. According to the view that sRNAs are the footprints of RBPs, *Chlamydomonas* sRNA data sets were examined in this study for the presence of chloroplast sRNAs. The well-studied *mbb1* mutant and wild-type *Chlamydomonas* were used to determine the relationship between MBB1 and the sRNAs found in the region of the MBB1 targets.

# 2. Material and methods

## 2.1 Material

### 2.1.1 Plant material and bacterial strain

*Nicotiana tabacum* lines of Hemagglutinin (HA)-tagged *matK* at N or C terminus (HA:MatK or MatK:HA) and control lines were prepared by Reimo Zoschke (Zoschke et al. 2010)

Bacterial strain TOP10 from Invitrogen and SURE from Agilent Technologies were used for plasmids propagation.

Total *Chlamydomonas* RNA samples from wild-type and *mbb1* mutant strains were kindly provided by Prof. Michel Goldschmidt-Clermont.

### 2.1.2 Antibodies and oligos

Mouse anti–HA antibody (Sigma, H3663) and mouse anti–HA-HRP (Sigma, H6533) antibody were used as 1:500 dilutions for Western blot.

See Appendix 1 for oligos used in this study.

### 2.1.3 Plasmids

PGL1 vector was kindly provided by Prof. Bernhard Grimm. pGW1 and pAV6 (Verhounig et al. 2010) vectors were kindly provided by Prof. Ralph Bock.

### 2.1.4 Plant culture material and medium

Soil for plant culture was from Einheits Erde

MS-Medium liquid: 0.44% (w/v) Murashige & Skoog Medium (Duchefa) 0.05% (w/v) MES, 3% (w/v) sucrose; adjust to pH 5.8 with KOH

MS-Medium plate: MS-Medium liquid with 0.5% (w/v) plant agar (Duchefa)

MG plates: same as MS plates, but with 16 g/L glucose instead of sucrose

Co-Cultivation medium: 2MS plates with 0.2 mg/L NAA and1 mg/L BAP

Callus induction medium: 2MG plates with 0.2 mg/L NAA, 1 mg/L BAP and appropriate antibiotics (2 mg/L Basta for pGL1 vector)

Root induction medium: 2MG plates with 0.1 mg/L NAA and appropriate antibiotics (2 mg/L Basta for pGL1 vector)

MS-Medium with Spectinomycin: MS-Medium with Spectinomycin (500 µg/ml, Duchefa)

RMOP-Medium: MS-Medium with spectinomycin, N6-Benzyladenine (1 mg/l, Sigma) and 1-Naphthaleneacetic acid (0.1 mg/l, Sigma) (Svab et al. 1990)

### 2.1.5 Bacteria medium

LB-medium: LB-medium 3.5% (W/V, Sigma)

LB-medium with agar: LB-medium with agar 3.5% (w/v, Sigma)

SOB-Medium: 2% (w/v) Trypton, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, adjust to pH 7.0 with NaOH, sterilize by autoclaving; add autoclaved 10 mM $MgCl_2$ before use (Hanahan 1983)

SOC-Medium: SOB-Medium add 20 mM Glucose and 10mM $MgCl_2$ (Hanahan 1983)

YEB medium: 5 g/L beef extract, 1 g/L yeast extract, 5 g/L peptone, 5 g/L sucrose, 2 mM/L $MgSO_4$, adjust to pH 7.0

### 2.1.6 Buffers and Solutions

Church hybridization buffer: 7% SDS, 0.5 M Sodiumphosphate (pH 7.0), 1 mM EDTA

Co-IP buffer: 0.15 M NaCl, 20 M Tris-HCl (pH 7.5), 2 mM $MgCl_2$, 0.5% NP-40 (v/v)

2× CTAB: 2% CTAB (w/v), 100 mM Tris-HCl (pH 8.0), 20 mM EDTA (pH 8.0), 1.4 M NaCl

Denaturing polyacrylamide gel (in TBE): 8 M Urea, 1× TBE, 15% Acrylamide, 0.5 ‰ APS and 5 µl TEMED for each gel

Dot blot sample buffer: 66% deionized Formamide, 21% Formaldehyde, 1.3× MOPS (pH 8.0)

ECL detection solution: 100 mM Tris-HCl pH 8.5, 1.25 mM Luminol in DMSO, 0.45 mM p- Coumaric acid in DMSO, 0.01% $H_2O_2$

EX buffer: 0.2 M KAc, 30 mM HEPES-KOH (pH 8.0), 10 mM $MgAc_2$, 2 mM DTT, before use, add 1× protease inhibitor (Protease Inhibitor Cocktail Tablets, Roche), 0.4 mM PMSF, 0.1 µg/ml Aprotinin

Extraction buffer (for polysome analysis): 200 mM Tris-HCl, 0.2 M KCl, 0.2 M Sucrose, 35 mM $MgCl_2$, 25 mM EGTA, 1% Triton X-100, 2% Polyoxyethylene-10, before use, add 71 ‰ 2-Mercaptoethanol, 0.12 mg/ml Chloramphenicol, 0.5 mg/ml Heparin

Extraction buffer without detergent (for polysome analysis): Extraction buffer without Triton X-100, Polyoxyethylene-10 and 2-Mercaptoethanol

Formaldehyde: 36.5-38% (w/v) Formaldehyde, 10-15% (w/v) Methanol (premixed, Sigma)

Grinding buffer: 50 mM HEPES-KOH (pH 8.0), 330 mM Sorbitol, 2 mM EDTA, 1 mM $MgCl_2$, 1 mM $MnCl_2$, 0.25% BSA, 1.5 mM Sodium ascorbate

Homogenization buffer: 10% sucrose, 0.1 M Tris-HCl (pH 7.2), 5 mM EDTA, 5 mM EGTA, add fresh 2 µg/ml Aprotinin, 40 mM β-mercaptoethanol and 2 mM PMSF

HS buffer: 50 mM HEPES-KOH (pH 8.0), 330 mM Sorbitol

Methylene blue: 0.3 M NaAc (pH 5.2), 0.03% (w/v) Methylene Blue

10× MOPS: 2 M MOPS, 0.8 M NaAc, 0.1 M $Na_2EDTA$, adjust to pH 7.0

Northern sample buffer: 65% deionized Formamide, 22% Formaldehyde, 13% 10× MOPS (pH 7.0)

Northern loading buffer: 50% Glycerol, 10% 10× MOPS (pH 7.0), 10 mM EDTA (pH 8.0), 0.25% (w/v) Bromphenol Blue, 0.25% (w/v) Xylene cyanol; sterile filtered

Northern running buffer: 88.5% $H_2O$, 10% 10× MOPS (pH 7.0), 1.5% Formaldehyde

Northern transfer buffer: 5× SSC

10× polysome gradiant salts: 0.4 M Tris (pH 8.0), 0.2 M KCl, 0.1 M $MgCl_2$

Ponceau stain solution: 30% (w/v) Trichloroacetic acid, 30% (w/v) 5-Sulfosalicylic acid, 2% (w/v) Ponceau S (Merck)

3× protein loading buffer: 30% Glycerol, 15% 2-MeEtOH, 7% (w/v) SDS, 200 mM Tris-HCl (pH 6.8), 0.25% (w/v) Bromphenol Blue

5× protein transfer buffer: 1 M Glycin, 125 mM Tris

1×protein transfer buffer: 20% 5× transfer buffer, 20% Methanol

Resolving gel: 13% Acrylamid, 375 mM Tris-HCl (pH 8.8), 0.1% (w/v) SDS, 0.1% (w/v) APS, 0.05% TEMED

10× SDS running buffer: 1.92 M Glycin, 250 mM Tris-Base, 1% (w/v) SDS

20× SSC: 3 M NaCl, 0.3 M Sodium citrate, adjust to pH 7.0 with HCl

Stacking gel: 5% Acrylamid, 130 mM Tris-HCl (pH 6.8), 0.1% (w/v) SDS, 0.1% (w/v) APS, 0.1% TEMED

50× TAE: 2 M Tris-Acetate (pH 8.0), 50 mM EDTA (pH 8.0)

10× TBE: 1.1 M Tris, 900 mM Boric acid, 25 mM EDTA, adjust to pH 8.3 with HCl

10× TBST: 0.5 M Tris-HCl (pH 7.5), 1.5 M NaCl, 1% Tween 20

Sterilization buffer: 15% DanKlorix (Colgate-Palmolive), 0.03% Tween 20

### 2.1.7 Providers of chemicals and equipments

| | |
|---|---|
| Abcam | Abcam Inc., Cambridge, MA, USA |
| Ambion | Invitrogen Life Technologies GmbH, Darmstadt, Germany |
| Amersham Biosciences | Amersham Biosciences Europe GmbH, Freiburg, Germany |
| Applied Biosystems | Applied Biosystems Inc., Foster City, CA, USA |
| Beckman | Beckman Coulter Inc., Fullerton, CA, USA |
| Bio-Rad | Bio-Rad Laboratories, Hercules, CA, USA |
| Biozym | Biozym Scientific GmbH., Hessisch Oldendorf, Germany |
| Calbiochem | Merck Chemicals GmbH., Hessen, Germany |
| Carl Roth | Carl Roth GmbH & Co. KG, Karlsruhe, Germany |
| Colgate-Palmolive | Colgate-Palmolive GABA GmbH, Hamburg, Germany |
| Duchefa | Duchefa Biochemie B.V., Haarlem, Niederlande |
| Einheits Erde | Einheitserde- und Humuswerke Gebr. Patzer GmbH & Co. KG, Sinntal-Jossa, Germany |
| Epicentre | Epicentre Biotechnologies, Madison, USA |
| Franz Eckert GmbH | Franz Eckert GmbH, Waldkirch, Germany |
| GE Healthcare | GE Healthcare Germany, Munich, Germany |
| Harnishmacher | Harnischmacher Labortechnik, Germany |
| Hartmann-Analytics | Hartmannnanlytic GmbH., Braunschweig, Germany |
| Heraeus | Heraeus, Hanau, Germany |
| Millipore | Millipore Corp., Bedford, MA, USA |
| Miltenyi Biotec | Miltenyi Biotec GmbH., Bergisch Gladbach, Germany |
| NEB | New England Biolabs, Ipswich, MA, USA |
| peqlab | peqlab Biotechnologie GmbH, Erlangen, Germany |
| QIAGEN | QIAGEN GmbH, Hilden, Germany |
| Retsch | Retsch GmbH, Haan, Deutschland |
| Roche | Roche Diagnostics GmbH, Mannheim, Deutschland |
| Roth | C. Roth GMBH & Co, Karlsruhe, Deutschland |
| Sigma | Sigma-Aldrich Corporation, St. Louis, MO, USA |
| Thermo | Thermo Fisher Scientifc Inc. Waltham MA, USA |
| Whatman | Whatman Paper, Maidstone, Großbritannien |

### 2.1.8 Softwares

| Programm | supplier or website |
|---|---|
| Acrobat | Adobe Professional |
| BLAST | http://www.ncbi.nlm.nih.gov/BLAST |
| Mfold | http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1.cgi |
| Oligo caculator | http://www.basic.northwestern.edu/biotools/oligocalc.html |
| TM caculator | Thermo scientific<br>http://www.thermoscientificbio.com/webtools/tmc/ |
| Quantity One | Bio-Rad |
| Imagelab | Bio-Rad |
| SnapGene | GSL Biotech |
| CLC workbench | CLC Bio |
| EndNote | Thomson Reuters |
| Office | Microsoft |

## 2.2 Methods

### 2.2.1 Cultivation of *Nicotiana tabacum*

*Nicotiana tabacum* plants were grown on soil under long day conditions (16 hours light, 8 hours dark) at 27°C with light intensities of approximately 300 µmol. $m^{-2}$. $s^{-1}$. For subsequent acquisition of 7-day-old tissue, seeds were sown on polyamide nets (mesh size 500 µM, Franz Eckert GmbH) on soil. Both seed sowing and seedling harvesting were performed at 10:00 am.

### 2.2.2 Chloroplast isolation and stroma extraction

The procedure of chloroplast isolation from tobacco was modified from Voelker (Voelker and Barkan 1995) and was performed at 4 °C. Seedlings of 7-day-old tobacco (without roots) were taken from 3-4 plates (d = 14 cm) and were homogenized with a waring blender in 350 ml grinding buffer, once for 5 s at low speed, and twice more for 5 s at the high speed setting. Next, the mixture was filtrated through one layer of MicroCloth (Calbiochem) and centrifuged at 1000 g for 6 min and, after which the pellet was resuspended in 1-2 ml of HS buffer. After adding 3-4 volumes of HS buffer, it was

transferred to corex tubes, and centrifuged at 1000 g for 6 min. The supernatant was discarded again, 200-400 µl EX buffer was added to the pellet, and the mixture was transferred to 1.5 ml eppendorf tube and squeezed through syringes (needle: 0.55 × 25 mm) for ~40 times in order to crack chloroplasts. Finally, cracked chloroplasts were centrifuged at 30,000 g for 30 min to separate the membrane and stroma fractions. The stroma fraction was used for immunoprecipitation (IP) of MatK. For longer storage, glycerol was added to the stroma fraction to a final concentration of 10% and samples were stored at -80 °C.

### 2.2.3 Co-Immunoprecipitation (co-IP) of MatK and rpL32

Co-IP was performed at 4 °C. For each MatK co-IP, two volumes of co-IP buffer and 5 µg of mouse anti–HA antibody (Sigma) were added to the stroma fraction containing approximately 200 µg of protein. The stroma-antibody mixture was vertical rotated at 12 rpm for 1 hour, after which 50 µl Magnetic Beads (Life Technologies) were added and the reaction was rotated for one more hour. Next, the beads were pelleted on a magnetic rack, and the supernatant was taken for RNA extraction and Western blot. Finally, IP pellet was washed 3 times with co-IP buffer, and 200 µl EX buffer was added to the IP pellet before storage or RNA extraction. The same fraction (1/50) was aliquoted from both IP supernatant and pellet for Western blot.

GFP tagged rpL32 was immunoprecipitated with µ MACS column (Miltenyi Biotec). Firstly, the MACS anti-GFP microbeads were equilibrated at room temperature with 200 µl extraction buffer with detergents. Next, 200 mg of frozen leaf material was homogenized before mixing with 1330 µl of the extraction buffer. This was followed by centrifugation for 5 min at 13,200 g and 4 °C to remove the cell debris and nuclei. Aliquots were taken from the supernatant for RNA extraction (200 µl) and Western blot (10 µl), while the remaining quantity was used for IP (~1130 µl). 1/20 volume of 10% sodium deoxycholate (SDC) was added to the remaining fraction and incubation was performed on ice for 5 min. After centrifugation for 15 min at 13,200 g at 4 °C to remove insoluble material, 40 µl of anti-GFP µbeads was added to the supernatant, followed by incubation for one hour at 4 °C, while shieding the reaction from light. The reaction mixture was loaded onto pre-equilibrate column and the flow-through was collected. The colume was washed 3 times (first, with 500 µl of extraction buffer with detergents; second, with 300 µl of extraction buffer with detergents; third, with 300 µl of extraction buffer without detergents). The column was eluted with elution buffer that was preheated

to 95 °C (100 µl elution buffer for each IP; 20 µl was first added and incubated for 5 min before adding further 80 µl). Both the flow-through and the elution fractions were aliquoted for Western blot and RNA extraction.

### 2.2.4 DNA and RNA extraction

The isolation of total DNA from tobacco followed the CTAB method (Stewart and Via 1993). Plant material was homogenized together with a pre-cooled metal ball (d = 5 mm) in 2 ml tube using the Retsch apparatus. Next, 2× CTAB was added to the homogenate (1 ml of CTAB for approx. 300 mg homogenate). The mixture was incubated at 60 °C for 30 min. After extracting DNA with Chloroform/Isoamyl alcohol (24:1, Roth), DNA was pelleted with Isoproponal and washed with 70% ethanol.
Total RNA was extracted using TRIzol (Invitrogen) following the manufacturer's protocol. RNA was digested using Ambion DNase I (RNase-free, Life Technologies) and purified with Phenol/Chloroform/Isoamylalcohol (Roth).

### 2.2.5 Determination of DNA and RNA concentration

The concentrations of DNA and RNA were determined by optical density using a spectrophotometer Nanodrop 1000 (peqLab) at 260 nm. The integrity of DNA was verified by electrophoresis on 1% agarose gels. The integrity of rRNA bands was verified by electrophoresis on 1% denaturating agarose gels containing 1.7 M formaldehyde. The DNA and RNA samples were stored at -20 °C and -80 °C, respectively.

### 2.2.6 Reverse transcription

Reverse transcriptions were carried out using the SuperScript III Reverse Transcription Kit (Life Technologies) (for 3' RACE, 5' RACE and genotyping RT-PCR) or the QuantiTect Reverse Transcription Kit (Qiagen) (for qRT-PCR).

### 2.2.7 Polymerase chain reaction (PCR) analysis and cloning

For genotyping and probe fragments preparation, PCR was performed with DNA polymerase from Qiagen. For cloning, PCR was performed with Phusion High-Fidelity DNA Polymerase (Thermo). PCR products were analyzed by agarose gel electrophoresis and nanodrop.

### 2.2.8 Real-Time quantitative polymerase chain reaction (qPCR)

DNase digested RNA was reverse transcribed with a mixture of gene-specific reverse primers. SYBR Green PCR Core Reagent set (Applied Biosystems) and 7500 Fast Real-Time PCR System (Applied Biosystems) were used for performing the qPCR. The reaction was performed in MicroAmp Fast Optical 96-Well Reaction Plates (0.1 ml) and covered with MicroAmp Optical Adhesive Film (Applied Biosystems).

### 2.2.9 5' RACE (Rapid amplification of cDNA ends) and 3'RACE

*Chlamydomonas* chloroplast RNA was used in the 5' and 3' RACE.

For 5' RACE, the Rumsh RNA oligo (5'-GUGAUCCAACCGACGCGACAAGCU AAUGCAAGANNN-3') was ligated to the chloroplast RNA by T4 RNA ligase I (Life Technologies). The ligated product was reverse transcribed using SuperScript III Reverse Transcription Kit with Hexa/Nona random primer following the manufacturer's instruction. The resulting cDNA was amplified by PCR using Phusion High-Fidelity DNA Polymerases with the primer pair Rumsh1 and the gene-specific primers listed in Appendix 1. The PCR product was cloned using the pJect PCR cloning kit (Fermentas). The positive colonies were identified by colony PCR and were sequenced with Pjet1.2fwd and Pjet1.2rev primers.

For 3' RACE, the SRA 3' RNA oligo (TCGTATGCCGTCTTCTGCTTG) was ligated to the chloroplast RNA by T4 RNA ligase I. The ligated product was reverse transcribed using SuperScript III Reverse Transcription Kit with the adapter RT primer. The resulting cDNA was amplified by PCR using Phusion High-Fidelity DNA Polymerases with the primer pair adapter PCR primer and gene-specific primer listed in Appendix 1. The following cloning and sequencing procedures were the same as 5' RACE.

### 2.2.10 Western blot analysis

For the total protein sample preparation, leaves were ground in liquid nitrogen and homogenized in 250 µl of homogenization buffer. Protein concentration was measured using the Bradford method (Bradford 1976). Protein was separated by SDS-PAGE, and blotted onto nitrocellulose membrane (GE Healthcare). Blots were stained with ponceau to assess protein quality and control loading. Blots were blocked with 4% skimmed milk (Sigma) in 1× TBST and were probed with antibody. Detection was performed using the

ECL detection solution. The signal was imaged with ChemiDoc XRS (Bio-Rad) system and quantified with Imagelab software (Bio-Rad).

### 2.2.11 Polysome fractionation

Polysome fractionation of leaf tissue was performed following previously described protocol (Barkan 1998). First, 0.5 g of tissue was ground to a fine powder in liquid nitrogen, before adding 1 ml of the extraction buffer. The sample was further ground until thawed, after which the homogenized plant material was forced through a glass wool plug in a 2 ml syringe to remove debris. The flow through fraction was incubated on ice for 10 min and was centrifuged at 21,000 g for 5 min at 4 °C. Next, 1/20 volume of 10% Sodium deoxycholate (SDC) was mixed with the supernatant and incubated on ice for 5 min. After the mixture was centrifuged at 21,000 g for 5 min at 4 °C, 0.5 ml of the supernatant was loaded on to 4.4 ml of analytical sucrose gradients and was centrifuged at 275,000 g for 30 min at 4 °C without forced stop. The resulting sucrose gradients were fractionated into 12 factions (0.41 ml per fraction) from top to bottom and were transferred to 12 tubes with 50 µl 5% SDS/0.2 M EDTA. RNA was extracted from each fraction with Phenol/Chloroform/Isoamyl alcohol (P/C/I, 25:25:1, Roth).

Stock sucrose solutions containing 15%, 30%, 40%, and 55% sucrose were prepared, ensuring that each contained the indicated concentration of sucrose and 1× polysome gradient salts, 500 µg/ml heparin, 100 µg/ml chloramphenicol, and 25 µg/ml cycloheximide. Next, 1.1 ml of the 55% sucrose solution was pipeted into the bottom of a 5 ml ultracentrifuge tube, and was frozen at -80 °C. After this, 1.1 ml of the 40% solution was pipeted on top of it and was frozen again. The same process was followed for the 30% and 15% sucrose solutions.

### 2.2.12 Northern blot analysis

#### 2.2.12.1 RNA sample preparation and gel electrophoresis

The RNA samples were mixed with 2.5 volumes of Northern sample buffer. 10× RNA loading buffer was added after denaturing the RNA samples at 75 °C for 15 min. 2 µl of RiboRuler High Range RNA Ladder (Fermentas) was parepared in the same way as the RNA sample. The gels were run in the horizontal electrophoresis chamber (bsb11 biotech) for 2 hours at 100 V.

RNA Agarose Gel (1.0%, w/v): 1.5 g molecular biology agarose (Bio-Rad) was cooked in 109 ml of $H_2O$. Once it was cooled to 65 °C, 15 ml of 10× MOPS buffer (pH 8.0) and 26 ml of Formaldehyde were added and the gel was poured.

#### 2.2.12.2 RNA transfer

Electrophoretically separated RNA was transferred onto Hybond-N (Amersham Biosciences) membranes by capillary blotting using standard protocols. Next, the blot was crosslinked in a UV chamber (Bio-Rad) irradiate once with 150 mJ/cm$^2$. Finally, the blot was stained with Methylene blue solution, in order to assess the RNA quality and control loading.

#### 2.2.12.3 Probe labeling

Strand specific DNA probe was labeled with $\alpha^{32}$P-dCTP as follows. First, a mixture of 100 ng template (PCR product), 2 µl of 10 mM reverse primer, 5 µl of 10× Klenow buffer and $H_2O$ (add $H_2O$ to a total volume of 40 µl) was heated to 95 °C for 10 min. Subsequently, 3 µl of mixC (10 mM of each dATP, dTTP and dGTP), 2 µl of Klenow fragment 10 U/µl) and 5 µl (50 µCi) of γ 32P-dCTP were added. The reaction mixture was incubated at 37 °C for 10 min, before adding 4 µl of dNTP mix (0.25 mM of each dATP, dTTP, dGTP and dCTP) and incubating again for 10 min. Ultimately, the probe was purified using illustra MicroSpin G-50 column (GE Healthcare) to remove the uncooperated radioactivity following the manufacturer's protocol. Next, the purified probe was denatured for 10 min at 95 °C before hybridization.
Oligo probe was end-labeled by γ32P-ATP, whereby 40 pM oligo, 2 µl 10× buffer A, 1 µl Polynucleotide Kinase (PNK) and 5 µl γ32P-ATP (50 µCi) were mixed, after which $H_2O$ was added to 20 µl. Finally, the mixture was incubated at 37 °C for 10 min and was purified using MicroSpin G-25 column.

#### 2.2.12.4 Hybridization and detection

Hybridization was carried out overnight in Church buffer at 55 °C for ssDNA probe and 45 °C for oligo probe. The membrane was washed once in 0.5× SSC / 0.1% SDS (10 min) and twice in 0.2x SSC / 0.1% SDS (15 min each) at the bybridization temperature, before being exposed to a phosphoimager plate. The plate after exposure was scanned by

Personal Molecular Imager (Bio-Rad) and quantified using Quantity One software (Bio-Rad).

### 2.2.13 Dot blot analysis

For each dot, RNA was denatured in 80 µl of the sample buffer for 15 min at 75 °C, before being cooled on ice and mixed with 30 µl of 20× SSC. Next, N+ hybond membrane (12 cm× 9 cm) was moistened with water and soaked in 20× SSC for 10 min. Once Bio-dot SF Microfiltration Apparatus (Bio-Rad) was assembled and vacuum was applied, the wells were washed with 400 µl of 5× SSC, and 100 µl of the denatured RNA samples were loaded into each well. Once RNA samples were dried, the wells were washed again with 400 µl of 5× SSC, vacuum was applied for further 5 min and the membrane was cross-linked in a UV chamber irradiate once with 150 mJ/cm$^2$.

### 2.2.14 Plastid transformation

Stable transformation of plastids was performed with modifications according to the protocols of Svab (Svab and Maliga 1993) and Okuzaki (Ayako and Yutaka 2012). Sterile tobacco leaves from plants cultured on MS were cut into 0.5 cm× 0.5 cm squares before being placed onto MS plates with the lower leaf side up. They were cultured overnight and transformed using the Biolistic PDS-1000 / He Particle Delivery System (1,100 psi, L2 = 6 cm, 10 shots per construct, Bio-Rad) with gold particles (0.6 uM, Bio-Rad) loaded with the transformation plasmids. Spectinomycin resistant calli were regenerated on RMOP medium (2-4 rounds), rooted on MS medium and transferred to soil for seeds production.

### 2.2.15 Agrobacterium-mediate nuclear transformation of *Nicotiana tabacum*

The overexpression vector was transformed into *Nicotiana tabacum* via an agrobacterium-mediated method (Horsch 1985). Midribs were removed from sterile tobacco leaves, while the remaining leaf pieces were cut into approximately 1 cm× 1 cm squares before being transferred to petri dishes, where they were placed upside down. Next, a thin layer of sterile MS liquid medium was added to cover the leaf pieces, which were poked uniformly with a needle. The leaf pieces were subsequently transferred to the Agrobacterium suspension plates and incubated for 10 min, before being transferred onto co-cultivation plates that were over-laid with filter paper. The leaf pieces were placed upside down and remained in close contact with the filter paper. The plates were sealed

with parafilm and incubated in darkness at 24 °C for 3 days. Finally, the leaf pieces were transferred onto callus induction plates and were subcultured bi-weekly to fresh medium. The plates were incubated in dim light at 22 °C until shoots could be excised.
Shoots were excised from calli with forceps and were inserted in the MS basal medium with selective agent for rooting at 22 °C and 18/6 photoperiods with light intensity of 60-80 μmol. $m^{-2}$. $s^{-1}$. When they reached adequate size, the plantlets (comprising both shoots and roots were transferred to soil for acclimatization. The established seedlings were transferred to the greenhouse to set seeds.

### 2.2.16 RNA library construction and sequencing

RNA was extracted with Phenol/Chloroform from the IP pellet of HA tagged and control lines. Next, the RNA samples were proceeded for library construction using ScriptMiner Small RNA-seq library preparation Kit (Epicentre). Next, the libraries were sequenced (performed by Dr. Wei Chen's group, The Max Delbrück Center for Molecular Medicine, Germany).
Two additional libraries were prepared using Ion Total RNA-seq Kit (Life Technologies), which were sequenced using Ion Proton system (Life Technologies).

### 2.2.17 Reads mapping

The reads were quality trimmed and mapped to the *Nicotiana tabacum* chloroplast genome by CLC workbench (Version 6.0.1) with the repeat region masked. The following parameters were used: Mismatch cost = 2; Insertion cost = 3; Deletion cost = 3; Length fraction = 0.5; Similarity fraction = 0.8. Local alignment was applied.

### 2.2.18 SRNA data analysis

*Chlamydomonas* sRNA data was retrieved from Ibrahim et al. (Ibrahim et al. 2010), available as a GEO Series NCBI: GSE17815, including reads from four different libraries. Two of these libraries had been prepared from a knockout line of the terminal nucleotidyltransferase, while the remaining two are from the knockdown line of a tryptophan synthase β subunit. Three further libraries were available as the GEO Series GSE32457 and had been prepared from a wild-type *Chlamydomonas* strain grown under three different conditions, namely: normal conditions, phosphate starvation, or sulphate starvation. SRNA sequences with more than one read in each sample were mapped to the chloroplast genome (NC_005353) using CLC Genomics Workbench (Version 6.0.1),

allowing one mismatch in the core sequence. As a threshold, clusters of at least three overlapping sequences were counted, and each cluster should contain more than 10 reads. The core sequence of each cluster was defined as those nucleotides represented in at least 50% of all reads within the cluster. Only the clusters located in the non-coding regions were considered.

### 2.2.19 Low molecular weight (LMW) RNA enrichment and sRNA Northern blot

SRNAs were enriched according to Lu (Lu et al. 2007) from the total RNA prepared in Prof. Michel Goldschmidt-Clermont laboratory. RNA was resolved on 15% denaturing polyacrylamide gels containing 8 M urea before being transferred to Hybond-N nylon membranes (GE Healthcare). Membrane-bound RNAs were chemically crosslinked by 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC), as described by Pall and Hamilton (Pall and Hamilton 2008). Moreover, DNA oligonucleotides were end-labeled with 50 μCi γ32P-ATP (Hartmann-Analytics) using Polynucleotide kinase (Thermo Scientific) and were used as probes. Hybridization was carried out following the same protocol as for the Northern blot described before. The membranes were stripped by incubating in 0.5% SDS at 60 °C for 1 hour.

# 3. Results

## 3.1 MatK targets identification and regulation analysis

### 3.1.1 Pipeline of RIP-seq experiment

The described bacterial maturases only splice their home intron. In contrast, using RIP-Chip, the chloroplast maturase MatK has been shown to be associated with seven RNA targets (*trnK, trnV, trnI, trnA, rps12-in2, rpl2* and *atpF*) (Zoschke et al. 2010). Findings of the study using oligo tilling arrays indicate that MatK associates with multiple domains of the *trnK* intron（Zoschke et al. 2010）. Using the same method, DIV of *trnI* and *trnA* was also found to be associated with MatK, while no association was found with the domains of *atpF, trnV, rpl2* and *rps12-in2*. This result is most likely caused by the limited sensitivity of Chip hybridization using tilling arrays (Neumann 2011).

High-throughput sequencing is a highly sensitive method. Taking the advantage of the transplastomic plants that express a C terminus HA-epitope tagged MatK (+HA) and non–tagged control line ( – HA) of *Nicotiana tabacum* (Zoschke et al. 2010), RNA coprecipitated with MatK was analyzed by RNA-seq with the aim of narrowing down the MatK binding sites on RNA (Figure 3). The input control used in this study was from an RNA-seq project with total RNA of *Nicotiana tabacum* (unpublished data).

As it was shown that the MatK protein reaches its highest level at 7-day-old tobacco (Zoschke et al. 2010), chloroplasts were isolated from the seedlings of this age. Furthermore, due to the fact that MatK is a soluble protein in the chloroplast stroma (Zoschke et al. 2010), the stroma fraction of chloroplast was separated from the membrane fraction to enrich the MatK protein. Consequently, the IP was performed with the stroma fraction.

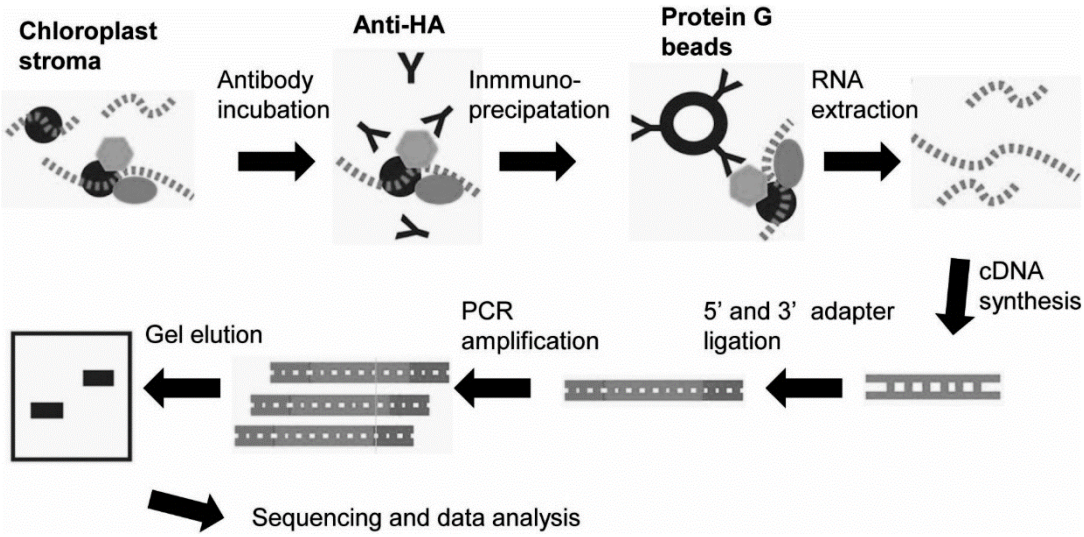



**Figure 3: Workflow of RIP-seq used in this study.** Chloroplasts were isolated from 7-day-old seedlings of HA tagged (+HA) and non-tagged (–HA) tobacco, stroma fraction was separated and used in immunoprecipitation (IP) with HA antibody. RNA from IP pellet was extracted with Phenol/Chloroform, precipitated with ethanol and used for library preparation. The libraries were sequenced and the reads were mapped to the tobacco chloroplast genome (NC_001879) for further data analysis. Modified from Zhao *et, al* (Zhao et al. 2010).

Immunoprecipitation of MatK was verified by Western blot analysis. As can be seen, a band with an apparent molecular weight of ~55kDa was detected in Input and Pellet fractions (Figure 4). The position of this band is in accordance with those reported in previous immunological analysis of MatK (Liere and Link 1995; Barthet and Hilu 2007; Zoschke et al. 2010). The RNA from the IP pellet was extracted next. The quality and quantity of the extracted RNA was analyzed using capillary electrophoresis. According to the report of the fragment analyzer (AdvanCE), the sizes of RNA from the IP pellet vary from 100 nt to 1500 nt (see Appendix 2). Two libraries were constructed using the sRNA preparation kit (Epicentre) and were sequenced by employing the Illumina HiSeq sequencing system. To recover the long transcripts in the RNA sample, the RNA was digested with RNase III before the preparation of two additional libraries for Ion Torrent sequencing. After digestion, the RNA population was fragmented to length of approximately 100 nt, and this size is compatible with the specifications of the Ion Total RNA-seq library preparation kit (Life Technologies) (see Appendix 2).

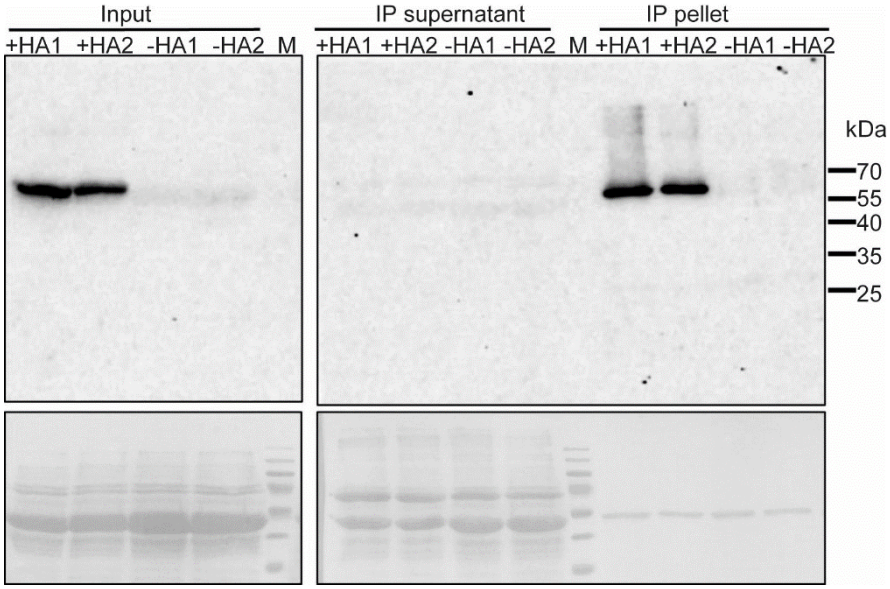



**Figure 4: Inmmunoprecipitation of the HA tagged MatK from transplastomic tobacco lines.** Chloroplasts were isolated from 7-day-old +HA and –HA tobacco seedlings, HA-tagged MatK was immunoprecipitated from chloroplast stroma with an HA antibody (Sigma) and magnetic beads (Life Technologies). Protein samples of input, IP supernatant and IP pellet were separated by the SDS-PAGE gel and blotted to a nitrocellulose membrane. HA-specific signals were detected with an HA antibody conjugated to horseradish peroxidase (HA-HRP, Sigma) and developed by ECL detection solution (upper panel). Ponceau staining was used as loading control (lower panel).

### 3.1.2 Multiple transcripts are associated with MatK

To gain an overview of the RNA species associated with MatK, the output reads of the sequencing were trimmed and mapped to the tobacco chloroplast genome (NC_001879). The read coverage at each genomic position was extracted and the relative enrichment was calculated as the ratio of +HA to either –HA or total RNA.

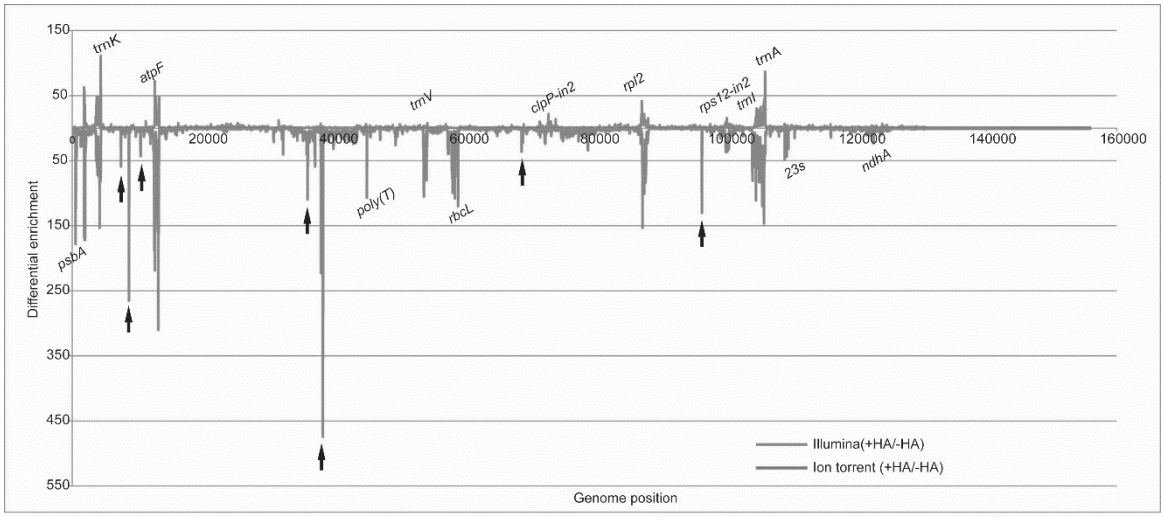



**Figure 5: Overall mapping of MatK targets identified by RIP-seq.** The libraries were prepared with these RNA samples from the IP pellet and were sequenced. Reads from each library (+HA and –HA tagged MatK-IP pellets) were mapped to the chloroplast genome (NC_001879), respectively. The ratio of mapped reads between +HA and –HA was calculated and plotted to the chloroplast genome. The arrowheads indicate the enriched intronless tRNAs. Illumina: upper curve, Ion torent: lower curve.

The relative enrichment was defined by the ratio of +HA to control. The -HA control was used for eliminating the unspecific binding of RNA to the antibody and beads during IP. The threshold of relative enrichment was set as five folds for a minimal length of 50 nt. Using the Illumina RIP-seq method, seven of the eight chloroplast group IIA intron-containing transcripts (with the exception of *trnV* ) could be identified. While six group IIA-containing intron transcripts are in line with the targets identified by the RIP-Chip method (Zoschke et al. 2010). The MatK home intron, *trnK*, is the most strongly enriched transcript. Additionally, a group IIA intron, *clpP-in2* and a group IIB intron, *ycf3-in2* (6.3 folds in the peak) were also found (Figure 5. upper curve). According to the Ion Torrent sequencing results, all seven group IIA targets previously identified by RIP-Chip were enriched while the *clpP-in2* was not enriched. However, the very abundant transcripts *psbA* and *rbcL* and some intronless tRNAs and rRNAs such as 16s rRNA, were also found (Figure 5. lower curve).

The expression levels of individual genes are very different in a certain transcriptome (Zhelyazkova et al. 2012b; Freeberg et al. 2013). The enrichment of some transcripts might be caused by their higher expression compared to other transcripts. Thus, in order to eliminate the unspecific signal caused by differential expression of the genes, the relative enrichment of +HA to the input (total RNA) was also calculated. According to the results of the Illumina sequencing experiment, the overall pattern did not change, with the

exception of the emergence of *trnV*, several intronless tRNAs and very abundant transcripts, such as 16s and 23s rRNAs (Appendix 3 and Appendix 6). The Ion Torrent sequencing experiment yielded, with the exception of rRNAs, the six group IIA intron-containing RNAs and some rRNAs, as well as group IIB intron *ndhA-in*, while most of the intronless RNAs obtained in the calculation with –HA disappeared (Appendix 3 and Appendix 6).

### 3.1.3 Multiple sites association and preferential domains of MatK binding to its targets

Here the focus was solely on the group IIA targets, which were consistently enriched in both sequencing methods. When zooming in on the mapping results, multiple peaks could be identified in each transcript, most of which are in the intron region. For example, for *atpF*, the major peaks are in the intron region, while only a very short exonic fragment adjacent to the intron region is enriched (*atpF* in Figure 6).

For the transcripts with short exons, such as *trnK*, enrichment was also found in the exon regions, especially in the Illumina experiment (Figure 6). This may be due to the incomplete endogenous degradation of transcripts by endo- and/or exonucleases. Interestingly, the entire *trnK* intron region is enriched, with the exception of the *matK* ORF region, indicating an exclusion of the maturase ORF region in the maturase-intron association. An enrichment of *matK* start codon region was also observed, indicating an association between the MatK protein and its own start codon region (Figure 6).

The mapping patterns differ for these two sequencing experiments. First, a higher relative enrichment value was obtained in the Ion Torrent than in the Illumina sequencing experiment. For example, for *atpF*, the highest relative enrichment value yielded by Illumina is 72.5, while Ion Torrent sequencing produced the highest value of 310 (Figure 6 and Appendix 6). Second, the peaks that emerged in Ion Torrent sequencing are generally broader than those obtained by the Illumina sequencing. For example, the first two major peakes of *trnI* found in the Illumina sequencing fused to one broad peak in the Ion Torrent sequencing (Figure 6). Both of aforementioned differences are likely caused by the inclusion of relatively long transcripts in the Ion Torrent sequencing.

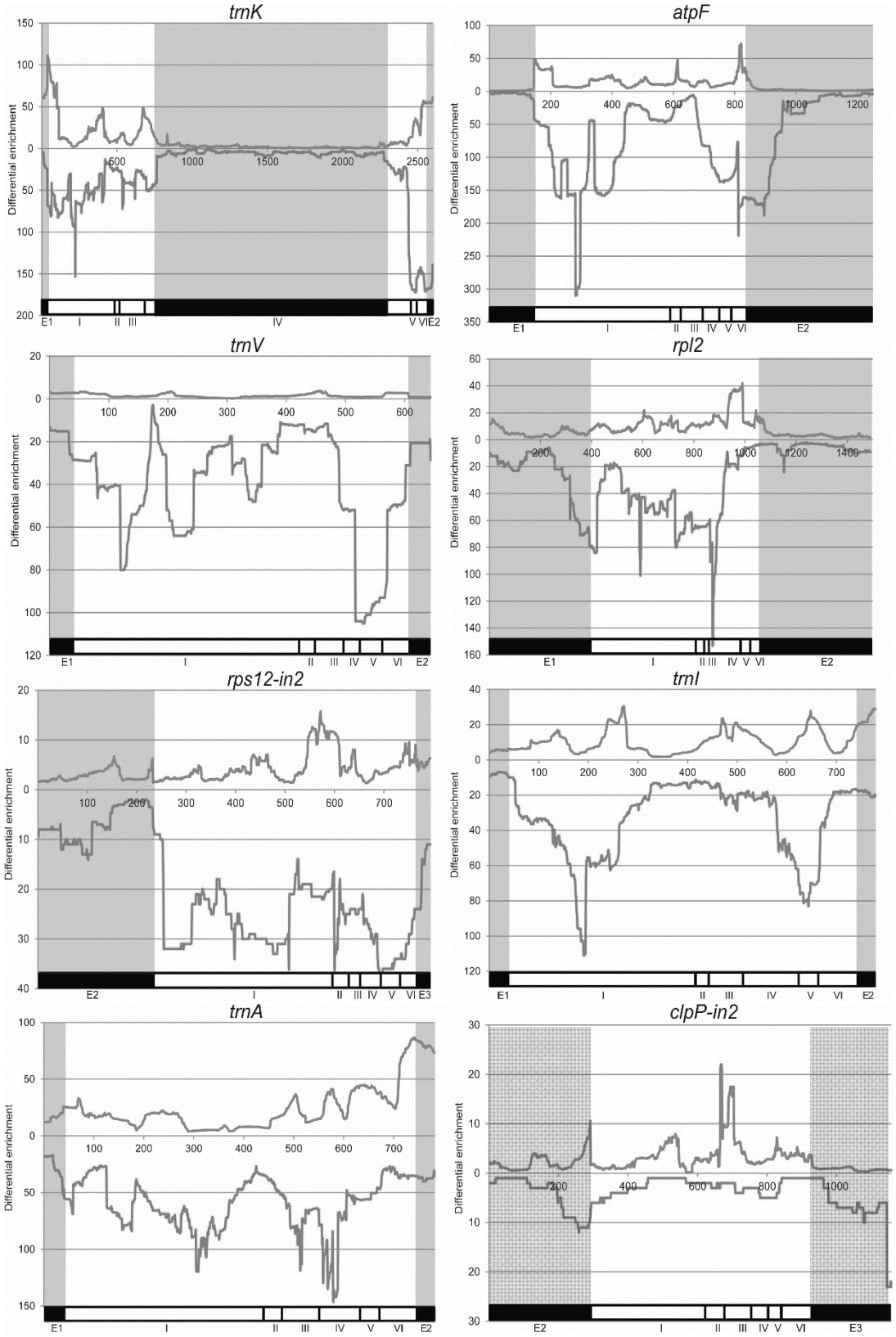

trnK
Differential enrichment
E1 I II III IV V VI E2
atpF
Differential enrichment
E1 I II III IV V VI E2
trnV
Differential enrichment
E1 I II III IV V VI E2
rpl2
Differential enrichment
E1 I II III IV V VI E2
rps12-in2
Differential enrichment
E2 I II III IV V VI E3
trnI
Differential enrichment
E1 I II III IV V VI E2
trnA
Differential enrichment
E1 I II III IV V VI E2
clpP-in2
Differential enrichment
E2 I II III IV V VI E3

**Figure 6: Zoom in of the MatK targets identified by RIP-seq.** The reads from two libraries (+HA and –HA tagged MatK) were mapped to the chloroplast genome (NC_001879), respectively, and the ratio of the mapped reads between +HA and －HA was calculated (Y-axis) and plotted to the chloroplast genome. The targets present in both libraries were examined and the mapping graphs for the respective gene region are shown. The blue curve in the upper part and the red curve in the lower parts of each graph represent the enrichments derived from the Illumina and Ion Torrent sequencings methods, respectively. The bar underneath the graph shows the distribution of exons (filled black blocks) and domains of group II intron (empty blocks) defined in a previous study (Michel et al. 1989). In the graphs, the exon regions are also marked by a grey shadow. *ClpP-in2* is also shown here, even though it is considered an unspecific target, as only one peak in its exon2 is consistently revealed in both sequencing experiments.

Group II introns in chloroplast are not conserved in terms of primary sequences. However, they share the secondary structure defined by the six domains. When the enrichment is connected with the domain regions, the DI and DVI are preferentially associated with maturase in all the targets found by Illumina experiment (Table 1). On the other hand, DI and DIV and DV are preferentially associated with the maturase, as shown in the Ion Torrent experiment (Table 2). In both experiments, the main peaks are mostly in DI. However, some peaks are not consistently present. For example, the highest enrichment is located in DIV and DV for *rpl2* in the Illumina sequencing, but DV is not enriched in the Ion Torrent sequencing (Table 1 and Table 2).

**Table 1 Relative enrichment of each domain of MatK associated RNAs by Illumina sequencing.**

| Illumina | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| RNA | E1 | DI | DII | DIII | DIV | DV | DVI | E2 |
| trnK | | | | | | | | |
| atpF | | | | | | | | |
| trnV | | | | | | | | |
| rpl2 | | | | | | | | |
| rps12-2 | | | | | | | | |
| trnI | | | | | | | | |
| trnA | | | | | | | | |

For each MatK associated RNA found by Illumina sequencing (Figure 6), the highest point of the relative enrichment for each domain was selected. According to these values, the enrichment of domains was ordered from the highest (darkest) to the lowest (lightest).

**Table 2: Relative enrichment of each domain of MatK associated RNAs by Ion Torrent sequencing.**

| Ion Torrent | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| RNA | E1 | DI | DII | DIII | DIV | DV | DVI | E2 |
| trnK | | | | | | | | |
| atpF | | | | | | | | |
| trnV | | | | | | | | |
| rpl2 | | | | | | | | |
| rps12-2 | | | | | | | | |
| trnI | | | | | | | | |
| trnA | | | | | | | | |

For each MatK associated RNA found by Ion Torrent sequencing (Figure 6), the highest point of the relative enrichment for each domain was selected. According to these values, the enrichment of domains was ordered from the highest (darkest) to the lowest (lightest).

A dot blot experiment was carried out to confirm the results of RIP-seq. The RNA was prepared the in same way as was the RNA used in RIP-seq. More specifically, two oligos for *atpF* and two oligos for *trnV* were selected for hybridization (Appendix 4A). The result of the dot blot experiment were in good agreement with those of the the result of RIP-seq, with the exception of a probe for *trnV*. Although trnV1 was slightly enriched (1.36 folds) in the dot blot, it was enriched in the Illumina but not in the Ion Torrent sequencing (Appendix 4 B).

### 3.1.4 Association of MatK to its targets is regulated

In its life cycle, RNA is dynamically associated with RBP. This association is regulated by the developmental stages and environmental conditions, such as nutrition (Tam et al. 2010; Freeberg et al. 2013) and light (Danon and Mayfield 1991; Lisitsky and Schuster 1995). Extant studies have shown that MatK is associated with seven targets. However, whether this association is static or dynamic remains to be determined. The protein and the RNA of MatK reach their highest level at 7 days and 25 days after imbibition, respectively (Hertel et al. 2013). Here, the binding of MatK to its targets was examined in these two developmental stages of tobacco (7 days and 25 days).

Chloroplasts were isolated from two tobacco lines with an HA tag at either the N terminus (HA:MatK) or the C terminus of MatK (MatK:HA, same as +HA in the RIP-seq analysis).

These two lines were considered as biological replicates. IP with HA antibody was performed with the stroma fraction of chloroplasts. RNA samples from both supernatant and pellet fractions were used for the dot blot. The blots were hybridized with an intron probe of each target (Figure 7A).

The relative enrichment ratio was defined by the fraction of each target in the sum of seven targets and was shown in a pie chart for the 7-day and 25-day stages, respectively (Figure 7B). In the 7-day-old tobacco, *trnA* is the most strongly enriched intron, followed by the *trnK* intron, while *rps12-in2* is the least enriched intron. By contrast, in the 25-day-old tobacco, *trnK* is the most strongly enriched intron, followed by *atpF* intron, with *trnV* as the least enriched (Figure 7B). A ratio-to-ratio value was calculated between the relative enrichment ratios at 7 days and 25 days. This value clearly shows the dynamic change. For example, for the MatK home intron *trnK*, the ratio-to-ratio value is 1, indicating that the proportion of MatK associated *trnK* intron does not change from 7 to 25 days. In addition, the *trnV* proportion decreased to less than half the initial amount (0.4 fold), and the most dramatic change was noted for *atpF*, for which the proportion increased 3 folds in the 25-day-old tobacco (Figure 7C). Surprisingly, these changes of association do not correlate either with the transcription rate or the steady-state level of RNA in these two stages (Hertel et al. 2013), indicating a regulation of association during development.

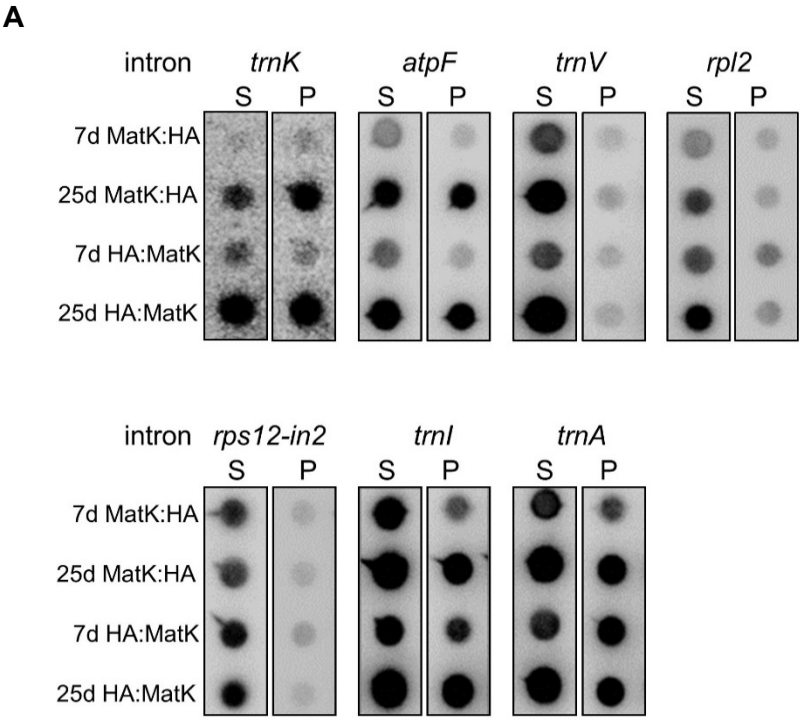

A
intron
trnK
atpF
trnV
rpl2
S
P
7d MatK:HA
25d MatK:HA
7d HA:MatK
25d HA:MatK
intron
rps12-in2
trnI
trnA
B
Ratio 7 d
trnA, 0.31
trnK, 0.28
atpF, 0.09
trnV, 0.05
rps12-in2, 0.04
trnI, 0.12
rpl2, 0.11
Ratio 25 d
trnA, 0.18
trnK, 0.27
trnI, 0.18
atpF, 0.25
rps12-in2, 0.04
rpl2, 0.06
trnV, 0.02
C
Ratio 25 d/Ratios 7 d
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
trnK
atpF
trnV
rpl2
rps12-in2
trnI
trnA

**Figure 7: Association between MatK and its targets in two developmental stages of *Nicotiana tabacum*.** (A) Dot-blot hybridization of RNAs that coimmunoprecipitate with HA tagged MatK. Here, 1/8 of the RNA from one IP was applied to replicate dot blots, and was hybridized with intron probes. P: RNA from pellet fraction, S: RNA from supernatant fraction. HA:MatK: HA tagged at the N terminus of MatK. MatK:HA: HA tagged at the C terminus of MatK. (B) The enrichment fraction of each MatK targets. The enrichment ratios between the pellet and the supernatant signal of the dot blot were calculated. The enrichment factions displayed in pie charts were calculated by the enrichment ratio of each target divided by the sum ratio of all seven targets. (C) The ratio-to-ratio indicates the dynamic association of RNA-MatK during development. For each target, the ratio-to-ratio value was calculated from the enrichment fraction ratio between the 25d and 7d data. There is an increased association of RNA-MatK in 25-day compare to 7-day old tobacco when the value is greater than 1, and *vice versa*.

### 3.1.5 Precursors are the predominante transcripts associated with MatK

Group II intron splicing consists of two reversible transesterification reactions and results in ligated exons and an excised lariat RNA. After splicing, the bacterial maturase remains tightly bound to the excised intron and presumably uses most or all of the interactions among the group II intron domains to stabilize the ribozyme structure and reverse splice the intron into DNA-a procedure also known as intron mobility (Wank et al. 1999). Thus far, the RNA species bound by MatK are still unknown. This leads to the need to ascertain whether the species-precursor, spliced intron or mature RNA-are associated with MatK and identify the one that is overrepresented. To meet this aims, primers located in different regions were used to discriminate these three RNA species associated with MatK by RT-quantitative PCR (RT-qPCR). This approach was chosen as primers in exon1 (Ex1f) and intron (Inr) can detect precursors; primers in intron (Inf) and exon2 (Ex2r) can detect precursor and lariat; and primers in exon1 and exon2 can be used for detecting mature RNA (Figure 8A).

The fold change was calculated by +HA relative to the –HA line. The findings indicate that the amplicon spans at the exon-intron junction is more abundant than that at the intron-exon junction and at the mature RNA. In the case of *trnV*, the exon-intron amplicon is 5.6 times greater than that of the control, while the fold change for the exon-exon amplicon is 2.1 (Figure 8B), indicating that the MatK is associated with the precursor, while its association with the mature RNA is almost imperceptible. The exon-intron amplicon represents the precursor while the intron-exon amplicon represents both the precursor and the lariat RNA with the second exon. In theory, the intron-exon

amplicon should be greater than or equal to the exon-intron amplicon. However, the findings obtained here contradict this assumption. A similar pattern was observed for *trnA* (Figure 8B).

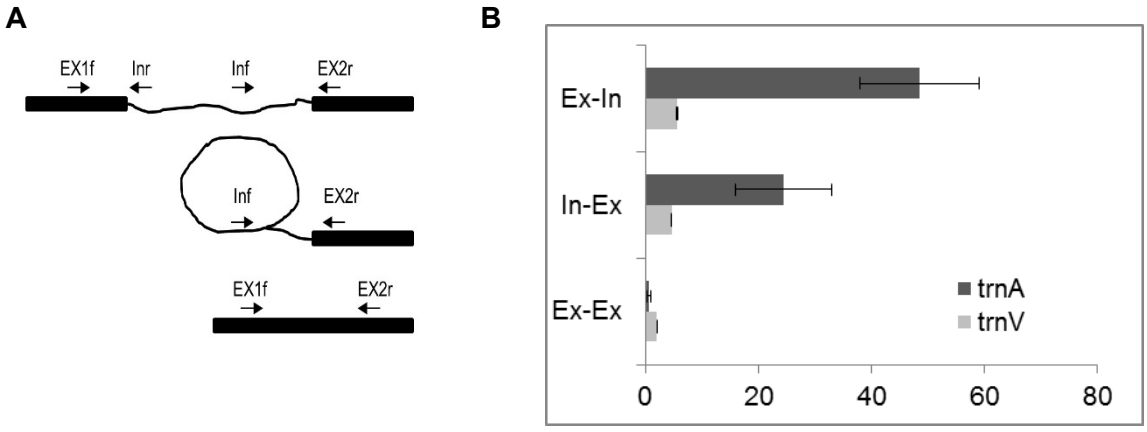



**Figure 8: Association of *trnA* and *trnV* transcripts with MatK.** IP was performed with HA antibody in the +HA and –HA tagged MatK tobacco lines. RNA from IP-pellet was treated with DNase I and reverse transcribed with gene-specific primers. (A) Schematic diagram of the primer positions on RNAs. QPCR was performed with primers spanning exon-intron (Ex-In), intron-exon junction (In-Ex) or both adjacent exons (Ex-Ex) of the transcripts. (B) MatK associated transcripts species. The fold change was calculated by +HA compare to the –HA line.

### 3.1.6 Polysome fractionation indicated that both the *trnK* precursors and free intron serve as the translational template for MatK

Bacterial maturase LtrA was shown to be autoregulated by binding to its own start codon region (Singh et al. 2002). A feedback regulation was also suggested for the chloroplast MatK (Hertel et al. 2013). However, the mechanism for the MatK regulation presently remains unknown. One possible autoregulation models based on the location of MatK posits that the *trnK* precusor serves as the translational template, and splicing of the template decreases the translation. In order to test this hypothesis, the MatK translational templates had to be identified. The RNA species containing *matK* ORF are potential MatK translational template. Northern blot analysis in previous study identified three *matK* ORF-containing transcripts species, namely *psbA-trnK* precursor, *trnK* precursor and *trnK* free intron (Hertel et al. 2013). If the *trnK* free intron is not the translational template, the simple *matK* autoregulation model would be proved. Due to the fact that the translational-active transcripts are associated with polysome complex, whereby the population of such transcripts loaded with polysomes can be size-fractionated by sucrose

density gradient. Consequently, the translational-active transcripts can be separated from the untranslational RNAs.

In the sucrose gradient, the more ribosomes are associated with transcripts, the heavier the transcripts are, and the deeper they migrate. Using the sucrose gradient, the total tobacco extract was separated into 12 fractions and Northern blot was performed with the RNA from these fractions. Using the probe in the *matK* ORF (Figure 9A), presence of signals of three bands could be observed in all the RNA samples, indicating that these three transcripts migrate to all the fractions, including the densest ones (Figure 9B). The uppermost band with a size of ~4kb is corresponding to the *psbA-trnK* transcript, while the mainband with a size of ~2.5kb corresponds to *trnK* precursor, and the last transcript, which is a little smaller than pre-*trnK*, is the *trnK* free intron (Figure 9B).

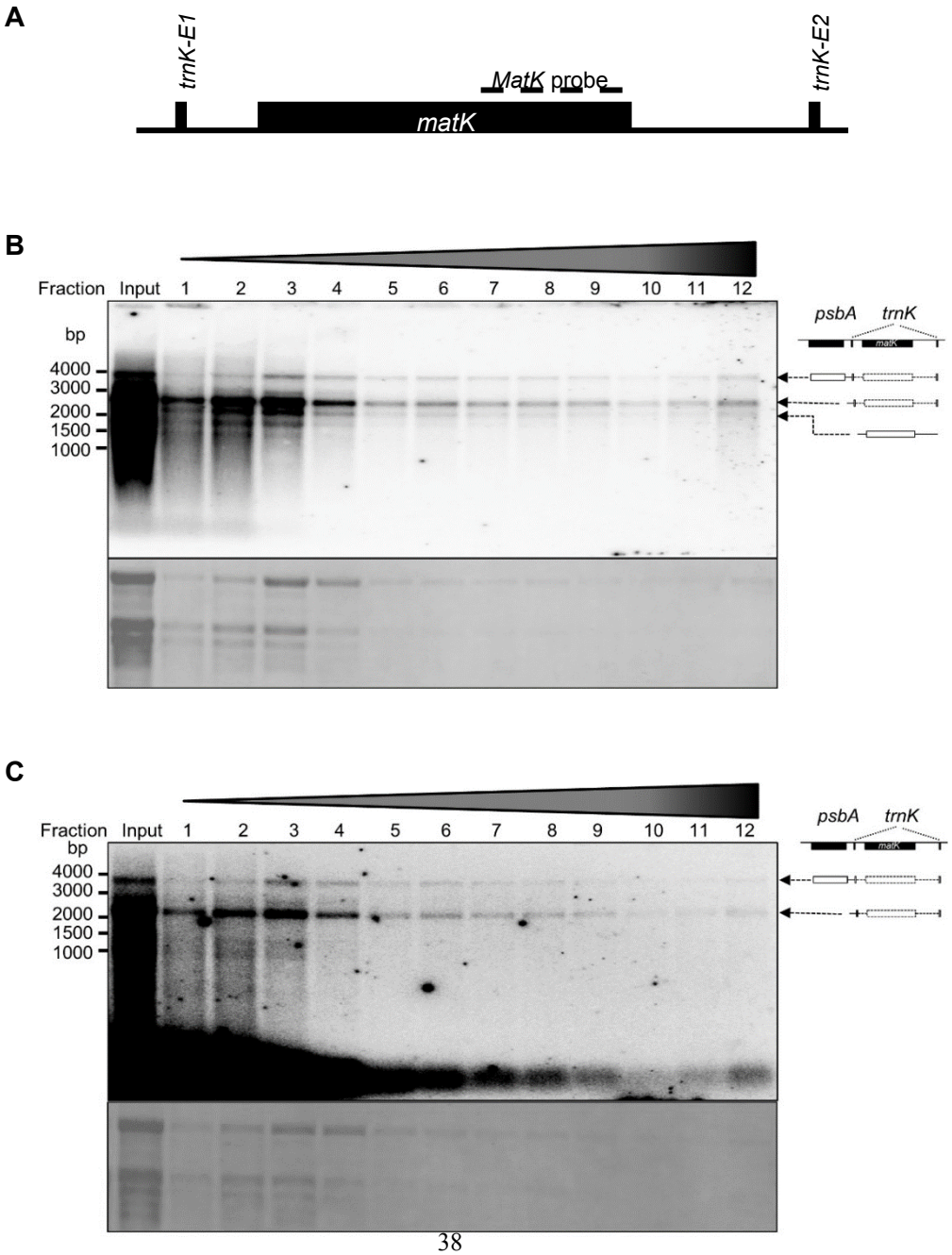


**Figure 9: Polysome analysis of transcripts for *matK* translation.** The total extract from 7-day-old tobacco was size-fractionated into 12 fractions by sucrose gradient (with the top fraction denoted as sample 1 and the bottom fraction as sample 12). The RNA was extracted from each fractions, and was separated by 1% denature agarose gel and blotted onto nylon membrane. (A) Schematic diagram of the probe locations. Dashed line: location of *matK* ORF probe. (B) The membrane was hybridized with single strand DNA (ssDNA) probe of partial *matK* ORF. (C) The membrane was hybridized with a mixture of oligo probes covering *trnK* exon1 and exon2.

To confirm the identification of these transcripts, another Northern blot was performed with the RNA samples from the same preparation, and was hybridized with a mixture of oligo probes located in the *trnK* exon1 and exon2 (Figure 9A). The ~4kb *psbA-trnK* and ~2.5kb pre-*trnK* transcripts detected by the *trnK* exons probes show the same migration pattern as these transcripts detected by the *matK* ssDNA probe (Figure 9B and 9C). However, when the *trnK* exon probes were applied, the free intron could not be detected due to the probe location. Additionally, tRNA(K) occurs when detecting with exon probes (Figure 9C, see the bottom of the membrane), whereby a small amount migrates to the bottom fractions, e.g., fraction 12 (Figure 9C), indicating an active function of tRNA(K) in the translation.

As the polysome fractionation Northern blot identified all *matK* ORF containing transcripts as translational template, it is important to ascertain if there is a preference between these templates. Taking advantage of a GFP tagged rpL32 tobacco line, dot blot can quantify the amount of translational-active transcripts relative to the endogenous total transcripts. Using μ MACS column (Miltenyi Biotec), rpL32 containing complexes were separated from the remaining cell extracts as elution fraction. The transcripts associated with these complexes are translational-active RNAs. However, due to the detection of tRNA(K) by the *trnK* exon probe, the quantified results cannot indicate the proportion usage of different transcripts (Appendix 5).

### 3.1.7 Overexpression *matK* in nucleus

During the coevolution of endosymbionts and their host, large numbers of genes were transferred into the host nucleus genome, including many maturases. The only exception was *matK* in chloroplast (de Longevialle et al. 2010). This scenario was justified as follows: (1) As MatK is toxic for the nuclear RNAs, it may splice some of the nuclear RNAs, which form a structure similar to group II intron of MatK targets, or (2) MatK needs to be rapidly and tightly autoregulated; however, autoregulation requires the

transcripts and protein to be in the same location. In this study, attempts were made to express *matK* from the nucleus, and subsequently knock out the chloroplast *matK*.

A vector named prmh was constructed for the nuclear transformation using the backbone of pGL1 vector that is derived from pGPTVbar (Becker et al. 1992). Prmh was constructed by insterting *matK* (*Nicotiana tabacum*) with 35s promoter, transit peptide and HA tag into the *Sma*I restriction site of pGL1 (Figure 10B).

Here, 35s promoter was used for overexpression and rbcs transit peptide was attached at the *matK* 5' end to ensure the chloroplast location for the translated MatK. Basta was used for the plant selection, as tobacco lines transformed with prmh and pGL1 are resistant to Basta (Figure 10A). The HA tag at 3' end was used to produce an epitope tagged MatK protein for detection (Figure 10B). This vector was transformed into wild-type *Nicotiana tabacum* mediated by agrobacteria. The empty pGL1 vector was transformed as a control. Seedlings were first proved to contain the insertion by genotyping PCR with two pairs of primers, one of which was located in promoter and transit peptide region, with the other pair in the *matK*:HA ORF (Figure 10B). RT-PCR was carried out for detecting the transcription of MatK. Total RNA samples were reverse transcribed with gene-specific primer, and a control experiment without the adding of reverse transcriptase was also conducted. With the matKseqfor1 and HArev primer pair, the expected 532 bp product can be amplified. On the other hand, the 3'matKseqrev1and HArev primer pair allowed amplifying the 183 bp product (Figure 10C). This result indicates that the inserted fragment were transcribed in plants p1, p2 and p4. Western blot with total protein extract was performed with the RT-PCR positive plants, whereby the pGL1 plant was used as negative control, and the HA tagged rpoA plant (Finster et al. 2013) was used as positive control. The results revealed that no signal can be detected in any nuclear transformed plants (Figure 10D), indicating either no expression or low expression of MatK.

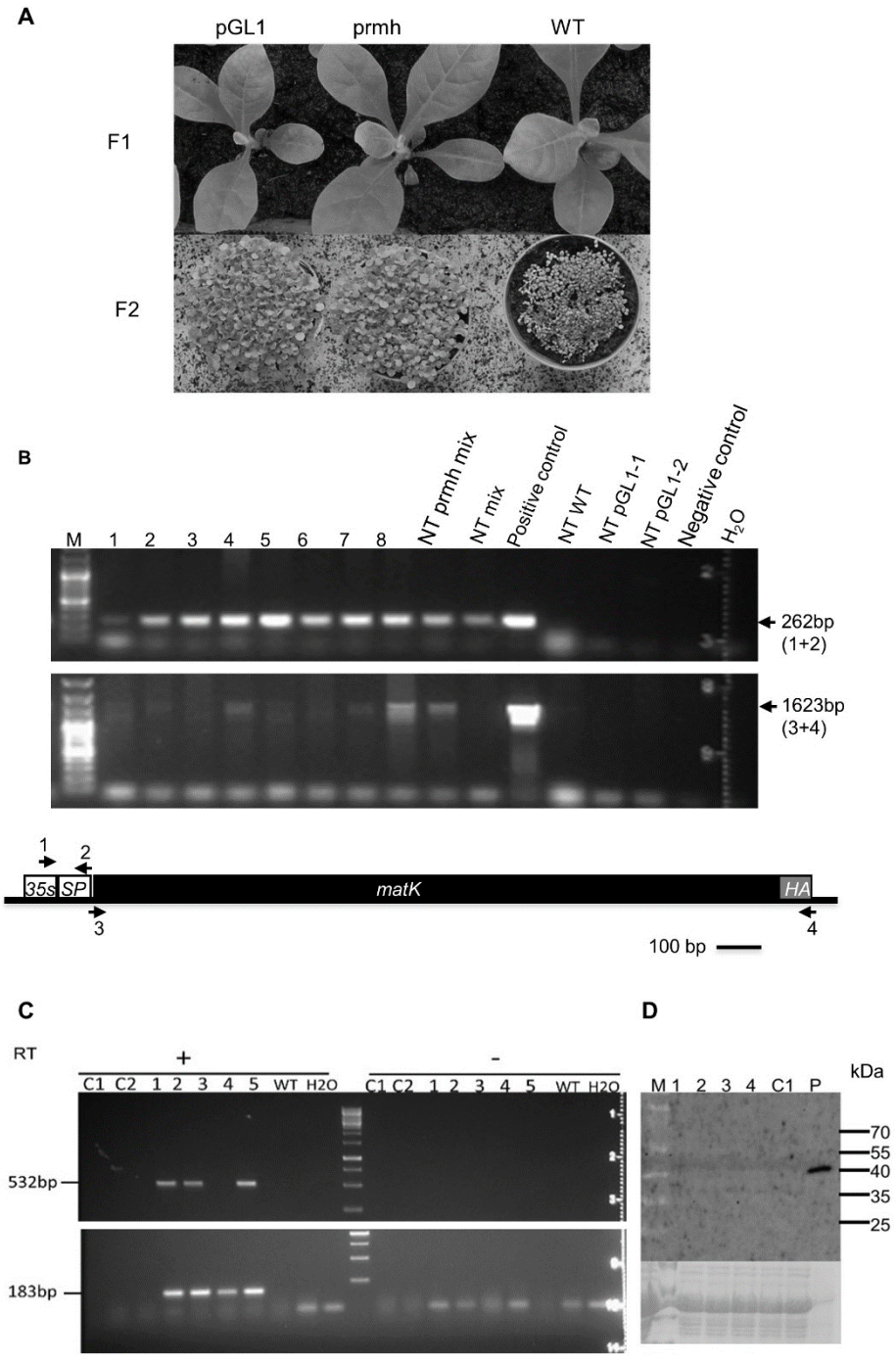

A
pGL1
prmh
WT
F1
F2
B
M 1 2 3 4 5 6 7 8
NT prmh mix
NT mix
Positive control
NT WT
NT pGL 1-1
NT pGL 1-2
Negative control
H2O
262bp
(1+2)
1623bp
(3+4)
1
2
35s
SP
matK
HA
3
4
100 bp
C
RT
+
-
C1 C2 1 2 3 4 5 WT H2O
C1 C2 1 2 3 4 5 WT H2O
532bp
183bp
D
M 1 2 3 4 C1 P
kDa
70
55
40
35
25

**Figure 10: Nuclear transformation of MatK in *Nicotiana tabacum*.** Chloroplast targeting expression vector prmh was transformed to nucleus by agrobacteria-mediated method. An empty vector pGL1 was also transformed as a control. (A) The F1 generation of nuclear transformants are indistinguishable from wild-type tobacco under standard growth condition (upper panel). Prmh and pGL1 transformants of F2 generation are resistant to Basta. 7-days old tobacco seedlings were sprayed with Basta and grown till 10-days old (lower panel). (B) The positive transformants (tobacco lines No. 1-8) were identified by genotyping PCR. Upper panel: PCR performed with primers located in the promoter and signal peptide region (primer pair 1 and 2), lower panel: PCR performed with primers located in the *matK*:HA ORF region (primer pair 3 and 4). The names of primers used are as follows, 1: 35s; 2: rbcssprev1, 3: matKsmalfor2; 4: HArev. (C) RT-PCR indicates the transcription of inserted fragments. DNase-treated total RNA was reverse transcribed with RT (+) or without RT (-) by a gene-specific reverse primer (HArev), and was amplified with primers located at the inserted fragments (HArev and 3'matKseqrev resulted a product of 183bp, HArev and matKseqfor1 resulted a product of 532bp). (D) Western blot indicates no expression of MatK protein. Total protein was separated with SDS-PAGE gel and blotted onto a nitrocellulose membrane. HA antibody was used for the detection, an HA tagged rpoA line (Finster et al. 2013) was used as a technical positive control. SP: signal peptide; RT: reverse transcriptase; M: Marker; C1 and C2: control lines transformed with pGL1.

### 3.1.8 Plastid transformation for *matK* overexpression

In chloroplasts, MatK is translated at a very low level, probably under a strict regulation (Barthet and Hilu, 2007). The RIP-seq showed that MatK binds to other structure-similar RNAs, indicating the importance of maintaining the low level expression. In this study, overexpression of MatK was carried out to examine the regulation of MatK. To avoid the potential toxic effects of MatK in chloroplast, vectors were constructed with an expression inducible element-Riboswitch (Verhounig et al. 2010). Two constant expression vectors with the T7g10 translation element (Ruhlman et al. 2010) were also prepared. The sequence used for MatK expression was either *matK* (*NTmatK*) of *Nicotiana tabacum* or synthetic *matK* (*AmatK*) that was codon optimized. AadA cassette was used for plant selection against spectinomycin and the HA tag at the 3' end of MatK was used for protein detection and purification. Four vectors were prepared in total, namely: pRSNT*matK*, pRSA*matK*, pT7g10NT*matK* and pT7g10A*matK* (Figure 11). They were used in stable transformation of tobacco plastids mediated by biolistic delivery systems (Svab et al., 1990; Ayako and Yutaka 2012).

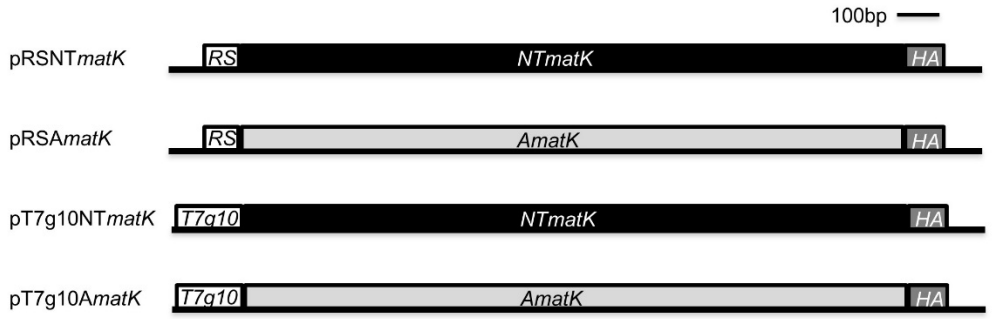



**Figure 11: Schematic diagram of vectors used for MatK overexpression in *Nicotiana tabacum* plastid.** Tobacco *matK* (*NTmatK*) and synthetic *matK* (*AmatK*) were either expressed constantly (pT7g10NT*matK* and pT7g10A*matK*) or inducible expressed by Riboswitch (pRSNT*matK*, pRSA*matK*). The 3' HA tag was used for protein detection and purification.

The first generation of plants (F1) was genotyped by PCR and Southern blot analysis, revealing that two independent lines generated from pRSA*matK* were homoplastomic, while only one heteroplastomic line was generated from pT7g10AmantK (Ranzini 2014). The analysis on the RNA splicing of the homoplastomic plants is still ongoing. The leaf pieces from the heterplastomic plants were cultured for further regeneration. At the same time, as segregation may occur in the seeds, the plantlets of the heteroplastomic line were transferred to soil after rooting for seed production.

## 3.2 SRNAs suggest the Mbb1 binding sites at 5'UTR of *psbB* and *psbH*

### 3.2.1 Chloroplast sRNAs identification within *Chlamydomonas* RNA-seq datasets

As sRNAs are distinct in functions but common in size, they are usually discovered by high-throughput sequencing of their cDNAs. Normally, the short RNAs are separated from total RNA by gel electrophoresis, after which they are eluted before being used as library template for sequencing (Malone et al. 2012). Datasets produced as outlined above include short RNAs derived from nuclear and organellar genomes when working with eukaryotes. In plants, these are the nuclear, plastid, and the mitochondrial genomes. Studies on sRNAs have been carried out in a wide variety of plant species including *Chlamydomonas* (Ibrahim et al, 2010).

Small-size RNAs such as miRNA and siRNA, were already shown to play diverse regulatory roles in the gene expression via different mechanisms (Huttenhofer et al.

2005; Lin et al. 2006; Kim et al. 2009; Lee et al. 2012; Duarte et al. 2014). Although most of these works focused on the nuclear or cytosolic sRNAs, other studies also identified sRNA members in organelles (Lung et al. 2006; Wang et al. 2011; Ruwe and Schmitz-Linneweber 2012; Hackenberg et al. 2013). Some of these organellar sRNAs are involved in plant stress response (Wang et al. 2011). In this study, the sRNAs in plastid of *Chlamydomonas* were identified and their roles in post-transcription regulation were examined.

Seven short RNA datasets in *Chlamydomonas*, originally designed to analyze miRNAs and siRNAs ((Ibrahim et al, 2010) and GEO series GSE32457), were used in the present study. Reads from all seven datasets were pooled and mapped to the chloroplast genome of *Chlamydomonas* (NC_005353). Identical reads were treated as one sequence, with each sequence potentially stands for up to hundreds of reads. A total of 17,400 unique chloroplast-specific sRNA sequences were obtained after mapping, which is equivalent to 2.2% of the total sRNA population. The enrichment of sequences across the chloroplast genome exhibited wide variations in local distribution in the tRNA, rRNA, mRNA regions as well as intergenic regions. The sRNA derived from *rrn* operon, tRNA and the other highly accumulated genes such as *psbA* and *rbcL* are the most abundant. In contrast, the sRNAs originating from other mRNA and intergenic regions were least prevalent but exhibited distinct isolated enrichment. As the non-coding regions are the potential binding sites for RNA-stabilizing proteins (Ruwe and Schmitz-Linneweber 2012; Zhelyazkova et al. 2012b), and were main focus of interest here, the remaining unrelated regions were excluded in further analysis. In total, 61 sRNA clusters were identified in this study, most of which were less than 300 bp away from their respective start or stop codon (Appendix 7).

### 3.2.2 SRNAs coincidence with transcript ends

In angiosperm, the binding of helical-repeat proteins can protect RNAs from degradation and generate sRNAs in the binding sites (Prikryl et al. 2011; Ruwe and Schmitz-Linneweber 2012; Zhelyazkova et al. 2012b). Consistent with this idea, in this study, chloroplast sRNAs tended to locate toward transcript ends. Moreover, in the analyzed datasets, five sRNAs map immediately at known transcript ends (Drapier et al. 1992; Sturm et al. 1994; Vaistij et al. 2000; Zerges et al. 2003; Somanchi et al. 2005) (Appendix 8). Four of these transcript ends were confirmed by 5' RACE (Figure 12A), and coincide with the sites where sRNAs emerge.

The sRNAs at 3'UTRs were frequently adjacent to a stem-loop structure, such as the sRNAs at *psbL* 3' and *trnR* 3'. *PsbL* sRNA 3' end was found to coincide with the 3' mRNA end identified using 3' RACE, and the stem-loop structure found was adjacent to this end (Figure 12B). Similarly, the 3' ends of *psbA* were mapped adjacent a stem loop (dG = -18.80 kcal/mol). However, no sRNA was mapped to this 3' end. Two sRNAs were found in the 5'UTR of *psbB* and *psbH*, respectively. Mbb1, the RBP which is contributing to the stabilization of these two transcripts, is proposed to bind to these two 5'UTR regions (Hammani et al. 2012) (Figure 12A).

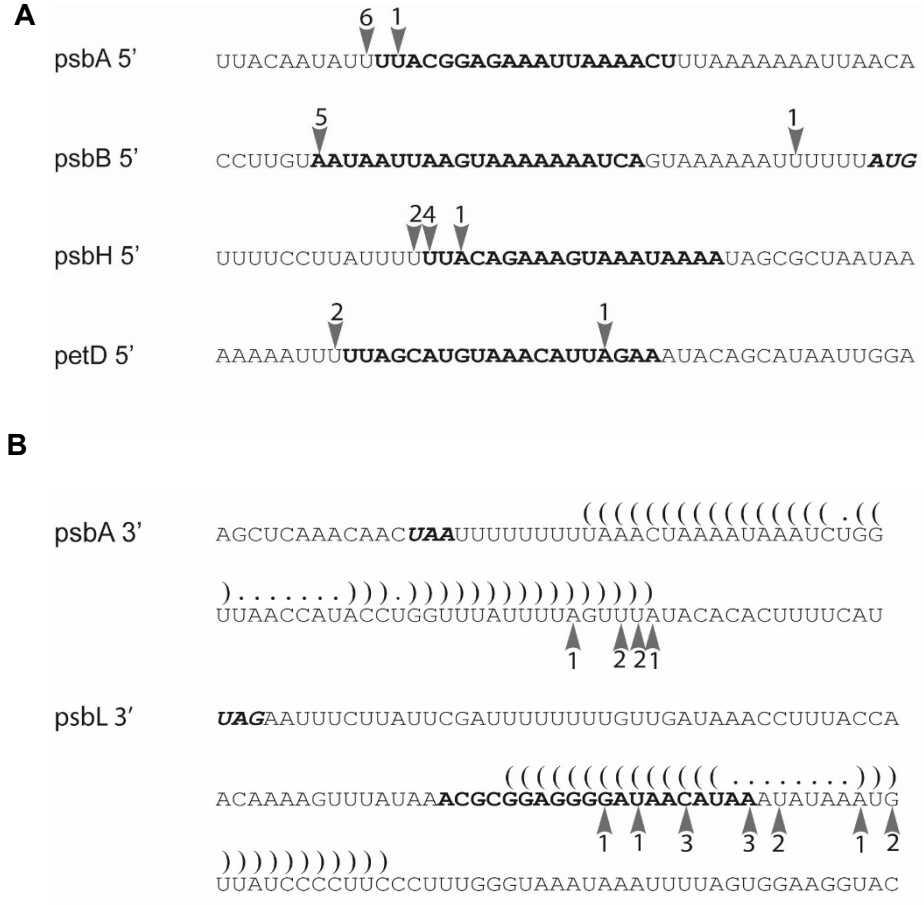



**Figure 12: SRNAs corresponding to the ends of chloroplast mRNAs in *Chlamydomonas*.** (A) The 5' ends of mRNA mapped by 5' RACE. Most clone ends coincide with the sRNA 5' ends. (B) The 3' ends of mRNA mapped by 3' RACE. *PsbA* 3' end coincides with a stem-loop structure (dG = -18.80 kcal/mol), while *psbL* 3' end coincides with the location of sRNA. The numbers above or below the arrowheads indicate the number of 5' or 3' RACE clones that terminate at the respective sites. The sequences corresponding to four chloroplast sRNAs (Appendix 8) are highlighted in bold. The start and stop codons are highlighted both in bold and italic. In the sequence with stem-loop structure, base-pairing nucleotides are indicated by matching parenthesized brackets while the unpaired positions are indicated by dots.

### 3.2.3 Absence of two sRNAs at UTR regions of Mbb1 target in the *mbb1* mutant

Binding of RBPs was shown to lead to the accumulation of sRNAs at the binding sites (Prikryl et al. 2011; Ruwe and Schmitz-Linneweber 2012; Zhelyazkova et al. 2012a). The wild-type and *mbb1* null mutant *Chlamydomonas* were used here to test whether Mbb1, the RBP that binds at the 5'UTR of *psbB* and *psbH,* is related to the accumulation of the sRNAs located at the respective regions. RNA gel blot hybridization was performed with low molecular weight (LMW) RNA fraction enriched from the total RNA. The aim of using this LMW fraction was to enrich the RNAs that with small molecular weight and thus increase the sensitivity of detection. Both of these sRNAs can be detected in wild-type with the sizes similar to the sequences found in deep sequence datasets (Figure 13A and Appendix 7). Moreover, individual sequences in each sRNA cluster exhibited variations in length, and the sharpness of the sRNA band in RNA gel blot could reflect this variation (Figure 13A). However, neither *psbB* nor *psbH* sRNAs can be detected in the mutant. The lack of these two sRNA in *mbb1* could be caused by a general loss of sRNA in this mutant. To exclude this possibility, the same blot was stripped and hybridized with probes for two control sRNAs: *petB* 5' and *psbH* 3'antisense (Figure 13A). Both control sRNAs were present in the mutant. It is obvious that the accumulation of *psbB* 5' and the *psbH* 5' sRNAs are dependent on the presence of Mbb1. And it has already been shown that the mRNA stability and/or processing of *psbB* and *psbH* transcripts are affected by Mbb1 (Monod et al, 1992). Interestingly, a sequence similarity between sRNAs of the *psbB* and *psbH* 5'UTRs was noted. More specifically, they share a core sequence: AAGUAAA (Figure13B), and the locations of this core sequence are ~7 nt downstream of 5' end for both transcripts (Figure 12).

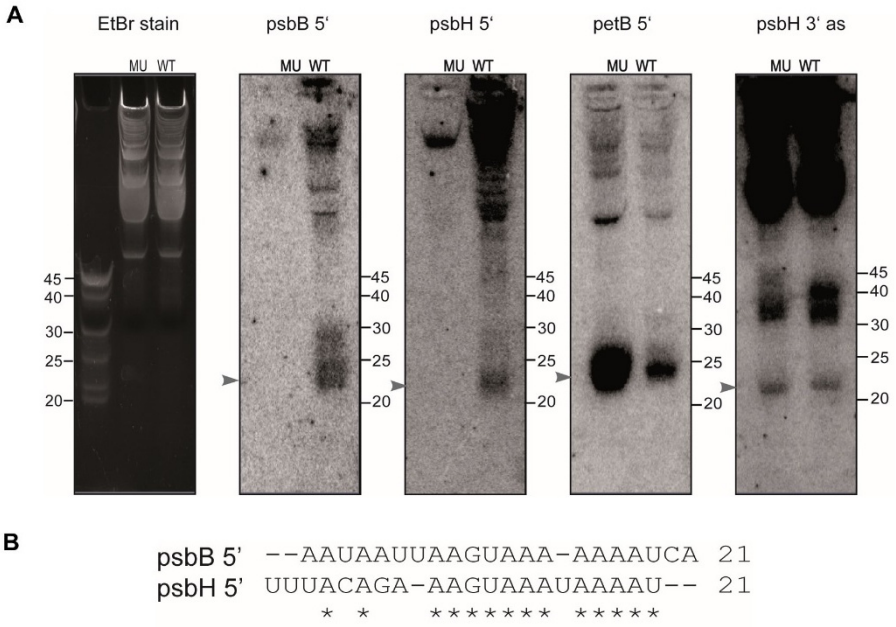



**Figure 13: SRNAs correspond to the 5' ends of chloroplast mRNAs in *Chlamydomonas***. (A) Autoradiograms of RNA blots from the WT and the *mbb1* mutant (MU). Low molecular weight (LMW) RNA was enriched from the total RNA and was resolved by 15% polyacrylamide gel containing 8M urea. A fluorescence image of the gel stained with ethidium bromide (EtBr) is shown as loading control (the leftmost panel). The probes are ssDNA oligonucleotides (<=25 nt) antisense to the sRNAs. The same membrane was repeatedly probed and stripped as described in Materials and Methods section. The radioactive probes used in hybridization are indicated above the panels. The arrowheads indicate the signal of the sRNAs. (B) Sequence alignment of the *psbB* 5' and *psbH* 5' sRNAs. SRNAs were aligned by CLUSTAL 2.1. The asterisks denote conserved nucleotides.

# 4. Discussion

## 4.1 Binding specificity of MatK to its RNA targets

Organelle group II introns have features not shared by their intron relatives. Unlike the prokaryotic group II introns, they are unable to perform self-splicing. Moreover, unlike the eukaryotic nuclear transcripts, they are not spliced via spliceosome. Thus far, a large number of nuclear-encoded splicing factors have been identified, which facilitating the splicing of group II introns (de Longevialle et al. 2010). Presently, MatK is the only presumed chloroplast encoded splicing factor, and is consistently found in the intron of *trnK* in most plants (http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi). The location of *matK* suggests its function in splicing the intron of *trnK*. Additionally, *matK* is retained and standing freely in several chloroplast genomes where *trnK* is lost (McNeal et al. 2009). MatK is associateed with the *trnK* as well as other group II introns ((Zoschke 2010) and this study), indicating its potential ability to splice other targets.

Binding of RBPs to RNA is the first step for facilitating their function. Typically, the binding specificity is achieved by recognition of a particular RNA sequence and/or structure (Draper 1999; Antson 2000; Hall 2002; Mougin et al. 2002; Singh 2002; Maris et al. 2005). A wide variety of experimental methods can be applied for identifying RNA ligands of RBPs. Next generation sequencing combined with bioinformatics analysis can uncover the more specific RNA structural contexts that are recognized by RBPs *in vivo* (Fukunaga et al. 2014).

### Structural requirements for RNA recognition by MatK

In this study, in addition to the enrichment of seven group IIA targets, which all show strong association with MatK, the signal for the group IIA intron *clpP-in2* is slightly above background. Furthermore, the two group IIB intron, *ycf3-in2* and *ndhA*, are enriched in one of the two RIP-seq results (Figure 5 and Appendix 5). These inconsistently associated introns have a very similar structure to the other introns targets (Michel et al. 1989), suggesting that RNA structure plays an important role in the RNA-MatK interaction.

Bacterial maturases such as LtrA bind their cognate intron (in this case *Ll.LtrB*) with high affinity and specificity. No other group II intron was found to bind to LtrA under stringent binding conditions (Saldanha et al. 1999). When the conditions were less stringent, LtrA

was also found to bind other group II introns, such as yeast mtDNA introns *aI2* and *aI5* and *E. coli IntB* (Saldanha et al. 1999).

As the substrate of maturase, group II introns possess only a few conserved nucleotides. With the exception of the relatively highly conserved and well characterized DV, the remaining conserved segments are short and scattered across the intron. For example, the consensus sequences at the 5' and 3' boundaries of introns are GUGYG and AY (Y = C, U), respectively. However, they are usually too short to define boundaries in genomes (Michel and Ferat 1995). In contrast, the secondary structures of the group II introns are highly conserved.

RNA structures can form cotranscriptionally and can thus provide rapid access to their protein ligands while not yet fully produced (Lai et al. 2013; Geary et al. 2014). Moreover, MatK may bind to RNAs when they are not yet fully folded during transcription or to introns accumulating in an unfolded state. In both cases, it could play a role in the further folding of its targets. MatK is likely to bind to some introns, including *clpP-in2*, since they are structurally related to the core group IIA targets. However, the affinity to these RNAs might be low (see low *clpP-in2* signal in Figure 6) due to structural deviations from the canonical group IIA intron scheme (Michel et al. 1989). Consequently, it is possible that MatK does not support splicing in these cases. Genetic studies exploring the splicing role of MatK for its various target introns are thus needed to answer this question.

**Proof for MatK being a true splicing factor is still missing**

MatK is presumed to splice multiple targets. However, the only evidence for the involvement of MatK in chloroplast splicing is that the seven group IIA introns (with the exception of *clpP-in2*) are not spliced in the plastid translation deficient barley mutant, while the group IIB intron in *trnG* was spliced efficiently (Hess et al. 1994; Hubschmann et al. 1996; Vogel et al. 1999). The intron in *trnV* was presumed to be among these seven MatK-dependent introns, although the data are hard to interpret due to the low signals obtained in Northern blot experiments (Vogel et al. 1999). In the present study, *trnV* association with MatK is also under the predefined threshold in one of the two RIP-seq experiments. It is possible that t*rnV* is not a true splicing target of MatK. However, *trnV* intron is lost in the plastid genome of the underground orchid *Rhizanthella gardneri*, where *matK* is also lost. As a result, MatK and *trnV* may lose contact due to a weak association.

Two introns bound by MatK in tobacco, group IIA intron *rpl2* and *clpP-in2*, are retained and are correctly spliced in *Rhizanthella gardneri* (Delannoy et al. 2011). The splicing of *clpP* requires a PPR protein in the moss *Physcomitrella patens* (Hattori et al. 2007). Thus, MatK could be redundant in the presence of such factors, or serve only non-essential functions in splicing of such introns. Similar argument could be applied to the *rpl2* intron. In other words, while MatK is able to bind to multiple RNAs, its splicing function is not presently fully understood. Thus, further manipulation of MatK is necessary for verifying its function in splicing. Since standard knock out approaches were unsuccessful (Sorefan et al. 2012), conditional knock out or overexpression of MatK should be attempted.

**Binding of MatK to its targets is developmentally regulated**

In this study, MatK was found to be dynamically associated with group II introns. More specifically, in a comparison of 7-day and 25-day-old tobacco (Figure 7), MatK association with *atpF* was found to increase 3 times in older tissue, while its association with *trnV* decreased two-fold. These differences in association are independent of accumulation levels of RNAs that are described by Hertel *et. al.* (Hertel et al. 2013). The splicing efficiency of intons could be affected as other studies have shown that splicing efficiency varies at different developmental stages. For example, differences were found in the ratio of spliced to nonspliced *atpF, petB, petD, rbl16* and *ycf3* in maize and *trnG* in mustard (Barkan 1989; McCullough et al. 1992; Liere and Link 1995). Additionally, the splicing of the group IIB intron in *ndhB* is impaired by high temperature (Karcher and Bock 2002).

The dynamic association between MatK and its targets indicates a regulation in splicing. This raised an important question of, whether MatK-dependent splicing is important for the accumulation of the final product of gene expression, the mature tRNA or the encoded protein. The MatK association with *trnA*, *trnV* and *trnI* varied during the development while all three corresponding mature tRNAs decreased (Hertel et al. 2013). This suggests that, at least in the developmental stages analyzed, MatK is not rate-limiting for these splicing events. Unfortunately, no protein expression data for the ORF containing targets of MatK such as *atpF* and *rps12* in tobacco, presently exists. Furthermore, it is clear that MatK is not the only factor that assists the splicing of these introns. The involvement of other nucleus-encoded splicing factors (Figure 2) is likely to cause the dynamic association between MatK and its targets.

**Association of MatK with intron-containing transcripts**

Bacterial maturases are able to bind to the precursor and lariat intermediate during splicing, and afterwards still bind to the excised lariat intron to facilitate the retrohoming activity (Bonen and Vogel 2001). For MatK it was unclear, whether the precursor or the spliced intron is associated with MatK. In this study, RT-qPCR was used to discriminate the different transcripts associated with MatK. Although detection of the lariat structure failed (Koehler 2013), precursor was found to be the predominant transcripts associated with MatK, while the mature transcripts were not associated (Figure 8B). The binding of MatK to the lariat RNA could be impaired due the evolutionary loss of the RT domain, which is known to be important for assisting lariat-based retrohoming (Cui et al. 2004). In the same assay, the 5' end of the precursor is more enriched than the 3' end for *trnA* (Figure 8B), indicating that associations of RNA to MatK occur at the 5' end of transcripts rather than at the 3' end. This association tendency could due to the uneven degradation of RNA at the 5' and 3' end during chloroplast isolation and co-IP. However, this observation is not supported by the RIP-seq data (*trnA* and *trnV* in Figure 6), where tendency of 5' end enrichment was not found. As a result, the most likely reason for this scenario is that the degraded short RNA fragments at 3' end detected by RIP-seq were excluded in the RT-qPCR.

**Enriched intronless transcripts in co-IP with MatK**

In this study, the intron ligands that were repeatedly found to co-precipitate with MatK are group IIA introns. In addition, some intronless transcripts, such as tRNAs, rRNAs and RNAs coding for photosystem subunits, were also found to be bound to MatK (Appendix 6). One reason for their presence in MatK-IP pellets could be their high abundance within the chloroplast transcriptome in comparison to the very low-expressed "MatK-bait" used in the co-IP. These RNAs could be an artifact of the co-IP procedure and would not be true MatK targets. Since it was found that re-association of unspecific RNAs to protein could occur after cell lysis (Mili and Steitz 2004).

## 4.2 The binding of MatK to different domains of its targets

**Adaptation of MatK to assist splicing of multiple introns**

The purified LtrA was shown to bind to RNA as dimers *in vivo* (Saldanha et al. 1999). Biochemical analysis suggests that the N-terminus of the RT domain of LtrA is essential for high affinity binding to the *LI. LtrB* intron, while domain X of LtrA is associated with the conserved regions of the catalytic core of the intron (Cui et al. 2004). However, MatK maintained only part of the C-terminal RT domain and the entire X domain. The lack of large parts of the RT domain, especially the N-terminal part, may lead to the loss of binding specificity of MatK. Potentially, this loss was a prerequisite for MatK to bind to multiple targets and assist splicing of multiple introns.

**Relationship between function of domain and its enrichment**

As shown in this study, DI of group II introns is associated with MatK for all its targets (the differences in the two sequencing results will be discussed later). In contrast, there is no consistent enrichment of DII and DIII across all target introns. This domain is also usually the largest and the most complex. DI interacts strongly with DV and DVI at the tertiary level (Boudvillain and Pyle 1998). It was assumed that DI delivers the molecular scaffold into its catalytic active structure (Qin and Pyle 1997). In plants, DII and DIII are usually short in length. Their contributions have been shown to be limited in structure formation and splicing efficiency in bacterial group II introns (Kwakman et al. 1989; Koch et al. 1992). DIV is variable in size in bacteria since it does not always retain an ORF. Extant studies on bacterial maturase mutants revealed the interaction between different regions of maturase and the domains of group II intron. For example, stable binding of the maturase to the intron requires binding to its DIV (Huang et al. 2003). However, 10% and 70% of residual splicing could still be detected *in vivo* for *L1. LtrB* and *cox1-in2* when their DIVs are missing, respectively (Huang et al. 2003; Cui et al. 2004). The residual splicing efficiency suggests importance of DIV-maturase interaction differs depending on the respective intron and/or maturase. For MatK, a strong enrichment of DIV was found for *rpl2* and *trnA*, whereas this domain is relative weakly enriched in other MatK target introns (Figure 6, Table 1 and Table 2). This scenario suggests that, even for the same maturase, the association of DIV differs among its intron ligands. RIP-seq data of *trnK* suggests that the *matK* ORF is not bound by the maturase; this observation is in line with the results of a *trnK*-oligo-tilling-array (Zoschke 2010). DV is the most conserved domain in both length (always 34 nt in plants) and sequence (Michel et al. 1989). It is relatively more enriched than other domains in the seven introns discussed in this work, which may be due to its essential role for catalytic activity

(Peebles et al. 1995; Konforti et al. 1998). DVI in the data reported here is also enriched in most of the introns found in Illumina RIP-seq results. It encompasses the branch "A" site and has been shown to interact with DI and DV (Dib-Hajj et al. 1993; Podar and Perlman 1999). Together, the MatK RIP-seq results indicate a preference of binding to different domains for its individual targets.

**Differences of domains enrichment between bacterial maturases and MatK**

Using quantitative binding assays and RNA footprinting experiments, RNA sites protected by bacterial maturase LtrA was found in DI, DII and DVI (Matsuura et al. 2001). As DII is not strongly enriched in MatK RIP-seq, the binding results reported here do not fully coincide with those pertaining to the bacterial maturase. This incongruence may be due to two possible reasons: (1) Bacterial maturase only serves its own intron but the plastid MatK serves multiple group II introns. Thus, there must be some differences in the recognition and binding mechanisms between these two types of maturases; (2) The RIP-seq results obtained in this study suggest the *in vivo* association. However, the available data for bacterial maturase binding were obtained from recombinant maturase and RNA substrate produced by *in vitro* transcription. Thus, as co-transcriptional folding of RNA was missing (Zemora and Waldsich 2010), the bacterial maturase is likely to make additional and/or different contacts with the RNA substrate and thus correct the RNA structure, changing it from the random-folded or even misfolded conformation.

**Different library preparation methods result in different association patterns**

Some differences were found for domains association between the two sequencing methods. With the exception of DI, consistent binding pattern could not be established for other domains. According to the Illumina RIP-seq results, DI and DVI are enriched for all target introns, while DV is enriched in three target introns in Ion Torrent sequencing. The differences are likely caused by the library preparation. The RNAs from the IP pellet were of various lengths, ranging from about 20 nt to more than 1 kb. Using the sRNA library preparation kit, only RNAs smaller than 80 nt can be retained, as the rest is excluded. As a result, the average read length for this library is 78 nt. In order to include all the RNA from IP pellet into the library, an RNaseIII digestion step was used to shorten the RNA in the library preparation for Ion Torrent sequencing. RNase III digests the RNA molecules at the positions where the double strand forms and thus, very useful for sequencing the entire transcriptome with abundant long transcripts. However, it is difficult to recover the

low abundant transcripts after the RNase III digestion. Additionally, RNase III digestion may be also a challenge for the group II introns, as their complex secondary structures are built based on the formation of double strand. Thus, a digestion-induced bias could be an issue, due to the nonrandom double strand formation. Moreover, it is very likely that MatK recognizes structure instead of sequence. As a result, digesting the RNA targets without crosslinking may release the RNA ligands from the RNA-protein complex. However, treatment with RNase is required for obtaining a higher resolution. RNase digestion combined with crosslink prior to IP may sharpen the peaks of enrichment shown in Figure 6 and eliminate the enrichment at the exon regions. RNA cloning bias, and ligation bias in particular, could also be the reason for the differences found between these two sequencing methods. Ligation bias can be introduced by the ligase and RNA secondary structure (Linsen et al. 2009; Hafner et al. 2011). For the group II intron RNAs with complex secondary structure, the ligation bias could be alleviated by correcting the adapters and/or via utilization of a thermo stable ligase at a higher temperature (Sorefan et al. 2012; Zhang et al. 2013).

Thus far, the minimal binding sites of many RBPs were characterized with RIP-seq in combination with the RNA-crosslink technique（Friedersdorf and Keene 2014; Munschauer et al. 2014; Webb et al. 2014）. Presently, there is no available deep sequencing data for the interaction of MatK to its targets. While the first trial without crosslink in this study narrowed down the binding site to the intron domains level. In order to obtain a higher resolution, it would be beneficial to conduct the crosslink of MatK to its targets, along with RNase digestion and size selection.

## 4.3 Regulation of MatK

MatK is the only maturase retained in the chloroplast, as all the other counterparts were transferred to the nucleus during evolution (de Longevialle et al. 2010). Its autoregulation is one possible reason for retention of MatK in chloroplasts. In other words, as the MatK protein could regulate its own transcription and/or translation, it would need to be encoded in *cis*. Supporting this idea of regulated MatK expression, a discrepancy of MatK at the RNA and protein level was found in 7-day and 25-day-old tobacco plants (Hertel et al. 2013).

**Feedback regulation in chloroplast**

In *Chlamydomonas*, feedback regulation in the expression of some ribosomal proteins, some tRNAs, tufA and genes encoding the subunits of the plastid RNA polymerase has been reported (Ramundo et al. 2013). Interestingly, a feedback regulation for MatK was suggested by the discrepancy between protein level and RNA level during development, and was further supported by bioinformatics analysis using additional data on the developmental expression of MatK targets (Hertel et al. 2013). Bacterial maturase LtrA is autoregulated via a stem-loop structure in the start codon region (Singh et al. 2002). A similar structure was found in the same region of the *matK* gene in maize and tobacco. Consequently, the autoregulation model used for LtrA was suggested to apply in this case (Zoschke et al. 2009). However, the following analysis of co-expression of MatK and the 5' stem-loop region in *E.coli* did not provide evidence of any interaction (Zoschke et al. 2009). While this may be due to the limitations of the *E.coli* system, it can also reflect the actual situation. Several attempts were also made to test the potential autoregulation in an *in vitro* translation system (Yukawa et al. 2007). However, no expression of the reporter gene was obtained when it is driven by various lengths of *matK* 5'UTR (unpublished data, prof. Masayuki Nakamura, personal communication). It is likely that the *matK* expression elements used in this system were too weak for producing detectable reporter gene products.

**A potential autoregulation mechanism for *matK***

A possible autoregulation mechanism can be achieved by the location of *matK* and the function of its protein. Chloroplast RNAs are usually organized as polycistronic transcripts. After processing of these precursors, different transcripts with variable ORF content are produced. Transcripts containing the *matK* ORF are: *psbA-trnK*, pre-*trnK* and *trnK*-in (see Figure 9). As MatK splices pre-*trnK* to produce *trnK*-in, if pre-*trnK* is the only or major translational template for MatK, an increased amount of MatK could decrease the template from which it is translated by splicing. As a result, translation of MatK would be decreased. In the maize *crp1* mutant lacking the processing of *petB* and *petD*, the polycistronic precursor accumulates to its normal level while monocistronic forms of *petB* and *petD* are missing. PetB protein accumulates to the normal level in this mutant, while PetD protein largely decreases, suggesting that PetD is translated more efficiently from the monocistronic transcript (Barkan et al. 1994; Fisk et al. 1999). In the polysome analyses conducted in this study, no selection was found for a particular translational template of MatK. *psbA-trnK,* pre-*trnK* and *trnK*-in are all translated, in line

with the observation that maize PsbB is translated from all the transcripts containing its ORF (Barkan 1988). The initiation of translation requires the recognition of specific RNA elements within RNAs, including the recognition of Shine-Dalgarno like sequences for some chloroplast genes (Sugiura et al. 1998; Kozak 1999; Zerges 2000; Hirose and Sugiura 2004). It is thus likely that these *matK* transcripts comprise all the necessary elements since they mostly differ in the presence of the *trnK* exons. As a result, there is no discrimination in the selection of translational templates as evidenced by the analysis presented here. Thus, it is unlikely that autoregulation is driven by choice of the unspliced, rather than the spliced, *matK* RNAs.

**Misexpression of MatK is toxic**

Manipulation of protein expression is an essential method for studying protein function. However, all the previous attempts to knock out or point mutate *matK* had failed (Drescher 2003b; Zoschke et al. 2010). In this work, no detectable amount of MatK was found in nuclear transformants that were generated to overexpress MatK ectopically. In addition, thus far, any attempt to constantly express MatK in transplastomic plant produced heteroplastomic plants only, while further selection experiments for segregation are still ongoing. Nevertheless, the available data suggests the essential nature of *matK*. Potentially, MatK is able to target other RNAs aside of its canonical intron targets, which could lead to problems in the expression of such promiscuous targets when MatK is present in large quantities. One potential issue could arise because MatK interferes with the splicing of non-target introns, i.e. group IIB introns. In order to test this premise, a truncated form of MatK lacking domain X could be expressed, which would be expected to loose its “toxicity”. Such truncation approaches have indeed shown a high protein expression in plants (Akama and Takaiwa 2007; Wu et al. 2014). Moreover, an inducible MatK expression system was chosen here, as it might help to control the timing and/or protein levels during the specific critical stages of plant development (in particular in early chloroplasts development in meristematic tissue). This, in turn, could minimize the detrimental effects of MatK overexpression. Riboswitch has been shown to be able to control the transcription and translation in chloroplast (Verhounig et al. 2010) and, as a result, can be used in the inducible expression or knock out. Another option for the future would be to overexpress MatK in plant cell lines such as tobacco BY-2 (Bright Yellow - 2 of the tobacco plant) (Kato 1972) which would overcome the potential toxic effect of MatK in certain (young) tissues *in vivo*.

## 4.4 MatK, an intermediate between bacterial maturases and the spliceosome?

The splicing of each bacterial group II intron is catalyzed by the maturase encoded in the intron. By contrast, a large number of nuclear introns are spliced by a uniform splieosome machinery. These two splicing machineries are identical in terms of biochemical reaction steps (Sontheimer et al. 1999; Fica et al. 2013). There are also sequence similarities at the splice sites of the RNA substrates, as well as crystal structure similarities within the ribozyme and the spliceosome (Cech 1986; Toor et al. 2008; Keating et al. 2010; Marcia et al. 2013; Marcia and Pyle 2014). Based on these parallels, the catalytic group II introns are believed to be the ancestors of spliceosomal introns (Sharp 1985; Cech 1986). Sequence and crystal structure similarities between the Th/X domain of spliceosome core protein Prp8 and the X domain of group II intron encoded maturase in bacteria and fungi were also found (Dlakic and Mushegian 2011; Galej et al. 2013). However, there is still a large "gap" between them, raising more questions to be answered. For example: how did the one-intron specific maturase evolve into the multiple-intron serving spliceosome? What was the fate of the domains of the group II introns? Although a function of splicing is not yet fully understood, organelle maturase such as MatK or MatR may yield some valuable information, allowing the "gap" between bacterial maturase and the spliceosome to be narrowed.

The present study yielded several important findings. First, MatK is associated with multiple introns. This is similar to the nuclear spliceosome but different from the bacterial maturase. Second, similar to the bacterial maturase, MatK makes contact with multiple domains within group II introns, even though it interacts with different domains for each intron. Similarity in protein sequences of bacterial maturase and organellar maturase has already been reported (Neuhaus and Link 1987). Additionaly, four nuclear maturases, namely nMAT1-4, were identified in the plant genomes. They are closely related to group II intron maturase and facility the splicing of multiple mitochondrial mRNAs (Keren et al. 2009; Keren et al. 2012; Brown et al. 2014; Cohen et al. 2014). Moreover, in addition to the similarity in splicing mechanism, several group II intron domains, such as DV, can complement a spliceosome lacking certain U-RNAs (Jarrell et al. 1988; Dib-Hajj et al. 1993; Chin and Pyle 1995; Michels and Pyle 1995; Podar et al. 1995). Recently, sequence similarity between organellar intron *trnI* and U2 snRNA was found (Gert Weber, personal communication), this resembles the functional equivalence between snRNAs and domains of group II intron that have been previously described and experimentally tested (Hetzer

et al. 1997; Shukla and Padgett 2002). Prp8 was demonstrated to bind to U1, U2, U5, and U6 snRNAs, as well as the 5′ splice site, break point site and 3′ splice site region of intronic pre-mRNA (Li et al. 2013). This suggests that the involvement of multiple RNA segments in spliceosome is necessary to facilitate splicing (Grabowski et al. 1985; Pikielny and Rosbash 1986). Because MatK associates with multiple intron elements, it might be taken as a simplified model for the extremely versatile Prp8 protein, which makes contacts with multiple RNA segments and proteins during splicing.

## 4.5 SRNAs reveal the binding sites of proteins which are functioning in RNA processing and stability in *Chlamydomonas reinhardtii*

**Relationship between sRNAs and RBPs**

Nuclear-encoded proteins have been shown to play essential roles in the regulation of plastid gene expression in *Chlamydomonas* (Monod et al. 1992; Stampacchia et al. 1997; Boudreau et al. 2000; Tanaka et al. 2000; Auchincloss et al. 2002). Unlike the PPRs mostly found in higher plants, OPRs are extensively found in *Chlamydomonas* (Rahire et al. 2012), and an additional group, TPRs, is ubiquitous in bacteria, fungi, plants and animals (Blatch and Lassle 1999). In the present study, some of the sRNAs mapped to the UTRs of RNA and were particularly adjacent to the ends of transcripts (Appendix 6). This is in line with the fact that regulatory RBPs usually bind in the UTR regions of RNAs. Helical-repeat proteins form a large group among these RBPs in organellar RNA metabolism. RBP, and PPR proteins in particular, can generate short remnant RNAs by binding and protecting RNA from exonucleolytic degradation (Prikryl et al. 2011; Ruwe and Schmitz-Linneweber 2012; Zhelyazkova et al. 2012a). As a result, the sRNAs at the 5'UTR of the *psbA*, *psbD* and *petD* may represent the footprints of nuclear-encoded proteins Rbp63, Nac2 and MCD1 (Boudreau et al. 2000; Ossenbuhl et al. 2002; Murakami et al. 2005), respectively. Binding of Mbb1 at the 5'UTR of *psbB* and *psbH* also result in sRNAs.

In the *mbb1* mutant, the two sRNAs disappear while other sRNAs such as *petB* 5' and *psbH* 3' are still present, supporting the specific connection between sRNA and RBP. The remaining sRNAs in the list produced in this study (Appendix 7) are also interesting candidates for protein binding sites, especially for the nuclear-encoded PPRs or OPRs. Several systematic studies suggested that eukaryotic RBPs regulate sets of functionally

related RNAs (Keene and Tenenbaum 2002; Hogan et al. 2008). MBB1 bi-targetes the *psbB* and *psbH* RNAs that encode subunits of photosystem II, suggesting a correlation of regulation between these two subunits, possibly similar to the CES (epistasy of synthesis) process (Choquet et al. 1998). Indeed, PsbH protein was found locate close to CP47 in PSII complex and was shown to stabilize the binding of CP47 to the PSII reaction center in cyanobacteria (Komenda et al. 2002; Bumba et al. 2005; Iwai et al. 2006). Thus, sRNAs can be used in identifying RBP binding sites and, additionally, they may supply information of the RBP regulation.

**SRNAs undergo dynamic changes**

Several sRNAs were detected successfully by Northern blot. However, the sRNA for *petD* 5'UTR was not detectable when the same method was applied (data not shown). This difference may be caused by the nutrition conditions that the *Chlamydomonas* were subjected to. Indeed, it has been shown that the transcriptome of *Chlamydomonas* is connected with the copper and iron nutrition (Castruita et al. 2011; Urzica et al. 2012). Supporting this, in the present study, the number of reads for each sRNA varies in the libraries prepared from wild-type *Chlamydomonas* strain grown under different nutrition conditions (normal conditions, phosphate starvation, or sulphate starvation). Specifically, in the case of *psbH* 5' sRNA, the number of reads obtained under the three nutrition conditions are 25, 4 and 0, respectively, indicating a correlation between sRNA and nutrition for *Chlamydomonas*. Since these read numbers are rather low, verification by targeted sequencing of chloroplast sRNAs in Chlamydomonas needs to be carried out to corroborate these preliminary findings.

**Presence of abnormal sRNAs**

Three overlapping sRNAs were found to be mapping to the 5'UTR *psaC* transcript. However, using sRNA Northern blot, only a single sRNA with a size of 50 nt could be detected. The three sRNAs found by RNA-seq might, when combined, simply represent the longer sRNAs identified by classical Northerns blot (Appendix 8). This discrepancy could be due to the size selection of RNAs in the sequencing library. In the library analyzed in this study, RNAs exceeding 30 nt in length were excluded. As a result, some of the remaining sRNAs may be further degradation products of relatively longer sRNAs (i.e. the 50 nt sRNA detected by Northern blot). The existence of these three sRNAs could also mean that multiple proteins bind this region. However, these three overlapped

sRNAs can not be derived from the one single transcript. Moreover, the likehood that they are produced by the same RBP is low, as they are not conserved in sequence. Thus, they may have originated from different RBPs that capture part of the *psaC* population at the 5'UTR.

**Roles of sRNAs**

Eventhough the sRNAs found in the UTR regions of plastid RNAs were shown to be caused by RBP and/or stem-loop structure (Ruwe and Schmitz-Linneweber 2012; Zhelyazkova et al. 2012b), they might also play regulatory roles rather than only being the mere "junk" of the degradation products. Firstly, the sRNAs could compete for the RBPs with their respective long transcripts and could thus regulate their stability and/or translation. Secondly, by selectively binding to the complex that contains a protein factor, such as PPR, sRNAs may guide the regulatory roles of the protein complex. Thirdly, as small potentially mobile molecules, sRNAs could mediate the crosstalk between the organelle and nucleus if they are able to go through the barrier of organelle membrane. Finally, some sRNAs are antisense to their respective long transcripts. As a result, an interaction between them may assistant the regulation of target RNAs.

# 5. References

Akama K, Takaiwa F. 2007. C-terminal extension of rice glutamate decarboxylase (OsGAD2) functions as an autoinhibitory domain and overexpression of a truncated mutant results in the accumulation of extremely high levels of GABA in plant cells. *J Exp Bot* 58: 2699-2707.

Antson AA. 2000. Single-stranded-RNA binding proteins. *Curr Opin Struct Biol* 10: 87-94.

Asakura Y, Galarneau E, Watkins KP, Barkan A, van Wijk KJ. 2012. Chloroplast RH3 DEAD box RNA helicases in maize and Arabidopsis function in splicing of specific group II introns and affect chloroplast ribosome biogenesis. *Plant Physiol* 159: 961-974.

Auchincloss AH, Zerges W, Perron K, Girard-Bascou J, Rochaix JD. 2002. Characterization of Tbc2, a nucleus-encoded factor specifically required for translation of the chloroplast psbC mRNA in Chlamydomonas reinhardtii. *The Journal of cell biology* 157: 953-962.

Ayako O, Yutaka T. 2012. Improvement of the plastid transformation protocol by modifying tissue treatment at pre- and post-bombardment in tobacco. *Plant Biotechnology* 29: 307-310.

Ayliffe MA, Scott NS, Timmis JN. 1998. Analysis of plastid DNA-like sequences within the nuclear genomes of higher plants. *Mol Biol Evol* 15: 738-745.

Baginsky S, Gruissem W. 2002. Endonucleolytic activation directs dark-induced chloroplast mRNA degradation. *Nucleic Acids Res* 30: 4527-4533.

Barkan A. 1988. Proteins encoded by a complex chloroplast transcription unit are each translated from both monocistronic and polycistronic mRNAs. *EMBO J* 7: 2637-2644.

Barkan A. 1989. Tissue-dependent plastid RNA splicing in maize: transcripts from four plastid genes are predominantly unspliced in leaf meristems and roots. *Plant Cell* 1: 437-445.

Barkan A. 1998. Approaches to investigating nuclear genes that function in chloroplast biogenesis in land plants. in *Methods in Enzymology*, pp. 38-57.

Barkan A, Rojas M, Fujii S, Yap A, Chong YS, Bond CS, Small I. 2012. A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet* 8: e1002910.

Barkan A, Walker M, Nolasco M, Johnson D. 1994. A nuclear mutation in maize blocks the processing and translation of several chloroplast mRNAs and provides

evidence for the differential translation of alternative mRNA forms. *EMBO J* 13: 3170-3181.

Barthet MM, Hilu KW. 2007. Expression of matK: functional and evolutionary implications. *American journal of botany* 94: 1402-1412.

Becker D, Kemper E, Schell J, Masterson R. 1992. New plant binary vectors with selectable markers located proximal to the left T-DNA border. *Plant Mol Biol* 20: 1195-1197.

Blatch GL, Lassle M. 1999. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *BioEssays : news and reviews in molecular, cellular and developmental biology* 21: 932-939.

Blocker FJ, Mohr G, Conlan LH, Qi L, Belfort M, Lambowitz AM. 2005. Domain structure and three-dimensional model of a group II intron-encoded reverse transcriptase. *Rna* 11: 14-28.

Bollenbach TJ, Stern DB. 2003. Secondary structures common to chloroplast mRNA 3'-untranslated regions direct cleavage by CSP41, an endoribonuclease belonging to the short chain dehydrogenase/reductase superfamily. *J Biol Chem* 278: 25832-25838.

Bonen L, Vogel J. 2001. The ins and outs of group II introns. *Trends in genetics : TIG* 17: 322-331.

Boudreau E, Nickelsen J, Lemaire SD, Ossenbuhl F, Rochaix JD. 2000. The Nac2 gene of Chlamydomonas encodes a chloroplast TPR-like protein involved in psbD mRNA stability. *EMBO J* 19: 3366-3376.

Boudvillain M, Pyle AM. 1998. Defining functional groups, core structural features and inter-domain tertiary contacts essential for group II intron self-splicing: a NAIM analysis. *Embo J* 17: 7091-7104.

Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry* 72: 248-254.

Braukmann T, Kuzmina M, Stefanovic S. 2013. Plastid genome evolution across the genus Cuscuta (Convolvulaceae): two clades within subgenus Grammica exhibit extensive gene loss. *J Exp Bot* 64: 977-989.

Brown GG, Colas des Francs-Small C, Ostersetzer-Biran O. 2014. Group II intron splicing factors in plant mitochondria. *Front Plant Sci* 5: 35.

Bumba L, Tichy M, Dobakova M, Komenda J, Vacha F. 2005. Localization of the PsbH subunit in photosystem II from the Synechocystis 6803 using the His-tagged Ni-NTA Nanogold labeling. *Journal of structural biology* 152: 28-35.

Butcher SE, Brow DA. 2005. Towards understanding the catalytic core structure of the spliceosome. *Biochemical Society transactions* 33: 447-449.

Carignani G, Groudinsky O, Frezza D, Schiavon E, Bergantino E, Slonimski PP. 1983. An mRNA maturase is encoded by the first intron of the mitochondrial gene for the subunit I of cytochrome oxidase in S. cerevisiae. *Cell* 35: 733-742.

Castruita M, Casero D, Karpowicz SJ, Kropat J, Vieler A, Hsieh SI, Yan W, Cokus S, Loo JA, Benning C et al. 2011. Systems biology approach in Chlamydomonas reveals connections between copper nutrition and multiple metabolic steps. *Plant Cell* 23: 1273-1292.

Cech TR. 1986. The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell* 44: 207-210.

Chin K, Pyle AM. 1995. Branch-point attack in group II introns is a highly reversible transesterification, providing a potential proofreading mechanism for 5'-splice site selection. *RNA* 1: 391-406.

Choquet Y, Stern DB, Wostrikoff K, Kuras R, Girard-Bascou J, Wollman FA. 1998. Translation of cytochrome f is autoregulated through the 5' untranslated region of petA mRNA in Chlamydomonas chloroplasts. *Proc Natl Acad Sci U S A* 95: 4380-4385.

Cohen S, Zmudjak M, Colas des Francs-Small C, Malik S, Shaya F, Keren I, Belausov E, Many Y, Brown GG, Small I et al. 2014. nMAT4, a maturase factor required for nad1 pre-mRNA processing and maturation, is essential for holocomplex I biogenesis in Arabidopsis mitochondria. *Plant J* 78: 253-268.

Cousineau B, Smith D, Lawrence-Cavanagh S, Mueller JE, Yang J, Mills D, Manias D, Dunny G, Lambowitz AM, Belfort M. 1998. Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell* 94: 451-462.

Cui X, Matsuura M, Wang Q, Ma H, Lambowitz AM. 2004. A group II intron-encoded maturase functions preferentially in cis and requires both the reverse transcriptase and X domains to promote RNA splicing. *J Mol Biol* 340: 211-231.

D'Andrea LD, Regan L. 2003. TPR proteins: the versatile helix. *Trends in biochemical sciences* 28: 655-662.

Danon A, Mayfield SP. 1991. Light regulated translational activators: identification of chloroplast gene specific mRNA binding proteins. *EMBO J* 10: 3993-4001.

de Longevialle AF, Small ID, Lurin C. 2010. Nuclearly encoded splicing factors implicated in RNA splicing in higher plant organelles. *Mol Plant* 3: 691-705.

Delannoy E, Fujii S, Colas des Francs-Small C, Brundrett M, Small I. 2011. Rampant gene loss in the underground orchid Rhizanthella gardneri highlights evolutionary constraints on plastid genomes. *Mol Biol Evol* 28: 2077-2086.

Deshpande NN, Bao Y, Herrin DL. 1997. Evidence for light/redox-regulated splicing of psbA pre-RNAs in Chlamydomonas chloroplasts. *RNA* 3: 37-48.

Dib-Hajj SD, Boulanger SC, Hebbar SK, Peebles CL, Franzen JS, Perlman PS. 1993. Domain 5 interacts with domain 6 and influences the second transesterification reaction of group II intron self-splicing. *Nucleic Acids Res* 21: 1797-1804.

Dlakic M, Mushegian A. 2011. Prp8, the pivotal protein of the spliceosomal catalytic center, evolved from a retroelement-encoded reverse transcriptase. *Rna* 17: 799-808.

Drager RG, Girard-Bascou J, Choquet Y, Kindle KL, Stern DB. 1998. In vivo evidence for 5'-->3' exoribonuclease degradation of an unstable chloroplast mRNA. *Plant J* 13: 85-96.

Draper DE. 1999. Themes in RNA-protein recognition. *J Mol Biol* 293: 255-270.

Drapier D, Girard-Bascou J, Wollman FA. 1992. Evidence for Nuclear Control of the Expression of the atpA and atpB Chloroplast Genes in Chlamydomonas. *Plant Cell* 4: 283-295.

Drescher A. 2003. ycf1, ycf14 und RNA-Edierung: Untersuchungen an im Lauf der Plastidenevolution neu hinzu gewonnenen Genen und Eigenschaften. LMU München, München.

Duarte FV, Palmeira CM, Rolo AP. 2014. The Role of microRNAs in Mitochondria: Small Players Acting Wide. *Genes* 5: 865-886.

Enyeart PJ, Mohr G, Ellington AD, Lambowitz AM. 2014. Biotechnological applications of mobile group II introns and their reverse transcriptases: gene targeting, RNA-seq, and non-coding RNA analysis. *Mob DNA* 5: 2.

Felder S, Meierhoff K, Sane AP, Meurer J, Driemel C, Plucken H, Klaff P, Stein B, Bechtold N, Westhoff P. 2001. The nucleus-encoded HCF107 gene of Arabidopsis provides a link between intercistronic RNA processing and the accumulation of translation-competent psbH transcripts in chloroplasts. *Plant Cell* 13: 2127-2141.

Fica SM, Tuttle N, Novak T, Li NS, Lu J, Koodathingal P, Dai Q, Staley JP, Piccirilli JA. 2013. RNA catalyses nuclear pre-mRNA splicing. *Nature* 503: 229-234.

Finster S, Eggert E, Zoschke R, Weihe A, Schmitz-Linneweber C. 2013. Light-dependent, plastome-wide association of the plastid-encoded RNA polymerase with chloroplast DNA. *Plant J* 76: 849-860.

Fisk DG, Walker MB, Barkan A. 1999. Molecular cloning of the maize gene crp1 reveals similarity between regulators of mitochondrial and chloroplast gene expression. *EMBO J* 18: 2621-2630.

Freeberg MA, Han T, Moresco JJ, Kong A, Yang YC, Lu ZJ, Yates JR, Kim JK. 2013. Pervasive and dynamic protein binding sites of the mRNA transcriptome in Saccharomyces cerevisiae. *Genome biology* 14: R13.

Fukunaga T, Ozaki H, Terai G, Asai K, Iwasaki W, Kiryu H. 2014. CapR: revealing structural specificities of RNA-binding protein target recognition using CLIP-seq data. *Genome biology* 15: R16.

Funk HT, Berg S, Krupinska K, Maier UG, Krause K. 2007. Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, Cuscuta reflexa and Cuscuta gronovii. *BMC plant biology* 7: 45.

Galej WP, Oubridge C, Newman AJ, Nagai K. 2013. Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature* 493: 638-643.

Garcia-Andrade J, Ramirez V, Lopez A, Vera P. 2013. Mediated plastid RNA editing in plant immunity. *PLoS pathogens* 9: e1003713.

Geary C, Rothemund PW, Andersen ES. 2014. RNA nanostructures. A single-stranded architecture for cotranscriptional folding of RNA nanostructures. *Science* 345: 799-804.

Gordon PM, Piccirilli JA. 2001. Metal ion coordination by the AGC triad in domain 5 contributes to group II intron catalysis. *Nat Struct Biol* 8: 893-898.

Grabowski PJ, Seiler SR, Sharp PA. 1985. A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell* 42: 345-353.

Hackenberg M, Huang PJ, Huang CY, Shi BJ, Gustafson P, Langridge P. 2013. A comprehensive expression profile of microRNAs and other classes of non-coding small RNAs in barley under phosphorous-deficient and -sufficient conditions. *DNA Res* 20: 109-125.

Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jungkamp AC, Munschauer M et al. 2010. PAR-CliP--a method to identify transcriptome-wide the binding sites of RNA binding proteins. *Journal of visualized experiments : JoVE*.

Hafner M, Renwick N, Brown M, Mihailovic A, Holoch D, Lin C, Pena JT, Nusbaum JD, Morozov P, Ludwig J et al. 2011. RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA* 17: 1697-1712.

Hall KB. 2002. RNA-protein interactions. *Curr Opin Struct Biol* 12: 283-288.

Hammani K, Barkan A. 2014. An mTERF domain protein functions in group II intron splicing in maize chloroplasts. *Nucleic Acids Res*.

Hammani K, Cook WB, Barkan A. 2012. RNA binding and RNA remodeling activities of the half-a-tetratricopeptide (HAT) protein HCF107 underlie its effects on gene expression. *Proc Natl Acad Sci U S A* 109: 5651-5656.

Hanahan D. 1983. Studies on transformation of Escherichia coli with plasmids. *J Mol Biol* 166: 557-580.

Hattori M, Miyake H, Sugita M. 2007. A Pentatricopeptide repeat protein is required for RNA processing of clpP Pre-mRNA in moss chloroplasts. *J Biol Chem* 282: 10773-10782.

Hausner G, Olson R, Simon D, Johnson I, Sanders ER, Karol KG, McCourt RM, Zimmerly S. 2006. Origin and evolution of the chloroplast trnK (matK) intron: a model for evolution of group II intron RNA structures. *Mol Biol Evol* 23: 380-391.

Hellman LM, Fried MG. 2007. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat Protoc* 2: 1849-1861.

Hertel S, Zoschke R, Neumann L, Qu Y, Axmann IM, Schmitz-Linneweber C. 2013. Multiple checkpoints for the expression of the chloroplast-encoded splicing factor MatK. *Plant Physiol* 163: 1686-1698.

Hess WR, Hoch B, Zeltz P, Hubschmann T, Kossel H, Borner T. 1994. Inefficient rpl2 splicing in barley mutants with ribosome-deficient plastids. *Plant Cell* 6: 1455-1465.

Hetzer M, Wurzer G, Schweyen RJ, Mueller MW. 1997. Trans-activation of group II intron splicing by nuclear U5 snRNA. *Nature* 386: 417-420.

Hirose T, Sugiura M. 2004. Multiple elements required for translation of plastid atpB mRNA lacking the Shine-Dalgarno sequence. *Nucleic Acids Res* 32: 3503-3510.

Hogan DJ, Riordan DP, Gerber AP, Herschlag D, Brown PO. 2008. Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS biology* 6: e255.

Horsch RB, Fry, J. E., Hoffman, N. L., Eichholtz, D., Rogers, S. G. 1985. A simple and general method for transferring genes into plants. *Science* 227: 1229-1231.

Huang HR, Chao MY, Armstrong B, Wang Y, Lambowitz AM, Perlman PS. 2003. The DIVa maturase binding site in the yeast group II intron aI2 is essential for intron homing but not for in vivo splicing. *Mol Cell Biol* 23: 8809-8819.

Hubschmann T, Hess WR, Borner T. 1996. Impaired splicing of the rps12 transcript in ribosome-deficient plastids. *Plant Mol Biol* 30: 109-123.

Huttenhofer A, Schattner P, Polacek N. 2005. Non-coding RNAs: hope or hype? *Trends in genetics : TIG* 21: 289-297.

Ibrahim F, Rymarquis LA, Kim EJ, Becker J, Balassa E, Green PJ, Cerutti H. 2010. Uridylation of mature miRNAs and siRNAs by the MUT68 nucleotidyltransferase promotes their degradation in Chlamydomonas. *Proc Natl Acad Sci U S A* 107: 3906-3911.

Iwai M, Katayama M, Ikeuchi M. 2006. Absence of the psbH gene product destabilizes the Photosystem II complex and prevents association of the Photosystem II-X protein in the thermophilic cyanobacterium Thermosynechococcus elongatus BP-1. *Photosynthesis research* 87: 313-322.

Jarrell KA, Dietrich RC, Perlman PS. 1988. Group II intron domain 5 facilitates a trans-splicing reaction. *Mol Cell Biol* 8: 2361-2366.

Karcher D, Bock R. 2002. Temperature sensitivity of RNA editing and intron splicing reactions in the plastid ndhB transcript. *Curr Genet* 41: 48-52.

Kato K, Matsumoto T.; Koiwai, A.; Mizusaki, S.; Nishida, K.; Noguchi, M.; Tamaki, E.;. 1972. Liquid suspension culture of tobacco cells. In: Terui G, ed. in *Fermentation technology today*, pp. 689–695. Osaka: Society of Fermentation Technology.

Ke J, Chen RZ, Ban T, Zhou XE, Gu X, Tan MH, Chen C, Kang Y, Brunzelle JS, Zhu JK et al. 2013. Structural basis for RNA recognition by a dimeric PPR-protein complex. *Nature structural & molecular biology* 20: 1377-1382.

Keating KS, Toor N, Perlman PS, Pyle AM. 2010. A structural analysis of the group II intron active site and implications for the spliceosome. *RNA* 16: 1-9.

Keene JD, Komisarow JM, Friedersdorf MB. 2006. RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nature protocols* 1: 302-307.

Keene JD, Tenenbaum SA. 2002. Eukaryotic mRNPs may represent posttranscriptional operons. *Mol Cell* 9: 1161-1167.

Keren I, Bezawork-Geleta A, Kolton M, Maayan I, Belausov E, Levy M, Mett A, Gidoni D, Shaya F, Ostersetzer-Biran O. 2009. AtnMat2, a nuclear-encoded maturase required for splicing of group-II introns in Arabidopsis mitochondria. *RNA* 15: 2299-2311.

Keren I, Tal L, des Francs-Small CC, Araujo WL, Shevtsov S, Shaya F, Fernie AR, Small I, Ostersetzer-Biran O. 2012. nMAT1, a nuclear-encoded maturase involved in the trans-splicing of nad1 intron 1, is essential for mitochondrial complex I assembly and function. *Plant J* 71: 413-426.

Kim VN, Han J, Siomi MC. 2009. Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol* 10: 126-139.

Koch JL, Boulanger SC, Dib-Hajj SD, Hebbar SK, Perlman PS. 1992. Group II introns deleted for multiple substructures retain self-splicing activity. *Mol Cell Biol* 12: 1950-1958.

Koehler F. 2013. Functional Analysis of Maturase K - A Chloroplast RNA Splicing Factor. Humboldt university of Berlin.

Komenda J, Lupinkova L, Kopecky J. 2002. Absence of the psbH gene product destabilizes photosystem II complex and bicarbonate binding on its acceptor side in Synechocystis PCC 6803. *Eur J Biochem* 269: 610-619.

Konforti BB, Liu Q, Pyle AM. 1998. A map of the binding site for catalytic domain 5 in the core of a group II intron ribozyme. *Embo J* 17: 7105-7117.

Konig J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J. 2011. iCLIP--transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. *Journal of visualized experiments : JoVE*.

Kozak M. 1999. Initiation of translation in prokaryotes and eukaryotes. *Gene* 234: 187-208.

Kwakman JH, Konings D, Pel HJ, Grivell LA. 1989. Structure-function relationships in a self-splicing group II intron: a large part of domain II of the mitochondrial intron aI5 is not essential for self-splicing. *Nucleic Acids Res* 17: 4205-4216.

Lai D, Proctor JR, Meyer IM. 2013. On the importance of cotranscriptional RNA structure formation. *RNA* 19: 1461-1473.

Lambowitz AM, Belfort M. 1993. Introns as mobile genetic elements. *Annual review of biochemistry* 62: 587-622.

Lee SH, Chung BH, Park TG, Nam YS, Mok H. 2012. Small-interfering RNA (siRNA)-based functional micro- and nanostructures for efficient and selective gene silencing. *Accounts of chemical research* 45: 1014-1025.

Li X, Zhang W, Xu T, Ramsey J, Zhang L, Hill R, Hansen KC, Hesselberth JR, Zhao R. 2013. Comprehensive in vivo RNA-binding site analyses reveal a role of Prp8 in spliceosomal assembly. *Nucleic Acids Res* 41: 3805-3818.

Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X et al. 2008. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 456: 464-469.

Liere K, Link G. 1995. RNA-binding activity of the matK protein encoded by the chloroplast trnK intron from mustard (Sinapis alba L.). *Nucleic Acids Res* 23: 917-921.

Lin SL, Miller JD, Ying SY. 2006. Intronic microRNA (miRNA). *Journal of biomedicine & biotechnology*: 26818.

Linsen SE, de Wit E, Janssens G, Heater S, Chapman L, Parkin RK, Fritz B, Wyman SK, de Bruijn E, Voest EE et al. 2009. Limitations and possibilities of small RNA digital gene expression profiling. *Nature methods* 6: 474-476.

Lisitsky I, Schuster G. 1995. Phosphorylation of a chloroplast RNA-binding protein changes its affinity to RNA. *Nucleic Acids Res* 23: 2506-2511.

Loiselay C, Gumpel NJ, Girard-Bascou J, Watson AT, Purton S, Wollman FA, Choquet Y. 2008. Molecular identification and function of cis- and trans-acting determinants for petA transcript stability in Chlamydomonas reinhardtii chloroplasts. *Mol Cell Biol* 28: 5529-5542.

Lu C, Meyers BC, Green PJ. 2007. Construction of small RNA cDNA libraries for deep sequencing. *Methods* 43: 110-117.

Lung B, Zemann A, Madej MJ, Schuelke M, Techritz S, Ruf S, Bock R, Huttenhofer A. 2006. Identification of small non-coding RNAs from mitochondria and chloroplasts. *Nucleic Acids Res* 34: 3842-3852.

Malone C, Brennecke J, Czech B, Aravin A, Hannon GJ. 2012. Preparation of small RNA libraries for high-throughput sequencing. *Cold Spring Harbor protocols* 2012: 1067-1077.

Marcia M, Pyle AM. 2012. Visualizing group II intron catalysis through the stages of splicing. *Cell* 151: 497-507.

Marcia M, Pyle AM. 2014. Principles of ion recognition in RNA: insights from the group II intron structures. *RNA* 20: 516-527.

Marcia M, Somarowthu S, Pyle AM. 2013. Now on display: a gallery of group II intron structures at different stages of catalysis. *Mob DNA* 4: 14.

Maris C, Dominguez C, Allain FH. 2005. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *Febs J* 272: 2118-2131.

Matera AG, Wang Z. 2014. A day in the life of the spliceosome. *Nat Rev Mol Cell Biol* 15: 108-121.

Matsuura M, Noah JW, Lambowitz AM. 2001. Mechanism of maturase-promoted group II intron splicing. *Embo J* 20: 7259-7270.

McCullough AJ, Kangasjarvi J, Gengenbach BG, Jones RJ. 1992. Plastid DNA in developing maize endosperm : genome structure, methylation, and transcript accumulation patterns. *Plant Physiol* 100: 958-964.

McNeal JR, Arumugunathan K, Kuehl JV, Boore JL, Depamphilis CW. 2007. Systematics and plastid genome evolution of the cryptically photosynthetic parasitic plant genus Cuscuta (Convolvulaceae). *BMC biology* 5: 55.

McNeal JR, Kuehl JV, Boore JL, Leebens-Mack J, dePamphilis CW. 2009. Parallel loss of plastid introns and their maturase in the genus Cuscuta. *PLoS One* 4: e5982.

Meierhoff K, Felder S, Nakamura T, Bechtold N, Schuster G. 2003. HCF152, an Arabidopsis RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast psbB-psbT-psbH-petB-petD RNAs. *Plant Cell* 15: 1480-1495.

Michel F, Dujon B. 1983. Conservation of RNA secondary structures in two intron families including mitochondrial-, chloroplast- and nuclear-encoded members. *Embo J* 2: 33-38.

Michel F, Ferat JL. 1995. Structure and activities of group II introns. *Annual review of biochemistry* 64: 435-461.

Michel F, Jacquier A, Dujon B. 1982. Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. *Biochimie* 64: 867-881.

Michel F, Umesono K, Ozeki H. 1989. Comparative and functional anatomy of group II catalytic introns--a review. *Gene* 82: 5-30.

Michels WJ, Jr., Pyle AM. 1995. Conversion of a group II intron into a new multiple-turnover ribozyme that selectively cleaves oligonucleotides: elucidation of reaction mechanism and structure/function relationships. *Biochemistry* 34: 2965-2977.

Mili S, Steitz JA. 2004. Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA* 10: 1692-1694.

Mills DA, McKay LL, Dunny GM. 1996. Splicing of a group II intron involved in the conjugative transfer of pRS01 in lactococci. *J Bacteriol* 178: 3531-3538.

Mohr G, Perlman PS, Lambowitz AM. 1993. Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. *Nucleic Acids Res* 21: 4991-4997.

Monod C, Goldschmidt-Clermont M, Rochaix JD. 1992. Accumulation of chloroplast psbB RNA requires a nuclear factor in Chlamydomonas reinhardtii. *Molecular & general genetics : MGG* 231: 449-459.

Moran JV, Mecklenburg KL, Sass P, Belcher SM, Mahnke D, Lewin A, Perlman P. 1994. Splicing defective mutants of the COXI gene of yeast mitochondrial DNA: initial definition of the maturase domain of the group II intron aI2. *Nucleic Acids Res* 22: 2057-2064.

Mougin A, Gottschalk A, Fabrizio P, Luhrmann R, Branlant C. 2002. Direct probing of RNA structure and RNA-protein interactions in purified HeLa cell's and yeast spliceosomal U4/U6.U5 tri-snRNP particles. *J Mol Biol* 317: 631-649.

Murakami S, Kuehnle K, Stern DB. 2005. A spontaneous tRNA suppressor of a mutation in the Chlamydomonas reinhardtii nuclear MCD1 gene required for stability of the chloroplast petD mRNA. *Nucleic Acids Res* 33: 3372-3380.

Nagy V, Pirakitikulr N, Zhou KI, Chillon I, Luo J, Pyle AM. 2013. Predicted group II intron lineages E and F comprise catalytically active ribozymes. *RNA* 19: 1266-1278.

Neuhaus H, Link G. 1987. The chloroplast tRNALys(UUU) gene from mustard (Sinapis alba) contains a class II intron potentially coding for a maturase-related polypeptide. *Curr Genet* 11: 251-257.

Neumann L. 2011. Functional analyses of the chloroplast maturase MatK. *master thesis*.

Nickelsen J. 2003. Chloroplast RNA-binding proteins. *Curr Genet* 43: 392-399.

Nickelsen J, Fleischmann M, Boudreau E, Rahire M, Rochaix JD. 1999. Identification of cis-acting RNA leader elements required for chloroplast psbD gene expression in Chlamydomonas. *Plant Cell* 11: 957-970.

Ogawa N, Biggin MD. 2012. High-throughput SELEX determination of DNA sequences bound by transcription factors in vitro. *Methods in molecular biology* 786: 51-63.

Ossenbuhl F, Hartmann K, Nickelsen J. 2002. A chloroplast RNA binding protein from stromal thylakoid membranes specifically binds to the 5' untranslated region of the psbA mRNA. *Eur J Biochem* 269: 3912-3919.

Pall GS, Hamilton AJ. 2008. Improved northern blot method for enhanced detection of small RNA. *Nature protocols* 3: 1077-1084.

Peebles CL, Perlman PS, Mecklenburg KL, Petrillo ML, Tabor JH, Jarrell KA, Cheng HL. 1986. A self-splicing RNA excises an intron lariat. *Cell* 44: 213-223.

Peebles CL, Zhang M, Perlman PS, Franzen JS. 1995. Catalytically critical nucleotide in domain 5 of a group II intron. *Proc Natl Acad Sci U S A* 92: 4422-4426.

Pfalz J, Bayraktar OA, Prikryl J, Barkan A. 2009. Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts. *Embo J* 28: 2042-2052.

Pikielny CW, Rosbash M. 1986. Specific small nuclear RNAs are associated with yeast spliceosomes. *Cell* 45: 869-877.

Podar M, Dib-Hajj S, Perlman PS. 1995. A UV-induced, Mg(2+)-dependent crosslink traps an active form of domain 3 of a self-splicing group II intron. *RNA* 1: 828-840.

Podar M, Perlman PS. 1999. Photocrosslinking of 4-thio uracil-containing RNAs supports a side-by-side arrangement of domains 5 and 6 of a group II intron. *Rna* 5: 318-329.

Prikryl J, Rojas M, Schuster G, Barkan A. 2011. Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proc Natl Acad Sci U S A* 108: 415-420.

Qin PZ, Pyle AM. 1997. Stopped-flow fluorescence spectroscopy of a group II intron ribozyme reveals that domain 1 is an independent folding unit with a requirement for specific Mg2+ ions in the tertiary structure. *Biochemistry* 36: 4718-4730.

Qin PZ, Pyle AM. 1998. The architectural organization and mechanistic function of group II intron structural elements. *Curr Opin Struct Biol* 8: 301-308.

Rahire M, Laroche F, Cerutti L, Rochaix JD. 2012. Identification of an OPR protein involved in the translation initiation of the PsaB subunit of photosystem I. *Plant J* 72: 652-661.

Rambo RP, Doudna JA. 2004. Assembly of an active group II intron-maturase complex by protein dimerization. *Biochemistry* 43: 6486-6497.

Ramundo S, Rahire M, Schaad O, Rochaix JD. 2013. Repression of essential chloroplast genes reveals new signaling pathways and regulatory feedback loops in chlamydomonas. *Plant Cell* 25: 167-186.

Ranzini G. 2014. Characterization of overexpression lines of the chloroplast maturase MatK. in *Biology*. HU Berlin.

Rousseau-Gueutin M, Ayliffe MA, Timmis JN. 2012. Plastid DNA in the nucleus: new genes for old. *Plant signaling & behavior* 7: 269-272.

Ruhlman T, Verma D, Samson N, Daniell H. 2010. The role of heterologous chloroplast sequence elements in transgene integration and expression. *Plant Physiol* 152: 2088-2104.

Ruwe H, Schmitz-Linneweber C. 2012. Short non-coding RNA fragments accumulating in chloroplasts: footprints of RNA binding proteins? *Nucleic Acids Res* 40: 3106-3116.

Saldanha R, Chen B, Wank H, Matsuura M, Edwards J, Lambowitz AM. 1999. RNA and protein catalysis in group II intron splicing and mobility reactions using purified components. *Biochemistry* 38: 9069-9083.

Sane AP, Stein B, Westhoff P. 2005. The nuclear gene HCF107 encodes a membrane-associated R-TPR (RNA tetratricopeptide repeat)-containing protein involved in expression of the plastidial psbH gene in Arabidopsis. *Plant J* 42: 720-730.

Schmitz-Linneweber C, Small I. 2008. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends Plant Sci* 13: 663-670.

SenGupta DJ, Zhang B, Kraemer B, Pochart P, Fields S, Wickens M. 1996. A three-hybrid system to detect RNA-protein interactions in vivo. *Proc Natl Acad Sci U S A* 93: 8496-8501.

Sharp PA. 1985. On the origin of RNA splicing and introns. *Cell* 42: 397-400.

Sharp PA. 1987. Splicing of messenger RNA precursors. *Science* 235: 766-771.

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K et al. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *Embo J* 5: 2043-2049.

Shukla GC, Padgett RA. 2002. A catalytically active group II intron domain 5 can function in the U12-dependent spliceosome. *Mol Cell* 9: 1145-1150.

Singh R. 2002. RNA-protein interactions that regulate pre-mRNA splicing. *Gene Expr* 10: 79-92.

Singh RN, Saldanha RJ, D'Souza LM, Lambowitz AM. 2002. Binding of a group II intron-encoded reverse transcriptase/maturase to its high affinity intron RNA binding site involves sequence-specific recognition and autoregulates translation. *J Mol Biol* 318: 287-303.

Small ID, Peeters N. 2000. The PPR motif - a TPR-related motif prevalent in plant organellar proteins. *Trends in biochemical sciences* 25: 46-47.

Somanchi A, Barnes D, Mayfield SP. 2005. A nuclear gene of Chlamydomonas reinhardtii, Tba1, encodes a putative oxidoreductase required for translation of the chloroplast psbA mRNA. *Plant J* 42: 341-352.

Sontheimer EJ, Gordon PM, Piccirilli JA. 1999. Metal ion catalysis during group II intron self-splicing: parallels with the spliceosome. *Genes & development* 13: 1729-1741.

Sorefan K, Pais H, Hall AE, Kozomara A, Griffiths-Jones S, Moulton V, Dalmay T. 2012. Reducing ligation bias of small RNAs in libraries for next generation sequencing. *Silence* 3: 4.

Stampacchia O, Girard-Bascou J, Zanasco JL, Zerges W, Bennoun P, Rochaix JD. 1997. A nuclear-encoded function essential for translation of the chloroplast psaB mRNA in chlamydomonas. *Plant Cell* 9: 773-782.

Stewart CN, Jr., Via LE. 1993. A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. *BioTechniques* 14: 748-750.

Stoppel R, Meurer J. 2011. The cutting crew - ribonucleases are key players in the control of plastid gene expression. *J Exp Bot* 63: 1663-1673.

Sturm NR, Kuras R, Buschlen S, Sakamoto W, Kindle KL, Stern DB, Wollman FA. 1994. The petD gene is transcribed by functionally redundant promoters in Chlamydomonas reinhardtii chloroplasts. *Mol Cell Biol* 14: 6171-6179.

Su LJ, Waldsich C, Pyle AM. 2005. An obligate intermediate along the slow folding pathway of a group II intron ribozyme. *Nucleic Acids Res* 33: 6674-6687.

Sugiura M, Hirose T, Sugita M. 1998. Evolution and mechanism of translation in chloroplasts. *Annual review of genetics* 32: 437-459.

Svab Z, Hajdukiewicz P, Maliga P. 1990. Stable transformation of plastids in higher plants. *Proc Natl Acad Sci U S A* 87: 8526-8530.

Svab Z, Maliga P. 1993. High-frequency plastid transformation in tobacco by selection for a chimeric aadA gene. *Proc Natl Acad Sci U S A* 90: 913-917.

Swisher JF, Su LJ, Brenowitz M, Anderson VE, Pyle AM. 2002. Productive folding to the native state by a group II intron ribozyme. *J Mol Biol* 315: 297-310.

Tam PP, Barrette-Ng IH, Simon DM, Tam MW, Ang AL, Muench DG. 2010. The Puf family of RNA-binding proteins in plants: phylogeny, structural modeling, activity and subcellular localization. *BMC plant biology* 10: 44.

Tanaka Y, Nishiyama Y, Murata N. 2000. Acclimation of the photosynthetic machinery to high temperature in Chlamydomonas reinhardtii requires synthesis de novo of proteins encoded by the nuclear and chloroplast genomes. *Plant Physiol* 124: 441-449.

Toor N, Hausner G, Zimmerly S. 2001. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *Rna* 7: 1142-1152.

Toor N, Keating KS, Taylor SD, Pyle AM. 2008. Crystal structure of a self-spliced group II intron. *Science* 320: 77-82.

Turmel M, Otis C, Lemieux C. 2005. The complete chloroplast DNA sequences of the charophycean green algae Staurastrum and Zygnema reveal that the chloroplast genome underwent extensive changes during the evolution of the Zygnematales. *BMC biology* 3: 22.

Turmel M, Otis C, Lemieux C. 2006. The chloroplast genome sequence of Chara vulgaris sheds new light into the closest green algal relatives of land plants. *Mol Biol Evol* 23: 1324-1338.

Urzica EI, Casero D, Yamasaki H, Hsieh SI, Adler LN, Karpowicz SJ, Blaby-Haas CE, Clarke SG, Loo JA, Pellegrini M et al. 2012. Systems and trans-system level analysis identifies conserved iron deficiency responses in the plant lineage. *Plant Cell* 24: 3921-3948.

Vaistij FE, Goldschmidt-Clermont M, Wostrikoff K, Rochaix JD. 2000. Stability determinants in the chloroplast psbB/T/H mRNAs of Chlamydomonas reinhardtii. *Plant J* 21: 469-482.

van der Veen R, Arnberg AC, van der Horst G, Bonen L, Tabak HF, Grivell LA. 1986. Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro. *Cell* 44: 225-234.

Verhounig A, Karcher D, Bock R. 2010. Inducible gene expression from the plastid genome by a synthetic riboswitch. *Proc Natl Acad Sci U S A* 107: 6204-6209.

Voelker R, Barkan A. 1995. Two nuclear mutations disrupt distinct pathways for targeting proteins to the chloroplast thylakoid. *EMBO J* 14: 3905-3914.

Vogel J, Borner T. 2002. Lariat formation and a hydrolytic pathway in plant chloroplast group II intron splicing. *Embo J* 21: 3794-3803.

Vogel J, Borner T, Hess WR. 1999. Comparative analysis of splicing of the complete set of chloroplast group II introns in three higher plant mutants. *Nucleic Acids Res* 27: 3866-3874.

Vogel J, Hubschmann T, Borner T, Hess WR. 1997. Splicing and intron-internal RNA editing of trnK-matK transcripts in barley plastids: support for MatK as an essential splice factor. *J Mol Biol* 270: 179-187.

Wang L, Yu X, Wang H, Lu YZ, de Ruiter M, Prins M, He YK. 2011. A novel class of heat-responsive small RNAs derived from the chloroplast genome of Chinese cabbage (Brassica rapa). *BMC genomics* 12: 289.

Wank H, SanFilippo J, Singh RN, Matsuura M, Lambowitz AM. 1999. A reverse transcriptase/maturase promotes splicing by binding at its own coding segment in a group II intron RNA. *Mol Cell* 4: 239-250.

Watanabe K, Lambowitz AM. 2004. High-affinity binding site for a group II intron-encoded reverse transcriptase/maturase within a stem-loop structure in the intron RNA. *Rna* 10: 1433-1443.

Wolfe KH, Morden CW, Palmer JD. 1992. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proc Natl Acad Sci U S A* 89: 10648-10652.

Wostrikoff K, Stern D. 2007. Rubisco large-subunit translation is autoregulated in response to its assembly state in tobacco chloroplasts. *Proc Natl Acad Sci U S A* 104: 6466-6471.

Wu L, Zhang D, Xue M, Qian J, He Y, Wang S. 2014. Overexpression of the maize GRF10, an endogenous truncated growth-regulating factor protein, leads to reduction in leaf size and plant height. *Journal of integrative plant biology*.

Yang J, Mohr G, Perlman PS, Lambowitz AM. 1998. Group II intron mobility in yeast mitochondria: target DNA-primed reverse transcription activity of aI1 and reverse splicing into DNA transposition sites in vitro. *J Mol Biol* 282: 505-523.

Yean SL, Wuenschell G, Termini J, Lin RJ. 2000. Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature* 408: 881-884.

Yeap WC, Namasivayam P, Ho CL. 2014. HnRNP-like proteins as post-transcriptional regulators. *Plant science : an international journal of experimental plant biology* 227C: 90-100.

Yoshida T, Furihata HY, Kawabe A. 2014. Patterns of genomic integration of nuclear chloroplast DNA fragments in plant species. *DNA Res* 21: 127-140.

Yukawa M, Kuroda H, Sugiura M. 2007. A new in vitro translation system for non-radioactive assay from tobacco chloroplasts: effect of pre-mRNA processing on translation in vitro. *Plant J* 49: 367-376.

Zemora G, Waldsich C. 2010. RNA folding in living cells. *RNA Biol* 7: 634-641.

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

Zerges W, Auchincloss AH, Rochaix JD. 2003. Multiple translational control sequences in the 5' leader of the chloroplast psbC mRNA interact with nuclear gene products in Chlamydomonas reinhardtii. *Genetics* 163: 895-904.

Zhang Z, Lee JE, Riemondy K, Anderson EM, Yi R. 2013. High-efficiency RNA cloning enables accurate quantification of miRNA expression by deep sequencing. *Genome biology* 14: R109.

Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, Song JJ, Kingston RE, Borowsky M, Lee JT. 2010. Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell* 40: 939-953.

Zhelyazkova P, Hammani K, Rojas M, Voelker R, Vargas-Suarez M, Borner T, Barkan A. 2012a. Protein-mediated protection as the predominant mechanism for defining

processed mRNA termini in land plant chloroplasts. *Nucleic Acids Res* 40: 3092-3105.

Zhelyazkova P, Sharma CM, Forstner KU, Liere K, Vogel J, Borner T. 2012b. The primary transcriptome of barley chloroplasts: numerous noncoding RNAs and the dominating role of the plastid-encoded RNA polymerase. *Plant Cell* 24: 123-136.

Zoschke R, Nakamura M, Liere K, Sugiura M, Borner T, Schmitz-Linneweber C. 2010. An organellar maturase associates with multiple group II introns. *Proc Natl Acad Sci U S A* 107: 3245-3250.

Zoschke R, Ostersetzer O, Börner T, Schmitz-Linneweber C. 2009. Analysis of the regulation of MatK gene expression. *Endocytobiosis Cell Res* 19 127-135.

# 6. Appendices

## Abbreviations

| | |
|---|---|
| % | percentage |
| µl | macroliter |
| A, T, G, C, U nucleic acid bases | adenosine, thymine, guanine, cytosine, uracil |
| bp | basepair |
| BSA | Bovine serum albumin |
| CR | *Chlamydomonas reinhardtii* |
| CRM domain | Chloroplast RNA splicing and ribosome maturation domain |
| DNA | deoxyribonucleic acid |
| DTT | Dithiothreitol |
| E.coli | *Escherichia coli* |
| EDC | 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide |
| EDTA | ethylenediaminetetraacetic acid |
| EGTA | ethyleneglycoltetraacetic acid |
| En domain | endonuclease domain |
| HA | hemagglutinin epitope |
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| HRP | horseradish peroxidase |
| HV | *Hordeum vulgare* |
| IEP | intron encoded protein |
| IP | immunoprecipitation |
| kb | kilobase |
| kDa | kilodalton |
| LB medium | Lysogeny broth medium |
| LMW | low molecular weight |
| M | molar |
| min | minute |
| mJ | millijoule |
| ml | mililiter |
| mM | milimolar |
| MOPS | 3-(N-morpholino)propanesulfonic acid |
| MS medium | Murashige and Skoog medium |
| mtDNA | mitochondrial DNA |
| mTERF | mitochondrial transcription termination factor |
| NEP | nuclear-encoded plastid RNA polymerase |
| ng | nanogram |
| NMD | nonsense-mediated mRNA decay |

| | |
|---|---|
| nt | nucleotide |
| NT | *Nicotiana tabacum* |
| $^{o}C$ | degree Celsius |
| OE | over expression |
| OPR | octatricopeptide repeat |
| ORF | open reading frame |
| PAGE | Polyacrylamide gel electrophoresis |
| PAR-Clip | Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation |
| PCR | polymerase chain reaction |
| PEP | plastid-encoded plastid RNA polymerase |
| pH | potential hydrogen, -log [H+] |
| pM | picomolar |
| PMSF | phenylmethylsulfonyl fluoride |
| PORR | Plant Organelle RNA Recognition |
| PPR | pentatricopeptide repeat |
| qPCR | quantitative PCR |
| RACE | Rapid amplification of cDNA ends |
| RBP | RNA-binding proteins |
| REMSAs | RNA electromobility shift assays |
| RIP-Chip | RNA immunoprecipitation and Chip hybridization |
| RIP-seq | RNA immunoprecipitation and sequencing |
| RNA | ribonucleic acid |
| RNase | ribonuclease |
| RNA-seq | RNA sequencing; transcriptome sequencing |
| RNP | ribonucleoprotein particles |
| rpm | round per minute |
| rRNA | ribosomal RNA |
| RS | riboswitch |
| RT | reverse transcriptase |
| RT-PCR | reverse transcription-PCR |
| SD region | Shine-Dalgarno region |
| SDC | sodium deoxycholate |
| SDS | sodium dodecyl sulfate |
| SELEX | Systematic Evolution of Ligands by Exponential Enrichment |
| snRNP | small nuclear ribonucleoprotein particles |
| SOB-Medium | Super-Optimal-Broth-Medium |
| SOC | Super Optimal broth with Catabolite repression |
| sRNA | small RNA |
| SS | splicing site |
| SSC | saline-sodium citrate |
| ssDNA | single strand DNA |
| ssRNA | single strand RNA |
| TAE | Tris-Acetate-EDTA |

| | |
|---|---|
| TBST | Tris Buffered Saline with Tween |
| TPR | tetratricopeptide repeats |
| tRNA | transfer RNA |
| UTR | untranslated region |
| UV | ultraviolet |
| w/v | mass/volume |
| ZM | *Zea mays* |

**Appendix 1: Oligonucleotides used this study**

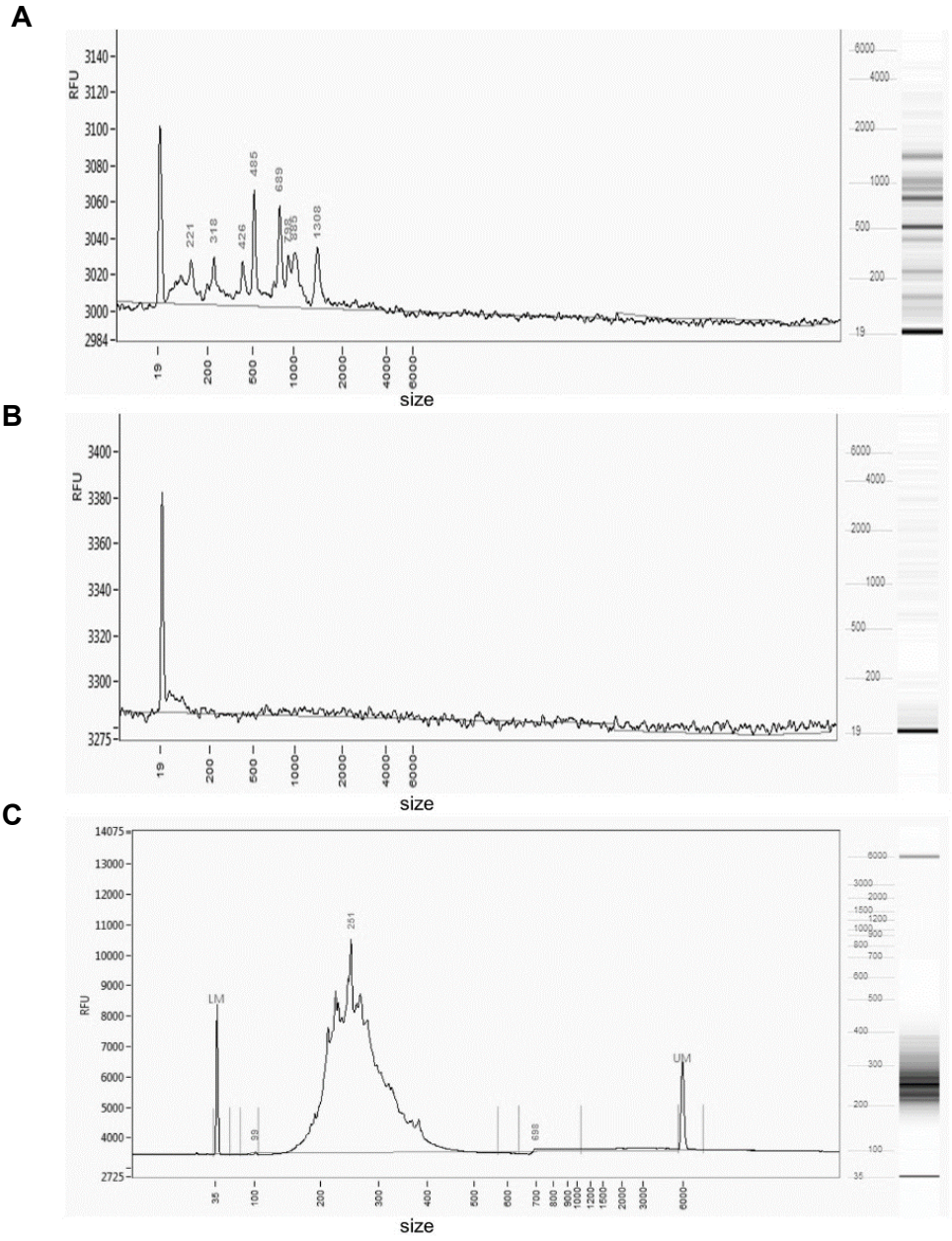
| Name | Sequence (5'->3') | orgamism | purpose |
|---|---|---|---|
| AtmatkBamHIr | GAGGATCCTTATTCATGATTGACCAAATCATTAAG | AT | *matK* expression |
| ATmatkseq1f | GTACTAATACCTTACCCCATC | AT | sequencing |
| ATmatkseq2f | GGAAGATGCATTCTGGCAAC | AT | sequencing |
| crpsbL3'race | TGGCTAGACCAAATCCAAATAAAC | CR | RACE |
| cratpA3'race | CCAATTAGCTCGTGGTGCTC | CR | RACE |
| crpetD5'race | CCCATAGCAGCTGGGTC | CR | RACE |
| crpetD3'race | CCAGAATGGTATTTCTACCCTG | CR | RACE |
| crpsbA5'race | GATGTCTACTGGCGGAGC | CR | RACE |
| crpsbA3'race | GGTAATCGGTATTTGGTTCACTG | CR | RACE |
| crpsbB5'race | CCAACCACCCCAAGATTGTG | CR | RACE |
| crpsbB3'race | GCTCGTAAAGCTCAATTAGGTG | CR | RACE |
| crpsbH5'race | CCTGCTTCTGAGTTAAGTGG | CR | RACE |
| crpsbH3'race | CAGGCTGGGGTACAACTG | CR | RACE |
| crpetB5'race | TGATGCGAAAGCTTCTGCTAC | CR | RACE |
| crpetB3'race | CATTACCATGGGACCAAGTTG | CR | RACE |
| cprrn75race | CGTATCGCCGGTATCTGC | CR | RACE |
| cr222Eoligo | AACCTTGTAATAATTAAGTAAAAAAATCAGTAAAAACCTGTCTC | CR | RNase protection |
| psbH5oligo | CCTTATTTTTTACAGAAAGTAAATAAAATAGCGCTCCTGTCTC | CR | RNase protection |
| psbB anti oligo | TACTGATTTTTTTACTTAATTATT | CR | sRNA Northern |
| psbH anti oligo | CTATTTTATTTACTTTCTGTAAA | CR | sRNA Northern |
| crpetD antioligo | AATATGGTTTAGCCGTTCCGAAAAGTTTTTT | CR | sRNA Northern |
| crpsbA antioligo | TTTTTAAAGTTTTAATTTCTCCGTAAAAT | CR | sRNA Northern |
| psbH3 oligo | TGCCACTGGCCTTCCGTTAAGAT | CR | sRNA Northern |
| psbBcds antioligo | ATTTACCGAATTCAACTTGGTCATTAAT | CR | sRNA Northern |
| psbH3 antioligo | TGCCACTGCCGAATATAAATATGGTT | CR | sRNA Northern |
| petD 5T7 | AAAAAACTTTTCGGAACGGCTAAACCATATCCTGTCTC | CR | sRNA Northern |
| petB 5 antioligo | ATTGTGACATGACCATTAGGCTTTC | CR | sRNA Northern |
| psbH3 antisl | ATCTTAACGGAAGGCCAGTGGCAGTTGGCGGTGCCACTGCC | CR | sRNA Northern |
| crpetDanti5end | ATGCTGTATTTCTAATGTTTACATGCTAAA | CR | sRNA Northern |
| petD51T7 | AAAAATTTTTAGCATGTAAACATTAGAAATACAGCATAACCTGTCTC | CR | *In vitro* transcription |
| petD52T7 | TGGCTTTATAAAATAAAAAACTTTTCGGAACGGCTAAACCATATTTATTCCTGTCTC | CR | *In vitro* transcription |
| crpsbBf | ATGGGTTTACCTTGGTATCGTG | CR | PCR |
| crpsbHf | ATGGCAACAGGAACTTCTAAAG | CR | PCR |
| crpsbHr | GCTAAAGTTTCCCAACTCATAG | CR | PCR |
| crpsaC1as | CATATATTTCTCTCCTCATAC | CR | oligo probe |
| crpsaC2as | CAAAAAGAAGATTGAGAATCGACTT | CR | oligo probe |

| | | | |
|---|---|---|---|
| crpsaC3as | TATGTCATCTCCATATCAAAAA | CR | oligo probe |
| crpsaC4as | TGTGCTAAATATGTCATCTCCATA | CR | oligo probe |
| Rumsh1 | TGATCCAACCGACGCGAC | Artificial | 5‘ RACE |
| AdapterRTPrimer | CAAGCAGAAGACGGCATA | Artificial | 3' RACE |
| AdapterPCRPrimer | CAAGCAGAAGACGGCATACG | Artificial | 3' RACE |
| PJET1.2fwd | CGACTCACTATAGGGAGAGCGGC | E.coli | cloning sequencing |
| PJET1.2rev | AAGAACATCGATTTTCCATGGCAG | E.coli | cloning sequencing |
| hisNdeIexf | CGTCATATGAAACATCACCATCACCATCAC | histag | *matK* expression |
| HvmatkBamHIr | GAGGATCCTTAATTAAGAGGGTTGACCAGG | HV | *matK* expression |
| HVmatkseq1f | GAAATCTTGGTTCAACTCCTTC | HV | sequencing |
| HVmatkseq2f | GGAAAGGCAATTCTTGCATC | HV | sequencing |
| atpF1 | CGAAACTCCCGGCAGATGGCCAGTGGCCCAAAGAAACGAAAGAATCGGTT | NT | dot blot |
| atpF2 | ATACACCAAGCACTACACTTAGATTTATTGGATTTGTTGCTAAAATATCG | NT | dot blot |
| matK sma1 for | GTAGCCCGGGATGGAAGAAATCCAAAG | NT | *matK* and HA cloning |
| HA rev | TCATTAAGCATAATCAGGTACATCGTATGGATATCC | NT | *matK* and HA cloning |
| Rbcs SP for | ATGGCTTCCTCTGTTCTTTCCTC | NT | Rbcs SP cloning |
| Rbcs SP rev | CTGCATGCATTGCACTCTTCC | NT | Rbcs SP cloning |
| matk smaI for2 | GTAGCCCGGGATGGAAGAAATCCAAAGATATTTACAGC | NT | cloning |
| matK ol rev1 | GATAATTCGCCAGATCATTGATACA | NT | cloning |
| matK ol rev2 | CCAAATCCGACTTCTATATACTCC | NT | cloning |
| MatK seq for1 | GAAGTTCGATACCCTTGTTCC | NT | sequencing |
| matK seq rev1 | GGATGCCCTAATACGGTAC | NT | sequencing |
| matk seq for2 | CGTCTTTCTACGGAACCAATC | NT | sequencing |
| matK seq rev2 | CGTTCAAGAAGGGCTCCA | NT | sequencing |
| MatK seq for3 | ACGATCAATTCATTCAACATTTCC | NT | sequencing |
| matK seq rev3 | CAGGGTAGGGTATTAGTATATC | NT | sequencing |
| MatK seq for4 | GCACTTGCTCATGATCATGG | NT | sequencing |
| trnK in1for | GTAGAAGCTTGCTGTAATACGACTCACTATAGGGTCTTCGGGTTGCTAACTCAACG | NT | autoregulation |
| trnK in1 ol | TCTTCGGGTTGCTAACTCAACGGTAGAGTACTCGGCTTTTAAGTGCGGC | NT | autoregulation |
| trnV 1 | GTGTAAACGAGTTGCTCTACCAACTGAGCTATAGCCCT | NT | dot blot |
| trnV 2 | AACTCCCTTGAAAAACATTGGCGCGCGTGTAAACGAGTTGCTCTACCAAC | NT | dot blot |
| trnV ex1f | TGGTAGAGCAACTCGTTTACACG | NT | qPCR |
| trnV inr | GGACCGAACTCTTTGTCAGG | NT | qPCR |
| trnV inf | AATGCATGTTGGGTCTTTGAA | NT | qPCR |
| trnV inr2 | CTGATTGTATGATGAACTCCCTTG | NT | qPCR |
| trnV ex2r | GGGCTATACGGACTCGAACC | NT | qPCR |
| trnA exf | TCTTGCAATTGGGTCGTTG | NT | qPCR |
| trnA inr | GCACGTTTCGGTCCTCTTC | NT | qPCR |

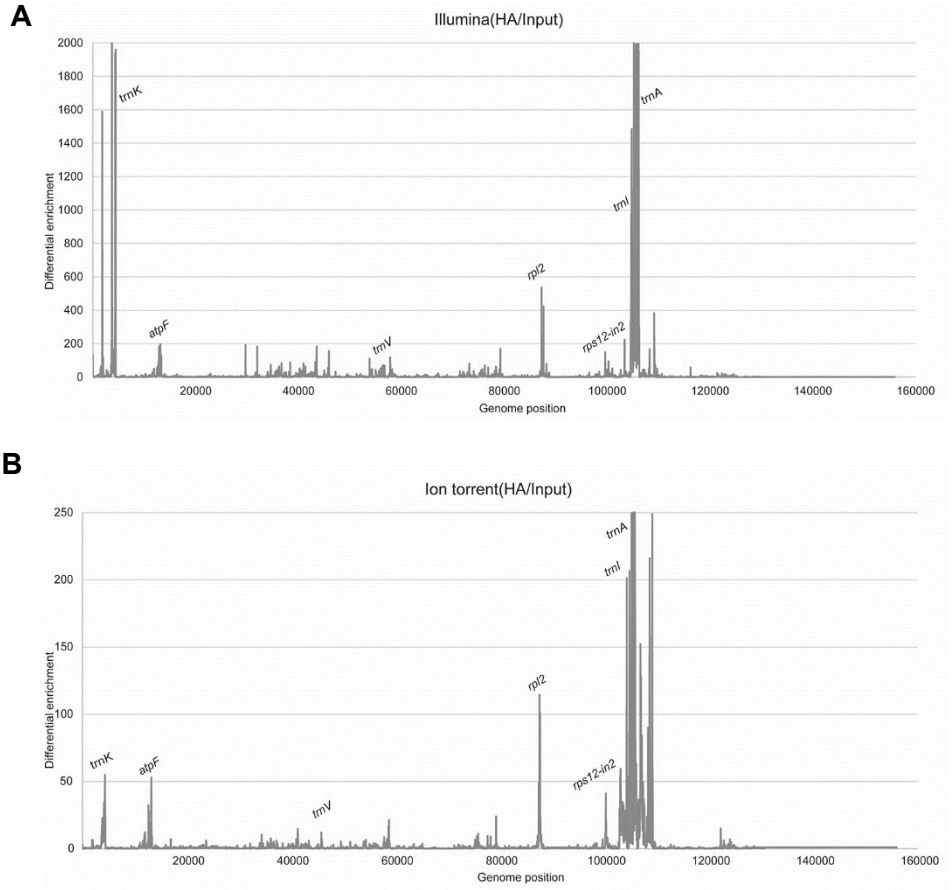
| | | | |
|---|---|---|---|
| trnA inf | TCTATCGTTGGCCTCTATGGTAG | NT | qPCR |
| trnA inf2 | TTGGAGAGCACAGTACGATGA | NT | qPCR |
| trnA exr | GACTCGAACCGCTGACATC | NT | qPCR |
| Ntmatkf | ATGGAAGAAATCCAAAGATATTTACA | NT | cloning |
| NTmatKseqr | TGTTGTTGCGATCTATCTGG | NT | cloning |
| NTmatKseqf | GAAGTCGGATTTGGTATTTGG | NT | cloning |
| clpP2inr2 | GAATTCGAAGTGCCATGCTA | NT | cloning |
| rpl2inr2 | ACTGAACTCAATCACTTGCTG | NT | cloning |
| rps12inr2 | GGTTAGCCATACACTTCACA | NT | cloning |
| trnIinr2 | GAGCACATTGAACTATCCATG | NT | cloning |
| trnKin5f | ACACATATGGATGAAGTGAGG | NT | PCR |
| trnKin5r | CGTTCCAAGAATTCTAACACG | NT | PCR |
| trnGf | GTTCGATTCCCGCTATCC | NT | PCR |
| seqNOS | TACATGCTTAACGTAATTCAACAG | vector | sequencing |
| 35Sprom | ATCCTTCGCAAGACCCTTCC | vector | sequencing |
| pURfor | GCTTGACCGCAACTTTGAC | Vector | *matK* expression |
| pURrev | CTTTGCTTCCAGATGTATGC | Vector | *matK* expression |
| pGEXf | GGGCTGGCAAGCCACGTTTGGTG | vector | sequencing |
| pGEXr | CCGGGAGCTGCATGTGTCAGAGG | vector | sequencing |
| PFLf | GATAACCATCTCGCAAATAAATAAG | vector | sequencing |
| PFLr | CAAGTAAAACCTCTACAAATGTG | vector | sequencing |
| ZmmatkBamHIr | GAGGATCCTTAATTAAGAGTAAGAGGATTCACC | ZM | *matK* expression |
| ZMmatkseq1f | CCGGAGTCAAGATGTTCC | ZM | sequencing |
| ZMmatkseq2f | CTTACCTTGTCAATTTCTCGC | ZM | sequencing |
| ZMmatKseqr | GCCTTTCCTTGATATCGAAC | ZM | Sequencing |
| pavHAr | CCTTAATTGAATTTCTCTAGAGCCTCATTAAGCATAATCAGGTACATC | Artificial | cloning |
| Riboswitchr | ATGATCCTCTCCACGAGAG | Artificial | cloning |
| Prrnr | GGATCCTCCCAGAAATATAGC | Artificial | cloning |
| HA5ANTmatkr | CATAATCAGGAACATCATAAGGATACTGGTAGTTTGCCAGGTC | Artificial | cloning |
| RSprrnr | GAATCAACGTAGGATCCTCC | Artificial | cloning |
| ScIPpsbAf | GCTGGAGCTCACCTTGGTTGACACGAGTATA | Artificial | cloning |
| BHIPpsbAr | AATTAAGAATCAACGTAGGATCCGGTAAAATCTTGGTTTATTTAATCATC | Artificial | cloning |
| aadaseqr | ACCTCTGATAGTTGAGTCGA | Artificial | cloning |
| Amatkseqf | GTTCTCGCATTTGGTATCTC | Artificial | cloning |
| SacIf | GAACAAAAGCTGGAGCTCG | Artificial | cloning |
| NTmatkRSr | GTAAATATCTTTGGATTTCTTCCATATGATCCTCTCCACGAGAG | Artificial | cloning |
| NTmatkPrrnr | GTAAATATCTTTGGATTTCTTCCATGGATCCTCCCAGAAATATAGC | Artificial | cloning |
| AmatkRSr | GTAACGCTGGATTTCTTCCATATGATCCTCTCCACGAGAG | Artificial | cloning |
| AmatkPrrnr | GTAACGCTGGATTTCTTCCATGGATCCTCCCAGAAATATAGC | Artificial | cloning |
| matKPpsbAr | GTAAATATCTTTGGATTTCTTCCATGGTAAAATCTTGGTTTATTTAATCATC | Artificial | cloning |
| AmatkPpsbAr | GTAACGCTGGATTTCTTCCATGGTAAAATCTTGGTTTATTTAATCATC | Artificial | cloning |

| | | | |
|---|---|---|---|
| NTmatkp7g10r | GTAAATATCTTTGGATTTCTTCCATGCTAGCCATATGTATATCTCCTTCT | Artificial | cloning |
| Amatkp7g10r | GTAACGCTGGATTTCTTCCATGCTAGCCATATGTATATCTCCTTCT | Artificial | cloning |
| ANTmatkseq1f | CCACCTGAATTACGTCCTG | Artificial | sequencing |
| ANTmatkseq2f | CACCTTCCTGCTCATGAAC | Artificial | sequencing |
| ANTmatkf | ATGGAAGAAATCCAGCGTTAC | Artificial | cloning |
| AZMmatkseq1f | CAGGAAGTCGAAGCAAGAG | Artificial | sequencing |
| AZMmatkseq2f | CTCCAAGAATGGGAAGATAG | Artificial | sequencing |
| trnfMf | TTCAAATCCTGTCTCCGCAA | Artificial | cloning |
| Amatkseqr | GTAGAGGAAATTGTGTTGCTG | Artificial | cloning |
| NtmatkBamHIr | GAGGATCCCTACTGGTAGTTTGCCAGG | Artificial | *matK* expression |
| Pawgfor | GTGATACCATGGGCAGAAG | Artificial | cloning |
| Pawgrev | GCTGAAAATCTTCTCTCATCC | Artificial | cloning |

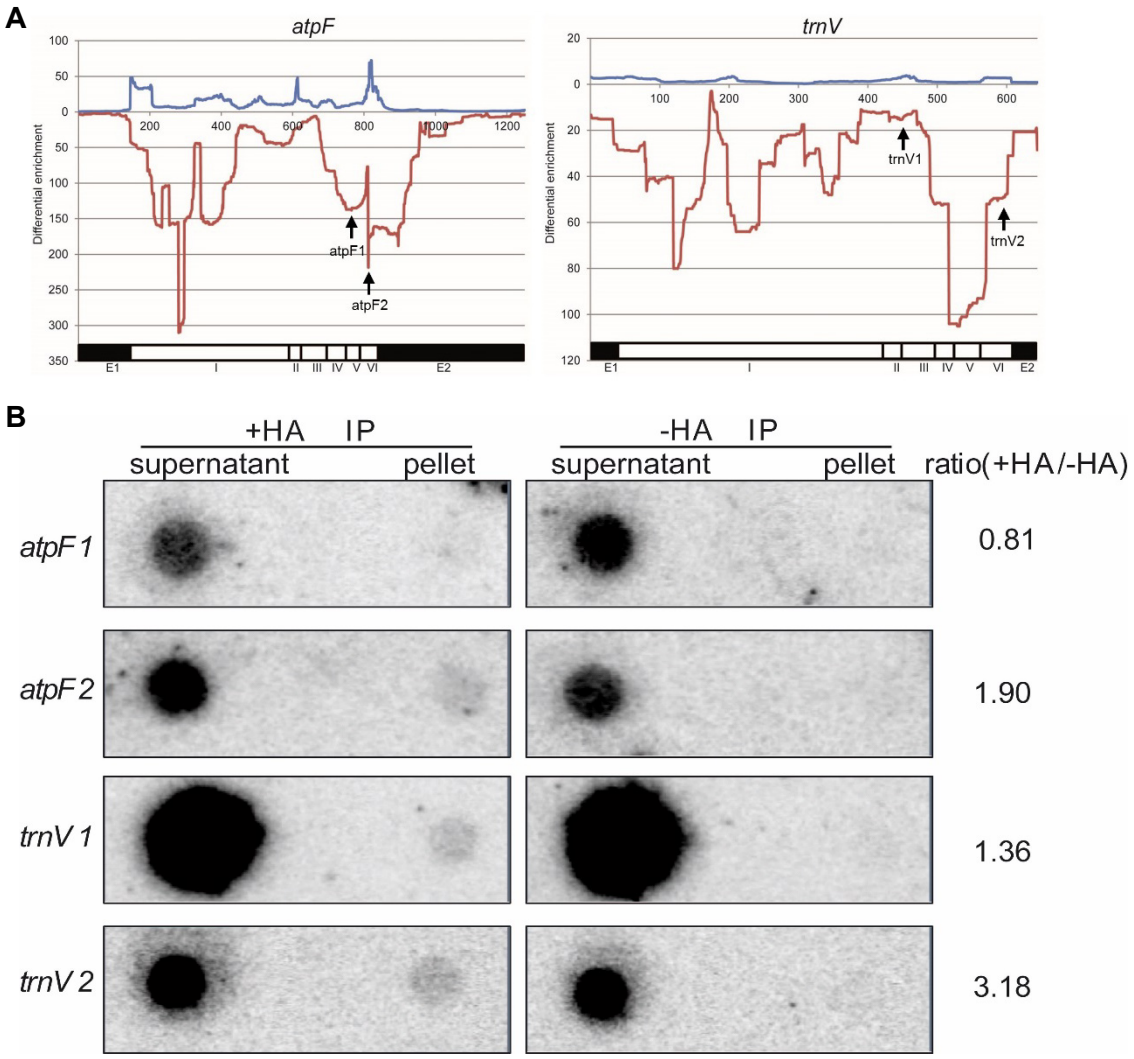
**Appendix 2 Analysis of immnuprecipated RNA before and after RNase III treatment.** The reports of measurement by fragment analyzer. (A) RNA samples from IP pellet of +HA tobacco. (B) RNA samples from IP pellet of –HA tobacco. (C) RNA sample from IP pellet of +HA, and digested with RNase III. Digested RNA sample from IP pellet of –HA was not measured due to the very small amount of RNA.



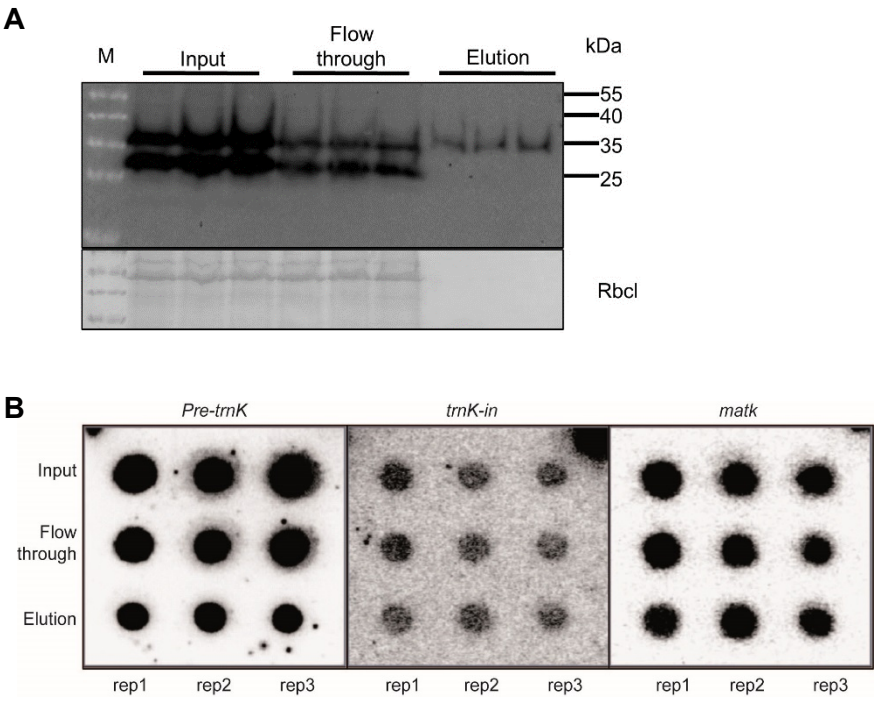
**Appendix 3 Overall view of MatK targets identified by RIP-seq.** RNA was extracted from the IP pellet +HA tobacco. With the RNA samples, libraries were prepared and sequenced with. Illumina and Ion Torrent sequencing systems. A total RNA library was used as input control. Reads from libraries were mapped to the chloroplast genome (NC_001879), respectively. The ratio of mapped reads between +HA and Input was calculated and plotted to the chloroplast genome for Illumina (A) and Ion Torrent (B) sequencing, respectively.



**Appendix 4 Confirmation of RIP-seq results by dot blot.** RNAs were extracted from IP supernatant or pellet and spotted to a nylon membrane. Oligo probes were radio-labeled and hybridized to the blots. (A) The oligos used in the hybridization. The positions of oligos were indicated with red arrows. (B) The signal of dots were quantified and the enrichment was calculated as IP pellet to supernatant. The enrichment ratios were calculated between IPs of +HA and –HA.



**Appendix 5 Dot blot analysis with RNA from ribosome IP.** (A) Inmmunoprecipatition of rpL32. GFP tagged rpL32 was inmmunoprecipited with μ MACS column (Miltenyi Biotec). Consequently, rpL32 associated RNA was eluted from the column. The protein samples of input, flow through and elution fractions were separated by SDS-PAGE gel and detected with GFP antibody. (B) Dot blot analysis of ribosome associated RNA. RNA from elution fraction was extracted and blotted onto nylon membranes. Probes located in the exon region of *trnK*, 5'border of *trnK* intron and *matK* ORF were radio-labeled and hybridized with the membranes.



**Appendix 6 Top hits with enrichment values in MatK RIP-seq**

| Illumina(+HA/–HA) | | | Illumina(+HA/input) | | | Ion Torrent(+HA/–HA) | | | Ion Torrent(+HA/input) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| top hits[1] | Position[2] | Value[3] | top hits | position | value | top hits | position | value | top hits | position | value |
| **trnK-in** | **4372** | 111.2 | **trnK-in** | **4359** | **31305.0** | trnfM | 38392 | 474.3 | 23s-4.5s rRNA | 109153 | 1063.3 |
| **trnA-in** | **106135** | 86.9 | **trnA-in** | **105542** | **4858.0** | **atpF-in** | **13177** | **310.0** | **trnA-in** | **105763** | **973.0** |
| **atpF-in** | **12638** | 72.5 | **trnI-in** | **105196** | **2946.5** | trnS | 8684 | 265.0 | **trnI-in** | **105196** | **497.5** |
| **rpl2-in** | **87244** | 41.8 | **rpl2-in** | **87252** | **536.8** | trnG | 38097 | 222.0 | 23s rRNA | 108631 | 215.0 |
| **trnI-in** | **104823** | 30.3 | 16s rRNA | 103382 | 224.1 | psbA-trnH | 500 | 178.0 | 16s rRNA | 104245 | 138.7 |
| clpP-*in2* | 72964 | 22.0 | **atpF-in** | 13095 | 197.3 | **trnK-in** | **1923** | **172.0** | **rpl2-in** | **87538** | **114.0** |
| **rps12-*in2*** | **100284** | 15.7 | unannotated | 29666 | 192.0 | **rpl2** | **87361** | **153.0** | **trnK-in** | **4304** | **55.0** |
| rpoC1-in1 | 23640 | 7.0 | unannotated | 43520 | 183.8 | **trnA-in** | **105974** | **145.5** | **atpF-in** | **13200** | **52.8** |
| rpoB | 25329 | 7.0 | petD-in | 79235 | 170.0 | trnL | 96461 | 130.0 | **rps12-*in2*** | **100267** | **41.0** |
| ycf3-*in2* | 44996 | 6.3 | 23s rRNA | 108253 | 167.3 | psbA | 538 | 124.0 | petD | 79192 | 24.0 |
| rpoC2 | 21229 | 6.0 | ycf3-in1 | 45886 | 156.0 | **trnI-in** | **104743** | **111.1** | rbcL | 58703 | 21.4 |
| | | | rps7 | 99606 | 149.0 | psbC | 36049 | 109.0 | ndhA-in | 122246 | 15.0 |
| | | | rbcL | 57814 | 117.3 | ycf3-*in2* | 45131 | 106.0 | atpA | 11981 | 12.0 |
| | | | **trnV-in** | **53797** | **110.6** | **trnV-in** | **53872** | **105.0** | ycf3-in1 | 45725 | 12.0 |
| | | | **rps12-*in2*** | **100269** | **95.0** | rbcL | 58267 | 99.0 | psaA | 41239 | 11.3 |
| | | | psaA | 43227 | 89.0 | 16s rRNA | 104244 | 80.8 | psbB | 75779 | 11.0 |
| | | | psbC | 36685 | 83.8 | trnQ | 7456 | 59.0 | unannotated | 34329 | 10.5 |
| | | | psaB | 40924 | 83.0 | trnS | 37183 | 59.0 | psaB | 41181 | 10.3 |
| | | | clpP-*in2* | 73209 | 79.5 | unannotated | 109438 | 43.9 | petB | 77597 | 9.5 |
| | | | 4.5s rRNA | 109296 | 78.8 | trnR | 10467 | 43.0 | psbC | 36058 | 7.7 |
| | | | psbD | 34582 | 73.0 | trnE | 32300 | 40.0 | **trnV-in** | **54268** | **6.6** |
| | | | atpB | 56475 | 72.0 | **rps12-*in2*** | **100159** | **37.0** | rpoC1 | 23660 | 6.0 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | psbB | 76216 | 69.9 | trnW | 68829 | 36.0 | ndhJ | 51245 | 5.0 |

1 Top 23 hits from each calculation were selected, the minimal enrichment ratio is equal to or greater than 5;

2 Chloroplast genomic position according to genbank entry NC_001879;

2 The value is the enrichment ratio of +HA to –HA or input;

3 The transcript names denoted in bold indicate that the hits are intron-containing and present in at least three caculations.

**Appendix 7 Short chloroplast RNAs in *Chlamydomonas reinhardtii***

| locus | 5' end[1] | 3' end[1] | core sequence[2] | Distance to start codon[3] | Distance to stop codon[3] | Stem-loop (dG[4]) |
|---|---|---|---|---|---|---|
| unannotated | 659 | 642 | ATCAGGTGGAAGCTTCCG | | | |
| petA 3' antisense | 4176 | 4158 | GAAGGGGTTTACTGATATC | | 325 | |
| petD 5'-1 | 6021 | 6041 | TTTAGCATGTAAACATTAGAA | -361 | | |
| petD 5'-2 | 6123 | 6141 | AACTTTTCGGAACGGCTAA | -259 | | |
| trnR1 3' | 7331 | 7348 | TAAGAGATTGTGGATTAC | | 139 | -13.2 |
| tufA 5' | 12453 | 12471 | TAAACCTGAAAAATTGGAT | -256 | | |
| tufA 3' | 14083 | 14104 | TCGACCATAGGTGAGGACAAAT | | 139 | |
| trnE1 3' | 14622 | 14640 | AGAACAGGAGGTAGTTTTT | | 105 | |
| trnC 3' | 14952 | 14969 | CGAGCTCTTCTGTTGTTC | | 82 | |
| rpl20 3' antisense | 16782 | 16800 | GTTCAAGCGGAGGATAAAT | | 128 | |
| trnW 5' | 17286 | 17268 | GTAAAAGTGGCTGGTTTAA | -40 | | |
| petB 3' | 19986 | 19969 | TACTTGCTTCGGATATAA | | 38 | |
| petB 5' | 20695 | 20674 | GAAAGCCTAATGGTCATGTCAC | -41 | | |
| trnG1-rps4 | 33506 | 33523 | TTAGAATCCAGGCATCTT | -93 | 91 | |
| unannotated | 57444 | 57423 | CGGACGAGGCGAGCTATCTTTA | | | |
| unannotated | 57942 | 57926 | CGCGGCGAGGTTGGGGC | | | |
| unannotated | 57992 | 57975 | TTTAGAGAGTACGGGATG | | | |
| rps14 5' antisense | 62789 | 62807 | ATAAAAGAAGGATGACTTG | -60 | | |
| psbM 3' antisense | 64139 | 64157 | GAAGACATAAGCGTGAACA | | 58 | |
| psbZ-psbM | 64404 | 64385 | TTTGTAACTCTAGACGGTAT | -103 | | |
| psbZ 5' antisense | 65407 | 65428 | TAAAGACCAAAATACTCTATCT | | 477 | |
| unannotated | 66263 | 66284 | AAGACTAAAGTGCCCAAGCTCA | | | |
| unannotated | 66761 | 66743 | AGGGGACGTCTGCTGTACG | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| ccsA 5' antisense | 68790 | 68808 | AAAGGGGAAGAGGCTTGTC | -231 | | |
| unannotated | 75118 | 75100 | ATGGGACGGATGGCATTTT | | | |
| psbH 3' antisense | 76604 | 76625 | TCTTAACGGAAGGCCAGTGGCA | | 143 | -16.7 |
| psbH 5' | 77067 | 77047 | TTTACAGAAAGTAAATAAAAT | -54 | | |
| psbB 5' | 81838 | 81818 | AATAATTAAGTAAAAAAATCA | -35 | | |
| rps9 5' antisense | 93937 | 93956 | TATGACTGATTTCGGCCAAT | -219 | | |
| trnM2-psbE 5' | 94895 | 94878 | AAAGCAGACAAATTGTTGAAAA | -59 | 80 | |
| psbF 5' | 102095 | 102116 | AACGAGTTAGCTTAATACAAAA | -53 | | |
| psbL 3' | 103083 | 103102 | ACGCGGAGGGGATAACATAA | | 74 | -21.1 |
| petG 5' | 103264 | 103284 | TCTTGAAGTGTGATGACTCCT | | | |
| petG 3' | 103725 | 103743 | TCTTCTATGGACGGAATAC | | 89 | |
| rpoC2 5' antisense | 107969 | 107987 | AGGAGACGGTAAGGTAGAG | | | |
| unannotated | 118991 | 119009 | GCGCAGAATCGAAGGGTAT | | | |
| psaB 5' | 119227 | 119245 | ACAGGATTATGGCGTAGTC | -374 | | |
| rbcL 5' | 123961 | 123943 | GATTATTTTAGGATCGTCA | | 44 | |
| atpA 5' | 124492 | 124510 | AGCAATCGGCGTCATAAAC | -280 | | -11.1 |
| atpA-psbI | 126304 | 126324 | AATTAAGTAGGAACTCGGTAT | -364 | 26 | -8.1 |
| atpH 5' | 128934 | 128952 | TTAGGAGGAAATACAATGA | -15 | | |
| atpH 3' | 129442 | 129461 | CTGTTTGGTATTGGCTAACG | | 264 | |
| ORF1995- atpB | 161954 | 161933 | AAAAATAAGCGTTAGTGAATAA | -314 | 167 | |
| psbD 5' | 175679 | 175660 | AATTTAACGTAACGATGAGT | -47 | | |
| ORF2971 3' | 186481 | 186503 | TTTAAGTGTTACAAAGAAATTGA | -421 | | |
| psbC 5' | 186938 | 186956 | TTATATGGGTGGACGTTAA | -87 | | |
| trnF-psaC | 189303 | 189320 | GTATGAGGAGAGAAATAT | -159 | 55 | |
| trnF-psaC | 189368 | 189391 | AAGTCGATTCTCAATCTTCTTTTT | | | |
| trnF-psaC | 189395 | 189414 | ATGGAGATGACATATTTAGC | -67 | 149 | |
| trnN 3' | 190887 | 190905 | ATCTTAGGCGTCTTTGGAC | | 54 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| rpoC1a 5' | 200832 | 200814 | TTCGGGCTGGCACGGCCAT | -267 | | |
| trnV 3' | 201023 | 201005 | ATTGCGGATGTATAAATTC | | 38 | |
| unannotated | 203152 | 203133 | AACTGGAGGCGAAAATTAGA | | | |
| unannotated | 14923(203499) | 14941(203517) | CCGGTAAACTCGCCCACAT | | | |
| unannotated | 1729(40427)<br>(153776) | 1710(40451)<br>(153752) | TCCGAAGGAGGGAGCAGGCA | | | |
| unannotated | 2238(14095)<br>(54810)(139393)<br>(176468) | 2262(14113)<br>(54792)(139411)<br>(176486) | AATAAATAAATTTGTCCTC | | | |
| unannotated | 35884(158319) | 35903(158300) | ATACGAATAGTAATACGGTT | | | |
| trnA-rrn7 | 41751(152452) | 41768()152435 | ATTAAGAGCGAATCAAAA | -78 | 381 | |
| psbA in 2 | 53489(140714) | 53471(140732) | CAGCTGGCGAGTCAATTGT | | | |
| psbA in 1 | 54778(139425) | 54761(139442) | AATGATCGAGCAGGACTT | | | |
| psbA 5' | 55449(138754) | 55431(138772) | TTTACGGAGAAATTAAAAC | -36 | | |

1 Chloroplast genomic position according to genbank entry NC_005353; in parentheses: positions multiple sites, mostly in inverted repeat IRA;

2 >50% of sequences in a cluster share this sequence;

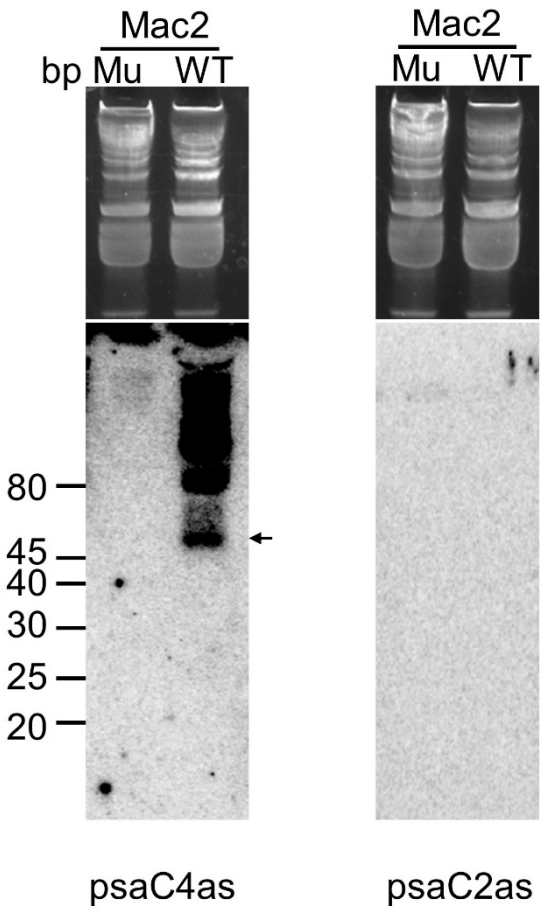
3 The start and stop codon next to the sRNA cluster is considered and only if it is no farther than 500 nt away;

4 Free energy was calculated using mfold.

**Appendix 8 SRNAs coincide with the transcripts ends**

| locus | core sequence | Distance to start codon | stability factor | UTR end | UTR end literature | stability factor literature |
|---|---|---|---|---|---|---|
| petD 5'-1 | TTTAGCATGTAAACATTAGAA | -361 | MCD1 | -361,-344 | Nancy R. Sturm,1994 MCB | Schinya Murakami 2005 NAR, Linda A. Rymarquis 2006 The Plant J. |
| tufA 5' | TAAACCTGAAAAATTGGAT | -256 | | -257,-256 | Gregg W. Silk 1993 PMB | |
| psbA 5' | TTTACGGAGAAATTAAAAC | -36 | RBP63 | -90, -36 | Jörg Nickelsen 1994 EMBO | Friedrich Ossenbühl 2002 EJB |
| psbH 5' | TTTACAGAAAGTAAATAAAAT | -54 | Mbb1 | -54,-53,-51,-78 | Clayton H. Johnson 1994 PMB | Fabián E. Vaistij 2000 PNAS |
| psbB 5' | AATAATTAAGTAAAAAAATCA | -35 | Mbb1 | -147,-35 | Fabián E. Vaistij 2000 The Plant J. | Fabián E. Vaistij 2000 PNAS |
| atpH 5' | TTAGGAGGAAATACAATGA | -15 | | -13 | Dominique Drapier 1998 Plant Physiology | |
| psbD 5' | AATTTAACGTAACGATGAGT | -47 | Nac2 | -74,-47 | Jörg Nickelsen 1994 EMBO | Christian Schwarz 2007 The Plant Cell |

**Appendix 9 Autoradiograms of an RNA blot of chloroplast RNA from the WT and the *Mac2* mutant.** The radioactive probes used in hybridization are indicated below the panels. The arrowheads indicate the signal of the sRNAs. LMW RNA was enriched from total RNA and resolved on a 15% polyacrylamide gel containing 8M urea. A fluorescence image (the upper part of each panel) from the gel stained with ethidium bromide is shown as loading control. The probes are single-stranded DNA oligonucleotides (<=25 nt) antisense to the sRNAs. The two blots were prepared from the same RNA sampless.



# Acknowlegments

So many people helped me in the past four years, they made my life in Berlin much easier than I thought. I would like to sincerely thank all of them.

I would like to thank my supervisor Prof. Christian Schmitz-linneweber for giving me this great chance to work in his group, for always being there to guide, encourage and support me with patience.

Hannes, Sabrina and Marlene gave me so much indispensable advice and cooperation, they have been supporting me both in my research and my daily life.

Reimo taught me how to work in a proper and efficient way. Julia, Adreas and Ayako are willing to help me with my experiments at any time.

Prof. Boener guided me to this group and supported me in the past years.

I would like to thank Tabea, much of my experimental work would have not been completed without her help. Boris, Kersten and Angelika also helped me with my experiments.

Cori, Conny and Jana have been organizing and maintaining our lab, making it a convenient place for working. Jan and Stephanie were of great help and saved a lot of time for me. Carola helped me dealing with the required documents and computer work. Arne and Gongwei always gave me necessary support.

I thank all the members in our lab for creating such a wonderful working atmosphere.

I thank Prof. Ostersetzer-Biran and Prof. Grimm for reviewing my dissertation, and Prof. Saumweber for serving in my defense committee.

I thank my friends in Berlin for accompanying and supporting during my difficult time. Thank my husband for accepting whatever I do and sharing life with me. Finally I want to thank my parents and my brothers, for encouraging me to follow my heart, for making me never worry about anything.

I can never list all the people who helped me, but I thank you all for making my time in Berlin so enjoyable.

# Curriculum Vitae

# Publications

Loizeau, K. #, **Y. Qu**#, S. Depp, V. Fiechter, H. Ruwe, L. Lefebvre-Legendre, C. Schmitz-Linneweber and M. Goldschmidt-Clermont 2014 Small RNAs reveal two target sites of the RNA-maturation factor Mbb1 in the chloroplast of Chlamydomonas. Nucleic Acids Res 42(5): 3286-3297.

Zoschke, R., **Y. Qu**, Y. O. Zubo, T. Borner and C. Schmitz-Linneweber 2013. Mutation of the pentatricopeptide repeat-SMR protein SVR7 impairs accumulation and translation of chloroplast ATP synthase subunits in Arabidopsis thaliana. J Plant Res 126(3): 403-414.

Hertel, S., R. Zoschke, L. Neumann, **Y. Qu**, I. M. Axmann and C. Schmitz-Linneweber 2013. Multiple checkpoints for the expression of the chloroplast-encoded splicing factor MatK. Plant Physiol 163(4): 1686-1698.

#Equal contribution

## Conference posters

Qu, Y., R. Zoschke, L. Neumann, S. Hertel, I. Axmann and C. Schmitz-Linneweber 2011 MatK as a CIS-Regulator for Chloroplast gene expression? Botanikertagung, Berlin, Germany

Qu, Y., C. Schmitz-Linneweber 2012 MatK RNA binding sites identified by RIP-Seq. ISE-G Tagung, Munich, Germany

Qu, Y., A. Weihe, G. Weber, M. Wahl, C. Schmitz-linneweber 2014 RIP-Seq analysis of the chloroplast maturase MatK. Conference of Molecular Biology of Plants. Dabringhausen, Germany

## Conference presentations

Qu, Y., Z., Reimo, and C. Schmitz-Linneweber 2012 Fine-mapping of the RNA binding sites of a promiscuous chloroplast intron maturase. Havel-Spree-Colloquium, Berlin Germany

# Selbstständigkeitserklärung

Hiermit erkläre ich, Yujiao Qu, die vorliegende Arbeit selbststständig verfasst zu haben. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt.

Yujiao Qu

Berlin, 16 December 2014