

**Etablierung einer auf RNA-Seq basierten Methode zur  
Charakterisierung von RNA-Polymerase-spezifischen  
Promotoren in Plastiden von *Arabidopsis thaliana***

**Dissertation**

zur Erlangung des  
Doktorgrades der Naturwissenschaften (Dr. rer. nat.)  
der  
Naturwissenschaftlichen Fakultät I - Biowissenschaften -

der Martin-Luther-Universität  
Halle-Wittenberg,

vorgelegt von

Frau M. Sc. Julia Rott  
geb. am 29.09.1988 in Ochansk

Gutachter:

1. Kristina Kühn
2. Christian Schmitz-Linneweber
3. Thomas Pfannschmidt

Datum der mündlichen Prüfung: 06.10.2021



Für meinen kleinen Tiziano.  
Du warst meine größte Motivation und Hilfe beim Schreiben,  
auch wenn du den Erfolg nicht mehr mit mir teilen konntest.



### SUMMARY

Transcription in plastids of higher plants is dependent on two different transcriptional machineries. One transcriptional activity, the plastid-encoded RNA polymerase (PEP), is encoded in the plastidial genome itself and derives from the bacterial ancestor. Another nuclear-encoded RNA polymerase (NEP) is similar to those of the T3/T7 phages. In plastids of dicotyledonous plants, the NEP activity is represented by two RNA polymerases, RPOTp and RPOTmp. RPOTp is exclusively present in plastids and RPOTmp is dually targeted to both plastids and mitochondria. PEP has been suggested to function in the transcription of photosynthesis genes in green tissue, whereas the NEPs supposedly transcribe housekeeping genes mostly in early non-green plant development. The individual roles of the RPOTp and RPOTmp and how they share their tasks with PEP remains to be clarified. So far only few plastidial transcription start sites (TSS) which are used by PEP or NEP are described using techniques like *Primer Extension*. A quantitative method which allows TSS mapping and a genome-wide analysis of TSS, such as RNA-Sequencing (RNA-Seq), is needed for the identification of differences in TSS usage.

This project was aimed to determine the function and transcriptional task division between RNA polymerases in the plastid by generating a genome-wide map of TSS usage in *Arabidopsis thaliana*. After a successful optimization of a differential RNA sequencing method (dRNA-Seq), a total of 121 TSS used by the different polymerases in the plastid were identified. The assignment of the TSS to RNA polymerase activities showed the individual function of RPOTp in the transcription of *ycf1* and genes for ribosomal proteins. Some additional TSS for photosynthesis genes were shown to be RPOTp dependent. RPOTmp displayed its main transcriptional activity during seed imbibition using only the previously described Pc promoter of the *rrn16* operon. In addition to their individual function, both RPOTp and RPOTmp have an overlapping function transcribing several genes like *clpP* and the *rpoB* operon. In comparison to RPOTp, the transcriptional role of RPOTmp in plastids is minor. RPOTmp is able to transcribe RPOTp dependent genes (e.g. *ycf1*) in the absence of RPOTp, however the transcript levels are lower than in the presence of RPOTp. A novel PEP function in the early developmental stages of *A. thaliana* is presented in this work. During imbibition and seed germination, PEP transcribes preferentially tRNAs and rRNAs. Several TSS for photosynthesis genes were found which are used by PEP before the seed is subjected to a light stimulus. This thesis presents important functions for both RPOTp and RPOTmp in plastid transcription. The polymerases display the NEP activity with individually specific as well as overlapping roles. The three polymerases PEP, RPOTp, and RPOTmp are active in non-green developmental stages as well as in seedlings and fulfill an overlapping function in both transcription of housekeeping genes and photosynthesis genes.

Keywords: chloroplast, plastidial transcription, RNA polymerase, NEP, PEP, dRNA-Seq, TSS

### ZUSAMMENFASSUNG

Plastiden enthalten ein eigenes Genom, welches von RNA-Polymerasen unterschiedlichen Ursprungs transkribiert wird. Dikotyle Pflanzen besitzen neben einer plastidkodierte RNA-Polymerase vom Eubakterientyp (PEP) zwei kernkodierte RNA-Polymerasen des Phagentyps (NEPs), die in Plastiden importiert werden. Hierbei ist RPOTp ausschließlich im Plastiden zu finden, während RPOTmp sowohl in Plastiden als auch in Mitochondrien importiert wird. Für PEP wurde die Hauptfunktion in grünem Gewebe in der Transkription der Photosynthesegene postuliert, während den beiden NEPs in früher Pflanzenentwicklung die Transkription einiger *housekeeping*-Gene zugeschrieben wurde. Jedoch sind die individuellen Rollen von RPOTp und RPOTmp sowie deren Zusammenarbeit mit PEP unzureichend erforscht. Das vorliegende Wissen beschränkt sich auf einige wenige beschriebene plastidäre Transkriptionsstartpunkte (TSS), die sich auf Methoden wie *Primer Extension* stützen. Quantitative Methoden für eine genomweite Kartierung von TSS, wie die Gesamt-Transkriptom-Shotgun-Sequenzierung (RNA-Seq), sind nötig, um Fragen über die TSS-Nutzung durch unterschiedliche RNA-Polymerasen zu beantworten.

Das Ziel der vorliegenden Arbeit war die Etablierung einer RNA-Seq basierten Methode zur Kartierung von Transkriptionsstarts zwecks der Charakterisierung der plastidären RNA-Polymerasen bezüglich ihrer Funktion im Modellorganismus *Arabidopsis thaliana*. Die erfolgreich optimierte Methode der differentiellen RNA-Sequenzierung (dRNA-Seq) ermöglichte die Identifizierung von 121 TSS, welche im Plastom zur Transkription genutzt werden. Die Analyse der TSS-Nutzung zeigte die individuelle Funktion von RPOTp in der Transkription von *ycf1* und Genen für ribosomale Proteine. Zusätzlich wurden RPOTp-abhängige TSS für Photosynthesegene identifiziert. Als einzige strikt RPOTmp-abhängige TSS wurde der bereits bekannte Pc-Promotor des *rrn16*-Operons identifiziert und die Hauptfunktion der Polymerase während der Imbibierung festgestellt. Neben den individuellen Rollen zeigten RPOTp und RPOTmp eine überlappende Funktion in der Transkription einiger Gene wie *clpP* und des *rpoB*-Operons. Insgesamt ist die Rolle von RPOTmp in plastidärer Transkription im Vergleich von RPOTp eher gering. Jedoch ist RPOTmp in Abwesenheit von RPOTp in der Lage, RPOTp-abhängige Gene wie *ycf1* zu transkribieren, auch wenn keine Transkriptmengen wie in Wildtyp erreicht werden können. Für PEP wurde eine noch unbekannt Funktion während der Imbibierung und früher Samenkeimung gezeigt, bei der PEP präferentiell rRNAs und tRNAs transkribiert. Zusätzlich wurden auch TSS für Photosynthesegene gefunden, welche schon in Dunkelheit von PEP zur Transkription genutzt werden. Die vorliegende Arbeit zeigt, dass sowohl RPOTp als auch RPOTmp wichtige Funktionen in der plastidären Transkription erfüllen und gemeinsam die NEP-Transkriptionsaktivität mit überlappender wie spezifischer Rolle darstellen. Die Polymerasen PEP, RPOTp, RPOTmp sind in nicht-grünen Pflanzenstadien und Keimlingen aktiv und nehmen eine größere überlappende Rolle in der Transkription von *housekeeping*- und Photosynthesegenen ein, als bisher angenommen.

Schlagworte: Chloroplast, plastidäre Transkription, RNA-Polymerase, NEP, PEP, dRNA-Seq, TSS

## INHALTSVERZEICHNIS

|  |           |
|--|-----------|
| Summary .....  | I         |
| Zusammenfassung.....   | II        |
| Inhaltsverzeichnis .....   | III       |
| Abbildungsverzeichnis.....   | VI        |
| Tabellenverzeichnis .....  | VIII      |
| <b>I Einleitung.....</b>   | <b>1</b>  |
| I.1 Organellen der Pflanzenzelle .....                                       | 1         |
| I.2 Das plastidäre Genom .....   | 1         |
| I.2.1 Organisation des plastidären Genoms von <i>A.thaliana</i> .....        | 2         |
| I.3 Plastidäre Transkription .....   | 3         |
| I.3.1 Plastidenkodierte Eubakterientyp-RNA-Polymerase (PEP) .....            | 4         |
| I.3.2 Kernkodierte Phagentyp-RNA-Polymerasen (NEP).....                      | 5         |
| I.3.3 Zusätzliche Phagentyp-RNA-Polymerase in Dikotyledonen .....            | 6         |
| I.4 Plastidäre Promotoren.....   | 7         |
| I.4.1 NEP-Promotoren.....  | 7         |
| I.4.2 PEP-Promotoren.....  | 9         |
| I.4.3 Funktion der RNA-Polymerasen während der Entwicklung der Pflanze.....  | 10        |
| I.5 Methoden zur Identifikation von Transkriptionsstarts .....               | 11        |
| I.6 Zielstellung der Arbeit .....  | 14        |
| <b>II Material und Methoden.....</b>   | <b>16</b> |
| II.1 Material und Anbieter .....   | 16        |
| II.1.1 Kits, Enzyme und Oligonukleotide .....                                | 16        |
| II.1.2 Pflanzenmaterial.....   | 16        |
| II.1.3 Pflanzenanzucht .....   | 17        |
| II.2 Methoden .....  | 19        |
| II.2.1 Isolierung von Nukleinsäuren .....                                    | 19        |
| II.2.2 Hochdurchsatz-Genotypisierung von <i>A. thaliana</i> Keimlingen ..... | 19        |
| II.2.3 Konzentrationsbestimmung von RNA und DNA .....                        | 20        |
| II.2.4 Aufreinigung und Fällung von RNA und DNA.....                         | 20        |
| II.2.5 Amplifikation von DNA mittels PCR.....                                | 20        |
| II.2.6 Agarosegelelektrophorese von RNA und DNA.....                         | 20        |
| II.2.7 Polyacrylamid-Gelelektrophorese (PAGE) von DNA .....                  | 21        |
| II.2.8 Northern Blot .....   | 21        |
| II.2.9 dRNA-Seq.....   | 23        |
| II.2.10 RNA-Seq.....   | 27        |
| II.2.11 Elektronenmikroskopische Untersuchung der Plastiden.....             | 27        |

## Inhaltsverzeichnis

---

|         |  |    |
|---------|--|----|
| II.2.12 | Lichtmikroskopische Untersuchung der Samenentwicklung.....                               | 27 |
| III     | Ergebnisse.....  | 29 |
| III.1   | Wachstumsanalyse der Mutanten.....   | 29 |
| III.2   | Optimierung des dRNA-Seq-Protokolls.....   | 31 |
| III.2.1 | Optimierung der Entfernung der prozessierten RNA.....                                    | 33 |
| III.2.2 | Nutzung eines UMI für einen quantitativen Vergleich.....                                 | 33 |
| III.2.3 | Optimierung der PCR-Zyklenzahl.....  | 34 |
| III.3   | Das mittels dRNA-Seq identifizierte plastidäre Transkriptom von <i>A. thaliana</i> ..... | 35 |
| III.3.1 | Erfolgreiche Detektion bekannter TSS.....  | 37 |
| III.3.2 | Plastomweite Identifikation von TSS.....   | 39 |
| III.4   | Charakterisierung der plastidären Promotoren.....  | 40 |
| III.4.1 | Methoden zur Klassifikation der TSS.....   | 40 |
| III.4.2 | TSS-Zuweisung nach Abhängigkeit von RNA-Polymeraseaktivität.....                         | 42 |
| III.4.3 | Erschwerte Zuordnung von TSS im späten Stadium.....                                      | 42 |
| III.5   | Validierung der TSS-Analyse mit RNA-seq.....   | 44 |
| III.6   | Funktion von RPOTp und RPOTmp in plastidärer Transkription.....                          | 44 |
| III.6.1 | Bekannte NEP-abhängige TSS konnten nicht identifiziert werden.....                       | 44 |
| III.6.2 | RPOTp hat eine genspezifische Funktion.....  | 45 |
| III.6.3 | RPOTp nutzt alternative TSS in Abwesenheit von PEP.....                                  | 51 |
| III.6.4 | Motivanalyse der RPOTp-abhängigen TSS.....   | 53 |
| III.6.5 | RPOTp und RPOTmp haben eine überlappende Funktion.....                                   | 53 |
| III.7   | Funktion von PEP in plastidärer Transkription.....                                       | 56 |
| III.7.1 | PEP-abhängige TSS für <i>ycf2</i> wurden detektiert.....                                 | 57 |
| III.7.2 | PEP transkribiert tRNAs und rRNAs während der Samenkeimung.....                          | 58 |
| III.8   | TSS-Nutzung in frühen Phasen der Samenkeimung.....                                       | 60 |
| III.9   | Untersuchung einer möglichen <i>rpoTp rpoTmp</i> Doppelmutante.....                      | 63 |
| III.9.1 | Segregationsanalyse der <i>rpoTmp rpoTmp RPOTp rpoTp</i> Mutante.....                    | 63 |
| III.9.2 | Phänotypische Untersuchung der Samen der <i>rpoTmp rpoTmp RPOTp rpoTp</i> Mutante...     | 63 |
| III.9.3 | Untersuchung der Embryonen der <i>rpoTmp rpoTmp RPOTp rpoTp</i> Mutante.....             | 64 |
| III.10  | Elektronenmikroskopische Untersuchung der Plastiden in weißen Mutanten.....              | 65 |
| IV      | Diskussion.....  | 68 |
| IV.1    | Erfolgreiche dRNA-Seq-Optimierung für quantitative Analyse der TSS-Nutzung in Plastiden  | 68 |
| IV.1.1  | Reduktion des Hintergrunds verbessert die eindeutige Festlegung der TSS.....             | 68 |
| IV.1.2  | Nutzung eines UMIs korrigiert die Genomabdeckung für PCR-Duplikate.....                  | 70 |
| IV.1.3  | Weitere mögliche Verbesserungen der Strategie.....                                       | 71 |
| IV.1.4  | Durch Sekundäreffekte erschwerte Klassifikation der TSS.....                             | 72 |
| IV.2    | Funktion der plastidären Phagentyp-RNA-Polymerasen.....                                  | 74 |
| IV.2.1  | RPOTp transkribiert <i>ycf1 (tic214)</i> .....   | 74 |
| IV.2.2  | RPOTp transkribiert Komponenten des Translationsapparats.....                            | 76 |



## Inhaltsverzeichnis

---

|        |  |     |
|--------|--|-----|
| IV.2.3 | RPOTmp transkribiert das <i>rrn16</i> -Operon vom Pc-Promotor während der Imbibition ..... | 77  |
| IV.2.4 | PEP-Aktivität ist von beiden Phagentyp-Polymerasen abhängig.....                           | 78  |
| IV.2.5 | RPOTp und RPOTmp transkribieren <i>clpP</i> .....  | 78  |
| IV.2.6 | In Abwesenheit von RPOTp nutzt RPOTmp die gleichen TSS.....                                | 79  |
| IV.2.7 | Variable Motive in den NEP-Promotoren.....   | 79  |
| IV.2.8 | Modell für die Funktionen der Phagentyp-Polymerasen.....                                   | 80  |
| IV.3   | Der embryothalen NEP-Doppelmutante fehlt die Transkription essentieller Gene .....         | 81  |
| IV.4   | Zusammenspiel von NEPs und PEP in der Transkription der Plastidengene.....                 | 82  |
| IV.5   | Rolle von PEP in der frühen Samenkeimung.....  | 83  |
| IV.5.1 | Regulation der PEP-Aktivität während der Pflanzenentwicklung .....                         | 84  |
| IV.6   | Schlusswort .....  | 85  |
| V      | Literaturverzeichnis .....   | 87  |
| VI     | Anhang.....  | 96  |
| VI.1   | Tabellen.....  | 96  |
| VI.2   | Abbildungen.....   | 120 |
| VI.3   | Bioinformatische Datenanalyse .....  | 124 |
|        | Danksagung.....  | 132 |
|        | Lebenslauf und Publikationsliste .....   | 133 |
|        | Eidesstattliche Erklärung .....  | 134 |

ABBILDUNGSVERZEICHNIS

Abbildung I.1: Organisation des Plastoms von *A. thaliana* ..... 3

Abbildung I.2: Lokalisierung der Phagentyp-RNA-Polymerasen ..... 6

Abbildung I.3: Aufbau der plastidären Promotoren..... 9

Abbildung I.4: Gegenwärtiges Modell der Funktionen der plastidenlokalisierten RNA-Polymerasen 10

Abbildung I.5: Schematische Darstellung der Protokolle für dRNA-Seq und Terminome-Seq..... 13

Abbildung II.1: Genkarten mit den Positionen der T-DNA-Insertionen in den Genen ..... 17

Abbildung II.2: Schematische Darstellung der bioinformatischen Analyse für dRNA-Seq..... 25

Abbildung II.3: Schematische Darstellung der bioinformatischen Analyse für RNA-Seq..... 27

Abbildung III.1: Entwicklung eines Samens von Wildtyp bis zum Keimlingsstadium,..... 30

Abbildung III.2: Wachstumsentwicklung von Wildtyp und Mutanten vom Samen bis zum Keimlingsstadium ..... 30

Abbildung III.3: Lichtmikroskopische Aufnahme von Keimlingen von Wildtyp und Mutanten..... 31

Abbildung III.4: Lichtmikroskopische Aufnahme von gekeimten Samen von Wildtyp und Mutanten 31

Abbildung III.5: Schematische Darstellung des optimierten dRNA-Seq-Protokolls..... 32

Abbildung III.6: TEX-Behandlung von RNA..... 33

Abbildung III.7: Gesamtmenge der gemappten Reads und der PCR-Duplikate..... 34

Abbildung III.8: PCR-Amplifikation der cDNA-Bibliotheken ..... 35

Abbildung III.9: Karte des plastidären Genoms mit den Genomabdeckungen..... 36

Abbildung III.10: Identifizierte TSS für *clpP* und *trnV*..... 37

Abbildung III.11: Identifizierte TSS für *rrn16* ..... 38

Abbildung III.12: Identifizierte TSS für *rbcL* und *psbA*..... 38

Abbildung III.13: Aufteilung der 121 identifizierten TSS nach Genklassen ..... 40

Abbildung III.14: Statistisches Modell zur Klassifizierung der TSS nach Nutzung durch eine RNA-Polymerase ..... 41

Abbildung III.15: Zuordnung der TSS zu Polymeraseaktivitäten..... 42

Abbildung III.16: Nutzung der im späten Stadium nicht zugewiesenen 45 TSS im frühen und späten Stadium von Wildtyp ..... 43

Abbildung III.17: Nicht identifizierte TSS für *rps4*..... 45

Abbildung III.18: Schematische Darstellung der Gene mit RPOTp-abhängigen TSS ..... 45

Abbildung III.19: RNA-Seq-Analyse von *ycf1*..... 46

Abbildung III.20: Identifizierte TSS für *ycf1*..... 47

Abbildung III.21: Analyse des *rpl23*-Operons ..... 48

Abbildung III.22: Northern Blot-Analyse des *rpl23*-Operons ..... 49

Abbildung III.23: Analyse der Transkriptmengen von *rpl32* ..... 50

Abbildung III.24: Analyse der Transkriptmengen von *psbD*..... 51

Abbildung III.25: Analyse der Transkriptmengen von *atpB* ..... 52

Abbildung III.26: Identifizierte Promotormotive für das RPOTp-abhängige Cluster von TSS ..... 53

Abbildung III.27: Identifizierte Promotormotive für die nicht zugewiesenen 9 TSS ..... 54

Abbildung III.28: Analyse der Transkriptmengen von *clpP*..... 55

Abbildung III.29: Genomabdeckung aus der RNA-Seq-Analyse für die Gene *clpP*, *rps12*, *rpl20* und Antisense der Gene *rps18* und *rpl33*..... 55

Abbildung III.30: RNA-Seq-Analyse des *rpoB*-Operons ..... 56

Abbildung III.31: Aufteilung der 65 PEP-abhängigen TSS nach Genklassen..... 57

Abbildung III.32: Genomabdeckung aus der RNA-Seq-Analyse für Antisense *rpl16*, *rps3*, *rpl22*, *rps19*, *rpl2*, *rpl23*, *trnI* und die Gene *ycf2* sowie *ORF77* ..... 57

Abbildung III.33: Analyse der Transkriptmengen von *trnK*..... 58

Abbildung III.34: Identifizierte TSS für *trnK*..... 59

Abbildung III.35: Nutzung der PEP-abhängigen TSS in Wildtyp ..... 60

Abbildung III.36: Analyse der Transkription in sehr früher Samenentwicklung ..... 61

Abbildung III.37: Identifizierte TSS in der Samenentwicklung ..... 61

Abbildung III.38: Identifizierte TSS für *rrn16* während Stratifikation und Keimung..... 62

## Abbildungsverzeichnis

---

|  |     |
|--|-----|
| Abbildung III.39: Schoten von <i>rpoTmp-2 rpoTmp-2 RPOTp-2 rpoTp-2</i> Mutanten .....                            | 64  |
| Abbildung III.40: Embryonalentwicklung in Samen von <i>rpoTmp-2 rpoTmp-2 RPOTp-2 rpoTp-2</i> Mutanten .....      | 65  |
| Abbildung III.41: Elektronenmikroskopische Aufnahme von Plastiden aus Wildtyp und Mutanten...                    | 66  |
| Abbildung III.42: Elektronenmikroskopische Aufnahme von Plastiden aus 24 Tage alten <i>rpoTp-</i> Mutanten ..... | 67  |
| Abbildung IV.1: Funktion von RPOTp und RPOTmp in der Transkription der Plastidengene .....                       | 81  |
| Abbildung IV.2: Funktionen der plastidären RNA-Polymerasen .....   | 86  |
| Abbildung VI.1: Wachstumsanalyse der Mutanten .....  | 120 |
| Abbildung VI.2: Clusteranalyse mittels pamK für das frühe Stadium .....  | 120 |
| Abbildung VI.3: Clusteranalyse mittels pamK für das späte Stadium.....   | 121 |
| Abbildung VI.4: TSS-Nutzung in der sehr frühen Samenentwicklung .....  | 121 |
| Abbildung VI.5: RNA-Seq-Analyse in den Mutanten für das frühe und späte Stadium .....                            | 122 |
| Abbildung VI.6: Gegenüberstellung der Genomabdeckung für die 5'-Enden des ersten Introns von <i>clpP</i> .....   | 123 |
| Abbildung VI.7: Genomabdeckung für das <i>psbB</i> -Operon.....  | 123 |
| Abbildung VI.8: RNA-Seq-Analyse ausgewählter plastidenlokalisierter Kerngene .....                               | 124 |

## TABELLENVERZEICHNIS

|   |     |
|---|-----|
| Tabelle II.1: Liste der verwendeten Mutantenlinien.....   | 16  |
| Tabelle II.2: Erntezeitpunkte der keimenden Samen bzw. Keimlinge von Wildtyp und Mutanten .....   | 17  |
| Tabelle III.1: Segregationsanalyse einer <i>rpoTmp rpoTmp</i> RPOTp <i>rpoTp</i> Mutante . <b>Fehler! Textmarke nicht definiert.</b>                    |     |
| Tabelle III.2: Auszählung der Samen in Schoten von <i>rpoTmp rpoTmp</i> RPOTp <i>rpoTp</i> Mutanten.<br>..... <b>Fehler! Textmarke nicht definiert.</b> |     |
| Tabelle VI.1: Liste der Hersteller und Anbieter.....  | 96  |
| Tabelle VI.2: Liste der verwendeten Oligonukleotide und deren Anwendung.....  | 96  |
| Tabelle VI.3: Abkürzungsverzeichnis.....  | 97  |
| Tabelle VI.4: Liste der Genbezeichnungen.....   | 98  |
| Tabelle VI.5: Statistische Daten zu RNA-Seq für die Analyse von Wildtyp und Mutanten .....  | 99  |
| Tabelle VI.6: Differentielle Expressionsanalyse in Mutanten im frühen Stadium .....   | 100 |
| Tabelle VI.7: Differentielle Expressionsanalyse in Mutanten im späten Stadium .....   | 102 |
| Tabelle VI.8: Statistische Daten zu dRNA-Seq für die Analyse von WT und Mutanten .....  | 105 |
| Tabelle VI.9: Statistische Daten zu dRNA-Seq für die Analyse der frühen Samenentwicklung .....  | 106 |
| Tabelle VI.10: Liste der identifizierten TSS in Mutantenanalyse mit Referenzen zu beschriebenen<br>TSS.....   | 107 |
| Tabelle VI.11: Klassifizierung mit dem statistischen Modell im frühen Stadium.....  | 109 |
| Tabelle VI.12: Klassifizierung mit dem statistischen Modell im späten Stadium .....   | 111 |
| Tabelle VI.13: Klassifizierung der TSS mit pamK.....  | 114 |
| Tabelle VI.14: Liste der TSS nach zugewiesener Polymeraseabhängigkeit .....   | 116 |
| Tabelle VI.15: Liste der TSS in der sehr frühen Samenentwicklung.....   | 118 |

# I EINLEITUNG

## I.1 Organellen der Pflanzenzelle

Die Organellen höherer Pflanzen, Plastiden und Mitochondrien, erfüllen wichtige Aufgaben im Metabolismus der Pflanzenzelle. Die photosynthetisch aktive Form der Plastiden, der Chloroplast, nutzt das Sonnenlicht, um unter Freisetzung von Sauerstoff Kohlenhydrate wie Glukose herzustellen. So nimmt der Chloroplast eine zentrale Funktion im Metabolismus der Pflanzenzelle ein. Neben der Photosynthese können in Plastiden andere Syntheseprozesse stattfinden, wie die Synthese von Stärke, Fett- und Aminosäuren, und Pigmenten (Rolland *et al.*, 2018). Im Mitochondrium findet die zelluläre Atmung statt, bei der mithilfe der oxidativen Phosphorylierung ATP synthetisiert wird. Pflanzliche Mitochondrien sind auch an anderen Prozessen, wie z. B. der Aminosäurebiosynthese beteiligt (Day, 2004). Plastiden und Mitochondrien sind aus Endosymbiosen entstanden, bei der freilebende Prokaryoten (Cyanobakterien als Vorläufer von Chloroplasten und Purpurbakterien von Mitochondrien) von einer Wirtszelle aufgenommen worden sind (Gray, 1999; McFadden, 2014). Im Laufe der Koevolution von Wirtszelle und Endosymbionten wurden große Teile der bakteriellen Genome in den Kern transferiert, wodurch nur ein geringer Anteil des Ursprungsgenoms erhalten blieb (Rousseau-Gueutin *et al.*, 2012). So besitzen die Organellen immer noch Teile des eubakteriellen Genoms, welcher im Vergleich zu ihren bakteriellen Vorfahren geschrumpft ist. Gleichzeitig sind viele Proteine, deren Gene kernkodiert sind, zu einem Bestandteil des Proteoms der Organellen geworden. Von den etwa 2000 bis 3000 Proteinen, die in Plastiden zu finden sind, werden nur 90 bis 100 im plastidären Genom kodiert (Jarvis and López-Juez, 2013; Zimorski *et al.*, 2014). Die sogenannte Eukaryotisierung wird auch durch die Transkriptionsmaschinerien der Organellen repräsentiert (Liere *et al.*, 2011).

## I.2 Das plastidäre Genom

Das organelläre Genom hat einen voll funktionsfähigen Apparat für DNA-Replikation und Genexpression. Das Plastidengenom (Plastom) kodiert außerdem für einen eigenen Transkriptionsapparat (Dobrogojski *et al.*, 2020). Das Plastom ist zirkulär aufgebaut und in DNA-Proteinkomplexen, den Nukleoiden, organisiert, die in reifen Chloroplasten an die Thylakoiden angeheftet sind (Krupinska *et al.*, 2013; Powikrowska *et al.*, 2014). Die meisten Gene sind, wie in Bakterien, in Operons organisiert, die als polycistronische mRNA transkribiert werden. Jedoch findet in Organellen im Gegensatz zu Prokaryoten eine posttranskriptionale Prozessierung der polycistronischen mRNA in oligo- oder monocistronische Transkripte statt (Barkan, 1988). Es gibt Hinweise dafür, dass die Transkription und die Translation in Organellen parallel stattfinden, wie es bei Prokaryoten üblich ist (Pfalz *et al.*, 2006). Die Organellen erwarben weitere eukaryotische Eigenschaften, welche die Koevolution von Organellen und Kern untermauern. Viele der im Plastom

kodierten Gene besitzen Introns und werden *cis*- und/oder *trans*-gespleißt (Barkan, 2011). Die mRNA wird ediert und posttranskriptional prozessiert. Trotz der im Vergleich zum Kern relativ kleinen Genome sind die Transkriptionsmaschinerien der Organellen komplex aufgebaut und besitzen sowohl kernkodierte wie auch plastidär kodierte Bestandteile (Kap. I.3).

### I.2.1 Organisation des plastidären Genoms von *A.thaliana*

Das Plastidengenom höherer Pflanzen hat eine Größe von 120 bis 170 kb und umfasst ca. 100-150 Gene. Parasitische Pflanzen, denen Gene für den Photosyntheseapparat fehlen, haben sogar noch kleinere Plastome (Wicke *et al.*, 2011). Das Plastom ist hoch polyploid und liegt in Mesophyllzellen schätzungsweise in 20-35 Kopien pro Plastid vor (Zoschke *et al.*, 2007). Die Replikation des Plastoms erfolgt unabhängig vom Kerngenom. Das zirkuläre Plastidengenom des dikotylen Modellorganismus *Arabidopsis thaliana* hat eine Größe von 154 478 bp (Abb. I.1). Die meisten Gene sind wie bei Prokaryoten in Operons organisiert und befinden sich sowohl auf dem Minus- wie auf dem Plusstrang. Das Plastom ist in den kleinen (*small single copy*, SSC, 17 780 bp lang) und großen (*large single copy*, LSC, 84 197 bp lang) Einzelkopiebereich eingeteilt, die durch zwei identische gegenläufige Sequenzabfolgen (*inverted repeats*, IRA und IRB, jeweils 26 264 bp lang) voneinander getrennt sind. Es beherbergt Gene, die für Komponenten des Photosyntheseapparats, Komponenten des plastidären Transkriptionsapparats, ribosomale RNAs und tRNAs kodieren (Dobrogojski *et al.*, 2020). Die *inverted repeats* verdoppeln das Vorkommen einer Anzahl von Genen, z. B. treten die Gene für die ribosomale RNA in zwei Kopien auf. Weiterhin gibt es offene Leseraster (*ORF*, *open reading frame*, oder *ycf*, *hypothetical chloroplast reading frame*), über deren Funktion noch diskutiert wird. Einige dieser offenen Leseraster (*ycf1*, *ycf2*) scheinen Proteine zu kodieren, für welche eine Funktion in der Translokation über die innere Plastidenmembran vorgeschlagen wird (Kikuchi *et al.*, 2013, 2018). Zu den proteinkodierenden Genen kommen etliche nichtkodierende RNAs hinzu (Hotto *et al.*, 2011; Zhelyazkova *et al.*, 2012b). Diesen nichtkodierenden RNAs wird zum Teil eine Funktion in der Regulation der Genexpression zugesprochen (Hotto *et al.*, 2012), andererseits scheinen einige kleine RNAs Degradationsprodukte zu sein, deren regulierende Funktion in Frage gestellt wird (Börner *et al.*, 2014; Ruwe and Schmitz-Linneweber, 2012; Zhelyazkova *et al.*, 2012b).

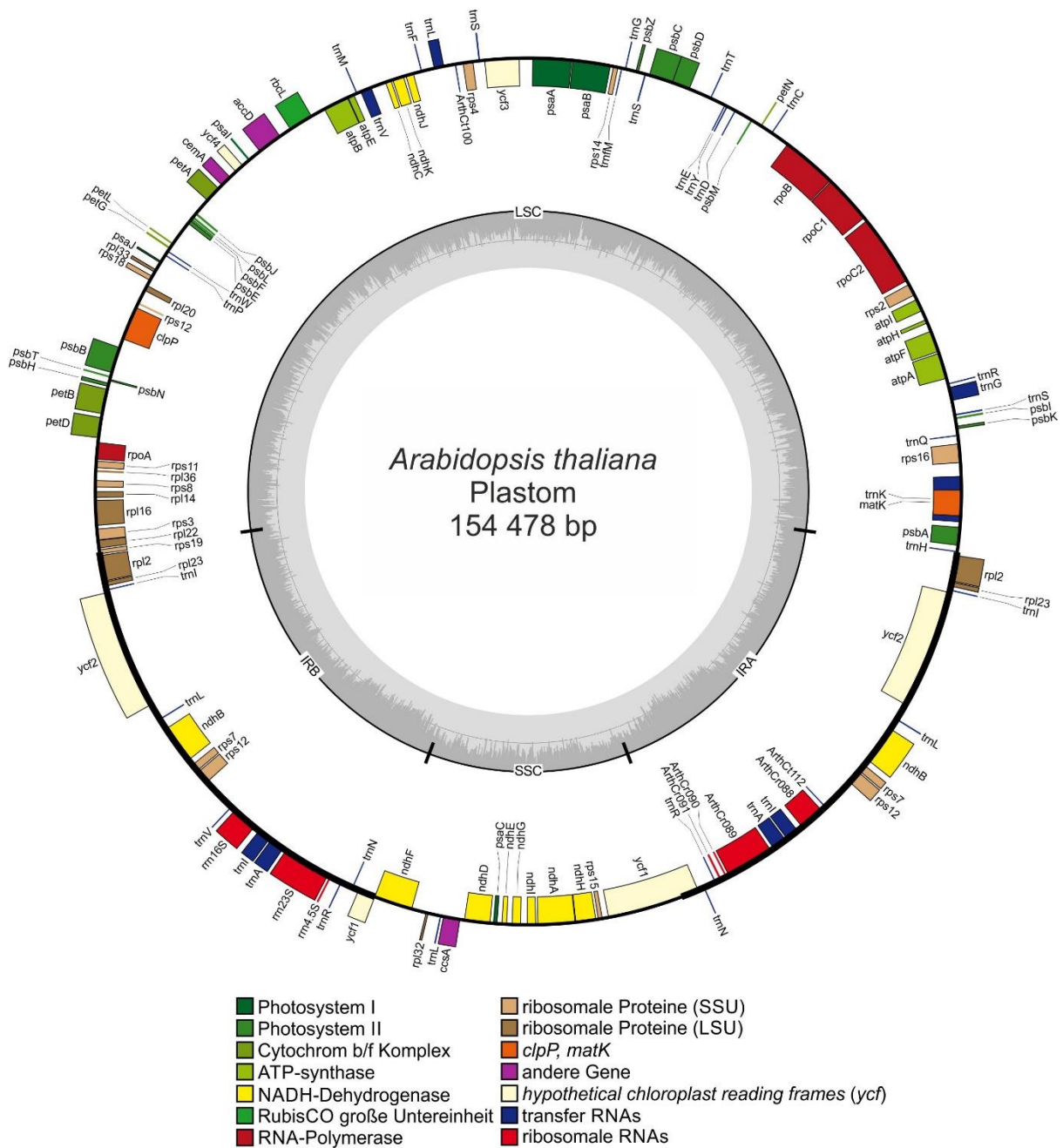


Abbildung I.1: Organisation des Plastoms von *A. thaliana* nach Sato *et al.*, 1999. Die Abbildung wurde mit OGDRAW generiert (Greiner *et al.*, 2019). Das Plastom ist zirkulär, Gene/Operons auf der Innenseite stellen den Minus-Strang dar, Gene/Operons auf der Außenseite stellen den Plus-Strang dar. SSU: Kleine Untereinheit, LSU: Große Untereinheit. (Für weitere Details zu Genen und Funktionen siehe Anhang Tab. VI.4).

### I.3 Plastidäre Transkription

Während das mitochondriale Genom in fast allen Eukaryoten (mit einer Ausnahme) von Phagentyp-RNA-Polymerasen transkribiert wird, ist die Situation in Plastiden komplexer. In Plastiden von fast allen Pflanzen findet man zwei Typen von RNA-Polymerasen, die Eubakterientyp-RNA-Polymerase, die von dem cyanobakteriellen Vorfahren stammt (PEP), sowie T3/T7-Phagentyp-RNA-Polymerasen (NEP) (Börner *et al.*, 2014). Parasitische Angiospermen haben ihre Eubakterientyp-RNA-Polymerase

sowie Gene des Photosyntheseapparats im Verlauf der Evolution verloren, ihr Plastom wird ausschließlich von Phagentyp-RNA-Polymerasen transkribiert (Liere *et al.*, 2011; Weihe *et al.*, 2012).

### I.3.1 Plastidenkodierte Eubakterientyp-RNA-Polymerase (PEP)

Die plastidenkodierte Eubakterientyp-RNA-Polymerase (*plastid-encoded plastid RNA polymerase*, PEP) stammt ursprünglich vom bakteriellen Endosymbionten und wurde im Laufe der Koevolution behalten. Sie ist der RNA-Polymerase von *Escherichia coli* sehr ähnlich, auch die Organisation der Gene für die Untereinheiten der Polymerase ist in Plastiden und Bakterien konserviert (Igloi and Kössel, 1992). Die plastidenkodierte RNA-Polymerase PEP besteht aus 5 Untereinheiten  $\alpha$ ,  $\beta$ ,  $\beta'$ , und  $\beta''$ , welche das *core*-Enzym ( $\alpha_2\beta\beta'\beta''$ ) mit der katalytischen RNA-Polymeraseaktivität bilden. Die vier Gene für die fünf Untereinheiten (*rpoA*, *rpoB*, *rpoC1*, *rpoC2*) sind in zwei Operons organisiert (Abb. I.1). Während *rpoB*, *rpoC1* und *rpoC2* in einem gemeinsamen Operon zu finden sind, ist *rpoA* in einem Operon mit Genen für ribosomale Proteine kodiert (Sugiura, 1992). Für die Promotorerkennung und Transkriptionsinitiation wird das *core*-Enzym durch assoziierte Proteine (PAPs, *PEP associated proteins*) und pTACs und Transkriptionsfaktoren (Sigma-Faktoren) vervollständigt (Abb. I.4) (Chi *et al.*, 2015; Pfalz and Pfannschmidt, 2013). PAPs, pTACs und Sigma-Faktoren sind kernkodiert, sodass die Aktivität von PEP vom Kern gesteuert wird. Um Promotorstrukturen erkennen und binden zu können, benötigt der PEP-*core*-Komplex Prokaryoten-ähnliche Sigma-Faktoren (Chi *et al.*, 2015; Lerbs-Mache, 2011; Schweer *et al.*, 2010). In *A. thaliana* sind 6 Sigma-Faktoren bekannt (Kap. I.4.2). Plastidäre PEP-*Knockout*-Mutanten in Tabak zeigen einen weißen bis gelblichen Phänotyp und eine Hemmung der Plastidenentwicklung (Allison *et al.*, 1996). Auch sekundäre PEP-*Knockdown*-Mutanten in *A. thaliana*, denen ein kernkodiertes Protein fehlt, welches für die Funktion des PEP-Transkriptionskomplexes nötig ist, haben einen ähnlichen Phänotyp (Chateigner-Boutin *et al.*, 2008; Pfalz *et al.*, 2006). Dieses bestätigt, dass eine intakte PEP(-Funktion) für die korrekte Chloroplastenentwicklung essentiell ist. Der PEP-Komplex existiert in mehreren Formen und kann als lösliches oder unlösliches Protein isoliert werden (Pfalz and Pfannschmidt, 2013). In der unlöslichen Form ist das Protein an DNA gebunden und wird als Transkriptionsaktives Chromosom (TAC, *transcriptionally active chromosome*) bezeichnet. Der TAC-Komplex besteht aus mehr als 50 Proteinen und ist *in-vitro* in der Lage, RNA nach Zugabe von Nukleotiden zu synthetisieren (Pfannschmidt *et al.*, 2015). Das lösliche PEP-Protein kann in zwei Formen gefunden werden. Während in grünem Gewebe ein PEP-Komplex mit vielen assoziierten Proteinen (PEP-A) vorkommt, welches keine Inhibierung durch Rifampicin zeigt, findet man in Etioplasten nur das rifampicinsensitive *core*-Enzym (PEP-B) (Pfannschmidt and Link, 1997, 1994). Es wurde daraufhin vermutet, dass PEP-B den minimalen Transkriptionskomplex aus dem *core*-Enzym und einem Sigma-Faktor bildet (Lysenko and Kuznetsov, 2005). Nach neuesten Forschungsergebnissen (Ji *et al.*, 2020) wurde jedoch ein voll assemblierter PEP-Komplex in Proplastiden und Etioplasten von *A. thaliana* nachgewiesen, welches die Transkription von PEP-abhängigen Genen in Dunkelheit und sehr früher Samenentwicklung ermöglicht.



### I.3.2 Kernkodierte Phagentyp-RNA-Polymerasen (NEP)

Im Gegensatz zu prokaryotischen und eukaryotischen RNA-Polymerasen bestehen die RNA-Polymerasen von T3/T7-Bakteriophagen aus einem einzelnen Polypeptid und können Promotorstrukturen ohne Hilfe von zusätzlichen Faktoren erkennen und Transkription initiieren (Cheetham and Steitz, 2000). Kernkodierte organelläre RNA-Polymerasen sind diesen T3/T7-Phagenpolymerasen ähnlich (Weihe *et al.*, 2012). Die ersten Hinweise dafür, dass neben einer Eubakterientyp-RNA-Polymerase PEP eine zusätzliche Polymeraseaktivität existiert, wurden geliefert, als in den Plastiden von ribosomenlosen Reis- und Gerstenblättern Transkription nachgewiesen werden konnte (Hess *et al.*, 1993). Es wurde vermutet, dass eine weitere plastidär lokalisierte RNA-Polymerase existiere, deren Gene im Kern kodiert sein müssen. In Tabak wurden in Abwesenheit einer Untereinheit der eubakteriellen RNA-Polymerase vermehrt *housekeeping*-Gene transkribiert, wobei diese alternative Transkriptionsaktivität nicht die typischen *E. coli*  $\sigma^{70}$ -Promotoren genutzt hat (Allison *et al.*, 1996). Aus Spinat wurde ein 110 kDa großes Polypeptid isoliert, welches eine RNA-Polymeraseaktivität besaß und den T7-Bakteriophagen-Promotor erkennen konnte (Lerbs-Mache, 1993). Die Nukleotidsequenz dieser identifizierten RNA-Polymerase zeigte eine hohe Ähnlichkeit zu der Nukleotidsequenz der Phagentyp-RNA-Polymerasen (Hedtke *et al.*, 1997).

Phagentyp-RNA-Polymerasen werden von einer Familie von kernkodierten *RPOT*-Genen repräsentiert. In dikotylen Pflanzen wie *A. thaliana* findet man drei Vertreter dieser Familie: *RPOTp*, *RPOTm*, *RPOTmp*. *RPOTp* wird zu den Plastiden geleitet, *RPOTm* wird in Mitochondrien importiert und *RPOTmp* in beide Organellen importiert. (Abb. I.2) (Azevedo *et al.*, 2006; Hedtke *et al.*, 2000). Den bis jetzt untersuchten Gräsern (*Oryza sativa*, *Zea mais*, *Hordeum vulgare*, *Triticum aestivum*) fehlt das in beide Organellen importierte *RPOTmp*-Enzym und es wird angenommen, dass dieses für alle Monokotyledonen zutreffen könnte (Liere *et al.*, 2011). Phylogenetische Analysen zeigen, dass die beiden plastidär lokalisierten Phagentyp-RNA-Polymerasen (*nuclear-encoded plastid RNA polymerase*, NEP) durch eine Genduplikation der mitochondrialen Phagentyp-RNA-Polymerase entstanden sind. Diese Genduplikation fand nach der Trennung der Angiospermen von den Gymnospermen statt (Yin *et al.*, 2010). Untersuchungen zur Expression der NEPs stellen fest, dass alle Phagentyp-RNA-Polymerasen eine sehr niedrige Expression in allen Geweben aufweisen (Emanuel *et al.*, 2006). Während *RPOTp* die maximale Expression in grünem Gewebe hat, ist die Expression von *RPOTmp* in Meristemen und jungen Zellen ohne grüne Plastiden am höchsten. Von allen Phagentyp-RNA-Polymerasen zeigte *RPOTm* die höchste Expression in allen Gewebetypen außer in ausdifferenzierten Blättern (Emanuel *et al.*, 2006).

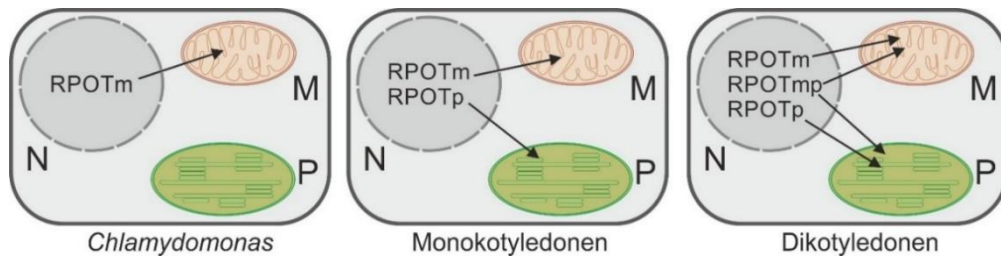


Abbildung I.2: Lokalisierung der Phagentyp-RNA-Polymerasen. Die Proteine werden im Nucleus (N) kodiert und nach der Translation im Cytoplasma in Mitochondrien (M) oder Plastiden (P) oder in beide Organellen importiert. Der Import ist durch die Pfeile angezeigt. Algen wie *Chlamydomonas* besitzen nur die mitochondrienlokalisierte RNA-Polymerase RPOtm. Gräser und vermutlich alle Monokotyledonen besitzen zwei Gene für Phagentyp-RNA-Polymerasen, welche jeweils in Mitochondrien bzw. Plastiden aktiv sind. Die Dikotyledonen haben eine zusätzliche Phagentyp-RNA-Polymerase RPOtmp, welche in beide Organellen importiert wird. Modifiziert nach Liere *et al.*, 2011.

Die NEP-Aktivität in Plastiden dikotyler Pflanzen wird durch RPOtp und RPOtmp repräsentiert. Die Anwesenheit beider Enzyme in Plastiden wurde durch spezifische Antikörper nachgewiesen (Azevedo *et al.*, 2008). Der Zusammenhang zwischen RPOtp und der NEP-Aktivität wurde in Überexpressionstudien von RPOtp in Tabak und *Arabidopsis* deutlich (Liere *et al.*, 2004). Hier wurde die erhöhte Nutzung von NEP-Promotoren nachgewiesen und RPOtp als eine NEP-Aktivität identifiziert. Eine *rpoTp-Knockout*-Mutante zeigt einen Phänotyp mit starker Wachstumshemmung und chlorotischen Blättern, die gezackt sind (Hricová *et al.*, 2006). Der Chlorophyllgehalt ist im Vergleich zum Wildtyp reduziert. Eine drei Wochen alte Mutante besitzt weniger und kleinere Chloroplasten, die keine Stärke akkumulieren und weniger ausgeprägte Thylakoidstrukturen zeigen, als der Wildtyp (Hricová *et al.*, 2006). Im Gegensatz dazu zeigt eine *rpoTmp-Knockout*-Mutante nur leichte phänotypische Veränderungen, das Wachstum ist im Vergleich zum Wildtyp verzögert und die Pflanzen zeigen eine Blattmorphologie mit eingedrehten Blättern (Baba *et al.*, 2004; Kühn *et al.*, 2009).

### I.3.3 Zusätzliche Phagentyp-RNA-Polymerase in Dikotyledonen

Dikotyle Pflanzen besitzen eine weitere Phagentyp-RNA-Polymerase RPOtmp in Plastiden, die neben dem Import in Plastiden zusätzlich in Mitochondrien importiert wird (Azevedo *et al.*, 2006; Hedtke *et al.*, 2000). Da eine Mutante mit fehlender RPOtmp-Aktivität eine verminderte Abundanz der respiratorischen Komplexe I und IV zeigte, wurde RPOtmp eine Funktion bei der Transkription der Gene dieser Komplexe in Mitochondrien zugesprochen (Kühn *et al.*, 2009). RPOtm wird als zentrale RNA-Polymerase betrachtet, da eine Mutante mit fehlender RPOtm-Aktivität lethal ist (Kühn *et al.*, 2009). Eine Komplementation der *rpoTm*-Mutante mit *RPOtmp* unter der Kontrolle eines *RPOtm*-Promotors konnte den lethalen Phänotyp nicht umkehren (Tan *et al.*, 2010). Auch im Plastiden konnte RPOtmp bis jetzt keine wichtige Funktion zugeordnet werden und es wurde nur ein Promotor charakterisiert, welcher von RPOtmp als Transkriptionsstartstelle genutzt wird (*Prrn16-139*, Pc-Promotor) (Courtois *et al.*, 2007) und ein weiterer Promotor *PclpP-58* mit potentieller RPOtmp-Abhängigkeit beschrieben (Liere *et al.*, 2004; Swiatecka-Hagenbruch *et al.*, 2008) (Kap. I.4.1). Da jedoch die meisten NEP-Promotoren auch in Abwesenheit von RPOtp genutzt werden, lässt sich der Schluss ziehen, dass auch RPOtmp in die plastidäre Transkription involviert ist (Swiatecka-

Hagenbruch et al., 2008). Bei Abwesenheit einer der beiden plastidären Phagentyp-RNA-Polymerasen wurde eine erhöhte Expression des Gens für die jeweils andere Polymerase beobachtet (Baba *et al.*, 2004; Hricová *et al.*, 2006), was auf eine überlappende Rolle dieser Polymerasen in der Transkription hindeuten könnte (Tarasenko et al., 2016). Durch die erhöhte Expression des intakten Gens könnte die Pflanze versuchen, das Fehlen der anderen Polymerase auszugleichen, um die für die korrekte Plastidentwicklung nötigen Transkriptlevel einiger Gene zu gewährleisten. Komplementationstudien der *rpoTnp*-Mutante geben Hinweise darauf, dass RPOTmp eine wichtige Rolle in Mitochondrien, aber nicht in Plastiden spielt (Tarasenko et al., 2016). Hier führte die Komplementation der RPOTmp-Aktivität in Plastiden zur Wiederherstellung der Transkription des *rrn16*-Operons vom RPOTmp-abhängigen Pc-Promotor. Der Phänotyp der *rpoTnp*-Mutante wurde jedoch durch die Komplementation nicht geändert (Tarasenko et al., 2016). Die Komplementation der *rpoTnp*-Mutante mit der mitochondrialen Zielsequenz dagegen führte zu einem wildtypischen Aussehen der Pflanze. Auch die mitochondriale Transkription in dieser Komplementationslinie entsprach dem Wildtyp (Tarasenko *et al.*, 2016). Welchen Vorteil das Vorkommen der zusätzlichen Phagentyp-RNA-Polymerase RPOTmp für die Pflanze hat, konnte bis jetzt nicht eindeutig geklärt werden. Der Studie von Hricová *et al.* (2006) zufolge zeigt eine Doppelmutante, welcher sowohl RPOTp wie auch RPOTmp fehlen, chlorotische Kotyledonen und ist nach dem Keimlingsstadium in ihrer Entwicklung gehemmt. Das wurde von den Autoren als Hinweis auf eine teilweise überlappende Funktion der beiden Phagentyp-RNA-Polymerasen interpretiert.

### I.4 Plastidäre Promotoren

Nach Hajdukiewicz *et al.*, 1997 werden plastidäre Gene bzw. Operons in drei Klassen eingeteilt: Klasse I-Gene/Operons haben PEP-Promotoren, Klasse II beinhaltet sowohl NEP- als auch PEP-Promotoren und in Klasse III sind Gene/Operons, die ausschließlich NEP-Promotoren aufweisen, eingeteilt. Diese Einteilung wurde jedoch nur mit wenigen bis dahin identifizierten Promotoren gemacht und heute ist bekannt, dass die meisten Gene sowohl NEP- als auch PEP-Promotorstrukturen aufweisen (Krause *et al.*, 2000; Legen *et al.*, 2002; Liere *et al.*, 2011; Zhelyazkova *et al.*, 2012b). Einige wenige Gene konnten als Klasse III-Gene identifiziert werden. So gelten die Gene *ycf2*, *accD* und *rpoB* als exklusiv von NEP transkribiert (Hajdukiewicz *et al.*, 1997; Hess *et al.*, 1993; Swiatecka-Hagenbruch *et al.*, 2008, 2007). Als exklusiv PEP-abhängige Gene (Klasse I) werden vor allem Gene für die Photosynthese, wie *psbA*, betrachtet (Liere *et al.*, 1995; Shen *et al.*, 2001; Swiatecka-Hagenbruch *et al.*, 2007; Zhelyazkova *et al.*, 2012b). Die meisten plastidären Gene und Operons haben multiple Promotoren, Gene mit nur einem Promotor sind eher selten (Liere and Börner, 2007).

#### I.4.1 NEP-Promotoren

NEP-Promotoren haben eine Länge von etwa 20 bp und umfassen den Transkriptionsstartpunkt (TSS) (Liere *et al.*, 2011; Liere and Maliga, 1999). Die bisher charakterisierten NEP Promotoren wurden in

drei Gruppen unterteilt (Abb. I.3 B). Der ersten Gruppe (Typ I) ist ein konserviertes YRTa-Motiv gemeinsam (Y= T/C, R=A/G) (Börner *et al.*, 2014; Liere and Maliga, 1999). Dieses Motiv ist für die Promotorerkennung entscheidend und befindet sich stromaufwärts (-15 bis +5) des TSS, welcher die Position +1 zugewiesen wird. Die Transkriptionsstartpunkte (*transcription start site*, TSS) werden im Folgenden so bezeichnet: P für Promotor gefolgt vom Gennamen und der Position des initiiierenden Nukleotids im Verhältnis zum Translationsstartpunkt oder zur reifen rRNA. Als Typ Ia-NEP-Promotoren mit YRTa-Motiv wurden in *Arabidopsis* *PaccD-252*, *Prps4-151*, *PatpB-318* und *Pycf1-104* identifiziert (Swiatecka-Hagenbruch *et al.*, 2007). Einige NEP-Promotoren zeigen zusätzlich zu dem YRTa-Motiv eine zweite konservierte Sequenz, die GAA-Box (ATAN<sub>0,1</sub>GAA), die sich etwa 18-20 bp stromaufwärts des YRTa-Motivs befindet, und werden dem Typ Ib zugeordnet (Kapoor and Sugiura, 1999; Yagi and Shiina, 2014). In *Arabidopsis* werden zu Typ Ib die Promotoren *PaccD-172*, *PrpoB-300*, *Prps15-139* und *Pycf1-39* (Swiatecka-Hagenbruch *et al.*, 2007) gezählt. Das Vorkommen diesen Promotortyps kann jedoch nicht für alle Pflanzenarten bestätigt werden und es wird vermutet, dass Monokotyledonen Typ Ib-Promotoren nicht besitzen, da eine genomweite Analyse der TSS in Gerste keine GAA-Motive identifiziert hat (Zhelyazkova *et al.*, 2012b). Die Rolle dieses Motivs ist noch nicht eindeutig geklärt, da bis jetzt nur für das *PatpB-289* in Tabak eine Funktionalität der GAA-Box gezeigt wurde (Xie and Allison, 2002).

Dem Typ II der NEP-Promotoren fehlen sowohl das YRTa-Motiv als auch das GAA-Motiv und sie werden als „Nicht-Konsensus“ NEP-Promotoren bezeichnet, da bisher keine Gemeinsamkeiten zwischen ihnen gefunden werden konnten (Kapoor *et al.*, 1997; Kapoor and Sugiura, 1999; Sriraman *et al.*, 1998). Zu diesem Typ werden die gut charakterisierten Promotoren *PclpP-58* (in *Arabidopsis*, in *Nicotiana PclpP-53*) und *Prn16-139* (Pc-Promotor) gezählt (Swiatecka-Hagenbruch *et al.*, 2007). Speziell der Pc-Promotor, welcher einen von zwei Promotoren des *rrn16*-Operons darstellt, ist ein interessantes Beispiel für einen Promotor des Typ II, welcher explizit von RPOTmp erkannt wird. In *Arabidopsis* scheint dieser Promotor während der Imbibierung und frühen Samenkeimung aktiv zu sein (Courtois *et al.*, 2007). Es wird vermutet, dass RPOTmp speziell die Typ II-Promotoren erkennt (Swiatecka-Hagenbruch *et al.*, 2008) und so auch für die Transkription von *clpP* verantwortlich ist. *PclpP-58* ist in Mono- und Dikotyledonen sowie in Koniferen und Lebermoosen konserviert, wird in Reis und *Chlamydomonas* jedoch nicht zur Transkription verwendet (Sriraman *et al.*, 1998). Wird die *clpP*-Promotorsequenz aus Reis in Tabakplastiden eingeführt, kann diese von der Tabak-NEP erkannt werden (Sriraman *et al.*, 1998). Da sowohl Monokotyledonen als auch *Chlamydomonas* die RPOTmp-Aktivität fehlt (Kap. I.3.2), wird das als Grund für die fehlende Nutzung des *PclpP-58* in diesen Arten angeführt (Liere *et al.*, 2011). Untersuchungen von Mutanten mit fehlender RPOTp-Aktivität oder erhöhter RPOTp-Aktivität stellten fest, dass die Nutzung des *clpP*-Promotors nicht verändert ist (Liere *et al.*, 2004; Swiatecka-Hagenbruch *et al.*, 2008), was wiederum für die Beteiligung von RPOTmp an der Transkription von *clpP* spricht. Gegen die Vermutung, dass RPOTmp *clpP* transkribiert, sprechen

Untersuchungen des Transkriptoms der *rpoTmp*-Mutante, welche keine Reduktion der *clpP*-Transkriptmengen feststellen (Baba *et al.*, 2004; Kühn *et al.*, 2009; Tarasenko *et al.*, 2016). Ergebnisse anderer Studien lassen vermuten, dass *clpP* von RPO<sub>TP</sub> transkribiert wird (Courtois *et al.*, 2007). In *in-vitro*-Transkriptionsassays konnte weder für RPO<sub>TP</sub> noch für RPO<sub>TP</sub> eine Transkription vom *clpP*-Promotor nachgewiesen werden (Kühn *et al.*, 2007), sodass die Frage, welcher Phagentyp-Polymerase *clpP* transkribiert, weiterhin offen bleibt.

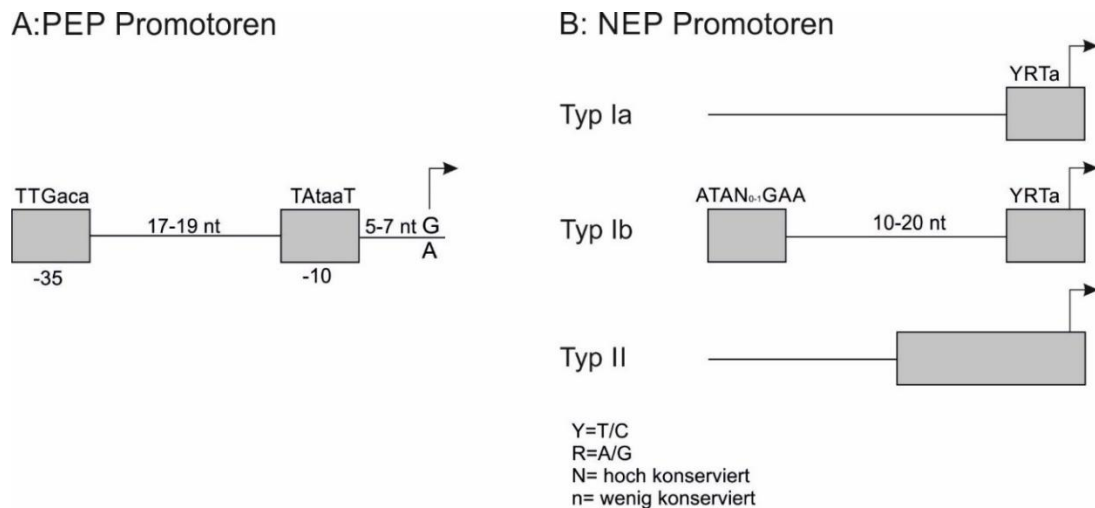


Abbildung I.3: Aufbau der plastidären Promotoren, nach Börner *et al.*, 2014; Liere & Börner, 2007. A: Schematischer Aufbau von PEP-Promotoren. B: Schematischer Aufbau von NEP-Promotoren. Der Transkriptionsstartpunkt ist durch den gebogenen Pfeil dargestellt. Die konservierten Motive sind als graue Boxen dargestellt, die konservierte Nukleotidsequenz ist über den Boxen abgebildet. Nt: Nukleotide.

#### I.4.2 PEP-Promotoren

Der PEP-Komplex erkennt die für Eubakterien typischen Promotormotive des  $\sigma^{70}$ -Typs. Die Konsensusmotive befinden sich stromaufwärts der TSS und werden als -35 (5'-TTGaca-3') und -10 (5'-TataaT-3')-Boxen bezeichnet (Abb. I.3 A) (Liere and Maliga, 2001; Shiina *et al.*, 2005; Yagi and Shiina, 2014). Als typische PEP-Promotoren können *PpsbA-77*, *PpsaA-188*, *Prrn16-112*, *PatpB-520* (Swiatecka-Hagenbruch *et al.*, 2007) und *PatpH-413* (Zghidi *et al.*, 2007) genannt werden. Da die in Gerste identifizierten Promotorregionen nur zum Teil einen oder beide der Konsensusmotive aufwiesen (Zhelyazkova *et al.*, 2012b), wird vermutet, dass die beiden Elemente für einen funktionellen plastidären Promotor nicht essentiell sein könnten (Börner *et al.*, 2014).

Für die Promotorerkennung benötigt der PEP-Komplex einen Sigma-Faktor, welcher den Promotor erkennen und die Aufspaltung der doppelsträngigen DNA an der Promotorregion initiieren kann (Davis *et al.*, 2017). Die Regulation der Sigma-Faktoren kann unter anderem über die Phosphorylierung erfolgen, hierbei kann die Phosphorylierung sowohl einen positiven wie auch einen negativen Effekt auf die Aktivität des Sigma-Faktors und die Transkription haben (Schweer *et al.*, 2010). Die Sigma-Faktoren haben entwicklungspezifische sowie genspezifische Funktionen. Die Faktoren SIG2 und SIG6 sind vor allem an der Transkription während der frühen Entwicklung des Chloroplasten beteiligt (Hanaoka

*et al.*, 2005; Ishizaki *et al.*, 2005; Loschelder *et al.*, 2006). SIG2 scheint für die Transkription der tRNA-Gene, jedoch nicht für Photosynthesegene verantwortlich zu sein (Kanamaru *et al.*, 2001), während SIG6 in der frühen Chloroplastenentwicklung die Transkription von Photosynthesegenen steuert (Ishizaki *et al.*, 2005; Ji *et al.*, 2020). Für SIG1 und SIG5 gibt es Hinweise, dass sie eher in reifen Plastiden die Genexpression regulieren. SIG5 wird besonders unter Stressbedingungen exprimiert und erkennt den Blaulicht-induzierten Promotor des *psbD*-Operons (Tsunoyama *et al.*, 2004). Für SIG3 und SIG4 wird angenommen, dass beide eine allgemeine Rolle haben (Ortelt and Link, 2014; Schweer, 2010).

#### I.4.3 Funktion der RNA-Polymerasen während der Entwicklung der Pflanze

Zunächst wurde angenommen, dass die NEPs die prädominante Polymeraseaktivität in frühen, nicht-grünen Stadien der Pflanzenentwicklung darstellen und Komponenten des Translationsapparats und die Gene für die PEP-Untereinheiten transkribieren, während die PEP-Aktivität im Prozess der Chloroplastenentwicklung zunimmt und in grünem Gewebe die prädominante Transkriptionsaktivität bildet, in dem es vor allem Photosynthesegene transkribiert (Liere & Maliga, 2001) (Abb. I.4).

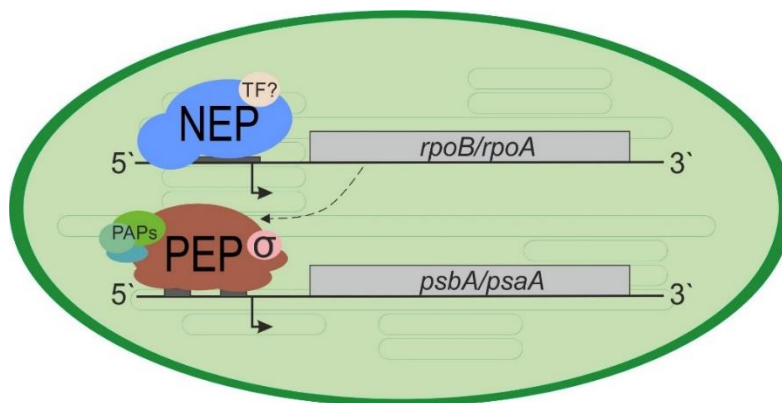


Abbildung I.4: Gegenwärtiges Modell der Funktionen der plastidenlokalisierten RNA-Polymerasen. PEP: Plastidenkodierte RNA-Polymerase, NEP: Kernkodierte RNA-Polymerase, hier RPOTp und RPOTmp. NEPs sind vor allem in die Transkription von *housekeeping*-Genen, die für die plastidkodierte Untereinheiten der PEP kodieren (*rpoA*, *rpoB*), involviert, während PEP vor allem Photosynthesegene transkribiert (*psbA*, *psaA*). Transkriptionsfaktoren (TF) für NEP sind nicht bekannt. PEP erfordert zusätzliche Faktoren (PEP-assoziierte Proteine. PAPs) sowie Sigma-Faktoren zur korrekten Promotorerkennung und Transkription. Modifiziert nach Liere *et al.*, 2011.

Mittlerweile gibt es Hinweise darauf, dass PEP und die NEPs parallele Transkriptionsaktivitäten aufweisen (Pfannschmidt *et al.*, 2015). Beide Arten von RNA-Polymerasen sind als mRNA und als Proteine bereits im trockenen Samen nachweisbar (Demarsy *et al.*, 2006) und alle drei RNA-Polymerasen sind im imbibierten Samen aktiv (Demarsy *et al.*, 2012). Weiterhin scheint eine PEP-Aktivität für die effiziente Keimung notwendig zu sein (Demarsy *et al.*, 2006). Da einige tRNA-Gene von PEP transkribiert werden (Börner *et al.*, 2014; Kanamaru *et al.*, 2001; Liere and Maliga, 2001; Williams-Carrier *et al.*, 2014), könnte PEP eine wichtige Rolle in der frühen Pflanzenentwicklung haben. Während der Entwicklung der Blätter sind beide Arten von Polymerasen aktiv, und PEP- sowie NEP-Promotoren werden in Kotyledonen und Rosettenblättern genutzt (Zoschke *et al.*, 2007). Auch

der kürzliche Nachweis eines aktiven PEP-Komplexes in Proplastiden und Etioplasten (Ji *et al.*, 2020) deutet auf eine parallele Aktivität beider Polymerasenarten hin und unterstützt die Rolle von PEP in der sehr frühen Samenentwicklung. Da in Gerste für einige Photosynthesegene auch NEP-Promotoren gefunden wurden (Zhelyazkova *et al.*, 2012b), könnte RPOTp auch in späten Entwicklungsstadien der Pflanze an der Transkription der Photosynthesegene beteiligt sein. Auch in dikotylen Pflanzen wurden für Photosynthesegene wie *atpB* und *atpI* NEP-Promotoren identifiziert (Swiatecka-Hagenbruch *et al.*, 2007). In vier Tage alten Keimlingen der *rpoTp*-Mutante wurde die Reduktion der Transkripte von *clpP*, *rpoB*, *rpoC1*, *accD* und *rps18* nachgewiesen (Hricová *et al.*, 2006). Dies weist auf die wichtige Rolle von RPOTp in früher Pflanzenentwicklung hin und gibt Anlass zur Vermutung, dass *rpoB* und *rpoC1* durch RPOTp transkribiert werden (Hricová *et al.*, 2006).

### I.5 Methoden zur Identifikation von Transkriptionsstarts

Bei den Methoden zur Identifizierung von Transkriptionsstarts in Organellen unterscheidet man zwischen Methoden, die zu einer gezielten Analyse von TSS führen, und Methoden, die eine genomweite Analyse der TSS ermöglichen. Die Methoden basieren auf dem Prinzip der Identifizierung des 5'-Endes eines Primärtranskripts. Ein Primärtranskript trägt ein 5'-Triphosphat und stellt das erste Produkt der Transkription, die Prä-mRNA, dar und wird zur reifen mRNA prozessiert, welche einen 5'-Monophosphat trägt. Da die bakteriellen und organellären Primärtranskripte die TSS umfassen, lassen sich vom 5'-Ende die TSS sowie mögliche Promotorregionen ableiten.

*Primer Extension* ist eine der Methoden zur gezielten Analyse von 5'-Enden und wird ebenfalls dazu benutzt, um die Menge einer gegebenen mRNA abzuschätzen (Boorstein and Craig, 1989; Carey *et al.*, 2013; Triezenberg, 1992). Die 5'-Enden von Transkripten werden bestimmt, indem ein, meist radioaktiv markierter, komplementärer DNA-Primer an die mRNA-Fragmente angelagert und eine reverse Transkription bis zum 5'-Ende der entsprechenden RNA durchgeführt wird. Anschließend können die Fragmente elektrophoretisch aufgetrennt und mittels einer Autoradiographie visualisiert werden (Boorstein & Craig, 1989; Triezenberg, 1992). Da bei dieser Methode nicht zwischen primären und prozessierten Transkripten unterschieden wird, kann sie mit *5'RACE* (*5'-rapid amplification of cDNA ends*) kombiniert werden, um die tatsächliche TSS zu bestimmen (Ghulam *et al.*, 2012). Bei *5'RACE* wird der Unterschied zwischen einem Primärtranskript mit 5'-Triphosphat und einem prozessierten Transkript mit 5'-Monophosphat ausgenutzt (Bensing *et al.*, 1996; Matteau and Rodrigue, 2015; Scotto-Lavino *et al.*, 2007). Da nur der letztere an RNA-Oligonukleotide ligiert und anschließend ein komplementärer Primer angelagert werden kann, wird nur das prozessierte Transkript in cDNA umgeschrieben und in der nachfolgenden PCR amplifiziert. Die Primärtranskripte können nur nach einer Behandlung mit TAP (*Tobacco Acid Pyrophosphatase*) ligiert werden, sodass nach einer TAP-Behandlung sowohl primäre wie auch prozessierte Transkripte detektiert werden. Ohne TAP-Behandlung werden nur prozessierte Transkripte identifiziert. Der Vergleich der +TAP und -TAP

Proben gibt Aufschluss über das Primärtranskript, welches die TSS enthält. Ein weiterer Vorteil dieser Methode gegenüber *Primer Extension* ist die anschließende Möglichkeit der Sequenzierung der Produkte, die Methode ist im Vergleich zu *Primer Extension* jedoch nicht quantitativ. *In-vitro-capping* kann ebenfalls zur Identifizierung von organellären TSS verwendet werden. Die 5'-Triphosphate der Primärtranskripte werden mittels einer Guanylyltransferase mit einem radioaktiv markierten 5'-Cap versehen. Nachdem die markierten Transkripte mit komplementären RNA-Sonden hybridisiert und einzelsträngige RNAs mit einer RNase verdaut werden, können die 5'-Enden der ausgewählten Transkripte identifiziert werden (Vera and Sugiura, 1992).

Die beschriebenen Methoden sind durch die obligate Verwendung eines transkriptspezifischen Primers limitiert und so können nur für prädefinierte Regionen der Transkripte TSS identifiziert werden. Der Abbruch der reversen Transkription, welcher bei diesen Methoden oft beobachtet wird, führt zu einem hohen Rauschen, was die Identifizierung des gesuchten Transkripts erschwert (Matteau and Rodrigue, 2015; Triezenberg, 1992). Es wurde ebenfalls berichtet, dass die genutzten Reversen Transkriptasen zusätzliche Nukleotide im Syntheseprozess an die cDNA anhängen können, was die Identifizierung des 5'-Endes verfälschen könnte (Chen and Patton, 2001).

So sind auf einer Hochdurchsatzsequenzierung basierte genomweite Analysen besser dazu geeignet, um eine quantitative simultane Untersuchung einer großen Anzahl von TSS zu ermöglichen und Unterschiede in der Nutzung dieser z.B. in Mutanten vergleichen zu können, da sie eine größere Auflösung und höhere Sensitivität zeigen. Die erste genomweite Analyse der Transkriptionsstartseiten wurde mittels der differentiellen Hochdurchsatzsequenzierung (*differential RNA-Seq*, dRNA-Seq) für Bakterien durchgeführt (Sharma *et al.*, 2010). Bei dieser Methode werden die Primärtranskripte angereichert, indem die prozessierten Transkripte mittels einer Terminator-Exonuklease (+TEX) degradiert werden. Als Hintergrundbibliothek dient eine unbehandelte Probe (-TEX). Beide Proben müssen mit TAP behandelt werden, damit bei der Herstellung der cDNA-Bibliothek der 5'-Adapter ligiert werden kann. Die sequenzierten cDNA-Bibliotheken werden anschließend auf das Genom gemappt und der Vergleich der +TEX-Bibliothek und der -TEX-Bibliothek lässt die TSS identifizieren. Für eine TSS sieht man eine Anreicherung der 5'-Enden in der +TEX-Probe, wobei die -TEX-Probe eine geringe bis fehlende Genomabdeckung an der gleichen Genomposition aufweist (Abb. I.5 A). Die Methode der dRNA-Seq wurde bereits erfolgreich für pflanzliche Proben angewendet und für die Identifizierung von PEP- und NEP-abhängigen TSS in Gerste genutzt (Zhelyazkova *et al.*, 2012b).



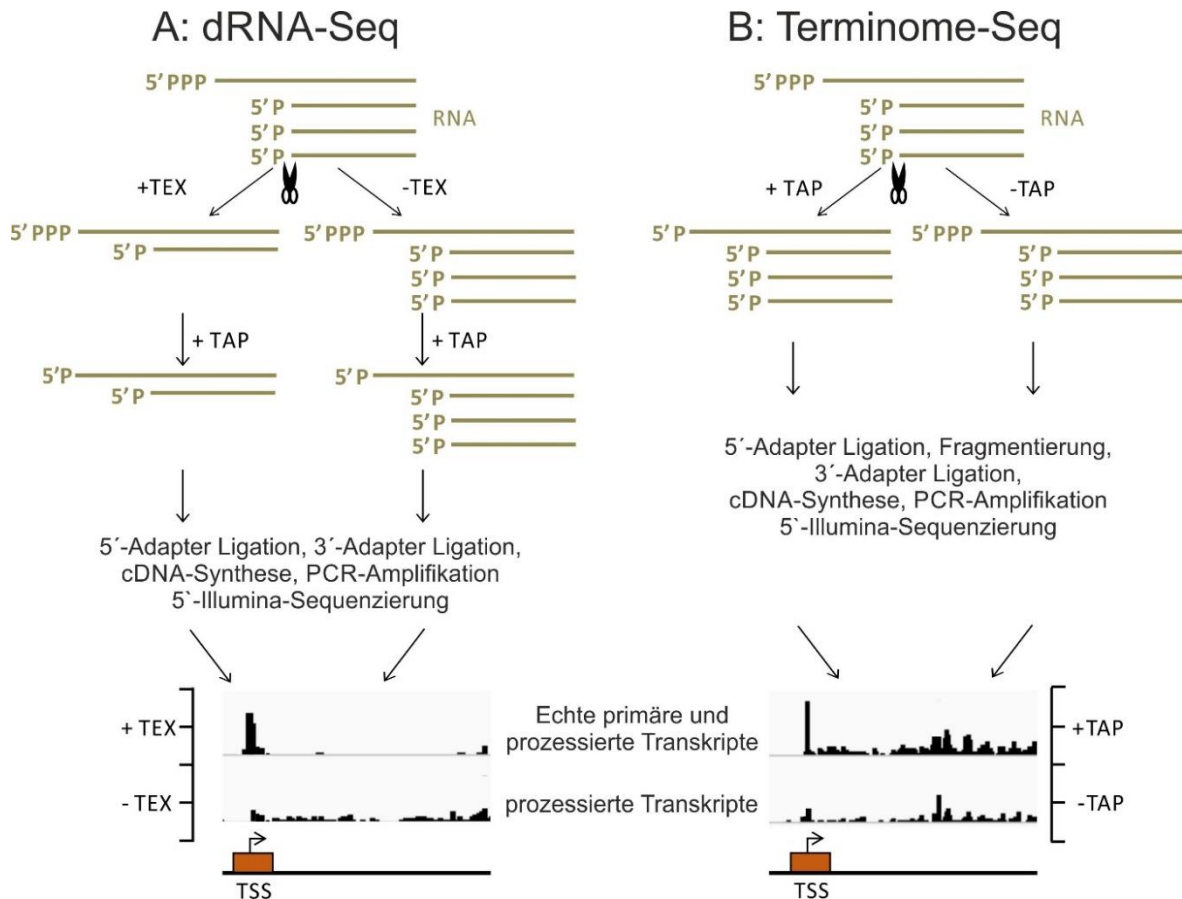


Abbildung I.5: Schematische Darstellung der Protokolle für dRNA-Seq und Terminome-Seq. Das Scherensymbol zeigt die Prozessierungsstelle des prozessierten Transkripts an. Die braune Box zeigt die Promotorregion mit der TSS, welche als gebogener Pfeil dargestellt ist. A: Bei dRNA-Seq (Sharma *et al.*, 2010) werden die Primärtranskripte angereichert, da die prozessierten Transkripte mittels Exonuklease TEX degradiert werden. Nach der Konvertierung der 5'-Triphosphate mit TAP, damit Adapter an die Transkripte ligiert werden können, werden die cDNA-Bibliotheken hergestellt (+TEX). Eine nicht mit TEX behandelte Probe wird als Kontrolle genutzt (-TEX). B: Bei Terminome-Seq (Castandet *et al.*, 2019) werden die Transkripte mit TAP konvertiert und nach einem modifizierten Protokoll mit einer Fragmentierung cDNA-Bibliotheken hergestellt (+TAP). Hier stellt die Kontrolle eine nicht mit TAP behandelte Probe dar (-TAP). Bei beiden Methoden wird der Vergleich beider cDNA-Bibliotheken für die TSS-Identifizierung genutzt. TSS: Transkriptionsstartpunkt, TEX: Terminator-Exonuklease, TAP: *Tobacco Acid Pyrophosphatase*.

Einen alternativen Ansatz zur genomweiten Kartierung der TSS stellt die Terminome-Seq dar (Castandet *et al.*, 2019). Im Gegensatz zur dRNA-Seq wird hier keine direkte Anreicherung der Primärtranskripte durchgeführt (Abb. I.5 B). Eine Probe wird mit TAP behandelt und so die 5'-Triphosphate der Primärtranskripte konvertiert, die nun an Adapter ligiert werden können. Die aus dieser Probe hergestellte Bibliothek enthält die primären und die prozessierten Transkripte (+TAP). Die Hintergrundbibliothek ist eine nicht mit TAP behandelte Probe (-TAP), welche nur die prozessierten Transkripte enthält. Wie bei dRNA-Seq zeigt der Vergleich der beiden Bibliotheken die TSS an.

Spezielle Varianten der RNA-Seq zur TSS-Kartierung sind Analysen von naszierender RNA, um auch Primärtranskripte zu identifizieren, die durch die schnelle Prozessierung in Proben mit Gesamt-RNA nicht detektierbar sind. Die Methode des GRO-Seq (*Global Run-On-Sequencing*) wird dazu genutzt, und die Transkriptionsaktivität sowie die Position und Orientierung von RNA-Polymerasen bestimmen (Core *et al.*, 2008). Hier wird nach der Isolation von Nuklei in Anwesenheit von markierten

Nukleotiden eine *in-vitro*-Transkription durchgeführt; die markierten Transkripte werden anschließend über eine Affinitätsaufreinigung für die im Laufe des *Run-Ons* transkribierten Transkripte angereichert. Die Transkripte können nach der Sequenzierung auf das Kerngenom gemappt und so die globale Transkription dargestellt werden. Weitere Methoden, die auf dem gleichen Prinzip beruhen und für Nuklei adaptiert worden sind, sind NET-Seq (*native elongating transcript sequencing*) (Churchman and Weissman, 2011) und PRO-Seq (*precision nuclear Run-On and sequencing*) (Kwak *et al.*, 2012). Auf der GRO-Seq-Methode basieren die Methoden des GRO-Cap (Kruesi *et al.*, 2013) und 5'-GRO-Seq (Hetzl *et al.*, 2016), welche speziell die Identifizierung der 5'-Enden von naszierenden Transkripten ermöglichen, von denen sich die TSS ableiten lassen und eine genomweite Analyse der genutzten TSS möglich ist. Diese beiden Methoden nutzen eine etwas andere Strategie als Sharma *et al.* (2010) (Abb. I.5 A), um Primärtranskripte anzureichern und die prozessierten Transkripte zu reduzieren. Hier wird die komplette Probe mit TEX behandelt und anschließend geteilt. Für die +cDNA-Bibliothek wird eine Konvertierung der 5'-Triphosphate durchgeführt, während für die -cDNA-Bibliothek (Hintergrundbibliothek) keine Behandlung erfolgt (analog Abb. I.5 B). So enthält die sequenzierte Hintergrundbibliothek TEX-resistente Transkripte, während die +cDNA-Bibliothek primäre sowie TEX-resistente Transkripte enthält. Die Methode des 5'-GRO-Seq ist auch kürzlich für Mitochondrien (Saleh *et al.*, im Druck) und Chloroplasten (Saleh *et al.*, nicht publiziert) etabliert worden.

### I.6 Zielstellung der Arbeit

Die Transkription in Plastiden von dikotylen Pflanzen, bei der eine plastidenkodierte RNA-Polymerase (PEP) und zwei kernkodierte RNA-Polymerasen (RPOTp und RPOTmp) beteiligt sind, ist trotz der geringen Genomgröße hochkomplex. So fehlt bis jetzt eine vollständige Identifizierung der plastidären TSS mit der Charakterisierung dieser bezüglich ihrer Nutzung durch die drei RNA-Polymerasen. Es wurde kürzlich eine Übersicht mit plastidären Transkriptionsstartpunkten für *A. thaliana* vorgestellt, die mehr als 215 TSS zeigt (Castandet *et al.*, 2019), jedoch wurde keine Klassifizierung der TSS unternommen. Die meisten Studien konzentrieren sich nur auf den Vergleich zwischen NEP und PEP, ohne jedoch die beiden Phagentyp-Polymerasen RPOTp und RPOTmp separat zu betrachten. Nur wenige Studien zeigen RPOTmp als zusätzliche Transkriptionsaktivität im Plastiden (Baba *et al.*, 2004; Courtois *et al.*, 2007; Kühn *et al.*, 2009; Tarasenko *et al.*, 2016) und auch dort bleiben die individuellen Rollen von RPOTp und RPOTmp weitestgehend unerforscht. Bis dato gibt es keine Studie, die alle drei plastidären RNA-Polymerasen in ihrer Zusammenarbeit bei der Transkription des Plastoms betrachtet.

Das Ziel dieser Arbeit ist die Identifizierung der Spezifitäten der einzelnen plastidären RNA-Polymerasen PEP, RPOTp und RPOTmp im Hinblick auf die Transkription des Plastoms und der Promotornutzung in frühen Samenstadien und Keimlingen der Pflanze. Der Fokus liegt auf den beiden Phagentyp-Polymerasen RPOTp und RPOTmp, deren Zusammenspiel und individueller Funktion. Da nur zwei Promotoren beschrieben wurden, die RPOTmp-spezifisch sind (Courtois *et al.*, 2007; Liere

*et al.*, 2004; Swiatecka-Hagenbruch *et al.*, 2008), sollte auch die spezifische Rolle dieser RNA-Polymerase in der Transkription des Plastoms untersucht werden. In den grünen Pflanzenstadien sind einige NEP-abhängige Transkripte gering abundant (Liere *et al.*, 2011; Zoschke *et al.*, 2007), sodass zwei Strategien ausgearbeitet wurden, um die niedrig exprimierten NEP-abhängigen Transkripte zu identifizieren. Zum einen sollte neben dem Keimlingsstadium die Transkription in sehr frühen Entwicklungsstadien der Pflanze ohne reife Chloroplasten untersucht werden, in denen Transkripte von Photosynthesegenen noch nicht dominieren. Zum anderen sollten zusätzlich Mutanten mit einer sekundären PEP-Reduktion analysiert werden. Weiterhin sollte die Nutzung der TSS in Wildtyp während der Samenkeimung und Skotomorphogenese untersucht werden, um speziell die Rolle von PEP in der frühen Entwicklung der Pflanze zu identifizieren.

Es sollte eine vollständige genomweite Karte der plastidären Transkriptionsstartseiten erstellt und die Klassifizierung der TSS nach Nutzung durch die jeweilige RNA-Polymerase, PEP, RPOTp oder RPOTmp in *A. thaliana* vorgenommen werden. Um zu diesem Ziel zu gelangen, stand die Optimierung eines Protokolls für eine differentielle Hochdurchsatzsequenzierung (*differential RNA-Seq*, dRNA-Seq) im Vordergrund, welche nicht nur eine qualitative, sondern auch eine quantitative Analyse ermöglicht. Eine ähnliche Methode wurde bereits angewandt, um die TSS in Gerste zu charakterisieren (Zhelyazkova *et al.*, 2012b). Für eine erweiterte Analyse der Transkription und der Nutzung der Transkriptionsstartpunkte sollen auch NEP-Doppelmutanten untersucht werden. Da in einer Publikation eine Doppelmutante gezeigt worden ist (Hricová *et al.*, 2006), wurde der Versuch unternommen, diese durch Kreuzen der Einzelmutanten zu generieren und zu analysieren.

## II MATERIAL UND METHODEN

### II.1 Material und Anbieter

Soweit nicht anders angegeben, wurden alle Puffer und Lösungen mit destilliertem und deionisiertem Wasser (ddH<sub>2</sub>O) hergestellt (Thermo Scientific™ Barnstead™ GenPure™). Sofern nicht anders angegeben, wurden die Chemikalien von Carl Roth, Sigma-Aldrich, Merck und Roche bezogen. Die Liste der Hersteller und Anbieter ist im Anhang Tab. VI.1 aufgeführt.

#### II.1.1 Kits, Enzyme und Oligonukleotide

Die Enzyme wurden von Lucigen und Thermo Fisher Scientific™, die Kits von Thermo Fisher Scientific™ und Qiagen bezogen. Für die Herstellung der cDNA-Bibliotheken wurden die Kits NEBNext® Small RNA Library Prep Set for Illumina® (New England Biolabs) und NEXTFLEX® Small RNA-Seq Kit v3 (PerkinElmer) verwendet. Die Oligonukleotide und Primer wurden bei Biologio bestellt. Der modifizierte 5` Adapter wurde bei Sigma-Aldrich bestellt. Die Liste der Oligonukleotide und Primer befindet sich im Anhang Tab. VI.2.

#### II.1.2 Pflanzenmaterial

Als Wildtyp wurde *A. thaliana* des Ökotyps Columbia (Col-0) verwendet, im Folgenden als Wildtyp (WT) bezeichnet. Die *A. thaliana* T-DNA-Insertionslinien im Col-0-Hintergrund wurden bei Nottingham *Arabidopsis* Stock Centre (NASC) bestellt. Die Insertionslinien (Tab. II.1) wurden mittels Polymerase-Kettenreaktion (PCR) (Kap. II.2.5) und anschließender Gelelektrophorese genotypisiert, wobei die DNA-Extraktion, wie unter II.2.1.4 angegeben, durchgeführt wurde. Die Liste der Primer für die Genotypisierung befindet sich im Anhang Tab. VI.2. Die Insertionsstellen der T-DNA-Insertionen in den einzelnen Linien sind in Abb. II.1 gezeigt.

Tabelle II.1: Liste der verwendeten Mutantenlinien

| Linie                   | T-DNA-Insertion          | betroffenes Gen | Referenz                               |
|-------------------------|--------------------------|-----------------|--|
| <i>rpoTp-2</i>          | SALK_067191              | RPOTp           | Hricova <i>et al.</i> , 2006           |
| <i>rpoTmp-2</i>         | SALK_132842              | RPOTmp          | Kühn <i>et al.</i> , 2009              |
| <i>clb19-3</i>          | SALK_123752              | CLB19           | Chateigner-Boutin <i>et al.</i> , 2008 |
| <i>ptac2</i>            | SALK_075736              | pTAC2           | Pfalz <i>et al.</i> , 2006             |
| <i>rpoTp-2 rpoTmp-2</i> | SALK_067191, SALK_132842 | RPOTmp, RPOTp   | diese Arbeit                           |
| <i>rpoTp-3 rpoTmp-1</i> | SALK_09388, GABI_286E07  | RPOTmp, RPOTp   | diese Arbeit                           |

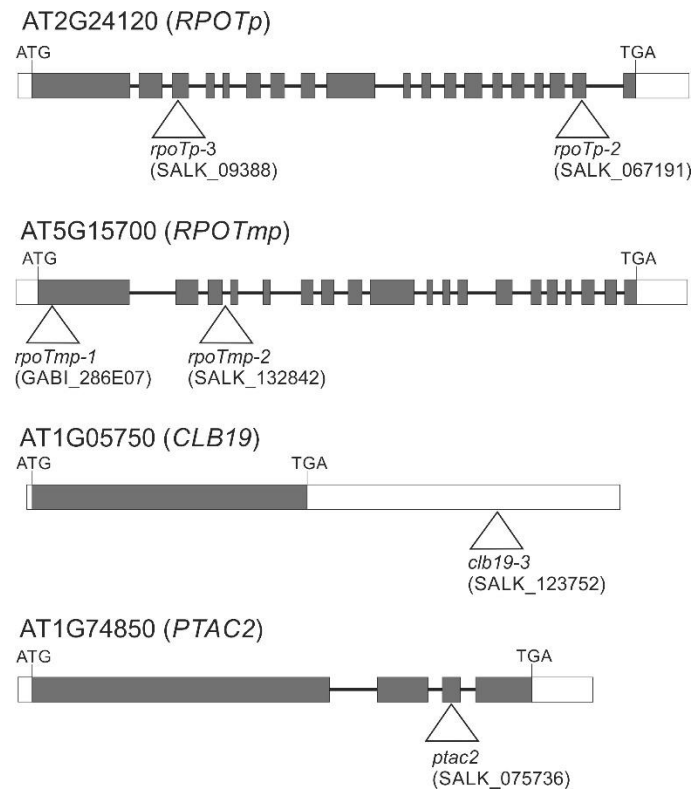


Abbildung II.1: Genkarten mit den Positionen der T-DNA-Insertionen in den Genen. Das Dreieck zeigt die Insertionsstelle der T-DNA an, in Klammern werden die Namen der Linien angegeben. ATG kennzeichnet den Translationsstart des Gens.

### II.1.3 Pflanzenanzucht

Die Untersuchung von Wildtyp und Mutanten erfolgte in zwei verschiedenen Entwicklungsstadien: Ein frühes Stadium, bei dem die Wurzel aus der Samenschale zu 0,5 mm herausragt (im Folgenden als frühes Stadium bezeichnet) und ein spätes Stadium, bei dem die ersten echten Blätter etwa 1 mm lang sind (im Folgenden als spätes Stadium bezeichnet). Als Referenz der Erntezeitpunkte diente die Wachstumsanalyse für *A. thaliana* in Boyes *et al.* (2001). Die Erntezeitpunkte für das frühe und späte Stadium sind in Tab. II.2 abgebildet.

Tabelle II.2: Erntezeitpunkte der keimenden Samen bzw. Keimlinge von Wildtyp und Mutanten. Für das frühe Stadium wurden die Samenplatten in Dauerlicht angezogen, für das späte Stadium unter Langtagbedingungen. Die Zeitmessung wurde nach dem Transfer der Anzuchtplatten ins Licht begonnen. Die Ernte für das späte Stadium erfolgte 5 Stunden nach Einschalten des Lichts. d: Tage, h: Stunden

| Genotyp       | Erntezeitpunkt<br>frühes Stadium | Erntezeitpunkt<br>spätes Stadium |
|---------------|----------------------------------|----------------------------------|
| Wildtyp       | 30 h                             | 7 d                              |
| <i>rpoTp</i>  | 48 h                             | 13 d                             |
| <i>rpoTmp</i> | 52 h                             | 10 d                             |
| <i>clb19</i>  | 40 h                             | --                               |
| <i>ptac2</i>  | --                               | 18 d                             |

Die Pflanzen wurden 5 h nach Einschalten des Lichts geerntet. Die Zeitmessung wurde im Moment des Transfers der Anzuchtplatten in die jeweiligen Lichtverhältnisse begonnen. Das frühe Stadium wurde unter Dauerlicht (24 h Licht) angezogen, während das späte Stadium unter Langtagbedingungen (16 h Licht/8 h Dunkel) angezogen wurde. Die Anzucht erfolgte auf MS-Medium bei 22 °C und 100 µE. Zur Oberflächensterilisation wurden die Samen für 5 min in 70 %igen EtOH mit 0,05 % Triton-X mit gelegentlichem Invertieren inkubiert, anschließend für jeweils 1 min in 70 %igem EtOH und 96 %igem EtOH inkubiert. Die Platten mit ausgebrachten Samen wurden für 72 h bei 4 °C in Dunkelheit stratifiziert und dann in die jeweiligen Lichtbedingungen transferiert.

Für die Generierung der Samen wurden die Pflanzen, wie oben beschrieben, auf Platten unter Langtagbedingungen angezogen und bei Erscheinen der ersten echten Blätter auf ein Gemisch aus Blumenerde und Vermikulit (2:1) umgesetzt. Nach dem Abblühen wurden die Pflanzen mit Papiertüten abgedeckt und eine Woche später nicht mehr gewässert. Nach drei weiteren Wochen ohne Bewässerung wurden die getrockneten Samen abgeerntet. Für die Segregationsanalyse wurden jeweils 100 Samen in Platten auf MS-Medium, wie oben beschrieben, angezogen und nach 14 Tagen, wie in Kap. II.2.2 beschrieben, genotypisiert.

Die potentiellen Doppelmutanten *rpoTp-2 rpoTmp-2* und *rpoTp-3 rpoTmp-1* (Tab. II.1) wurden durch Kreuzungen generiert und genotypisiert. Die Samen wurden, wie oben beschrieben, auf Platten angezogen und für die Untersuchung der Schoten nach 14 Tagen unter Langtagbedingungen auf Blumenerde transferiert und dort bis zur Schotenreife angezogen. Für die elektronenmikroskopische Analyse der Plastiden wurden Wildtyp und Mutanten *rpoTp* (SALK\_067191) und *ptac2* (SALK\_075736) wie oben beschrieben angezogen und im späten Stadium zu den in Tab. II.2 angegebenen Zeiten geerntet. Die *rpoTp*-Mutante für die elektronenmikroskopische Analyse wurde nach 14 Tagen unter Langtagbedingungen auf ein Gemisch aus Blumenerde und Vermikulit (2:1) transferiert und nach weiteren 10 Tagen für die Analyse geerntet.

Für die Untersuchung der frühen Samenentwicklung im Wildtyp wurde keine Oberflächensterilisation durchgeführt. Die untersuchten Entwicklungsstadien umfassten trockene Samen, in Kälte imbibierte Samen (geerntet nach 72 h bei 4 °C in Dunkelheit), etiolierte Keimlinge und das frühe Stadium, bei dem die Wurzel aus der Samenschale zu 0,5 mm herausragt (siehe oben). Für die etiolierten Keimlinge wurden Samen auf Platten ausgebracht, 72 h bei 4 °C stratifiziert, für 2 h bei 22 °C und 100 µE induziert und anschließend 6 Tage im Dunkeln bei 22 °C inkubiert. Die Ernte erfolgte im Dunkeln.

### MS-Medium (pH 5,7):

0,5 % MS

1 % Saccharose

0,05 % MES

0,5 % Agarose für Pflanzenanzucht

### II.2 Methoden

Die Sterilisation von Medien und Lösungen sowie die Inaktivierung von gentechnisch verändertem Pflanzenmaterial erfolgte durch Dampfsterilisation für 20 min bei 120 °C und 55 kPa im Autoklaven Varioklav Greenline 140S (HP Labortechnik).

#### II.2.1 Isolierung von Nukleinsäuren

##### II.2.1.1 Extraktion von RNA aus Samen mittels Kit

Die RNA-Extraktion aus gekeimten Samen erfolgte mit dem Plant RNA Isolation Aid und dem RNAqueous™ Total RNA Isolation Kit (beides Thermo Fisher Scientific™) nach Narsai *et al.* (2011).

##### II.2.1.2 Extraktion von RNA aus Samen und etiolierten Keimlingen

Die RNA-Extraktion aus trockenen, imbibierten und gekeimten Samen und etiolierten Keimlingen wurde mit dem Seeds&Siliques-Protokoll nach Oñate-Sánchez & Vicente-Carbajosa (2008) durchgeführt.

##### II.2.1.3 Extraktion von RNA aus Keimlingen

Die RNA-Extraktion aus Keimlingen wurde mit dem RNeasy Plant Mini Kit (Qiagen) nach Vorgaben des Herstellers durchgeführt.

##### II.2.1.4 Extraktion von genomischer DNA aus Blättern für Genotypisierung

Die Extraktion von DNA erfolgte nach einem modifizierten Protokoll von Edwards *et al.* (1991). Das in 500 µL Extraktionspuffer zerkleinerte Gewebe wurde mit 66 µL 10 % SDS vermischt, 10 min bei Raumtemperatur (RT) inkubiert und bei 14 000 ×g und RT für 10 min zentrifugiert. Der Überstand wurde mit 166 µL KaOAc vermischt und bei 14 000 ×g und RT für 15 min zentrifugiert. Der Überstand wurde mit 0,5 Vol. Isopropanol vermischt und 15 min auf Eis inkubiert. Anschließend wurde die DNA bei 14 000 ×g bei 4 °C für 15 min präzipitiert. In einem Waschschrift wurde das Pellet mit 800 µL 70 % EtOH bei 14 000 ×g bei 4 °C für 5 min zentrifugiert, anschließend getrocknet und in ddH<sub>2</sub>O gelöst.

##### DNA-Extraktionspuffer:

0.1 M Tris-HCl pH 7,5

0.05 M EDTA

0.5 M Natriumchlorid

1 % PVP-40

#### II.2.2 Hochdurchsatz-Genotypisierung von *A. thaliana* Keimlingen

MS-Platten mit jeweils 100 Samen wurden, wie unter II.1.3 beschrieben, angezogen. Die Herstellung von pflanzlichen Rohextrakten aus Keimlingen erfolgte nach 10 Tagen im Licht mittels der Saccharose-Präparation nach Berendzen *et al.* (2005). Die PCR wurde, wie unter II.2.5 angegeben, durchgeführt, wobei die Amplifikation auf 45 Zyklen verlängert wurde.

### Saccharose-Puffer:

50 mM Tris-HCl pH 7,5  
300 mM Natriumchlorid  
300 mM Saccharose

### II.2.3 Konzentrationsbestimmung von RNA und DNA

Die Konzentration von RNA wurde photometrisch mit dem NanoDrop ND-1000 (PiqLab) bestimmt. Die Intaktheit der RNA wurde in einem 1,5 %igen Agarosegel überprüft (II.2.6). Die Konzentration von DNA wurde in einem 1 %igen Agarosegel (II.2.6) mithilfe einer Lambda-DNA mit bekannter Konzentration abgeschätzt.

### II.2.4 Aufreinigung und Fällung von RNA und DNA

Die wie in Kap. II.2.1.1 und II.2.1.2 aus Samen und etiolierten Keimlingen isolierte RNA wurde mittels Phenol/Chloroform/Isoamylalkohol (25:24:1) Lösung mit pH 4 (Carl Roth) aufgereinigt. Bei der Aufreinigung mit Phenol/Chloroform/Isoamylalkohol wurde die Nukleinsäuremischung mit 1 Vol. Phenol/Chloroform/Isoamylalkohol versetzt, gevortext und bei 14 000  $\times g$  bei 4 °C für 8 min zentrifugiert. Der Überstand wurde mit 1 Vol. Chloroform vermischt, gevortext und bei 14 000  $\times g$  bei 4 °C für 8 min zentrifugiert. Der Überstand wurde mit 3 Vol. EtOH/NaOAc 30:1 pH 5,2 versetzt und ü. N. bei -20°C inkubiert. Die Nukleinsäuren wurden bei 14 000  $\times g$  bei 4 °C für 30 min pelletiert. Das Pellet wurde mit 800  $\mu L$  70 % EtOH bei 14 000  $\times g$  bei 4 °C für 5 min gewaschen, anschließend getrocknet und in ddH<sub>2</sub>O gelöst. Die wie in Kap. II.2.1.3 isolierte RNA aus Keimlingen wurde mit dem Qiagen RNeasy MinElute CleanUp Kit nach Herstellerangaben aufgereinigt.

Die Aufreinigung von DNA wurde mittels Phenol/Chloroform/Isoamylalkohol (25:24:1) Lösung mit pH 8 (Carl Roth) wie oben beschrieben durchgeführt.

### II.2.5 Amplifikation von DNA mittels PCR

Die Schmelztemperatur der Primer wurde mit Primer3Plus<sup>1</sup> berechnet. Die Polymerasenkettenreaktion (PCR, *polymerase chain reaction*) wurde mit der DreamTaq Polymerase (Thermo Fisher Scientific™) nach Herstellervorgaben durchgeführt. Ein 10- $\mu L$  Ansatz setzte sich wie folgt zusammen: 7  $\mu L$  ddH<sub>2</sub>O, 1  $\mu L$  10x Puffer, 0,1  $\mu L$  dNTPs (10  $\mu M$ ), je 0,5  $\mu L$  Vorwärts- und Rückwärts-Primer (10  $\mu M$ ), 0,04  $\mu L$  Dream Taq Polymerase, 1  $\mu L$  DNA (~100 ng). Die PCR lief unter folgenden Bedingungen ab: 94 °C 2,5 min; 35 Zyklen: 94 °C 30 s, 55 °C 30 s, 72 °C 1 min/1 kb; 72 °C 5 min.

### II.2.6 Agarosegelelektrophorese von RNA und DNA

Für die Agarosegelelektrophorese wurden die Nukleinsäuren mit 1/6 Vol. 6x Ladepuffer versetzt und in einem 1-1,5 % (w/v) Agarosegel mit 0,3 % Ethidiumbromid im 1x TAE Laufpuffer aufgetrennt. Für

---

<sup>1</sup> <http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>



DNA wurde ein Größenstandard aufgetragen (Thermo Fisher Scientific™ GeneRuler™ 1 kb DNA-Leiter oder 100 bp DNA-Leiter). Die Detektion erfolgte unter UV-Licht mit dem Geldokumentationssystem Quantum (Vilber).

1x TAE Puffer:

40 mM Tris  
20 mM Eisessig  
1 mM EDTA

6x Ladepuffer:

10 mM Tris-HCl (pH 7.6)  
60 mM EDTA  
0.03 % Bromophenolblau  
0.03 % Xylencyanol FF  
60 % Glycerin

### II.2.7 Polyacrylamid-Gelelektrophorese (PAGE) von DNA

Für die Analyse der Test-PCRs sowie der Größenselektion der cDNA-Bibliotheken wurden native Polyacrylamidgele verwendet. Die cDNA-Bibliotheken wurden mit 1/6 Vol. 6x Ladepuffer (New England Biolabs) versetzt und in 0,6x TBE-Laufpuffer bei 130 Volt für 80 min aufgetrennt, anschließend in 1x TBE mit 1:1000 SybrGold (Thermo Fisher Scientific™) gefärbt und unter UV-Licht mit dem Geldokumentationssystem Quantum (Vilber) sichtbar gemacht.

natives Polyacrylamidgel; 0,75 mm dick:

5 mL ddH<sub>2</sub>O  
1,5 mL 5x TBE  
0,95 mL Acrylamid/Bisacrylamid 40% (19:1)  
75 µL APS  
7,5 µL TEMED

5x TBE:

450 mM Borsäure  
450 mM Tris  
5 mM EDTA

### II.2.8 Northern Blot

Die Northern Blot Analyse wurde nur für das Keimlingsstadium durchgeführt. Im frühen Stadium werden keine eindeutigen Transkriptbanden detektiert und die mRNA erscheint diffus über die gesamte Länge der Gelspur verteilt. Es wird vermutet, dass hier die Detektion durch Sekundärstoffe gestört wird.

Gesamt-RNA von Wildtyp und Mutanten wurde wie unter II.2.1.3 beschrieben isoliert und in einem denaturierenden Formaldehyd-Agarosegel aufgetrennt. Für ein 10 x 15 cm großes Gel wurden 1,2 g Agarose in 70 mL ddH<sub>2</sub>O aufgekocht und nach Abkühlen auf etwa 60 °C mit 10 mL 10x MOPS sowie 20 mL 37 % Formaldehyd versetzt. 1,5 µg RNA wurden mit 2 Volumen RNA-Ladepuffer versetzt, bei 65 °C für 10 min denaturiert und in 1x MOPS für eine Stunde bei 35 V und weitere 2,5 Stunden bei 50 V elektrophoretisch aufgetrennt. Als Marker wurde der RiboRuler High Range RNA Ladder (Thermo Fisher Scientific™) verwendet. Die aufgetrennte RNA wurde auf Nylonmembranen (Amersham Hybond™ N+, GE Healthcare) mittels Kapillarblot transferiert (Sambrook and Russell, 2001). Die RNA wurde durch UV-Bestrahlung (120 mJ/cm<sup>2</sup>) auf der Membran fixiert. Die Beladung wurde mithilfe einer Methylenblau-Färbung überprüft. Der spezifische Nachweis der Transkriptmengen in Wildtyp und Mutanten erfolgte über DNA- bzw. RNA-Sonden, welche mit Digoxigenin (DIG) markiert worden sind. Für die Sonden-Hybridisierung wurden die Membranen in DIG Easy Hyb Buffer (Roche)

vorinkubiert. Die Sonde wurde 5 min bei 85 °C (RNA-Sonde) bzw. 10 min bei 95 °C (DNA-Sonde) denaturiert, 2 min auf Eis abgekühlt und zur vorinkubierten Membran im Puffer zugegeben. Die Hybridisierung erfolgte rotierend über Nacht bei 65 °C (RNA-Sonde) bzw. 50 °C (DNA-Sonde). Überschüssige ungebundene Sonde wurde in mehreren Waschschritten entfernt: 2x 5 min 2x SSC 0,1 % SDS bei RT; 2x 15 min 0,1x SSC 0,1 % SDS (RNA-Sonde) bzw. 0,5x SSC 0,1 % SDS (DNA-Sonde) bei 68 °C, 5 min Waschpuffer. Die Membran wurde mit 1x Block-Reagenz (Roche) in Maleinsäurepuffer für 1 h blockiert und anschließend mit 1x Block-Reagenz (Roche) in Maleinsäurepuffer und 1/10000 Vol. Anti-Digoxigenin-Antikörper Fab-Fragmente (Roche) für 1 h inkubiert. Überschüssiger Antikörper wurde mit 4x 8 min waschen mit Waschpuffer entfernt. Für die Detektion wurde die Membran 5 min in Detektions-Puffer gewaschen und mit CSPD (Roche) inkubiert. Die Detektion der Transkripte erfolgte mit dem Detektionssystem Fusion FX (Vilber).

### Herstellung von RNA-Sonden:

Für die Herstellung der RNA-Sonden wurde eine PCR mit genspezifischen Primern mit der Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific™) nach Herstellervorgaben durchgeführt. Der reverse Promotor trug zusätzlich zur genspezifischen Sequenz die T7-Promotorsequenz, die von der T7-Polymerase gebunden werden kann. Ein 10 µL Ansatz setzte sich wie folgt zusammen: 15 µL ddH<sub>2</sub>O, 5 µL 5x Puffer, 0,5 µL dNTPs (10 µM), je 1,25 µL Vorwärts- und Rückwärts-Primer (10 µM), 0,25 µL Phusion Polymerase, 2 µL DNA (~50 ng). Die PCR lief unter folgenden Bedingungen ab: 96 °C 2,5 min; 35 Zyklen: 99 °C 1 s, 55 °C 15 s, 72 °C 30 s/1 kb; 72 °C 7 min. Das PCR-Produkt wurde mit dem QIAquick PCR Purification Kit (Qiagen) aufgereinigt. Die Transkription mit gleichzeitiger DIG-Markierung erfolgte mit dem DIG RNA Labelling Kit (Sigma-Aldrich) nach Herstellerangaben mit anschließender Aufreinigung der RNA-Sonden mittels Phenol/Chloroform/Isoamylalkohol, wie unter II.2.4 angegeben.

### Herstellung von DNA-Sonden:

Für die Herstellung der DNA-Sonden wurde eine PCR mit genspezifischen Primern analog II.2.5 durchgeführt und mit dem QIAquick PCR Purification Kit (Qiagen) aufgereinigt. Markierung der PCR-Fragmente mit DIG erfolgte mit dem PCR DIG Probe Synthesis Kit (Sigma-Aldrich) nach Herstellerangaben. Die Primer für die Sondenherstellung sind dem Anhang Tab. VI.2 zu entnehmen.

#### RNA-Ladepuffer:

65 µL 10x MOPS  
320 µL Formamid  
110 µL 37 % Formaldehyd

#### Maleinsäurepuffer:

0,1 M Maleinsäure  
0,15 M Natriumchlorid  
pH 7,5

#### Waschpuffer:

Maleinsäurepuffer  
0,3 % Tween-20

#### 10x MOPS:

MOPS 0,2 M  
EDTA 5 mM  
Natriumacetat 50 mM

#### 20x SSC:

3 M Natriumchlorid  
0,3 M Natriumcitrat  
pH 7,0

#### Detektionspuffer:

0,1M Tris/HCl pH 9,5  
0,1M Natriumchlorid

### Methylenblau-Färbelösung:

0,5 M NaOAc pH 5,2

0,04 % Methylenblau

### II.2.9 dRNA-Seq

Für die Untersuchung der TSS-Nutzung in Wildtyp und Mutanten wurden pro Genotyp und Entwicklungsstadium 3 biologische Replikate untersucht. Für die Untersuchung der TSS-Nutzung während der Samenentwicklung in Wildtyp wurden jeweils 4 biologische Replikate geerntet.

#### II.2.9.1 Anreicherung der Primärtranskripte

Für die Anreicherung der Primärtranskripte wurde eine adaptierte Methode aus Kruesi *et al.* (2013) verwendet. Für die Analyse der TSS-Nutzung in Wildtyp und Mutanten wurde die RNA wie in Kap. II.2.1.1 und II.2.1.3 beschrieben isoliert. Für die Analyse der TSS während der Samenentwicklung wurde die RNA wie in Kap. II.2.1.2 beschrieben extrahiert. Die extrahierte RNA wurde mit Turbo DNase I (Thermo Fisher Scientific™) behandelt, um Kontaminationen mit genomischer DNA zu entfernen. Die Aufreinigung erfolgte wie in Kap. II.2.4 beschrieben. Anschließend wurden die Primärtranskripte in den Proben durch eine Behandlung mit der Terminator™ 5'-Phosphat-abhängigen Exonuclease (TEX, Lucigen) angereichert. Hierbei wurden jeweils 2 µg RNA für 5 min bei 75 °C denaturiert, um Sekundärstrukturen zu reduzieren und anschließend mit 4 U TEX behandelt. Nach einer Inkubation bei 30 °C für 60 min wurden weitere 4 U TEX dazugegeben und weitere 60 min bei 30 °C inkubiert. Die Aufreinigung erfolgte wie in Kap. II.2.4 beschrieben. Jede Probe wurde zweigeteilt. Eine Teilprobe wurde für 30 min mit 1 U 5'-Polyphosphatase (5'PPase, Lucigen) bei 37 °C behandelt. Die andere Teilprobe wurde nicht mit 5'PPase behandelt und nach Zugabe des 5'PPase-Puffers ebenfalls bei 37 °C für 30 min inkubiert. Diese Probe stellt die cDNA-Kontrollbibliothek dar. Die Aufreinigung der Proben erfolgte wie in Kap. II.2.4 beschrieben.

#### II.2.9.2 Herstellung von cDNA-Bibliotheken und Sequenzierung

Die Herstellung der cDNA-Bibliotheken für die Analyse der TSS-Nutzung in Wildtyp und Mutanten erfolgte mit dem NEBNext® Small RNA Library Prep Set for Illumina® Kit nach Herstellerangaben mit den folgenden Modifikationen. Als 5'-Adapter wurde statt dem Originaladapter ein modifizierter 5'-Adapter (Sequenz siehe Anhang Tab. VI.2) verwendet, welcher am 3'-Ende 8 randomisierte Nukleotide trägt, den sogenannten UMI (*unique molecular identifier*). Die Adapter wurden in einer 1:4-Verdünnung verwendet. Nach einer reversen Transkription wurden Test-PCRs in 10 µL-Ansätzen nach Kitprotokoll mit 18, 21 und 23 Zyklen durchgeführt, um die korrekte Zyklenzahl zu identifizieren. Die Primer wurden 1:5 verdünnt eingesetzt. Die PCR lief unter folgenden Bedingungen: 94°C 1 min; 18, 21 bzw. 23 Zyklen: 94 °C 15 s, 62 °C 30 s, 70 °C 1 min; 70 °C 7 min. Die Test-PCRs wurden in einem Polyacrylamidgel (Kap. II.2.7) aufgetrennt. Nach der Amplifikation der cDNA-Bibliotheken mit

identifizierter Zyklenzahl analog der o. g. Bedingungen wurden die PCRs über Phenol/Chloroform/Isoamylalkohol (Kap. II.2.4) aufgereinigt und in einem Polyacrylamidgel (Kap. II.2.7) aufgetrennt und gröbenselektiert. Es wurde ein Bereich zwischen 250 und 800 bp ausgeschnitten, um kleine RNAs und Adapterdimere abzureichern. Die ausgeschnittenen Gelstücke wurden zerkleinert, indem sie durch ein 0,5-mL-Reaktionsgefäß mit kleinem Loch 2 min bei  $14\,000 \times g$  bei RT zentrifugiert wurden. Das zerkleinerte Gel wurde mit Elutionspuffer bedeckt und rotierend ü. N. bei RT eluiert. Anschließend wurden die Gelstücke durch Zentrifugieren (3 min bei  $14\,000 \times g$  und RT) über Corning®-Costar®-Spin-X®-Zentrifugenröhrchen-Filter (Sigma-Aldrich) vom Eluat getrennt. Die DNA aus dem Eluat wurde ü. N. mit 3 Vol. EtOH/NaOAc 30:1 pH 5,2 und 20 µg Glykogen (Roche) bei  $-20\text{ °C}$  präzipitiert. Das Präzipitat wurde 30 min bei  $14\,000 \times g$  und  $4\text{ °C}$  pelletiert und das Pellet anschließend mit 80 % Ethanol durch rigoroses Vortexen gewaschen. Nach abschließender Zentrifugation für 30 min bei  $14\,000 \times g$  und  $4\text{ °C}$  wurde das Pellet getrocknet und in ddH<sub>2</sub>O gelöst. Die Sequenzierung wurde auf einer Illumina NextSeq500 Plattform durch die amedes Medizinische Dienstleistungen GmbH, Berlin durchgeführt.

Für die Analyse der TSS während der Samenentwicklung in Wildtyp wurde das Kit NEXTFLEX® Small RNA-Seq Kit v3 (PerkinElmer) nach Herstellerangaben mit den folgenden Modifikationen verwendet. Als 5'-Adapter wurde statt dem Originaladapter ein modifizierter 5'-Adapter verwendet (Anhang Tab. VI.2). Die Adapter wurden in einer 1:4-Verdünnung verwendet. Nach der reversen Transkription mit anschließender Aufreinigung mit Magnetkugeln nach Kitprotokoll wurden Test-PCRs in 10 µL-Ansätzen mit dem KAPAHiFi Kit (Roche) nach Herstellerangaben mit 18, 21, 24 und 29 Zyklen durchgeführt, um die korrekte Zyklenzahl zu identifizieren. Die PCR lief unter folgenden Bedingungen:  $98\text{ °C}$  45 s; 18, 21, 24 bzw. 29 Zyklen:  $98\text{ °C}$  15 s,  $62\text{ °C}$  30 s,  $72\text{ °C}$  30 s;  $72\text{ °C}$  2 min. Die Test-PCRs wurden in einem Polyacrylamidgel (Kap. II.2.7) aufgetrennt. Nach der Amplifikation der cDNA-Bibliotheken mit identifizierter Zyklenzahl nach Kitprotokoll wurden die PCRs über Phenol/Chloroform/Isoamylalkohol (II.2.4) aufgereinigt und in einem Polyacrylamidgel (Kap. II.2.7) aufgetrennt und gröbenselektiert. Es wurde ein Bereich von 250 bp bis 800 bp gröbenselektiert, die ausgeschnittenen Gelstücke wie oben beschrieben zerkleinert, mit Elutionspuffer bedeckt und rotierend ü. N. bei RT eluiert. Anschließend wurden die Gelstücke durch Zentrifugieren (3 min bei  $14\,000 \times g$  und RT) über Corning®-Costar®-Spin-X®-Zentrifugenröhrchen-Filter vom Eluat getrennt und das Eluat mit den cDNA-Bibliotheken einer zusätzlichen Aufreinigung mit Magnetkugeln nach Kit-Angaben unterzogen. Die Sequenzierung wurde auf einer Illumina NextSeq500 Plattform durch das Center for Molecular and Cellular Bioengineering (CMCB), Dresden durchgeführt.

### II.2.9.3 Bioinformatische Analyse der cDNA-Bibliotheken

Die Analyse der sequenzierten Daten erfolgte mit Hilfe des Galaxy Servers des TRR 175 des The Green Hub Consortiums (<http://141.20.62.236:8080/>) (Abb. II.2).

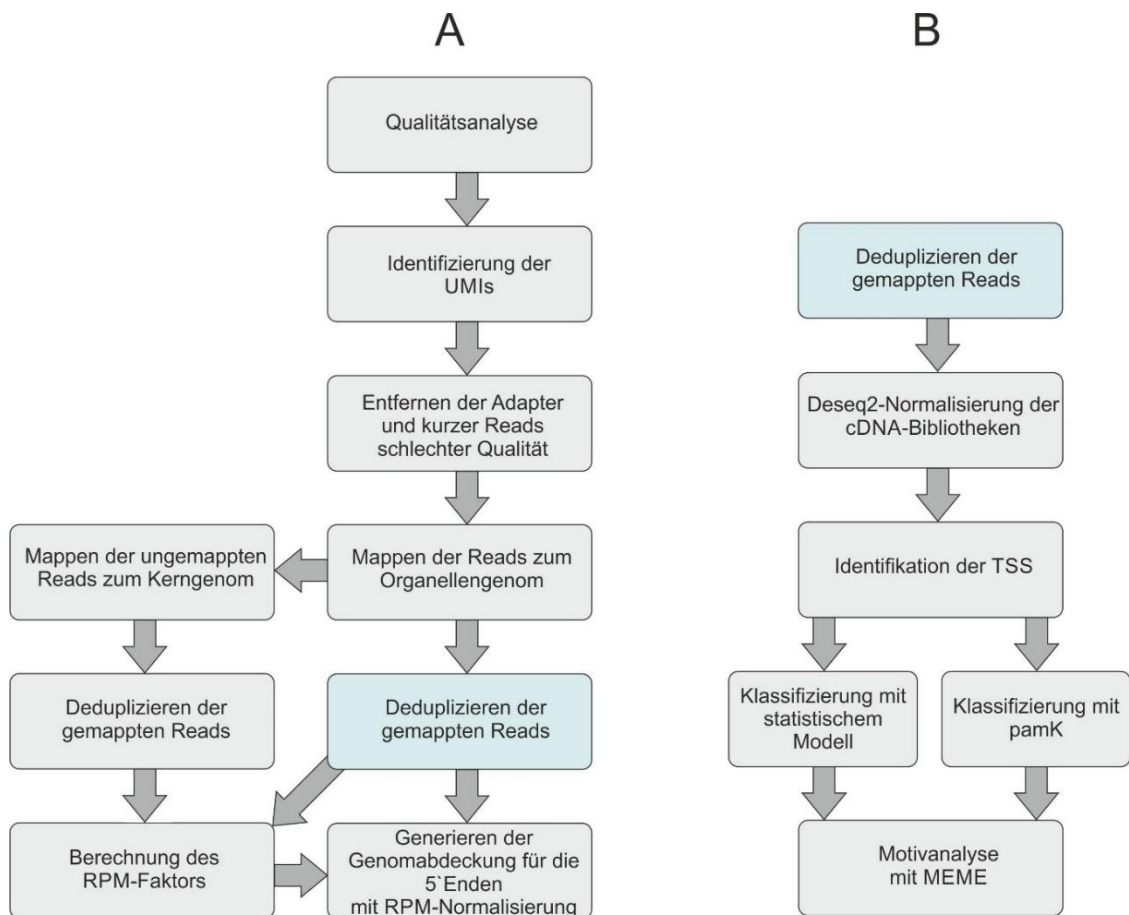


Abbildung II.2: Schematische Darstellung der bioinformatischen Analyse für dRNA-Seq. A: Nach der Qualitätsanalyse der Rohdaten wurden die UMIs (*unique molecular identifier*) markiert und Adapter sowie Reads mit schlechter Qualität entfernt. Nach dem Mappen zum Organellengenom wurden die Read-Duplikate entfernt. Für die Berechnung des RPM-Faktors (*reads per million*) wurden die restlichen Reads zum Kerngenom gemappt. Mit dem RPM-Normalisierungsfaktor wurden die Genomabdeckungen für die 5'-Enden der Reads generiert. Die Analyse wurde mit dem Galaxy Server durchgeführt. B: Schematische Darstellung der TSS-Prozessierung. Die deduplizierten zum Plastom gemappten Reads in den +5 PPase- und den -5 PPase-Bibliotheken wurden verglichen und die TSS identifiziert. Es wurde eine Klassifizierung mit zwei verschiedenen Methoden durchgeführt, wobei nur die +5 PPase- Bibliotheken für die Klassifizierung verwendet wurden. Die identifizierten Klassen der TSS wurden bezüglich ihrer Motive mit *MEME* untersucht.

Die Rohdaten wurden als fastqsanger-Format zusammen mit dem Referenzgenom hochgeladen. Nach der Qualitätskontrolle mit FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) wurden die UMIs mit Je-Clip (Girardot *et al.*, 2016) identifiziert und die UMI-Sequenz vom Readanfang entfernt. Anschließend wurden die Reads mit Trimmomatic (Bolger *et al.*, 2014) getrimmt, wobei die 3'-Adaptersequenzen entfernt wurden. Gleichzeitig wurden Readenden mit schlechter Qualität mit einer minimalen Phred-Qualitätsbewertung von 20 und Reads, die kürzer als 20 Nukleotide waren, eliminiert. Die Kartierung erfolgte sequenziell. Zunächst wurden die Reads mittels Bowtie2 (Langmead *et al.*, 2009) mit Standardeinstellungen end-to-end, -D 15 -R 2 -L 22 -i S,1,1.15 zum Organellengenom gemappt, hierbei wurden die ungemappten Reads in eine separate Datei geschrieben und diese im Anschluss mit o. g. Standardeinstellungen zum Kerngenom gemappt. Das Entfernen der Reads, welche als PCR-Duplikate mit dem gleichen UMI identifiziert wurden, erfolgte mit Je-MarkDuplicates (Girardot *et al.*, 2016) mit Standardeinstellungen ohne prädefinierte UMI-Liste. Die

Gesamtmenge der gemappten deduplizierten Reads wurde für eine Reads-pro-Million-Normalisierung (RPM, *reads per million*) verwendet. Es wurden normalisierte Genomabdeckungen mit Genome Coverage (Quinlan and Hall, 2010) für die 5'-Enden generiert, wobei die Genomabdeckung für die Plus- und Minustränge separat gemacht wurde. Die Genomabdeckungen wurden mit dem Integrated Genome Viewer (IGV) (Robinson *et al.*, 2011; Thorvaldsdóttir *et al.*, 2013) visualisiert. Für die statistische Auswertung wurde Samtools IdxStats (Li, 2011a, 2011b; Li *et al.*, 2009) genutzt. Die verwendeten Referenzgenome wurden der TAIR-Datenbank entnommen (<https://www.Arabidopsis.org/index.jsp>, TAIR10.1).

### II.2.9.4 TSS Identifikation und Klassifizierung

Für die Identifikation sowie Klassifikation der TSS (Abb. II.2) wurden die Datensätze anders als für die visuelle Darstellung im IGV (Kap. II.2.9.3) mittels DESeq2 normalisiert, wobei drei biologische Replikate gemittelt wurden. DESeq2 führt eine interne Normalisierung durch, bei der für jedes Gen und für alle Replikate das geometrische Mittel berechnet wird. Die Genomabdeckung für jedes Gen bzw. Genomposition wird durch diesen Mittelwert geteilt (Anders and Huber, 2010; Love *et al.*, 2014). Die Skripte für die Identifikation der TSS wurden von Jose Muino und Dermot Harnett (beide HU Berlin) zur Verfügung gestellt und sind dem Anhang VI.3 zu entnehmen. Für die anschließende Klassifizierung wurden nur die DESeq2-normalisierten Genomabdeckungen der +PPase-Bibliotheken und die Liste identifizierter TSS verwendet. Die Klassifizierung der TSS nach Nutzung durch die einzelnen RNA-Polymerasen wurde mit Hilfe eines statistischen Modells durchgeführt, welches sich an einem publizierten Modell orientiert (Weinholdt *et al.*, 2019). Das statistische Modell wurde durch Ivo Große (MLU Halle) entwickelt, durch Frederic Ehlers optimiert und unter Anleitung von Julia Rott angewendet. Die Skripte sind Anhang VI.3 zu entnehmen. Als zusätzliche Klassifikationsmethode wurde pamK benutzt (Gentle *et al.*, 1991), um die Ergebnisse der statistischen Analyse zu untermauern. Die Skripte für die Klassifikation mit pamK wurden von Jose Muino und Dermot Harnett (beide HU Berlin) zur Verfügung gestellt und von Frederic Ehlers unter Anleitung von Julia Rott angewendet. Nach der Identifikation der polymerasespezifischen TSS wurden die Motive, welche die TSS flankieren, analysiert. Für die Motivsuche wurden die Sequenzen extrahiert, die von der TSS, welcher die Position +1 zugeordnet wird, -40 stromabwärts und +5 stromaufwärts gehen. Die verwendeten TSS-spezifischen Sequenzen sind im Anhang Tab. VI.14 zusammengefasst. Für die Motivanalyse wurden Duplikatsequenzen, die aus den gleichen Genkopien im *A. thaliana*-Genom (siehe Kap. I.2.1) stammen, nur einmal verwendet. Die Motivsuche wurde mit MEME (*Multiple Expectation-Maximization for Motif Elicitation*) Version 5.3.0 durchgeführt, welches gemeinsame Motive in einem Cluster von Sequenzen finden kann (Bailey *et al.*, 2009). Die Suche wurde für die spezifischen Cluster der RPOTp-abhängigen TSS durchgeführt und nach Motiven mit einer Breite von 4 bis 8 Nukleotiden gesucht. Die folgenden Parameter wurden verwendet: `-dna -oc . -nostatus -time 18000 -mod oops -nmotifs 4 -minw 4 -maxw 8 -objfun classic -markov_order 0`.

### II.2.10 RNA-Seq

Für die Analyse wurden pro Genotyp und Entwicklungsstadium drei biologische Replikate untersucht. Die Herstellung der cDNA-Bibliotheken und die Sequenzierung wurden von Novogene Co., LTD durchgeführt. Die extrahierte RNA (Kap. II.2.1.1 und II.2.1.3) wurde mit Turbo DNase I (Thermo Fisher Scientific™) behandelt, um Kontaminationen mit genomischer DNA zu entfernen. Die Aufreinigung erfolgte wie in Kap. II.2.4 beschrieben. Anschließend wurde die Konzentration bestimmt, die rRNA abgereichert und die RNA zu einer Größe von 150 bis 200 bp fragmentiert. Die anschließende cDNA-Synthese erfolgte mit randomisierten Hexameren. Es folgten Endreparatur, Polyadenylierung, Adapterligation, Größenselektion und PCR-Amplifikation. Die cDNA-Bibliotheken wurden auf einer Illumina HiSeq Plattform sequenziert. Die bioinformatische Analyse ist in Abb. II.3 dargestellt und wurde durch Jose Muino (HU Berlin) durchgeführt. Nach der Qualitätsanalyse der Rohreads wurden die Reads mit Trimmomatic (Bolger *et al.*, 2014) getrimmt, wobei die 3'-Adaptersequenzen entfernt wurden. Gleichzeitig wurden Readenden von schlechter Qualität mit einer minimalen Phred-Qualitätsbewertung von 20 und Reads, die kürzer als 15 Nukleotide waren, eliminiert. Das Mappen erfolgte mit RNA STAR (Dobin *et al.*, 2013) mit den Einstellungen `--runThreadN 20 --outMultimapperOrder Random --outSAMtype BAM --alignIntronMax 3000 --outFilterIntronMotifs RemoveNoncanonical --outSAMstrandField intronMotif --outSAMattrIHstart 0` auf das komplette Genom. Das verwendete Referenzgenom wurde der TAIR-Datenbank entnommen (TAIR10.1). Die Genomabdeckungen wurden mit BamCoverage (Ramírez *et al.*, 2016) ohne Normalisierung für das komplette Genom berechnet und mit IGV visualisiert (Robinson *et al.*, 2011; Thorvaldsdóttir *et al.*, 2013). Anschließend wurde eine differentielle Expressionsanalyse zwischen den Mutanten und dem Wildtyp mit DESeq2-Normalisierung durchgeführt. Die signifikante differentielle Expression wurde mittels des Wald-Tests bestimmt.

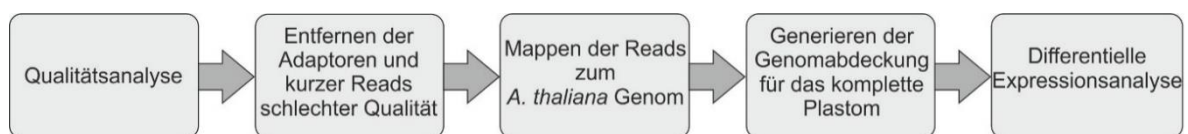


Abbildung II.3: Schematische Darstellung der bioinformatischen Analyse für RNA-Seq. Nach der Qualitätsanalyse der Rohdaten wurden die Adapter sowie Reads mit schlechter Qualität entfernt. Nach dem Kartieren zum *A. thaliana*-Genom wurden die Genomabdeckungen für das Plastom berechnet und eine differentielle Expressionsanalyse der Plastomgene durchgeführt.

### II.2.11 Elektronenmikroskopische Untersuchung der Plastiden

Die Untersuchung wurde von Dr. Gerd Hause und Fr. Simone Fraas (Abteilung Elektronenmikroskopie, Biozentrum MLU) durchgeführt. Die Vorbereitung und Analyse der Proben erfolgte wie in Frank *et al.* (2019) beschrieben.

### II.2.12 Lichtmikroskopische Untersuchung der Samenentwicklung

## Material und Methoden

---

Für die Untersuchung der Samen wurden die Schoten der Mutanten geöffnet und mithilfe des Zeiss Discovery.V12 Stereomikroskop (Carl Zeiss, Jena, Deutschland) visualisiert und fotografiert.

Für die Untersuchung der Embryonen wurden Samen aus der Schote herausgetrennt und auf Objektträger in Hoyer`s Medium (3,75 Gummi arabicum, 50 g Chloralhydrat, 2,5 mL Glycerin, 15 mL ddH<sub>2</sub>O) für 24 h entfärbt und fixiert. Anschließend wurden die Samen direkt mit dem Zeiss Axioskop Mikroskop (Carl Zeiss, Jena, Deutschland), ausgestattet mit einem Differentialinterferenzkontrast (DIC, *differential interference contrast*), visualisiert und fotografiert.



### III ERGEBNISSE

#### III.1 Wachstumsanalyse der Mutanten

*A. thaliana* besitzt drei plastidäre RNA-Polymerasen PEP, RPOTp und RPOTmp, deren Rollen und Zusammenspiel in der Transkription des Plastoms zu einem großen Teil unerforscht sind. Das Ziel dieser Arbeit war die Identifizierung der Funktionen dieser RNA-Polymerasen. Dazu wurden Mutanten mit fehlender/eingeschränkter RNA-Polymerasenaktivität analysiert. Es wurden *Knockout*-Linien mit fehlender RPOTp- bzw. RPOTmp-Aktivität sowie Linien mit reduzierter PEP-Aktivität für die Analyse ausgewählt. Als *rpoTp*-Mutante wurde die bereits charakterisierte Linie SALK\_067191 (*sca3-3, rpoTp-2*) verwendet (Hricová *et al.*, 2006). Als *rpoTmp*-Mutante diente die Linie SALK\_132842 (*rpoTmp-2*) (Kühn *et al.*, 2009). Da es bis vor Kurzem nicht möglich war, stabile Transformationen des Plastoms von *A. thaliana* vorzunehmen, wurden in der vorliegenden Arbeit Mutanten mit einer sekundären Reduktion der PEP-Aktivität verwendet. Eine geeignete Mutante, die die Reduktion vieler PEP-abhängiger Transkripte zeigte, ist *ptac2* (SALK\_075736) (Pfalz *et al.*, 2006). PTAC2 ist ein kernkodiertes Protein, welches als ein Teil des transkriptionsaktiven PEP-Komplexes identifiziert worden ist (Pfannschmidt *et al.*, 2015). Bei Abwesenheit des Proteins war die PEP-Aktivität stark reduziert und die Mutanten überlebten das Keimlingsstadium nicht (Pfalz *et al.*, 2006). Zusätzlich wurde die Mutante *clb19* (SALK\_123752) (Chateigner-Boutin *et al.*, 2008) in die Analyse mit einbezogen, welche Samen generieren kann. Dieser Mutante fehlt ein kernkodiertes PPR-Protein, welches für die Edierung der Transkripte *rpoA* und *clpP* nötig ist (Chateigner-Boutin *et al.*, 2008). Da das Transkriptomprofil dieser Mutante dem von *ptac2* ähnelte, weist die Mutante einen sekundären PEP-Defekt auf (Chateigner-Boutin *et al.*, 2008), sodass die Nutzung von *clb19* als PEP-Mutante berechtigt ist. Die T-DNA-Insertionsstellen in den einzelnen Linien sind in Abb. II.1 aufgeführt. Als Wildtyp-Kontrolle diente der Ökotyp Columbia-0, im Folgenden als Wildtyp bezeichnet.

Für die Untersuchung der TSS-Nutzung in diesen Mutanten und Wildtyp wurden zwei Zeitpunkte in der Pflanzenentwicklung ausgewählt. Eine phänotypische Analyse in *A. thaliana* diente als Orientierung (Boyes *et al.*, 2001). Untersucht wurden ein Stadium während der frühen Samenkeimung ohne reife Chloroplasten, wenn die Wurzel zu 0,5 mm aus der Samenschale herausragt, im Folgenden als frühes Stadium bezeichnet, und ein frühes Keimlingsstadium mit reifen Chloroplasten, wenn die ersten echten Blätter etwa 1 mm lang sind (Boyes *et al.*, 2001). Dieses Stadium wurde im Folgenden als spätes Stadium bezeichnet. Da *ptac2* keimlingslethal war und keine Samen generiert werden konnten, wurde diese Mutante nur für die Analyse des Keimlingsstadiums benutzt. Für das frühe Stadium der Samenkeimung wurde die Mutante *clb19* benutzt. Es wurden Wachstumsanalysen durchgeführt, um zu untersuchen, zu welchen Zeitpunkten die beiden Stadien in Wildtyp und den verschiedenen Mutanten erreicht werden und damit die jeweiligen Zeitpunkte für die Ernte abzuleiten. Dazu wurden jeweils 100 Samen auf Platten wie unter Kap. II.1.3 angegeben ausgebracht und unter

Langtagbedingungen angezogen. Die Entwicklung des Samens bis zum Keimlingsstadium wurde dokumentiert (Abb. III.1, Wildtyp; Anhang Abb. VI.1, Mutanten).

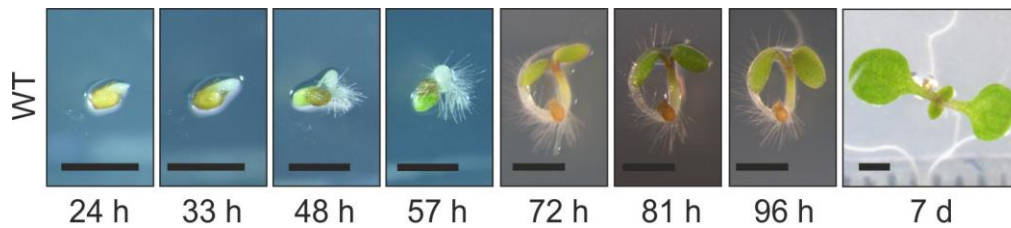


Abbildung III.1: Entwicklung eines Samens von Wildtyp bis zum Keimlingsstadium, angezogen auf MS-Medium mit Saccharose unter Langtagbedingungen. Die Zeitpunkte sind ab dem Transfer der Anzuchtplatten ins Licht angegeben. Die Maßstabsleiste zeigt 1 mm an. h: Stunden

In der Entwicklung bis zum Keimling wiesen die Mutanten eine deutliche Verzögerung im Wachstum im Vergleich zu Wildtyp auf (Abb. III.2). Die *rpoTmp*-Mutante erreichte das späte Stadium nach knapp 10 Tagen. Die beiden Mutanten *rpoTp* und *clb19* zeigten eine deutlich verzögerte Entwicklung und erreichten das späte Stadium nach durchschnittlich 13 bzw. 22 Tagen. Als PEP-Mutante wurde in der Wachstumsanalyse *clb19* verwendet, da diese Mutante Samen generiert. Für *ptac2* konnte keine Wachstumsanalyse durchgeführt werden, da diese Mutante keimlingslethal war. Hier wurden Samen mit dem Genotyp *PTAC2 ptac2* auf Platten ausgebracht und die weißen homozygoten Mutanten selektiert, da die heterozygoten Pflanzen einen grünen Phänotyp aufwiesen. Die Entwicklung von *ptac2* bis zum späten Stadium entsprach weitestgehend der Entwicklung von *clb19*.

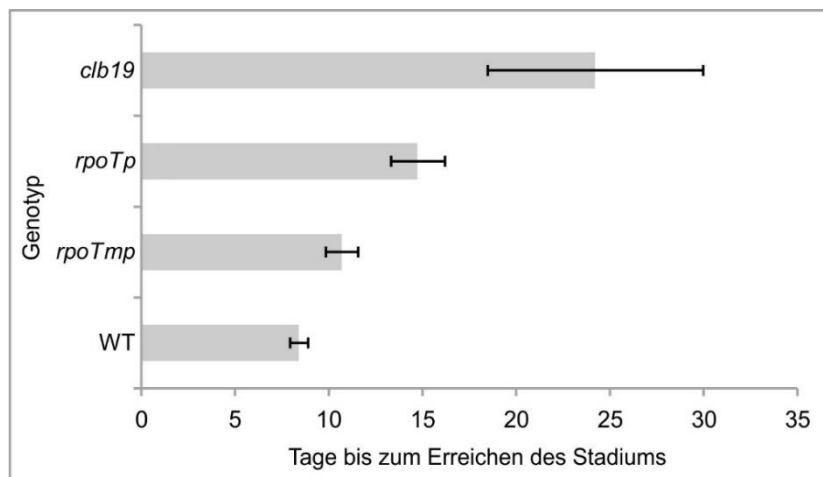


Abbildung III.2: Wachstumsentwicklung von Wildtyp und Mutanten vom Samen bis zum Keimlingsstadium unter Langtagbedingungen. Der Balken zeigt an, wann der jeweilige Genotyp das Keimlingsstadium erreicht hat. Untersucht wurden jeweils 100 Samen pro Genotyp. Der Fehlerbalken zeigt die Standardabweichung.

Der Phänotyp von Wildtyp und Mutanten zu den Zeitpunkten der Ernte ist in Abb. III.3 dargestellt. *rpoTmp* hatte als Keimling einen grünen Phänotyp, während die Mutanten *rpoTp* und *ptac2* eine weiße bis gelbliche Farbe aufwiesen (Abb. III.3).

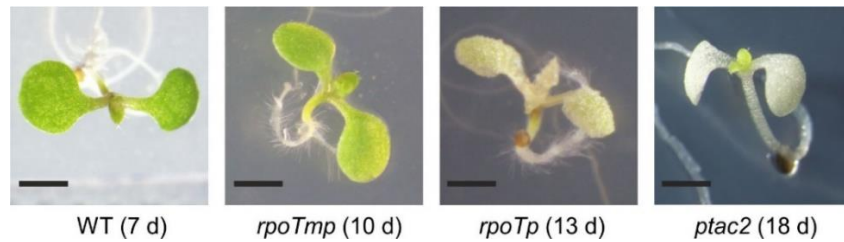


Abbildung III.3: Lichtmikroskopische Aufnahme von Keimlingen von Wildtyp und Mutanten, die nach Stratifikation unter Langtagbedingungen auf MS-Medium mit Saccharose angezogen wurden. In Klammern sind die jeweiligen Erntezeitpunkte ab dem Transfer der Anzuchtplatten ins Licht angegeben. Die Maßstabsleiste zeigt 1 mm an. d: Tage.

Die Ernte der Samen im frühen Stadium unter Langtagbedingungen konnte in den verschiedenen Genotypen nicht synchronisiert werden, sodass die Samen nicht 5 Stunden nach Einschalten des Lichts geerntet werden konnten. Die Samen für das frühe Stadium wurden schließlich unter Dauerlicht angezogen. Abb. III.4 zeigt die Zeitpunkte, welche unter Dauerlichtbedingungen zur Ernte bestimmt wurden. Im Vergleich zur Wildtyp-Kontrolle waren die Mutanten in ihrer Entwicklung verzögert.

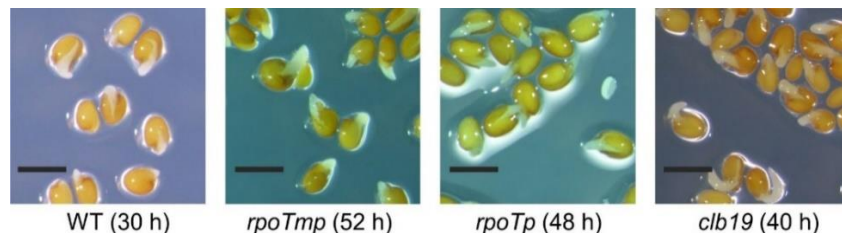


Abbildung III.4: Lichtmikroskopische Aufnahme von gekeimten Samen von Wildtyp und Mutanten, die nach Stratifikation unter Dauerlichtbedingungen auf MS-Medium mit Saccharose angezogen wurden. In Klammern sind die jeweiligen Erntezeitpunkte ab dem Transfer der Anzuchtplatten ins Licht angegeben. Die Maßstabsleiste zeigt 1 mm an. h: Stunden.

### III.2 Optimierung des dRNA-Seq-Protokolls

Die Methode der differentiellen RNA-Sequenzierung (dRNA-Seq) zur Identifizierung von Transkriptionsstartseiten wurde für Prokaryoten entwickelt (Sharma *et al.*, 2010) und erfolgreich für pflanzliche Proben angewandt, um die plastidäre Transkription in Gerste zu untersuchen (Zhelyazkova *et al.*, 2012b). Im Laufe dieser Arbeit wurde das publizierte Protokoll optimiert und an kleine Probenmengen, sowie für eine höhere Sequenziertiefe angepasst. Im Unterschied zum Protokoll aus Sharma *et al.* (2010) (Kap. I.5), bei dem als Kontrolle eine nicht mit TEX behandelte Probe verwendet wurde, wurde im Laufe dieser Arbeit analog zum Protokoll von Kruesi *et al.* (2013) die komplette Probe mit TEX-Enzym behandelt, um eine Reduktion des Hintergrunds zu erreichen (Abb. III.5). Als Kontrolle wurde eine  $-5'$ PPase-Bibliothek anstatt einer  $-TEX$ -Bibliothek benutzt. Die  $5'$ Polyphosphatase ( $5'$ PPase) ist eine Alternative zum TAP-Enzym (*Tobacco Acid Pyrophosphatase*) und ist laut Herstellerangaben nicht in der Lage, die  $5'$ -Cap-Strukturen von Kerntranskripten zu entfernen. Somit wurde durch die Nutzung der  $5'$ PPase die Kontamination der cDNA-Bibliotheken durch Kerntranskripte reduziert. Die meisten Protokolle nutzen zur Konvertierung der  $5'$ -Enden der Transkripte Enzyme wie TAP (Castandet *et al.*, 2019; Kruesi *et al.*, 2013; Sharma *et al.*, 2010; Zhelyazkova *et al.*, 2012b) oder RppH (*5'Pyrophosphohydrolase*) (Hetzl *et al.*, 2016), welche  $5'$ -

Caps entfernen können. Für eine Anwendung in Organellen ist das von Nachteil, da später auch Kerntranskripte in den cDNA-Bibliotheken enthalten sind. Bei Zhelyazkova *et al.* (2012b) wurde für die Reduktion der Kerntranskripte stattdessen eine Chloroplastenisolation durchgeführt. Die Kontrollbibliothek wurde genutzt, um prozessierte Transkripte und Fragmente von Transkripten zu identifizieren und sie von den Primärtranskripten zu unterscheiden. Die -5'PPase-Bibliothek erfüllte diesen Zweck, da dort Transkripte enthalten waren, die einen 5'-Monophosphat oder eine 5'-Hydroxygruppe tragen, an die der 5'Adapter ligiert werden kann. Der Vorteil dieser Methode war, dass auf eine vorhergehende rRNA-Entfernung verzichtet werden konnte, da mit TEX ein Großteil der prozessierten rRNAs degradiert wurde.

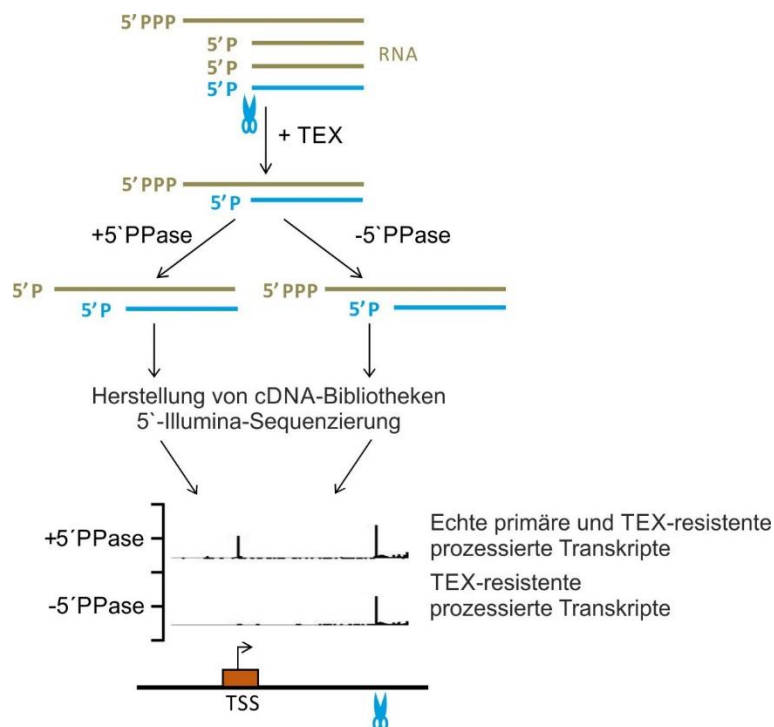


Abbildung III.5: Schematische Darstellung des optimierten dRNA-Seq-Protokolls. Der Pool von RNA wird mit dem Enzym TEX behandelt, welches Transkripte mit 5'-Monophosphaten degradiert. Nach Zerteilen der Proben werden die Transkripte in der einen Probe mit 5'PPase konvertiert (+5'PPase), die andere bleibt unbehandelt (-5'PPase). Beide Proben werden für die Herstellung von cDNA-Bibliotheken verwendet. Die sequenzierten Bibliotheken liefern im Vergleich zueinander die Informationen zu TSS, welche von Primärtranskripten stammen und in der +5'PPase-Bibliothek angereichert sind. Die Schere zeigt die Prozessierungsstelle der prozessierten Transkripte an. In blau ist ein TEX-resistentes Transkript dargestellt. TSS: Transkriptionsstartpunkt.

Für die Herstellung von cDNA-Bibliotheken zur Identifizierung von TSS kamen nur Kits in Frage, die auf die Herstellung von Bibliotheken aus kleinen RNAs optimiert sind. Diese Kits bieten die Möglichkeit, den 5'Adapter vor der reversen Transkription zu ligieren, da hier die Information wichtig ist, ob das Transkript ein prozessiertes oder primäres 5'-Ende besitzt. Im Laufe dieser Arbeit wurden die Kits NEBNext® Small RNA Library Prep Set for Illumina® (New England Biolabs) und NEXTFLEX® Small RNA-Seq Kit v3 (PerkinElmer) für die Herstellung von cDNA-Bibliotheken getestet. Beide Kits hatten den Nachteil, dass das Protokoll während der PCR und der Gelaufreinigung für kürzere Fragmente selektiert, sodass mögliche lange Transkripte nicht amplifiziert und/oder eluiert

werden können und es gleichzeitig zu einer vermehrten Amplifikation der kurzen Adapterdimere kommt. So wurden Strategien zur Vermeidung von Adapterdimerbildung in beiden Protokollen benutzt. Die im Laufe dieser Arbeit gewonnenen experimentellen Daten deuteten darauf hin, dass das Kit NEXTFLEX® Small RNA-Seq Kit v3 weniger Adapterdimere generiert, da freie nicht ligierte Adapter über eine Aufreinigung mit Beads entfernt werden.

### III.2.1 Optimierung der Entfernung der prozessierten RNA

Da die Vermutung vorlag, dass Sekundärstrukturen, wie sie rRNAs und tRNAs haben, von TEX schlechter abgebaut werden und so die Identifizierung der primären Transkripte in der größeren Menge der prozessierten Transkripte erschwert ist, wurden verschiedene Mengen des eingesetzten Enzyms getestet (Abb. III.6). Es wurde jeweils 1 µg RNA mit 1, 2 oder 4 Units TEX behandelt. Als Negativkontrolle diente eine nur mit TEX-Puffer inkubierte Probe. Mit steigender Konzentration des Enzyms wurde mehr prozessierte rRNA abgebaut, was anhand der verschwindenden 25S- und 18S-rRNA-Banden sichtbar war (Abb. III.6). So wurde die eingesetzte TEX-Menge im optimierten Protokoll im Vergleich zum Originalprotokoll um das Vierfache erhöht. In Zhelyazkova *et al.* (2012b) wurden pro 1 µg RNA 1 Unit TEX eingesetzt, im optimierten Protokoll 4 Units TEX je µg RNA.

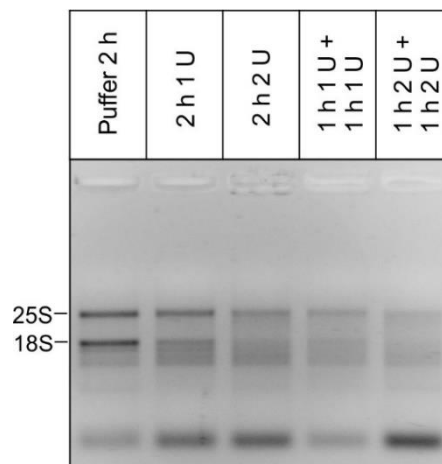


Abbildung III.6: TEX-Behandlung von RNA. 1,5 %iges Agarosegel mit Gesamt-RNA, die mit verschiedenen Mengen an TEX behandelt wurde. Mit steigender Menge an TEX-Enzym wird rRNA, die den Großteil der prozessierten RNA darstellt, abgebaut. Die Banden, die der 25S- und der 18S-rRNA entsprechen, sind links markiert. U: Unit(s), h: Stunde(n)

### III.2.2 Nutzung eines UMI für einen quantitativen Vergleich

Damit eine quantitative Analyse der Sequenzierungsdaten durchgeführt werden konnte, wurde ein modifizierter 5'-Adapter mit einem einzigartigen molekularen Identifikator (UMI, *unique molecular identifier*) verwendet. Der UMI bestand aus 8 randomisierten Nukleotiden, die die Identifikation der PCR-Duplikate ermöglichten. So konnten die Duplikate, die nur im Vorgang der PCR entstanden sind und deshalb dieselbe UMI-Sequenz tragen, eliminiert und die tatsächliche Anzahl der Transkripte miteinander verglichen werden (Fu *et al.*, 2018; Hong and Gresham, 2017; Kivioja *et al.*, 2012). Abb. III.7 zeigt als Beispiel die Gesamtmenge der gemappten Reads für jeweils ein Replikat eines Genotyps

der +5'PPase-cDNA-Bibliothek. Die Gesamtmenge wurde auf 100% gesetzt, sodass der Prozentsatz der PCR-Duplikate (hellgrau) prozentual zu den deduplizierten Gesamtreads (dunkelgrau) gezeigt wird. Die deduplizierten Gesamtreads stellen die Reads dar, die auf jeweils ein Ursprungs-RNA-Molekül zurückzuführen sind.

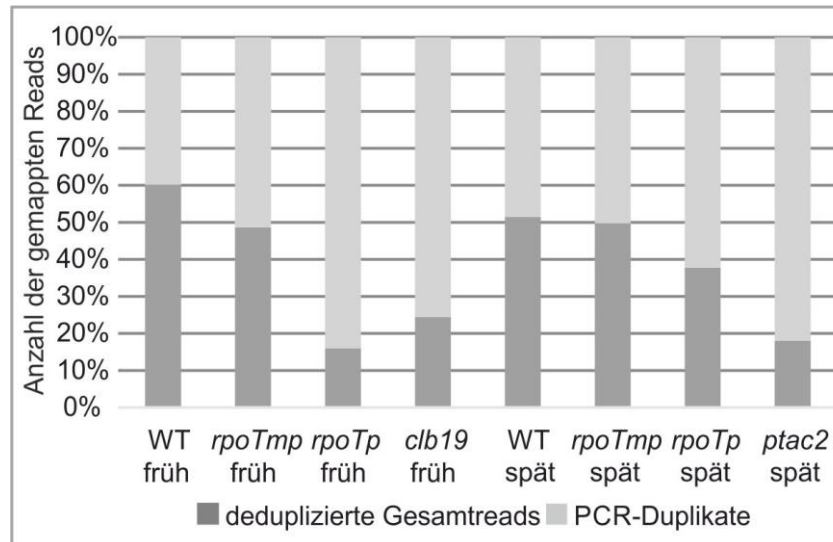


Abbildung III.7: Gesamtmenge der gemappten Reads und der PCR-Duplikate für jeweils ein Replikat eines Genotyps der +5'PPase-cDNA-Bibliothek. Die Menge der gemappten Reads wurde auf 100 % gesetzt und für jeden Genotyp die prozentualen Anteile der PCR-Duplikate sowie die deduplizierten Reads berechnet.

In allen cDNA-Bibliotheken wurde eine hohe Menge an PCR-Duplikaten festgestellt, die durchschnittliche PCR-Duplikatrate lag bei 64 %. Dieses lässt sich durch die niedrige Menge an Input-RNA erklären, mit welcher die Herstellung der cDNA-Bibliothek begonnen wurde. Eine Zusammenfassung der Menge der PCR-Duplikate für alle Replikate ist im Anhang Tab. VI.8 zu finden.

### III.2.3 Optimierung der PCR-Zyklenzahl

Für die Identifizierung der optimalen Menge und Zusammensetzung der PCR wurden die cDNA-Bibliotheken mit verschiedenen Zyklenzahlen amplifiziert und in einem Polyacrylamidgel verglichen (Abb. III.8). Die Herstellung dieser cDNA-Bibliothek erfolgte mit dem NEXTFLEX® Small RNA-Seq Kit v3, die Test-PCRs wurden mit dem KAPAHiFi Kit durchgeführt (Kap. II.2.9.2). In der beispielhaften Abbildung III.8 wurden 21 Zyklen als optimal gewählt. Hier war zum einen eine kontinuierlich diffuse Spur sichtbar, die zeigte, dass wenige rRNA-PCR-Produkte amplifiziert worden sind. Bei großer rRNA-Menge wären distinkte Banden sichtbar. Zum anderen waren die Adapterdimere noch als Bande bei etwa 135 nt sichtbar, welche bei höheren Zyklenzahlen nicht mehr sichtbar wurden, da die Adapterdimere Concatamere gebildet haben. Die unterbrochene Linie markiert den Bereich der cDNA-Bibliothek, der aus dem Gel eluiert und sequenziert wurde.

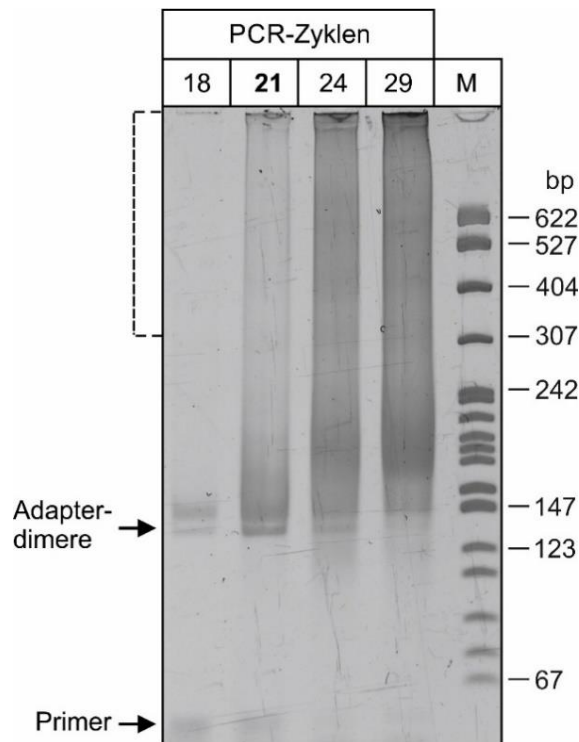


Abbildung III.8: PCR-Amplifikation der cDNA-Bibliotheken. 5 %iges Polyacrylamidgel mit cDNA-Bibliotheken, die mit verschiedener Zyklenzahl amplifiziert wurden. Primer, Adapterdimere und Marker sind gekennzeichnet. In diesem Beispiel sind 21 PCR-Zyklen als optimal gewählt worden. Der ausgeschnittene Gelbereich, aus dem die cDNA-Bibliothek für das Sequenzieren größenselektiert wurde, ist mit einer unterbrochenen Linie gekennzeichnet. M: DNA-Marker, bp: Basenpaare.

### III.3 Das mittels dRNA-Seq identifizierte plastidäre Transkriptom von *A. thaliana*

Mit dem optimierten dRNA-Seq-Protokoll wurden cDNA-Bibliotheken generiert, welche für die Identifizierung von TSS geeignet sind. Eine tabellarische Übersicht über die Sequenzierungsstatistiken der cDNA-Bibliotheken ist im Anhang Tab. VI.8 hinterlegt. Bei der Sequenzierung konnte eine ausreichende Genomabdeckung des Plastoms erreicht werden, um TSS zu detektieren (Anhang Tab. VI.8). Dazu konnten die 5'-Enden der gemappten Reads mit dem Integrated Genome Viewer (IGV) visualisiert werden (Robinson *et al.*, 2011; Thorvaldsdóttir *et al.*, 2013). Die parallele Visualisierung der Genomabdeckungen der +5'PPase- und der -5'PPase-Bibliotheken, die zur gleichen Probe gehörten (Abb. III.9), zeigte 5'-Enden, die TSS darstellen und 5'-Enden, welche offensichtlich von prozessierten Transkripten stammen. Wurden 5'-Enden mit ähnlicher Genomabdeckung in beiden Bibliotheken detektiert (Abb. III.9, grüner Pfeil), stellten diese TEX-resistente Transkripte dar. Wurde ein 5'-Ende in der +5'PPase- detektiert, aber nicht oder mit einer deutlich geringeren Genomabdeckung in der -5'PPase-Bibliothek (Abb. III.9, roter Pfeil), war das mit hoher Wahrscheinlichkeit ein Primärtranskript, welches von einer TSS stammte.





### III.3.1 Erfolgreiche Detektion bekannter TSS

Mittels der in dieser Arbeit optimierten dRNA-Seq-Methode konnten TSS, welche in früheren Publikationen mit anderen Methoden detektiert wurden, erfolgreich nachgewiesen werden. Abb. III.10 zeigt beispielhaft eine TSS, welche vor *clpP* lag und durch die Akkumulation der 5'-Enden in den +5' PPase-Bibliotheken deutlich wurde. Diese TSS wurde bereits als *PclpP-58* in *Arabidopsis* beschrieben (Swiatecka-Hagenbruch *et al.*, 2007) und somit kann bestätigt werden, dass das Protokoll erfolgreich für TSS-Kartierung angewandt werden kann. Die Menge der Reads an der Stelle der TSS in den verschiedenen Genotypen konnte dazu verwendet werden, um die Nutzung der jeweiligen TSS in jedem Genotyp zu bestimmen und zu identifizieren, welche RNA-Polymerase diese TSS für die Transkriptionsinitiation nutzt.

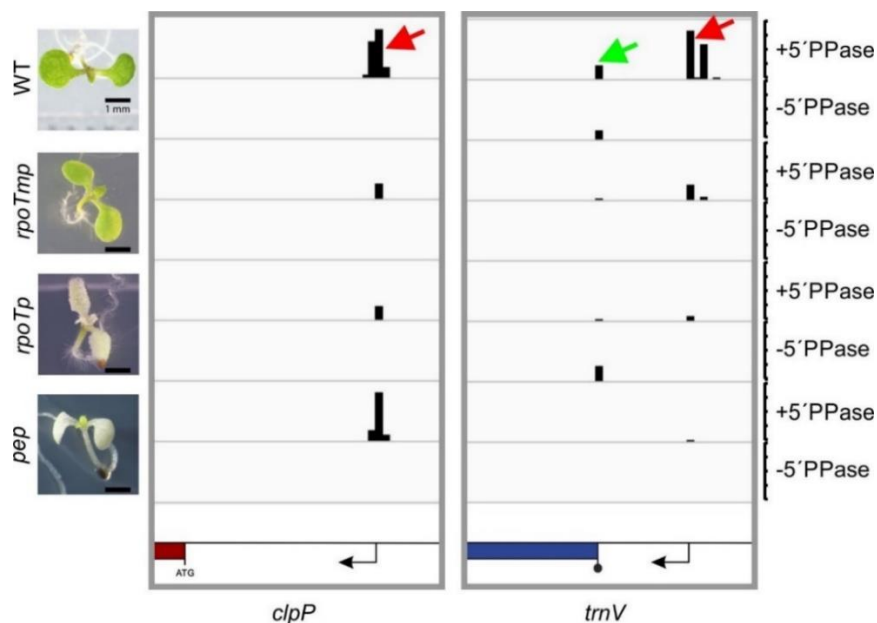


Abbildung III.10: Identifizierte TSS für *clpP* und *trnV*. Angezeigt sind die Genomabdeckungen der +5' PPase- sowie -5' PPase-cDNA-Bibliotheken für die 5'-Enden der Reads mit RPM-Normalisierung für Wildtyp und die RNA-Polymerase-Mutanten im späten Stadium. In dunkelrot ist schematisch ein Teil des *clpP*-Gens mit dem Startcodon ATG angezeigt (links), in dunkelblau ein Teil von *trnV* (rechts). Der rote Pfeil zeigt auf die Akkumulation von einem 5'-Ende, welches auf ein Primärtranskript zurückzuführen ist, der grüne Pfeil zeigt auf die Akkumulation von einem 5'-Ende, welches von einem prozessierten Transkript stammt. Der gebogene Pfeil deutet die TSS an. Die Höhe der Skala ist für alle Genotypen gleichgesetzt. Der Größenbalken zeigt 1 mm an.

Im Bereich der tRNAs wurden in den +5' PPase-Bibliotheken teilweise 5'-Enden detektiert, welche offensichtlich von prozessierten Transkripten stammen (Abb. III.10 rechts, grüner Pfeil). Das deutete darauf hin, dass tRNAs zum Teil TEX-resistent waren. Es war jedoch möglich, zwischen den TEX-resistenten prozessierten Transkripten und den Primärtranskripten zu unterscheiden (Abb. III.10 rechts, grüner und roter Pfeil). Zwei zuvor beschriebene TSS für das *rrn16*-Operon konnten mit der optimierten Methode ebenfalls detektiert werden (Abb. III.11). Die TSS *Prrn16-139* ist RPOTmp-abhängig (Pc-Promotor) (Courtois *et al.*, 2007), die TSS *Prrn16-112* (P2-Promotor) wird von PEP zur Transkription genutzt (Castandet *et al.*, 2019; Courtois *et al.*, 2007; Swiatecka-Hagenbruch *et al.*, 2007).

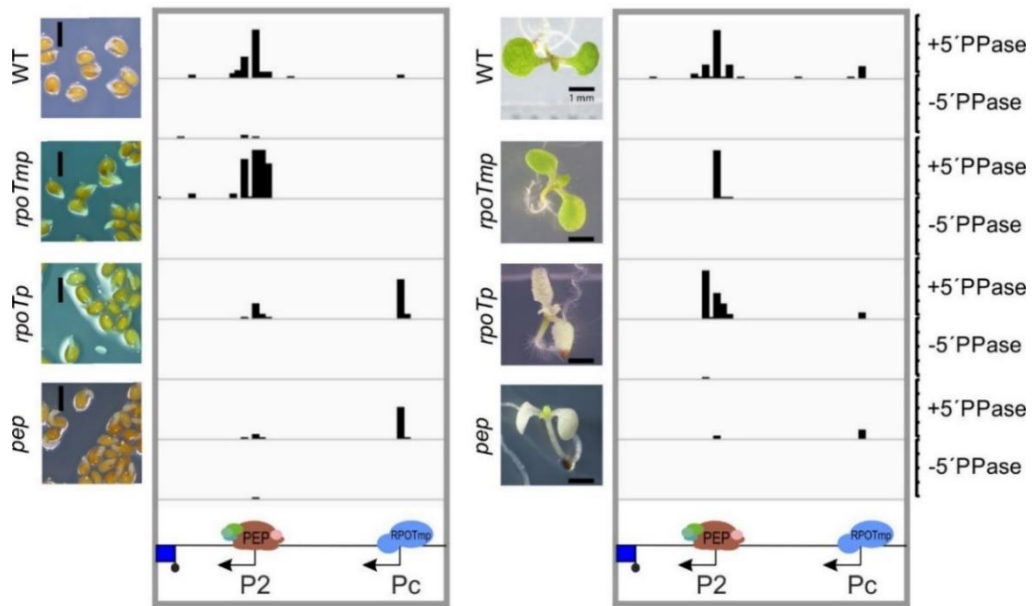


Abbildung III.11: Identifizierte TSS für *rrn16*. Angezeigt sind die Genomabdeckungen der +5'PPase- sowie -5'PPase-cDNA-Bibliotheken für die 5'-Enden der Reads mit RPM-Normalisierung für Wildtyp und die RNA-Polymerase-Mutanten in frühem (links) und spätem (rechts) Stadium. In dunkelblau ist schematisch ein Teil des *rrn16*-Gens gezeigt. Die RNA-Polymerasen sind in braun (PEP) bzw. in blau (RPOTmp) dargestellt. Die Höhe der Skala ist für alle Genotypen gleichgesetzt. Der Größenbalken zeigt 1 mm an.

Die Abhängigkeiten der TSS konnten anhand der Genomabdeckungen ebenfalls bestätigt werden, wobei im frühen Stadium in *rpoTp* eine sekundäre Reduktion der Nutzung der PEP-abhängigen TSS *Prrn16-112* (P2-Promotor) beobachtet wurde (Abb. III.11 links, P2). In der Analyse der frühen Stadien der Samenkeimung wurden auch TSS für Photosynthesegene in den Genomabdeckungen detektiert (Abb. III.12).

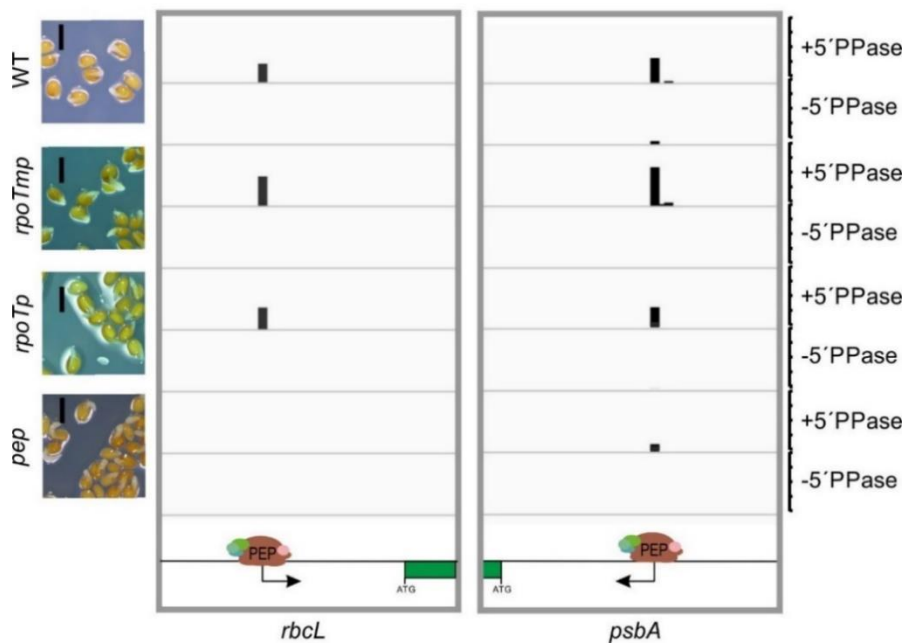


Abbildung III.12: Identifizierte TSS für *rbcL* und *psbA*. Angezeigt sind die Genomabdeckungen der +5'PPase- sowie -5'PPase-cDNA-Bibliotheken für die 5'-Enden der Reads mit RPM-Normalisierung für Wildtyp und die RNA-Polymerase-Mutanten

im frühen Stadium. In grün ist schematisch ein Teil des Gens mit dem Startcodon ATG gezeigt. Der Pfeil deutet die TSS an. Die Höhe der Skala ist für alle Genotypen gleichgesetzt. Der Größenbalken zeigt 1 mm an.

Für *psbA* und *rbcL* konnten zuvor beschriebene TSS *PpsbA*-77 und *PrbcL*-179 nachgewiesen werden (Hanaoka *et al.*, 2003; Liere *et al.*, 1995; Shen *et al.*, 2001; Swiatecka-Hagenbruch *et al.*, 2007; Zhelyazkova *et al.*, 2012b). Ausgehend aus der Höhe der Genomdeckung an der TSS kann für beide Gene die PEP-Abhängigkeit festgestellt werden.

### III.3.2 Plastomweite Identifikation von TSS

Im Gegensatz zur manuellen Identifizierung der TSS in der genomweiten Analyse des Transkriptom von Gerste (Zhelyazkova *et al.*, 2012b) wurden in dieser Arbeit bioinformatische Methoden verwendet, um eine unvoreingenommene Identifizierung der TSS zu gewährleisten und die Detektion von falsch-positiven TSS zu vermeiden. Für die Identifikation der TSS wurde die Anreicherung der 5'-Enden in der +5'PPase-Bibliothek im Vergleich zu der zugehörigen -5'PPase-Bibliothek untersucht, wobei die Daten mittels DESeq2 normalisiert wurden. Diese Methode wurde ausgewählt, da diese Art von Normalisierung für die Mengen der Gesamtreads der cDNA-Bibliotheken (Sequenziertiefe) sowie die Länge der Gene ausgleicht und die Verzerrung, die durch die RNA-Zusammensetzung entstehen könnte, reduziert. Diese Verzerrung der Daten kann auftreten, wenn zum Beispiel nur ein kleiner Teil von Genen in einer Probe, aber nicht in einer anderen, sehr hoch exprimiert ist (Anders and Huber, 2010; Love *et al.*, 2014). Zur Reduktion des Hintergrundrauschens und Stabilisierung der Daten wurden die Genomabdeckungen in den -5'PPase-Bibliotheken auf jeweils 5 Reads erhöht. TSS wurden als wahr akzeptiert, wenn sie in mindestens einem Genotyp als signifikant angereichert zwischen den +PPase und -PPase-Bibliotheken identifiziert wurden. Als Ausschlusskriterien wurde eine Falscherkennungsrate (FDR, *false discovery rate*) von kleiner als 0,05 und ein Unterschied zwischen der +PPase und der -PPase-Bibliothek von mindestens  $\log_2FC=1,5$  festgelegt. Jede TSS wurde dem Gen zugewiesen, welches sich dieser TSS im Genom auf dem gleichen Strang am nächsten befindet. Trotz der erhöhten Menge an TEX-Enzym wurde eine sichtbare Menge an rRNA in beiden Bibliotheken detektiert (Abb. III.9, in blau umrandet). Durch die Akkumulation der Reads im rRNA-Bereich wurden hier TSS identifiziert, welche offensichtlich falsch-positiv sind. Aus diesem Grund wurde bei der Identifizierung der TSS eine Modifizierung vorgenommen und die TSS aus der Analyse ausgeschlossen, welche innerhalb der rRNA-Gene liegen.

Die Liste der identifizierten TSS mit Referenzen zu bereits beschriebenen TSS befindet sich im Anhang Tab. VI.10. Die Nomenklatur der TSS wurde folgendermaßen angegeben: Genomposition, Strang, Zugehörigkeit zum Gen, z. B. 1522-PSBA. Da die bereits in der Literatur beschriebenen TSS üblicherweise die Position des initiiierenden Nukleotids im Verhältnis zum Translationsstartpunkt oder zur reifen rRNA angeben, wurde in der TSS-Liste für die Referenzen die Alternativschreibweise mit angegeben, sowie im Weiteren in Klammern genannt.

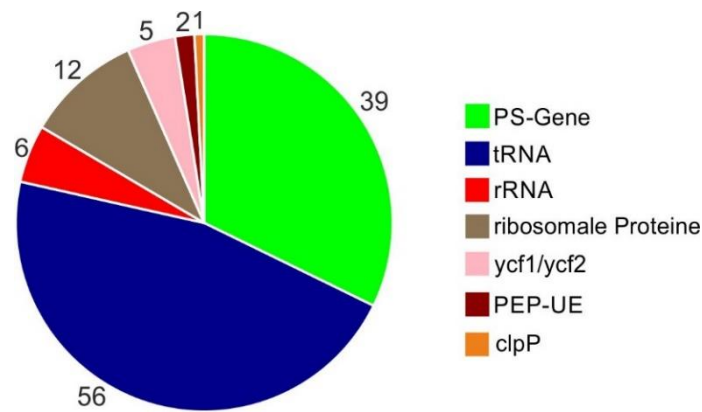


Abbildung III.13: Aufteilung der 121 identifizierten TSS nach Genklassen. Die Zahlen geben die Anzahl der TSS für jede Genart wieder. Die Identifizierung der TSS wurde von Frederic Ehlers unter Anleitung von Julia Rott und Dermot Harnett durchgeführt.

Insgesamt konnten 121 TSS identifiziert werden (Abb. III.13). 13 davon waren doppelt vorhanden, da sie zu den Genen in der *inverted repeat*-Region gehören. Das *A. thaliana*-Genom hat insgesamt 129 Gene, davon wurden für 69 Gene TSS gefunden, was etwa 54 % entsprach. Für 28 Gene (40%) wurde mehr als eine TSS detektiert. Die meisten TSS pro Gen (jeweils 5 TSS) wiesen *atpB* und *ndhF* auf. Die meisten TSS wurden für tRNAs (56 TSS) sowie photosytheserelevante Gene (39 TSS) gefunden. Hier wurden für tRNAs zwischen einer und vier TSS pro tRNA identifiziert. Im Gegensatz zu anderen Publikationen (Liere *et al.*, 2004; Zhelyazkova *et al.*, 2012b) wurden keine internen TSS für tRNAs detektiert. Für die meisten tRNAs lagen die identifizierten TSS stromaufwärts nah an der kodierenden Region.

### III.4 Charakterisierung der plästidären Promotoren

#### III.4.1 Methoden zur Klassifikation der TSS

Für die Klassifizierung der TSS, welche die Zuordnung zu einer (oder mehreren) Polymeraseaktivitäten erlauben würde, wurde ein statistisches Modell ausgearbeitet, welches sich an ein publiziertes Modell (Weinholdt *et al.*, 2019) anlehnt. Es wurden 8 theoretische Fälle als Möglichkeiten definiert, wie sich eine TSS in Abhängigkeit einer RNA-Polymerase verhalten könnte. Während die Fälle 2, 3 und 5 die Abhängigkeit der TSS von nur jeweils einer RNA-Polymerase (PEP, RPOTp bzw. RPOTmp) definierten, war die Transkription von den TSS in Fällen 4, 6 und 7 von zwei RNA-Polymerasen abhängig (Abb. III.14). In Fall 1 waren TSS definiert, die von keiner Polymeraseaktivität abhängig sind, während Fall 8 die Abhängigkeit der TSS von allen drei Polymerasen zeigte.

## Ergebnisse

| Polymerase         | Fall 1 | Fall 2 | Fall 3 | Fall 4 | Fall 5 | Fall 6 | Fall 7 | Fall 8 |
|--------------------|--------|--------|--------|--------|--------|--------|--------|--------|
| PEP                | —      | +      | —      | +      | —      | +      | —      | +      |
| RPOT <sub>p</sub>  | —      | —      | +      | +      | —      | —      | +      | +      |
| RPOT <sub>mp</sub> | —      | —      | —      | —      | +      | +      | +      | +      |

| TSS             | Fall 1 | Fall 2 | Fall 3 | Fall 4 | Fall 5 | Fall 6 | Fall 7 | Fall 8 |
|-----------------|--------|--------|--------|--------|--------|--------|--------|--------|
| 109231-TRNN.1   | 0,06   | 93,49  | 0,02   | 6,37   | 0,01   | 0,02   | 0,00   | 0,03   |
| 54781+RBCL      | 0,17   | 76,85  | 0,05   | 22,65  | 0,03   | 0,08   | 0,02   | 0,15   |
| 109366+TIC214   | 0,05   | 0,01   | 99,67  | 0,09   | 0,01   | 0,00   | 0,13   | 0,03   |
| 66089+PSAJ      | 1,28   | 9,43   | 0,54   | 87,63  | 0,15   | 0,36   | 0,15   | 0,46   |
| 137776-RRN16S.2 | 0,61   | 0,15   | 0,09   | 0,01   | 97,97  | 0,81   | 0,21   | 0,16   |
| 71940-CLPP1     | 16,17  | 7,13   | 11,03  | 37,84  | 4,24   | 4,70   | 5,13   | 13,76  |
| 20245-RPOC2     | 30,05  | 8,84   | 15,46  | 15,97  | 6,71   | 5,77   | 8,54   | 8,67   |

|  |          |
|--|----------|
|  | 0-29 %   |
|  | 30-49 %  |
|  | 50-59 %  |
|  | 60-69 %  |
|  | 70-79 %  |
|  | 80-100 % |

Abbildung III.14: Statistisches Modell zur Klassifizierung der TSS nach Nutzung durch eine RNA-Polymerase. Es wurden 8 Fälle unterschiedlicher Nutzung einer TSS von der jeweiligen Polymerase definiert. Das Plus zeigt eine Abhängigkeit, das Minus keine Abhängigkeit einer TSS von der jeweiligen RNA-Polymerase. Für jede TSS wurde für jeden der 8 Fälle eine *a-posteriori*-Wahrscheinlichkeit berechnet. Beispiele für die eindeutige sowie nicht eindeutige Zuweisung einiger TSS im frühen Stadium sind angegeben, wobei die Prozentwerte farbig unterlegt sind (unten). Rot: hohe Wahrscheinlichkeit für diesen Fall (60-100 %), blau: niedrige Wahrscheinlichkeit (0-59 %). Die Klassifizierung der TSS wurde von Frederic Ehlers unter Anleitung von Julia Rott und Ivo Große durchgeführt.

Für die Klassifizierung wurden die Genomabdeckungen für die zuvor identifizierten TSS-Positionen der +5 PPase-Bibliotheken mit DESeq2-Normalisierung verwendet. Es wurde ein separates Klassifizieren für alle TSS für das frühe und das späte Stadium durchgeführt. Für jede TSS wurde eine *a-posteriori*-Wahrscheinlichkeit in Prozent berechnet, mit welcher die TSS jedem der 8 Fälle zugeordnet werden kann. Anhand dieser Prozentwerte konnte die Entscheidung getroffen werden, ob eine TSS zu einem Fall eindeutig zugewiesen und so die Nutzung der TSS durch die verschiedenen RNA-Polymerasen identifiziert werden kann. Ein höher prozentualer Anteil zeigt darauf, dass eine TSS mit hoher Sicherheit in diesen Fall zugeordnet wird; ein niedriger Wert zeigt auf eine TSS, deren Zuordnung nicht möglich ist (Abb. III.14 unten). Die Liste mit den Wahrscheinlichkeiten für jeden Fall ist Anhang Tab. VI.11 und VI.12. zu entnehmen.

Unterstützend zum statistischen Modell wurden die Daten mittels der pamK-Methode geclustert (Gentle *et al.*, 1991). Die pamK-Methode wurde ausgewählt, weil sie ein unvoreingenommenes Clustern ohne die Festsetzung der Clusteranzahl ermöglichte und für Extremwerte ausglich. Auch hier wurden Genomabdeckungen für die zuvor identifizierten TSS-Positionen der +5 PPase-Bibliotheken mit DESeq2-Normalisierung verwendet, aus welchen die logarithmisch transformierten relativen Transkriptmengen in den Mutanten im Vergleich zum Wildtyp berechnet wurden. Mit pamK wurden die TSS anhand dieser Werte in Cluster zusammengefasst. Die berechneten relativen Transkriptmengen in den Mutanten im Vergleich zum Wildtyp sind in Anhang Tab. VI.13 zusammengefasst, die grafische Darstellung der Cluster in Anhang Abb. VI.2 und VI.3 zu finden.

### III.4.2 TSS-Zuweisung nach Abhängigkeit von RNA-Polymeraseaktivität

Die Annahmen für das Verhalten der TSS und die Fälle im statistischen Modell erwiesen sich größtenteils als korrekt, sodass die Abhängigkeit der TSS von Polymeraseaktivitäten identifiziert werden konnte. Die eindeutige Zuweisung einer TSS zu einem Fall wurde bei einer Wahrscheinlichkeit von mindestens 60 % vorgenommen. Die final zugewiesenen TSS mit Kennzeichnung der zugehörigen Polymerase sowie der umgebenden Sequenz sind dem Anhang Tab. VI.14 zu entnehmen. Die meisten TSS in beiden Stadien konnten eindeutig zugeordnet werden. Zu Fall 1 und 8 konnten keine TSS mit der genannten Wahrscheinlichkeit von mind. 60 % zugewiesen werden. Ebenso wurden zu Fall 7 keine TSS zugeordnet, was unerwartet war, da anhand von früheren Publikationen von einer teilweise überlappenden Funktion der beiden NEPs ausgegangen worden ist (Courtois *et al.*, 2007; Swiatecka-Hagenbruch *et al.*, 2008). Die eindeutige Zuordnung der TSS zu einer Polymerasenaktivität (Fälle 2, 3 oder 5) wurde akzeptiert, wenn die TSS in mindestens einem Stadium diesem Fall zugewiesen wurde. Insgesamt konnten in beiden Stadien 73 TSS eindeutig zu einer Polymeraseaktivität zugeordnet werden (Abb. III.15). Die meisten TSS wurden als PEP-abhängig identifiziert. Hier wurden 65 TSS detektiert, welche zumeist zu Photosynthesegenen und tRNAs gehörten. Als RPOTp-abhängig wurden 7 TSS identifiziert, die zu mehreren Genklassen, unter anderem zu ribosomalen Proteinen, gehörten. Der RPOTmp-Aktivität wurde eine TSS zugewiesen, welche vor einer rRNA lag.

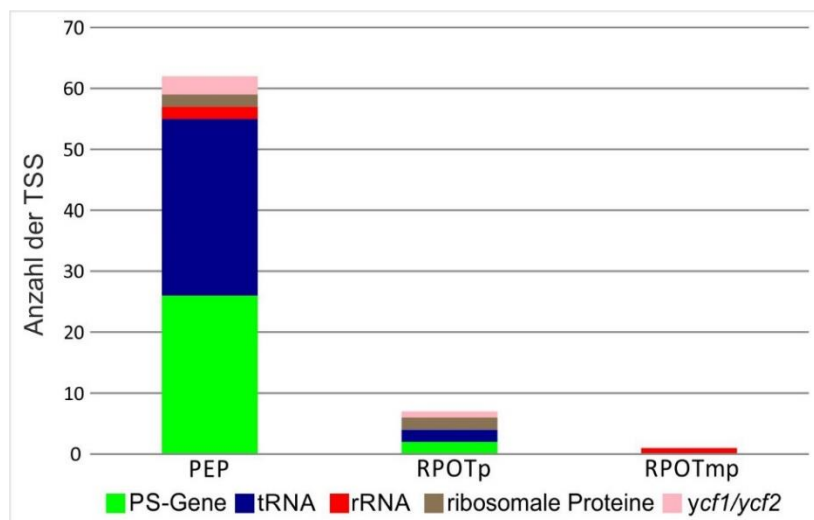


Abbildung III.15: Zuordnung der TSS zu Polymeraseaktivitäten. Die 73 mit dem statistischen Modell eindeutig zugewiesenen TSS in frühem und spätem Stadium eingeteilt nach Genklassen.

### III.4.3 Erschwerte Zuordnung von TSS im späten Stadium

Im späten Stadium konnten 45 der identifizierten TSS keinem Fall zugeordnet werden. Das betraf vor allem TSS für tRNAs und Gene für ribosomale Proteine. Da nur ab einer bestimmten Datentiefe TSS extrahiert werden können, wurde angenommen, dass diese zu tRNA und Genen für ribosomale Proteine gehörenden TSS in grünen Pflanzenstadien wenig genutzt werden und der quantitative Unterschied zur Nutzung der TSS in Wildtyp nicht groß genug ist, um zu einem polymerasespezifischen Fall

zugewiesen zu werden. Die Untersuchung der Nutzung dieser TSS in Wildtyp (Abb. III.16) konnte die Annahme größtenteils bestätigen. Die meisten TSS, welche im späten Stadium nicht zugewiesen werden konnten, zeigten eine höhere Genomabdeckung im frühen Stadium von Wildtyp. So wurde für diese TSS die Zuordnung im frühen Stadium akzeptiert.

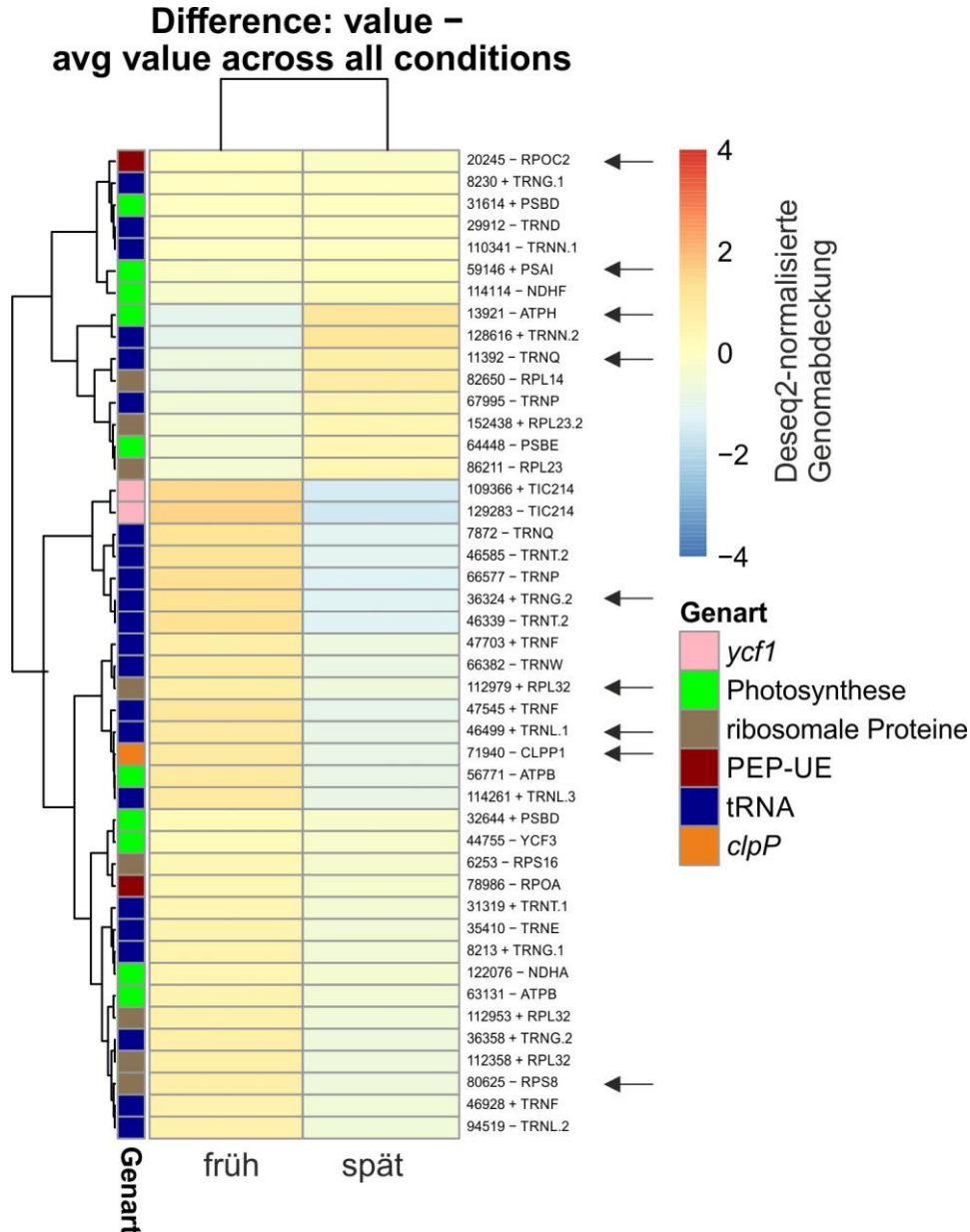


Abbildung III.16: Nutzung der im späten Stadium nicht zugewiesenen 45 TSS im frühen und späten Stadium von Wildtyp. Angezeigt sind die Genomabdeckungen der +5 PPase-cDNA-Bibliotheken mit DESeq2-Normalisierung. Pfeile markieren die TSS, welche in beiden Entwicklungsstadien nicht zugewiesen werden konnten (9 TSS). Die Analyse wurde durch Frederic Ehlers unter Anleitung von Julia Rott und Dermot Harnett durchgeführt.

Beim Hinzuziehen der pamK-Analyse zeigte sich, dass von den betroffenen 45 TSS im späten Stadium mehr als die Hälfte (25 TSS) in Cluster eingeordnet wurden, welche in keiner der Mutanten eine reduzierte Nutzung aufwiesen (Cluster 1 und 4, Anhang Abb. VI.3). Für insgesamt 9 TSS konnte das statistische Modell eine Zuordnung weder im frühen noch im späten Stadium erreichen. Die Nutzung

dieser TSS in Wildtyp (Abb. III.16, mit Pfeilen gekennzeichnet) zeigte, dass die meisten TSS im frühen Stadium intensiver als Transkriptionsstarts verwendet werden. Da keine TSS zu Fall 7 eingeordnet werden konnten, welcher die Abhängigkeit einer TSS von sowohl RPOT<sub>p</sub> als auch RPOT<sub>mp</sub> aufzeigt, wurde vermutet, dass bei diesen TSS die beiden NEPs eine überlappende Funktion haben könnten, da sie nach jetzigem Wissen die gleichen Promotormotive erkennen (Anhang Tab. VI.14, bezeichnet als „potenziell NEP“). Die pamK-Cluster-Analyse unterstützte diese Vermutung. 8 der 9 TSS fanden sich in einem möglichen Überlapp-Cluster im frühen oder späten Stadium (Anhang Abb. VI.2 Cluster 4, und Abb. VI.3, Cluster 6).

### III.5 Validierung der TSS-Analyse mit RNA-seq

Zur Validierung der Ergebnisse wurde zusätzlich zur TSS-Identifizierung eine RNA-Seq-Analyse durchgeführt, bei welcher die kompletten Transkriptome von Wildtyp und den RNA-Polymerase-Mutanten im frühen und späten Stadium untersucht wurden (Kap. II.2.10). Diese Methode wurde als eine Alternative zur RT-qPCR gewählt, da mit RNA-Seq monocistronische und polycistronische Transkripte separat analysiert werden können. Die statistischen Daten zur Sequenzierung sind Tab. VI.5 im Anhang zu entnehmen. Die Sequenzierungsdaten wurden für eine differentielle Expressionsanalyse verwendet, um den Einfluss der einzelnen TSS auf die Transkription eines Gens oder Operons zu untersuchen und die Klassifizierung der TSS nach Polymeraseaktivität mit einer alternativen Methode zu bestätigen (Anhang Tab. VI.6 und VI.7). Für die differentielle Expressionsanalyse wurde eine DESeq2-Normalisierung der Daten vorgenommen. Zusätzlich wurden aus den RNA-Seq-Daten Genomabdeckungen des Plastoms generiert, damit die Transkriptionsaktivität von verschiedenen TSS visualisiert und der Einfluss der Nutzung dieser TSS auf die Transkription untersucht werden konnte. Die grafische Darstellung der differentiellen Expressionsanalyse ist dem Anhang Abb. VI.5 zu entnehmen.

### III.6 Funktion von RPOT<sub>p</sub> und RPOT<sub>mp</sub> in plastidärer Transkription

#### III.6.1 Bekannte NEP-abhängige TSS konnten nicht identifiziert werden

In früheren Arbeiten wurden einige NEP-abhängige Promotoren mittels 5' RACE und *Primer Extension* identifiziert, welche im Laufe dieser Arbeit trotz der Optimierung des Protokolls für die Identifizierung von niedrig transkribierten NEP-abhängigen Genen nicht gefunden werden konnten. Für *rps4* wurde die TSS *Prps4-151* beschrieben (Swiatecka-Hagenbruch *et al.*, 2008, 2007). Die Visualisierung im Browser zeigte zwar eine Genomabdeckung in der +5' PPase-cDNA-Bibliothek, jedoch schien diese zu gering zu sein, um im Vergleich zur -5' PPase-cDNA-Bibliothek signifikant angereichert zu sein (Abb. III.17). Im späten Stadium von Wildtyp betrug die normalisierte Genomabdeckung lediglich 7 Reads. Im frühen Stadium konnte für diese TSS keine Genomabdeckung festgestellt werden. Für andere als NEP-abhängig bekannte TSS wie *PrpoB-300* und *PaccD-252* (Hess *et al.*, 1993; Liere *et al.*, 2004;



Swiatecka-Hagenbruch *et al.*, 2008) konnte keine Genomabdeckung in keinem der untersuchten Entwicklungsstadien detektiert werden.

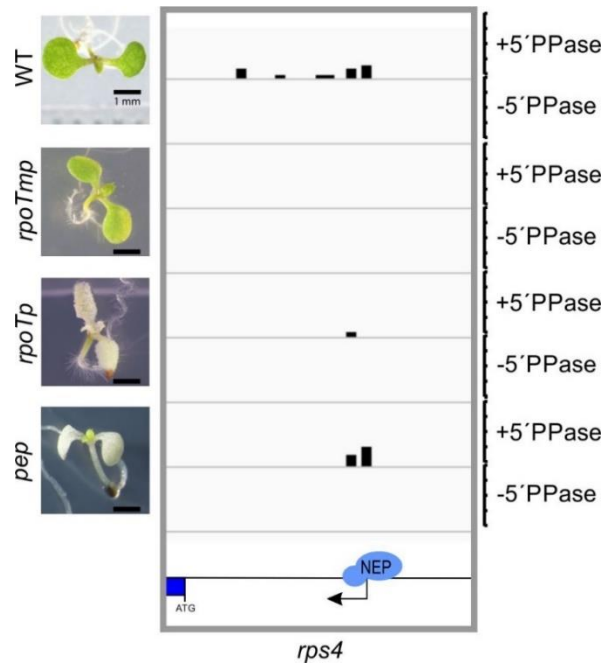


Abbildung III.17: Nicht identifizierte TSS für *rps4*. Angezeigt sind die +5' PPase- sowie -5' PPase-cDNA-Bibliotheken für die 5'-Enden der Reads mit RPM-Normalisierung für Wildtyp und die RNA-Polymerase-Mutanten im späten Stadium. In blau ist schematisch ein Teil des *rps4*-Gens mit dem Startcodon ATG angezeigt. Der Pfeil deutet die TSS an. Die Höhe der Skala ist in allen Genotypen gleichgesetzt. In blau ist eine Phagentyp-RNA-Polymerase abgebildet. Der Größenbalken zeigt 1 mm an.

### III.6.2 RPOTp hat eine genspezifische Funktion

Insgesamt konnten 7 TSS (ohne die doppelten TSS aus den *inverted repeats* zu betrachten) als RPOTp-abhängig identifiziert werden. Abb. III.18 zeigt die mit schwarzen Kreisen markierten TSS, welche von der RPOTp-Aktivität abhängig sind.

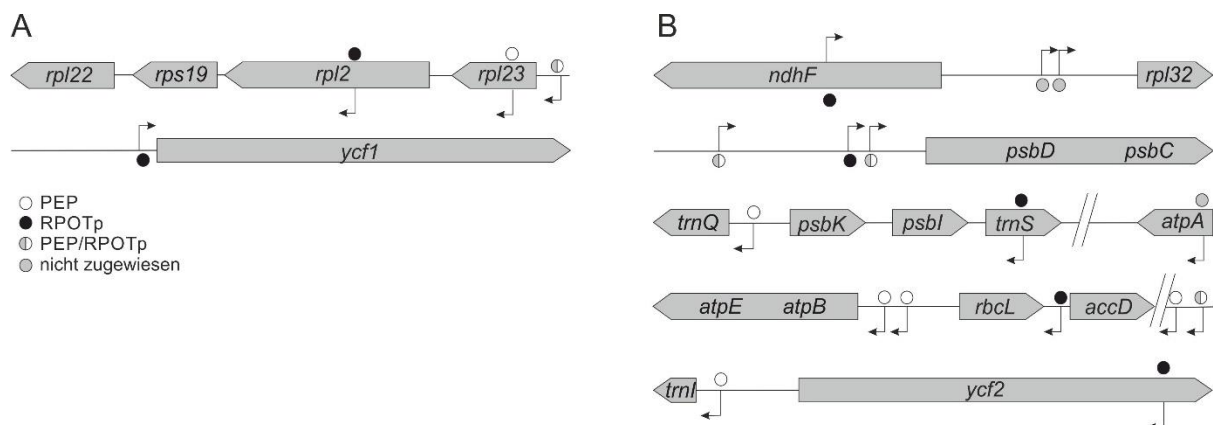


Abbildung III.18: Schematische Darstellung der Gene mit RPOTp-abhängigen TSS. A: Gene, die von einer exklusiven RPOTp-TSS transkribiert werden. B: Gene, die neben einer RPOTp-TSS auch andere TSS besitzen. Gene sind in als graue Balken dargestellt. Der Pfeil stellt die TSS dar. Die Kreise stellen die Abhängigkeit der TSS von einer Polymeraseaktivität dar: Die diagonalen Doppelstriche zeigen die Aussparung eines Genomteils auf. Die Abbildung ist nicht maßstabsgerecht.

Bis auf die TSS für *ycf1* (109366+TIC214.1 bzw. 129283-TIC214.2) und *rps19* bzw. für das identische Genomstück auf dem *inverted repeat* (85045-RPS19 und 153604+PSBK) (Abb. III.18 A) sind die anderen TSS nicht exklusiv für die gegebenen Gene/Operons. Die meisten Gene haben neben der RPOTp-abhängigen TSS auch eine oder mehrere zusätzliche TSS (Abb. III.18 B).

### III.6.2.1 Transkription von *ycf1* (*tic214*)

Die identifizierte TSS 109366+TIC214 wurde bereits als eine NEP-abhängige TSS *Pycf1*-39 beschrieben worden (Swiatecka-Hagenbruch *et al.*, 2008, 2007), es fehlte soweit jedoch eine eindeutige Zuordnung zur RPOTp- bzw. RPOTmp-Aktivität. Teilweise wurden in diesen Publikationen zusätzliche TSS *Pycf1*-34 und *Pycf1*-104 für dieses Gen gefunden. Das optimierte dRNA-Seq (Kap. III.2) identifizierte nur eine TSS 109366+TIC214 für *ycf1* und das statistische Modell wies im frühen Stadium die TSS mit hoher Wahrscheinlichkeit der RPOTp-Aktivität zu (Anhang Tab. Tab. VI.11). Die RNA-Seq-Analyse von *rpoTp* zeigte die signifikante Reduktion der Transkriptmengen nur im frühen Stadium, während im späten Stadium erhöhte Mengen detektiert worden sind (Abb. III.19).

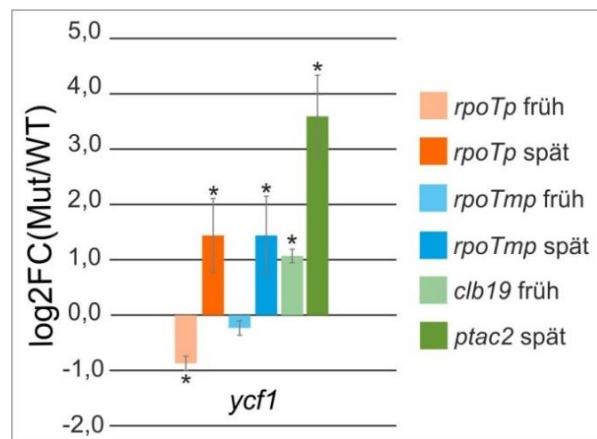


Abbildung III.19: RNA-Seq-Analyse von *ycf1* in den Mutanten für das frühe und späte Stadium dargestellt als logarithmierte relative Transkriptlevel bezogen auf WT mit DESeq2-Normalisierung. Der Fehlerbalken zeigt den Standardfehler an. \* Wald-Test mit p-Wert  $\leq 0,05$ . Die Expressionsanalyse wurde von Jose Muino durchgeführt.

Beim Anschauen der Genomabdeckung in Wildtyp im Bereich der identifizierten TSS zeigte sich ein deutlicher Unterschied der Transkriptmengen im frühen (Abb. III.20 A) und späten (Abb. III.20 B) Stadium. Die TSS wurde im nicht-grünen Gewebe stark zur Transkription genutzt, während im grünen Gewebe so gut wie keine Genomabdeckung an der TSS detektiert werden konnte. Im frühen Stadium war nur eine TSS sichtbar, im späten Stadium wurde in *rpoTp* eine alternative TSS genutzt (Abb. III.20 B, roter Pfeil). In der vorliegenden dRNA-Seq-Analyse wurde diese nicht als TSS identifiziert, entsprach jedoch der zuvor beschriebenen *Pycf1*-34, die als eine PEP-abhängige TSS dargestellt wird (Swiatecka-Hagenbruch *et al.*, 2008, 2007). Die Genomabdeckung deutete darauf hin, dass die zusätzliche TSS nur in Abwesenheit von RPOTp genutzt wird und in Wildtyp die *ycf1*-Transkription nur von der Aktivität von RPOTp abhängig ist.

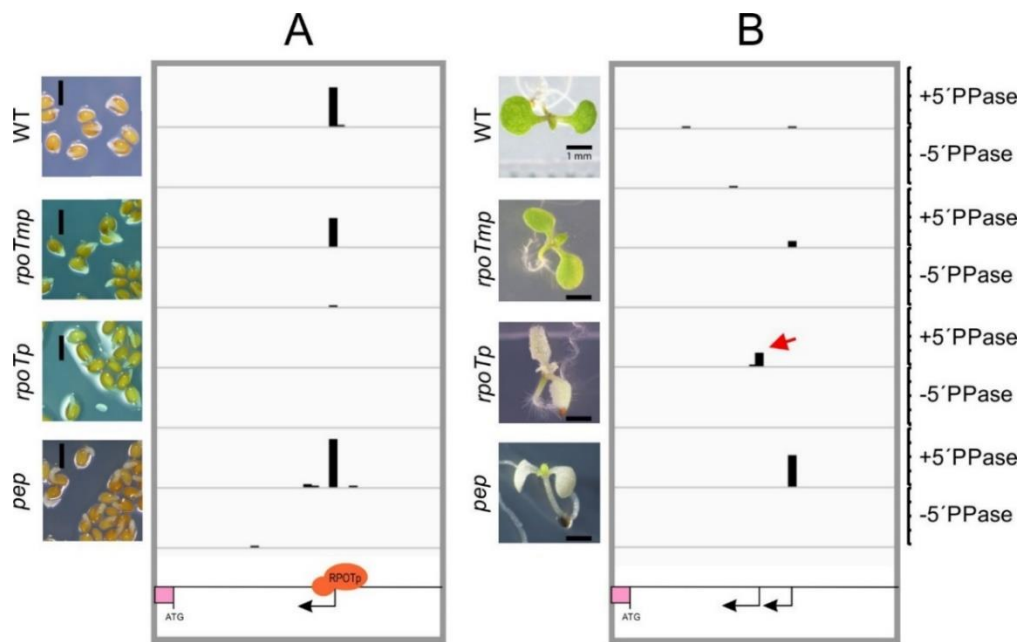


Abbildung III.20: Identifizierte TSS für *ycf1*. Angezeigt sind die +5'PPase- sowie -5'PPase-cDNA-Bibliotheken für die 5'-Enden der Reads mit RPM-Normalisierung für Wildtyp und die RNA-Polymerase-Mutanten im frühen (A) und späten (B) Stadium. In pink ist schematisch ein Teil des *ycf1*-Gens mit dem Startcodon ATG angezeigt. Der gebogene Pfeil deutet die TSS an, welche durch die Akkumulation der 5'-Enden der Reads an dieser Stelle identifiziert werden kann. Der rote Pfeil zeigt auf eine zusätzliche TSS, welche nur in *rpoTp* detektiert wird. Die Höhe der Skala ist in allen Proben gleichgesetzt. Der Größenbalken zeigt 1 mm an. Die RNA-Polymerase RPOtp ist in orange dargestellt.

### III.6.2.2 Transkription des *rpl23*-Operons

Da *rps19* innerhalb eines Operons mit Genen für ribosomale Proteine liegt, ist hier die Rolle von dem RPOtp-abhängigen Promotor vermutlich nicht bedeutend. Hier liegen vor der *rps19*-TSS zwei weitere TSS im Operon (Abb. III.21 B), der Transkriptionsstart für das komplette Operon in Wildtyp liegt laut der Genomabdeckung aus der RNA-Seq-Analyse vor *rpl23* (Abb. III.21 B). Es umfasst 11 Gene, 10 für ribosomale Proteine und ein Gen für die PEP-Untereinheit *rpoA*. Für das Operon wurden 6 TSS identifiziert. Die TSS, an der die Transkription des Operons beginnt (86211-RPL23), wurde als früh genutzte TSS nur im frühen Stadium zu Fall 4 (PEP/RPOtp) zugewiesen und könnte folglich entweder von beiden Polymerasen genutzt werden oder eine reine PEP-abhängige TSS mit sekundärer Reduktion der Nutzung in *rpoTp* sein. Im späten Stadium konnte die TSS zu keiner Polymeraseaktivität zugeordnet werden. Das gleiche Ergebnis lieferte die pamK-Klassifizierung (Anhang Abb. VI.2 und VI.3).

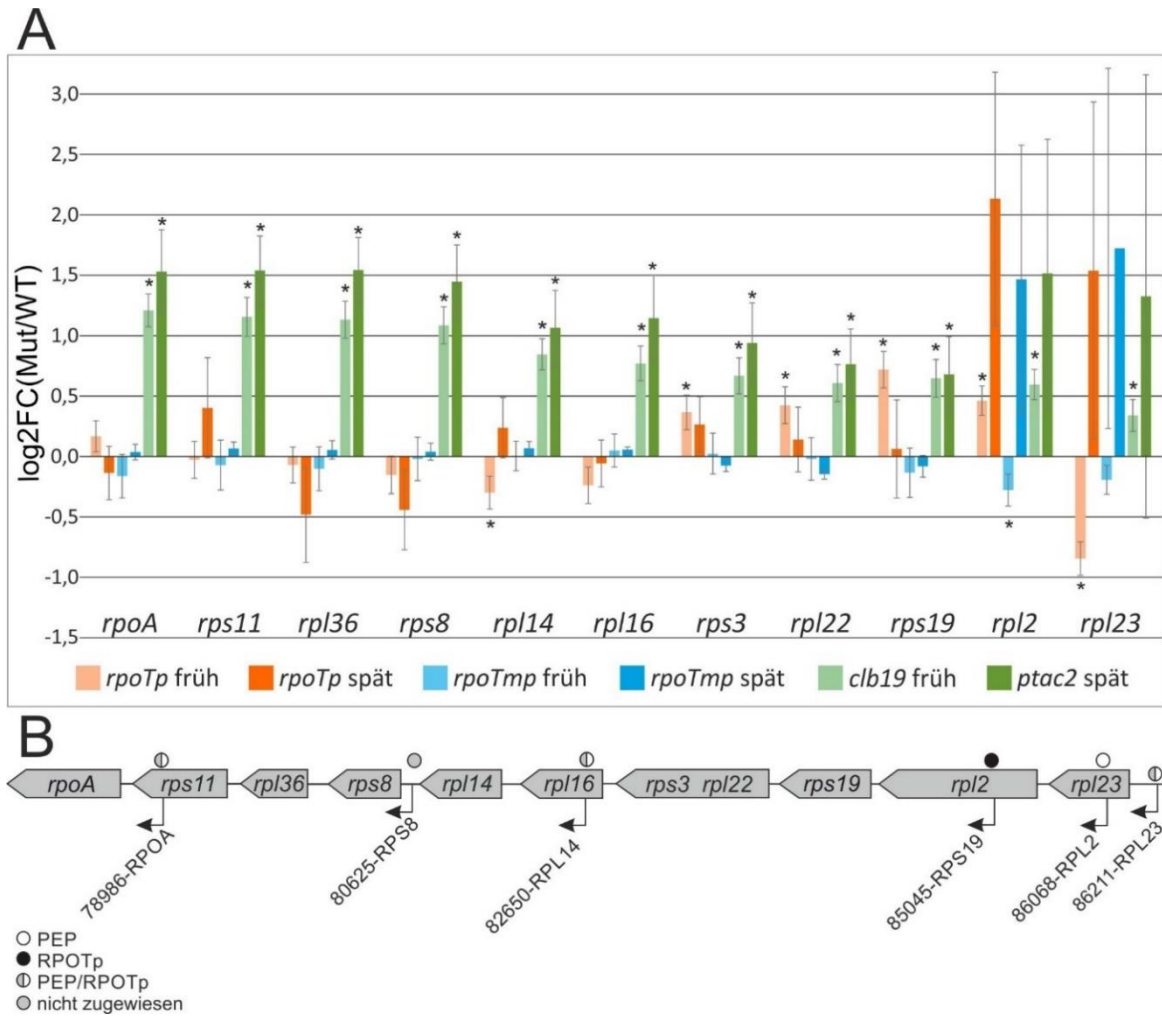


Abbildung III.21: Analyse des *rpl23*-Operons. A: RNA-Seq-Analyse von *rpl23* in den Mutanten für das frühe und späte Stadium dargestellt als logarithmierte relative Transkriptlevel bezogen auf WT mit DESeq2-Normalisierung. Der Fehlerbalken zeigt den Standardfehler an. \* Wald-Test mit p-Wert  $\leq 0,05$ . Die Expressionsanalyse wurde von Jose Muino durchgeführt. B: Schematische Darstellung des *rpl23*-Operons. Gene sind als graue Balken dargestellt. Die gebogenen Pfeile stellt die TSS dar, unter dem Pfeil ist die TSS benannt. Die Kreise stellen die Abhängigkeit der TSS von einer Polymesarenaktivität dar. Die Abbildung ist nicht maßstabsgerecht.

Die RNA-Seq-Analyse des Operons (Abb. III.21 A) war mit diesem Ergebnis nicht übereinstimmend. Für alle Gene des Operons war in beiden PEP-Mutanten die Expression erhöht, während beide NEP-Mutanten eine teilweise Reduktion aufwiesen, welche jedoch nicht konsistent für alle Gene galt. Die Schlussfolgerung war, dass dieses Operon nicht PEP-abhängig ist, obwohl eine PEP-abhängige TSS (86068-RPL2) identifiziert worden ist. Jedoch konnten die individuellen Rollen von RPOTp und RPOTmp nicht exakt bestimmt werden. Für die meisten Gene könnte hier eine überlappende Rolle vermutet werden, bei welcher sich die Polymerasen mehr oder weniger gut ergänzen können. Neben den identifizierten TSS könnten hier auch weitere TSS aktiv sein, welche nicht detektiert worden sind. Auch die unterschiedliche Nutzung der TSS im frühen und späten Stadium (auch durch unterschiedliche Polymerasen) ist denkbar, da in den NEP-Mutanten die Transkriptmengen einzelner Gene zum Teil nur in einem Stadium reduziert/erhöht sind.

Da für *rpl23* und *rpl2* die Erhöhung der Transkriptmengen in *ptac2* nicht signifikant war, konnte hier nicht eindeutig bestimmt werden, ob diese Mutante eine erhöhte Transkriptmenge aufweist. Als zusätzlicher Test wurde eine Northern Blot-Analyse des Operons mit einer *rpl23*-spezifischen Sonde durchgeführt (Abb. III.22). Die Beladung war anhand der Methylenblaufärbung für die Mutanten gleich, für Wildtyp etwas geringer, was eine leichte Schwächung des Signals zur Folge hatte. Die Größen der Transkripte passten zur Transkription von der identifizierten TSS 86211-RPL23 und zeigten reduzierte Mengen in *rpoTp*. Das deutete die Beteiligung von RPOTp bei der Transkription dieses Operons an, aber auch zu einem geringen Grad eine Kompensation von RPOTmp in dieser Mutante. Bei *rpoTmp* war keine Transkriptreduktion sichtbar, sodass hier die Aktivität von RPOTp alleine zur Transkription ausreichte. Die Schlussfolgerung hier wäre, dass in Wildtyp die Transkription des Operons alleine von der Aktivität von RPOTp abhängig ist und RPOTmp nur in Abwesenheit von RPOTp geringe Transkriptionslevel von der gleichen TSS produzieren kann. Die Reduktion des Transkripts der Größe 2,5 kb (a) (Abb. III.22) in *rpoTp* war nur schwach ausgeprägt. So kann auch spekuliert werden, dass in dieser Mutante ein Prozessierungsdefekt auftreten könnte.

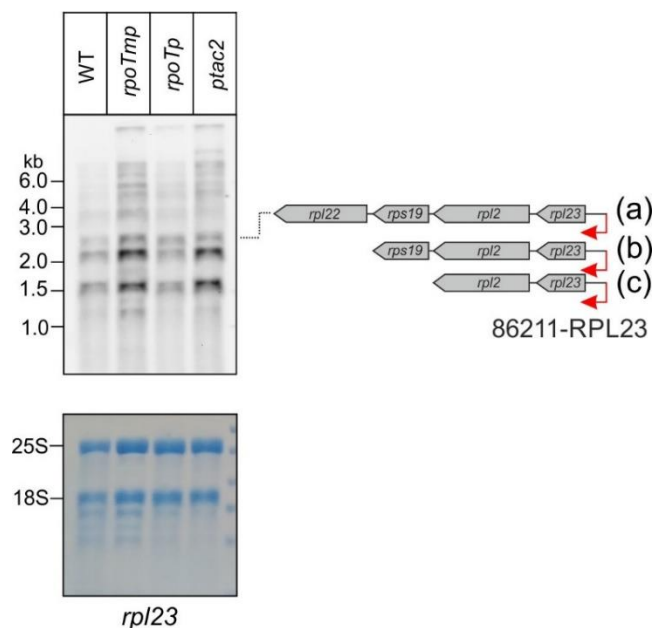


Abbildung III.22: Northern Blot-Analyse des *rpl23*-Operons für das späte Stadium von Wildtyp und Mutanten mit einer *rpl23*-spezifischen Sonde. Die Beladung ist durch eine Methylenblaufärbung gezeigt (unten). Die detektierten Signale können den verschiedenen Transkripten zugeordnet werden (Genkarten, rechts). Die roten Pfeile zeigen die TSS an. Auf der linken Seite sind die Markergrößen angezeigt. kb: Kilobasen

### III.6.2.3 Transkription von *rpl32*

Für *rpl32* wurden drei TSS identifiziert, von denen 112358+RPL32 als RPOTp-abhängig klassifiziert wurde (Abb. III.18). Die TSS 112953+RPL32 und 112979+RPL32 konnten mit dem statistischen Modell keinem Fall zugeordnet werden. Die pamK-Analyse wies 112358+RPL32 und 112979+RPL32 einem NEP-Cluster zu (Anhang Tab. VI.13). Die Ergebnisse des Northern Blots für *rpl32* (Abb. III.23 A) zeigten Transkripte, welche den identifizierten TSS entsprechen. Das Transkript mit der ungefähren

Größe von 2 kb (a) stammte aus der Transkription von 112358+RPL32, das kleinere Transkript bei etwa 1,4 kb (b) könnte von den TSS 112953+RPL32 oder 112979+RPL32, welche sehr nah aneinander liegen, starten. Obwohl nur 112358+RPL32 als RPOTp-abhängig klassifiziert wurde, waren beide Transkripte (a) und (b) in *rpoTp* reduziert (Abb. III.23 A). (c) stellte laut der Genomabdeckung aus der RNA-Seq-Analyse (nicht gezeigt) ein Transkript mit einem prozessierten 5'-Ende dar. Die Methylenblaufärbung zeigte hier eine etwas geringere Beladung für Wildtyp, trotzdem konnte der Blot das Ergebnis der TSS-Analyse bestätigen und die Abhängigkeit von *rpl32* von der RPOTp-Transkriptionsaktivität aufzeigen. Da in *rpoTp* nur eine Reduktion und kein Fehlen der Transkripte beobachtet wurde, ließ sich daraus auf eine geringe Aktivität von RPOTmp in dieser Mutante schließen, welche vermutlich in Wildtyp nicht stattfindet. Die RNA-Seq-Analyse zeigte für die beiden PEP-Mutanten *clb19* und *ptac2* eine signifikante Erhöhung der Transkriptmengen (Abb. III.23 B). Für *rpoTp* und *rpoTmp* konnten keine signifikanten Änderungen festgestellt werden, tendenziell war jedoch eine Reduktion im späten Stadium von *rpoTp* zu beobachten (Abb. III.23 B).

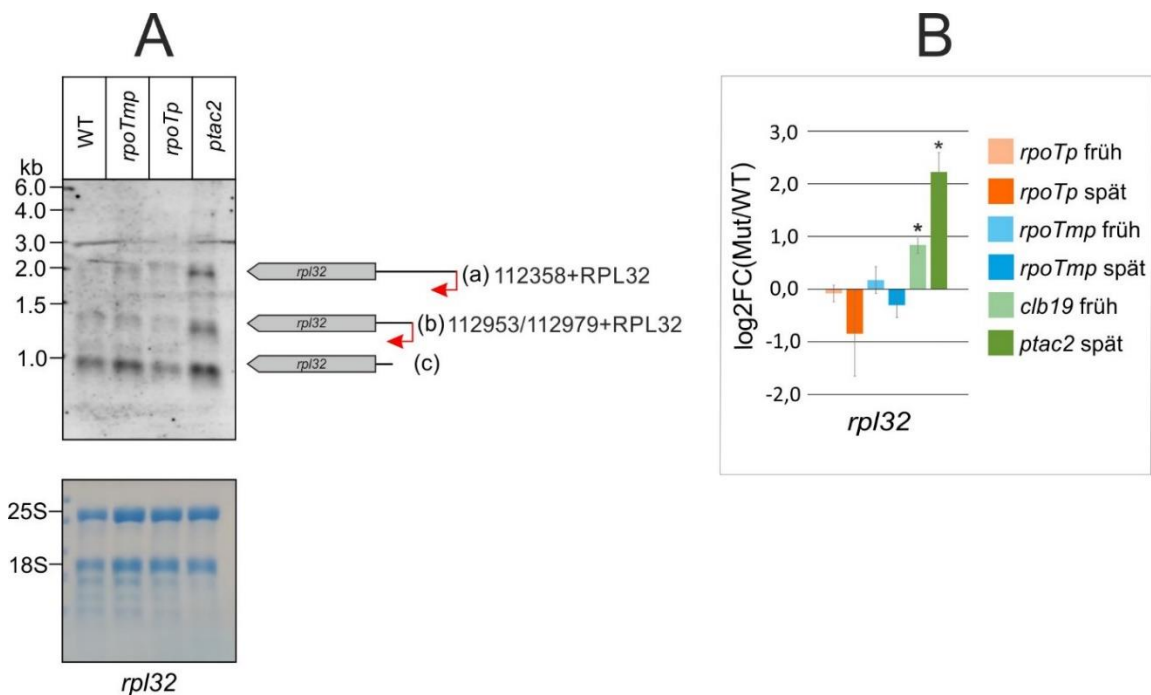


Abbildung III.23: Analyse der Transkriptmengen von *rpl32*. A: Northern Blot-Analyse für das späte Stadium von Wildtyp und Mutanten. Die Beladung ist durch eine Methylenblaufärbung gezeigt. Die detektierten Signale können den verschiedenen Transkripten zugeordnet werden (Genkarten, rechts). Die roten Pfeile zeigen die TSS an. Auf der linken Seite sind die Markergrößen angezeigt. kb: Kilobasen. B: RNA-Seq-Analyse in den Mutanten für das frühe und späte Stadium dargestellt als logarithmierte relative Transkriptlevel bezogen auf WT mit DESeq2-Normalisierung. Der Fehlerbalken zeigt den Standardfehler an. \* Wald-Test mit p-Wert  $\leq 0,05$ . Die Expressionsanalyse wurde von Jose Muino durchgeführt.

#### III.6.2.4 Transkription von *psbD*

Das statistische Modell identifizierte die TSS 32644+PSBD als RPOTp-abhängig. Die zwei weiteren TSS 31614+PSBD und 32672+PSBD wurden einer überlappenden Funktion von PEP und RPOTp (Fall 4) zugewiesen (Abb. III.18, Anhang Tab. VI.14). Die Northern Blot-Analyse von *psbD* zeigte in allen Mutanten reduzierte Transkriptmengen (Abb. III.24 A). Transkripte, welche aus der Transkription von

31614+PSBD (a) und 32644+PSBD bzw. 32672+PSBD (c) entstanden sind, konnten detektiert werden. Die beiden letzteren liegen sehr nah beieinander, sodass hier zur Richtigkeit der TSS-Klassifikation keine Aussage gemacht werden kann. Das Transkript der Größe von etwa 4,6 kb (a) war in *ptac2* stark und in *rpoTp* etwas geringer reduziert, auch die zugehörige TSS 31614+PSBD wurde als PEP/RPOTp-abhängig klassifiziert. Es ist jedoch wahrscheinlicher, dass diese TSS PEP-abhängig ist und *rpoTp* nur eine sekundäre Reduktion zeigt. Eine frühere Publikation identifizierte diese TSS (*PpsbD*-948) als PEP-abhängig (Shimmura *et al.*, 2017). Das Transkript mit der Größe von knapp 2,9 kb (b) stammte potentiell von der Transkription an einer TSS, welche in dieser Arbeit nicht identifiziert worden ist (blauer Pfeil), aber in vorigen Arbeiten gezeigt wurde (Castandet *et al.*, 2019; Hanaoka *et al.*, 2003). Das Transkript, welches auf die Transkription von den TSS 32644+PSBD bzw. 32672+PSBD zurückgeht, (c) zeigte eine Reduktion in allen Mutanten. Das Ergebnis der Northern Blot-Analyse deckte sich mit der RNA-Seq-Analyse (Abb. III.24 B). Es wurden leicht reduzierte Transkriptmengen in späten Stadien von *rpoTp* und *rpoTnp* detektiert, die Transkriptmengen von *ptac2* waren etwa um den Faktor 1,5 reduziert. Die Reduktion der Transkriptmengen in *rpoTnp* könnte darauf hindeuten, dass RPOTmp auch an der Transkription von *psbD* beteiligt ist.

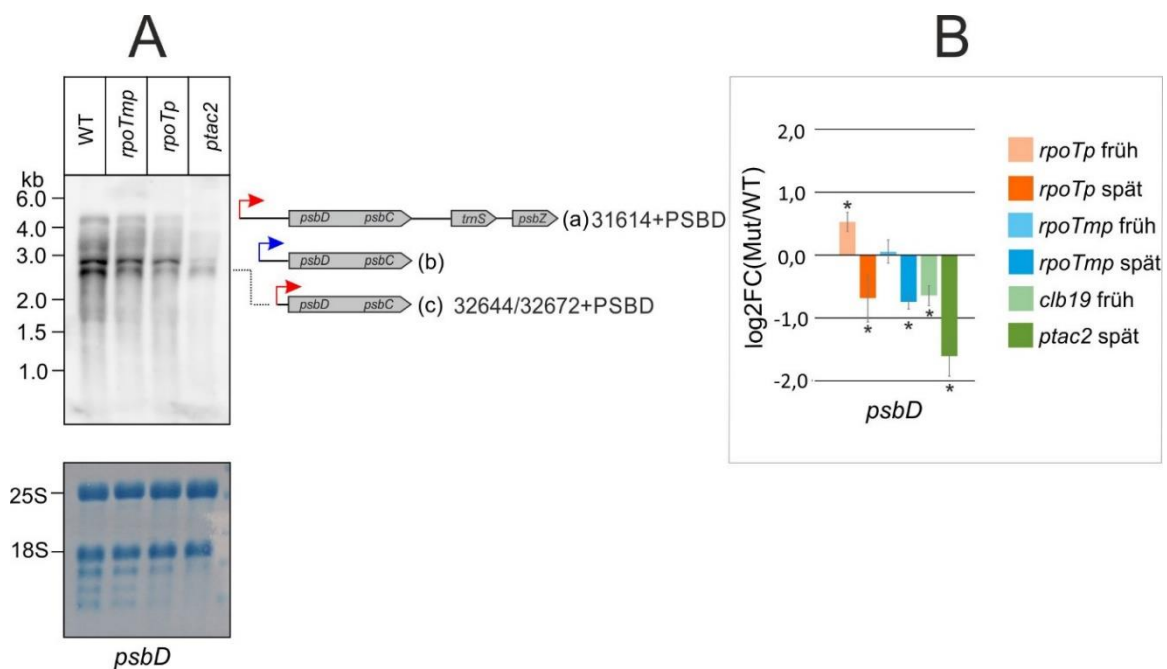


Abbildung III.24: Analyse der Transkriptmengen von *psbD*. A: Northern Blot-Analyse für das späte Stadium von Wildtyp und Mutanten. Die äquivalente Beladung ist durch eine Methylenblaufärbung gezeigt. Die detektierten Signale können den verschiedenen Transkripten zugeordnet werden (Genkarten, rechts). Die Pfeile zeigen die TSS an. Die roten Pfeile zeigen TSS, welche in dieser Arbeit identifiziert wurden, der blaue Pfeil auf der linken Seite sind die Markergrößen angezeigt. kb: Kilobasen. B: RNA-Seq-Analyse in den Mutanten für das frühe und späte Stadium dargestellt als logarithmierte relative Transkriptlevel bezogen auf WT mit DESeq2-Normalisierung. Der Fehlerbalken zeigt den Standardfehler an. \* Wald-Test mit p-Wert ≤ 0,05. Die Expressionsanalyse wurde von Jose Muino durchgeführt.

### III.6.3 RPOTp nutzt alternative TSS in Abwesenheit von PEP

Für das *atpB-atpE*-Operon wurden 6 TSS identifiziert, 5 davon liegen vor *atpB* (Abb. III.25 C). Nach der Klassifizierung mit dem statistischen Modell wurden drei der TSS als PEP-abhängig (54618-ATPB,

54671-ATPB, 62285-ATPB) und eine als RPOTp-abhängig (56771-ATPB) identifiziert (Anhang Tab. VI.14). Die TSS 54671-ATPB (*PatpB*-515) und 54618-ATPB (*PatpB*-467) wurden bereits als PEP-abhängig beschrieben (Ghulam *et al.*, 2013; Schweer *et al.*, 2006; Swiatecka-Hagenbruch *et al.*, 2007).

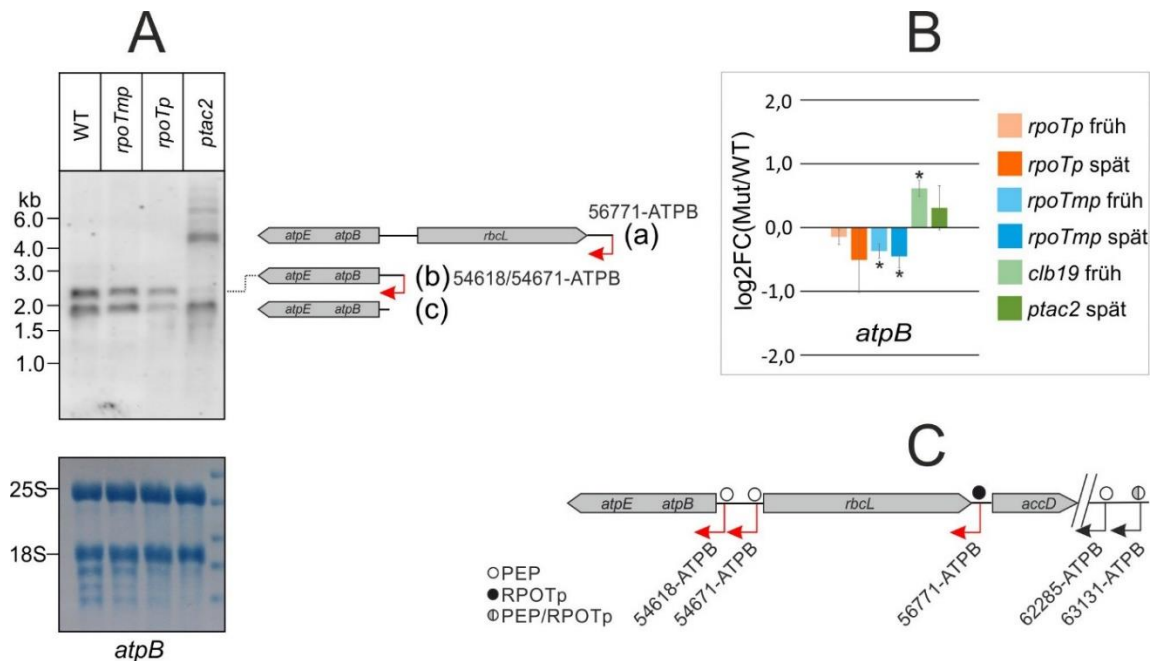


Abbildung III.25: Analyse der Transkriptmengen von *atpB*. A: Northern Blot-Analyse für das späte Stadium von Wildtyp und Mutanten. Die äquivalente Beladung ist durch eine Methylenblaufärbung gezeigt. Die detektierten Signale können den verschiedenen Transkripten zugeordnet werden (Genkarten, rechts). Die roten Pfeile zeigen die TSS an. Auf der linken Seite sind die Markergrößen angezeigt. kb: Kilobasen. B: RNA-Seq-Analyse in den Mutanten für das frühe und späte Stadium dargestellt als logarithmierte relative Transkriptlevel bezogen auf WT mit DESeq2-Normalisierung. Der Fehlerbalken zeigt den Standardfehler an. \* Wald-Test mit p-Wert  $\leq 0,05$ . Die Expressionsanalyse wurde von Jose Muino durchgeführt. C: Schematische Darstellung eines Abschnitts des Plastoms. Gene sind in als graue Balken dargestellt. Die Pfeile zeigen die TSS an, unter dem Pfeil ist die TSS benannt. Die roten Pfeile entsprechen den Pfeilen in A. Die Kreise stellen die Abhängigkeit der TSS von einer Polymesarenaktivität dar. Die diagonalen Doppelstriche zeigen die Aussparung eines Genomteils auf.

In Wildtyp wurden zwei Transkripte der Größen etwa 2,4 kb (b) und 1,9 kb (c) detektiert (Abb. III.25 A), wobei (b) von der Transkriptionsaktivität von den TSS 54618-ATPB (*PatpB*-467) oder 54671-ATPB (*PatpB*-515) stammte, während (c) vermutlich ein prozessiertes dicistronisches Transkript darstellte, welches zuvor ebenfalls detektiert wurde (Ghulam *et al.*, 2012). Die entsprechenden Transkripte waren auch in *rpoTmp* und *rpoTp* sichtbar, wobei die Transkriptmenge in *rpoTmp* mit dem Wildtyp vergleichbar war und in *rpoTp* für beide Transkripte reduziert war. Die Reduktion dieser Transkripte in *rpoTp* war vermutlich auf die Reduktion an PEP-Enzym in dieser Mutante zurückzuführen. In *ptac2* waren normale Level des 1,9-kb-Transkripts sichtbar (c), während das Transkript von 2,4 kb (b) kaum detektiert wurde. Stattdessen waren größere Transkripte (4,8 kb und etwa 6-7 kb) (a) zu sehen. Das 4,8 kb-Transkript (a) entsprach der Transkription von der identifizierten TSS 56771-ATPB. Diese TSS wurde bereits in *sig6*-Mutanten als „rettende“ TSS beschrieben, die bei reduzierter PEP-Aktivität durch fehlende Sigma-Faktoren aktiviert worden ist (Hanaoka *et al.*, 2005; Schweer *et al.*, 2006). Die RNA-Seq-Analyse (Abb. III.25 B) zeigte für die beiden NEP-Mutanten reduzierte Transkriptmengen, jedoch war die Reduktion nur in *rpoTmp* signifikant. Dieses Ergebnis



widerspricht den Ergebnissen der TSS-Identifizierung und des Northern Blots. In den PEP-Mutanten wurden erhöhte Transkriptmengen detektiert (Abb. III.25 B). Dieses ist vermutlich auf die Aktivität von RPOTp in diesen Mutanten zurückzuführen, sodass die fehlende Nutzung der TSS 54618-ATPB (*PatpB-467*) oder 54671-ATPB (*PatpB-515*) durch die Transkription von der TSS 56771-ATPB durch RPOTp maskiert wird.

### III.6.4 Motivanalyse der RPOTp-abhängigen TSS

Nach der Klassifizierung der TSS zu den Polymeraseaktivitäten sollten die Motive untersucht werden, welche von den Polymerasen zur Transkription genutzt werden. Es wurden die eindeutig zugewiesenen TSS für Fall 3 (RPOTp-abhängig) als Cluster von Sequenzen für die Suche verwendet (Anhang Tab. VI.14). Für RPOTmp-abhängige TSS wurde keine Motivanalyse durchgeführt, da hier nur eine einzige, zuvor bereits beschriebene TSS 100873+RRN16S.1 (Pc-Promotor) (Courtois *et al.*, 2007), identifiziert wurde. Es wurde für die 7 RPOTp-abhängigen TSS eine Motivanalyse mittels *MEME* (*Multiple Expectation-Maximization for Motif Elicitation*) durchgeführt und nach Motiven mit einer Breite von 4 bis 8 Nukleotiden gesucht. Das zuvor für die NEPs beschriebene YRTa-Motiv (Kap. I.5.1, Y: C/T, R: A/G) wurde in 4 Sequenzen (TSS 112358+RPL32, 109366+TIC214, 85045-RPS19, 146051+TRNI) an der Position 0 bis 3 Nukleotide stromaufwärts der TSS gefunden (Abb. III.26 B). Zusätzlich zeigten 2 Sequenzen (TSS 109366+TIC214, 85045-RPS19) stromaufwärts zum YRTa-Motiv eine mögliche GAA-Box mit einer Entfernung zur TSS von 21 bis 22 Nukleotiden (Abb. III.26 A). Eine Sequenz zeigte ein GAA-Motiv ohne identifiziertes YRTa-Motiv (TSS 32644+PSBD). Insgesamt war eine Tendenz erkennbar, bei welcher die Promotorregionen, die von RPOTp genutzt werden, ein YRTa-Motiv aufweisen. Die Tatsache, dass das YRTa-Motiv nicht überall detektiert wird, ist vermutlich auf die Variabilität in den Motiven und deren Position zurückzuführen, die von *MEME* nicht erkannt werden.

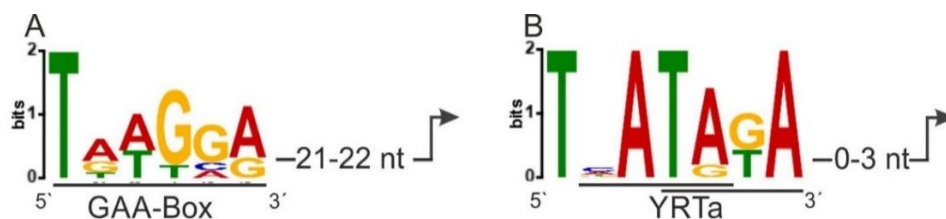


Abbildung III.26: Identifizierte Promotormotive für das RPOTp-abhängige Cluster von TSS. Die Motivsuche mittels *MEME* wurde auf die 7 RPOTp-abhängigen TSS angewandt. Die *MEME*-Ergebnisse wurden mit Rücksicht zur TSS-Position +1 interpretiert, der Abstand zur TSS (Pfeil) ist rechts angezeigt. Die Höhe der Nukleotide zeigt deren Häufigkeit an. Die Motive sind in 5' zu 3'-Richtung angegeben. Y: C/T, R: A/G.

### III.6.5 RPOTp und RPOTmp haben eine überlappende Funktion

Die 9 TSS, welche in beiden Entwicklungsstadien nicht zu einer Polymeraseaktivität zugewiesen werden konnten, könnten möglicherweise sowohl von RPOTp wie auch RPOTmp als Transkriptionsstart genutzt werden (Kap. III.4.3, Anhang Tab. VI.14, bezeichnet als „potentiell NEP“).

Eine Motivanalyse dieser TSS zeigte in 6 Sequenzen das YRTa-Motiv (Abb. III.27) 0 bis 5 Nukleotide von der TSS entfernt und unterstützte die Vermutung, dass diese TSS in überlappender Funktion von RPO<sub>Tp</sub> und RPO<sub>Tmp</sub> genutzt werden könnten. Für 71940-CLPP1, 20245-RPOC2 und 112979+RPL32 wurden keine YRTa-Motive detektiert. Der *clpP*-Promotor in Tabak wurde zuvor als Typ II-Promotor ohne YRTa-Motiv beschrieben, mit Erkennungssequenzen im Bereich -5 bis 25+ (Sriraman *et al.*, 1998). Die beiden anderen TSS könnten ebenfalls zu Typ II-Nicht-Konsensus-Promotoren gehören.

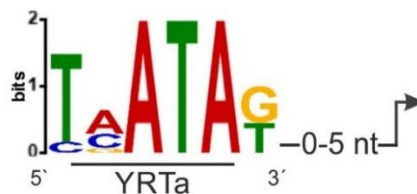


Abbildung III.27: Identifizierte Promotormotive für die nicht zugewiesenen 9 TSS. Die Motivsuche mittels *MEME* wurde auf 9 TSS angewandt, welche weder im frühen noch im späten Stadium einer Polymeraseaktivität zugewiesen werden konnten. Die *MEME*-Ergebnisse wurden mit Rücksicht zur TSS-Position +1 interpretiert, der Abstand zur TSS (Pfeil) ist rechts angegeben. Die Höhe der Nukleotide zeigt die Häufigkeit an. Die Motive sind in 5' zu 3'-Richtung angegeben. Y: C/T, R: A/G.

Die bereits beschriebene TSS 71940-CLPP für *clpP* (*PclpP*-53 in Tabak und *PclpP*-58 in *Arabidopsis*), die im Laufe dieser Arbeit bestätigt wurde, konnte nicht eindeutig zu einer Polymeraseaktivität zugeordnet werden. Zuvor wurde diese TSS *PclpP*-58 als NEP-anhängig identifiziert (Liere *et al.*, 2004; Swiatecka-Hagenbruch *et al.*, 2008, 2007). In einigen Arbeiten wurde für *clpP* auch eine zusätzliche PEP-abhängige TSS *PclpP*-115 in *A. thaliana* beschrieben (Swiatecka-Hagenbruch *et al.*, 2008, 2007), welche mit der optimierten dRNA-Seq-Methode nicht bestätigt werden konnte.

Der Northern Blot (Abb. III.28 A) zeigte, dass die Transkription von *clpP* nicht PEP-abhängig ist. Vielmehr wurden im Vergleich zu Wildtyp erhöhte Transkriptlevel in der PEP-Mutante *ptac2* beobachtet. Diese Beobachtung deckte sich mit anderen Publikationen, in denen erhöhte Mengen der NEP-abhängigen Transkripte in PEP-Mutanten detektiert wurden (Allison *et al.*, 1996; Hajdukiewicz *et al.*, 1997). Auch die RNA-Seq-Analyse bestätigte die erhöhten Transkriptmengen in *ptac2* (Abb. III.28 B). Die beiden NEP-Mutanten *rpoTmp* und *rpoTp* zeigten fast keine Unterschiede zu den Transkriptmengen in Wildtyp, lediglich für *rpoTp* konnte eine mäßige Reduktion des 1,5-kb-Transkripts (d) und des reifen 0,7-kb-Transkripts (e) beobachtet werden. In dieser Mutante wurden trotzdem sichtbare Transkriptmengen detektiert, sodass die Ergebnisse darauf hindeuten, dass *clpP* von RPO<sub>Tp</sub> und RPO<sub>Tmp</sub> transkribiert werden kann und die beiden NEPs hier eine überlappende Funktion haben. Dieses Ergebnis wurde auch von der RNA-Seq-Analyse unterstützt (Abb. III.28 B), welche unveränderte Mengen der Transkripte in *rpoTp* und *rpoTmp* zeigte.

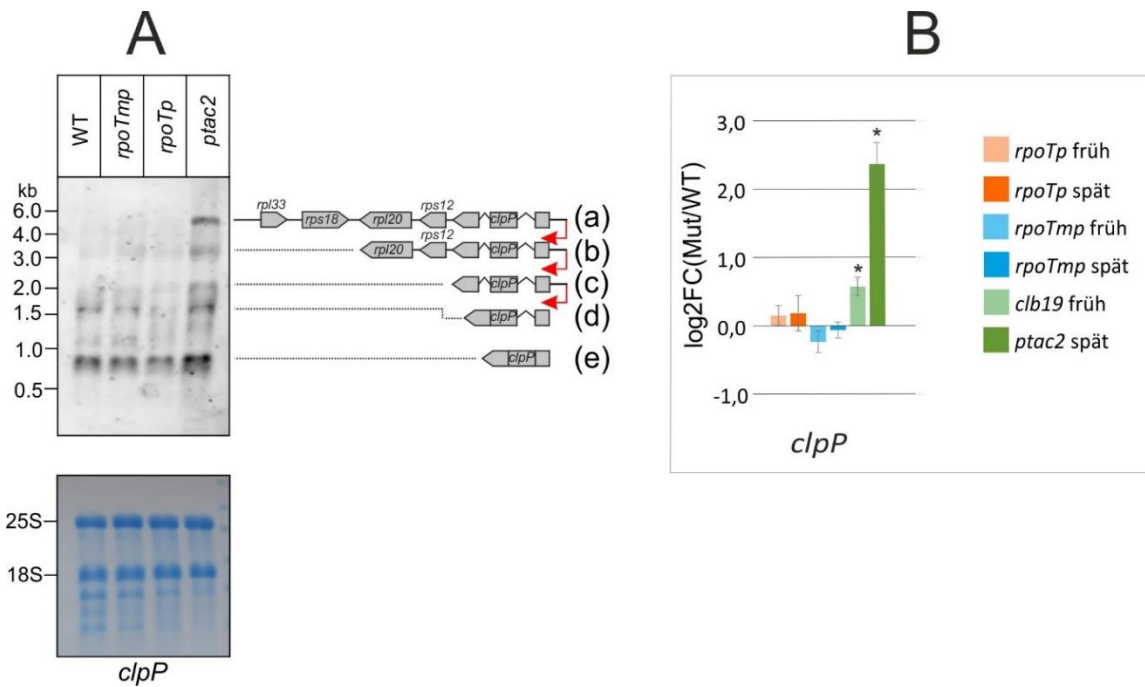


Abbildung III.28: Analyse der Transkriptmengen von *clpP*. A: Northern Blot-Analyse für das späte Stadium von Wildtyp und Mutanten. Die äquivalente Beladung ist durch eine Methylenblaufärbung gezeigt. Die detektierten Signale können den verschiedenen Transkripten zugeordnet werden (Genkarte, rechts). Die roten Pfeile zeigen die TSS an. Die Dächer deuten Introns an. Auf der linken Seite sind die Markergrößen angezeigt. kb: Kilobasen. B: RNA-Seq-Analyse in den Mutanten für das frühe und späte Stadium dargestellt als logarithmierte relative Transkriptlevel bezogen auf WT mit DESeq2-Normalisierung. Der Fehlerbalken zeigt den Standardfehler an. \* Wald-Test mit  $p\text{-Wert} \leq 0,05$ . Die Expressionsanalyse wurde von Jose Muino durchgeführt.

Das längere Transkript in *ptac2* mit der Größe von etwa 5 kb (a) fing bei einem Abgleich mit der Genomabdeckung aus der RNA-Seq-Analyse an der TSS 71940-CLPP an und umfasste die Transkripte der Gene *clpP*, *rps12*, *rpl20*, sowie Antisense-Transkripte von *rps18* und *rpl33* (Abb. III.29). Durch das Ausfallen einer weiter vorne liegenden PEP-abhängigen TSS war die Transkription von der TSS 71940-CLPP erhöht. In dieser Arbeit wurde die PEP-abhängige TSS 71416-RPS12 detektiert, deren fehlende Nutzung in *ptac2* das Vorkommen dieses Transkripts erklären würde (Abb. III.29).

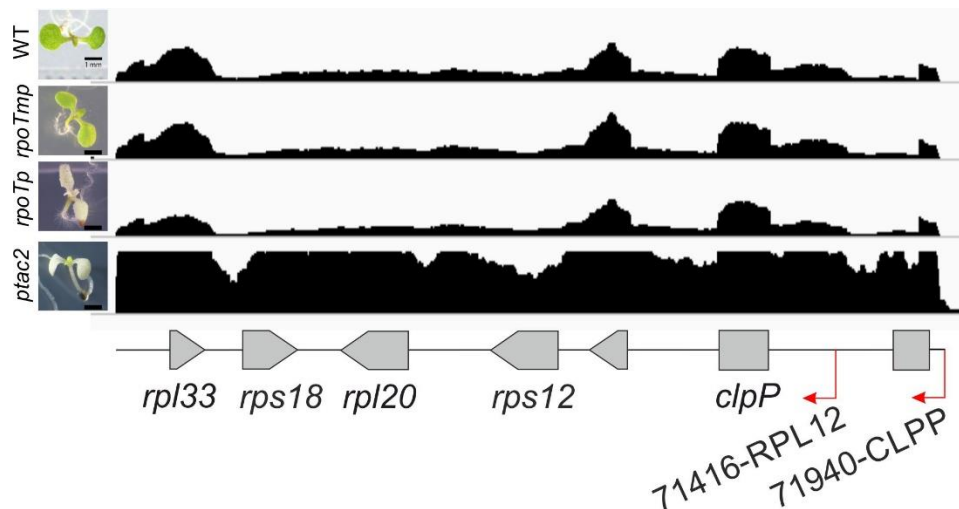


Abbildung III.29: Genomabdeckung aus der RNA-Seq-Analyse für die Gene *clpP*, *rps12*, *rpl20* und Antisense der Gene *rps18* und *rpl33* ohne Normalisierung für Wildtyp und Mutanten im späten Stadium. Die Pfeile zeigen die in diesem Bereich

identifizieren TSS an. Die Höhe der Skala ist in allen Genotypen gleichgesetzt. Die Kartierung wurde von Jose Muino durchgeführt.

Eine weitere Kernfrage dieser Arbeit war, von welcher Polymerase die Transkription der PEP-Untereinheiten abhängig ist. Da jedoch nur zwei TSS für die Gene der PEP-Untereinheiten identifiziert wurden (20245-RPOC2 und 78986-RPOA), welche mit dem statistischen Modell nicht eindeutig zu einer Polymeraseaktivität zugewiesen werden konnten, ist ausgehend von der TSS-Nutzung nicht klar, welche der Phagentyp-Polymerasen involviert ist. Die pamK-Klassifizierung konnte nur 20245-RPOC2 zu einem NEP-Cluster zuordnen (Anhang Tab. VI.13). So wurde die RNA-Seq-Analyse hinzugezogen, um diese Frage zu beantworten.

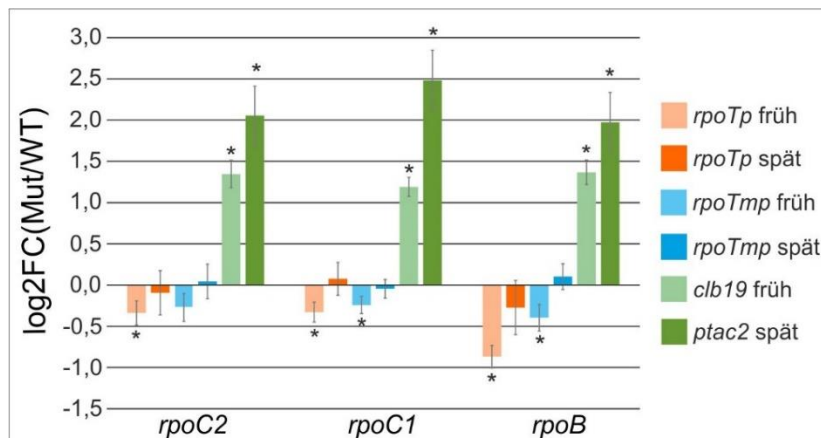


Abbildung III.30: RNA-Seq-Analyse des *rpoB*-Operons in den Mutanten für das frühe und späte Stadium dargestellt als logarithmierte relative Transkriptlevel bezogen auf WT mit DESeq2-Normalisierung. Der Fehlerbalken zeigt den Standardfehler an. \* Wald-Test mit p-Wert  $\leq 0,05$ . Die Expressionsanalyse wurde von Jose Muino durchgeführt.

Während in den PEP-Mutanten *clb19* und *ptac2* erhöhte Transkriptmengen der *rpo*-Gene zu beobachten waren, zeigten *rpoTp* und *rpoTmp* im frühen Stadium reduzierte Transkriptmengen (Abb. III.30). Die Transkriptmengen waren in beiden Mutanten reduziert, sodass hier vermutlich eine additive Funktion beider Phagentyp-Polymerasen vorliegt und beide Polymerasen vorhanden sein müssen, um Wildtyp-Level der Transkripte zu erreichen. Interessant ist, dass die signifikante Reduktion nur im frühen Stadium detektiert wurde, während im späten Stadium so gut wie keine Änderung der Transkriptabundanz festgestellt werden konnte.

### III.7 Funktion von PEP in plastidärer Transkription

Es wurden 65 TSS identifiziert, die in frühen und späten Stadien der Pflanzenentwicklung von PEP abhängig sind. Die Analyse der TSS in Abhängigkeit von der Genklasse zeigte, dass PEP in der Lage ist, neben Photosynthesegenen auch tRNAs, rRNAs und ribosomale Proteine zu transkribieren (Abb. III.31). Die zuvor beschriebene Funktion der PEP in der Transkription von Photosynthesegenen konnte in der vorliegenden Arbeit bestätigt werden. Es wurden unter anderem TSS für Gene wie *psbD* (Abb. III.24) und vorher bekannte TSS für *psbA* und *rbcL* (Abb. III.12) als PEP-abhängig identifiziert (Anhang VI.14).

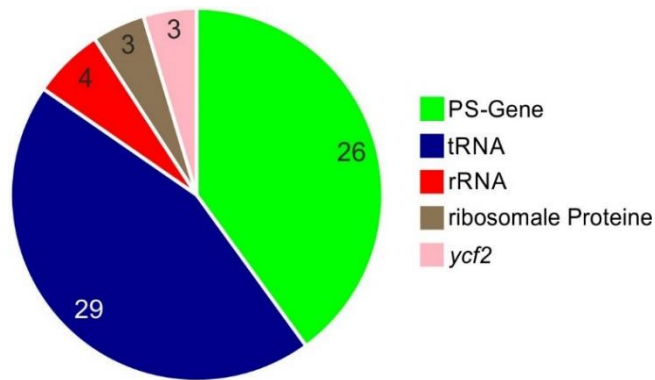


Abbildung III.31: Aufteilung der 65 PEP-abhängigen TSS nach Genklassen. Die Zahlen geben die Anzahl der TSS für jede Genart wieder. Die Klassifizierung der TSS mittels eines statistischen Modells wurde von Frederic Ehlers unter Anleitung von Ivo Große und Julia Rott durchgeführt.

### III.7.1 PEP-abhängige TSS für *ycf2* wurden detektiert

Es wurden drei PEP-abhängige TSS für *ycf2* identifiziert, gegensätzlich zu einer früheren Untersuchung in Tabak, die *ycf2* als ein Gen mit einer einzelnen NEP-abhängigen TSS (*Pycf2*-1577) beschrieben hat (Hajdukiewicz *et al.*, 1997). Im Laufe dieser Arbeit wurde diese TSS nicht detektiert. Die drei PEP-abhängigen TSS lagen mehr als 5 kb vom Gen entfernt (Abb. III.32, (a), (b), (c)) und schienen nur einen geringen Beitrag zur Transkription des Gens zu leisten. Laut der Genomabdeckung aus der RNA-Seq-Analyse startete die Transkription von *ycf2* von einer nicht identifizierten TSS (d), die sich im Bereich der Genomposition 86100 befand und etwa 300 Nukleotide stromaufwärts von der Translationsstartstelle entfernt war (Abb. III.32).

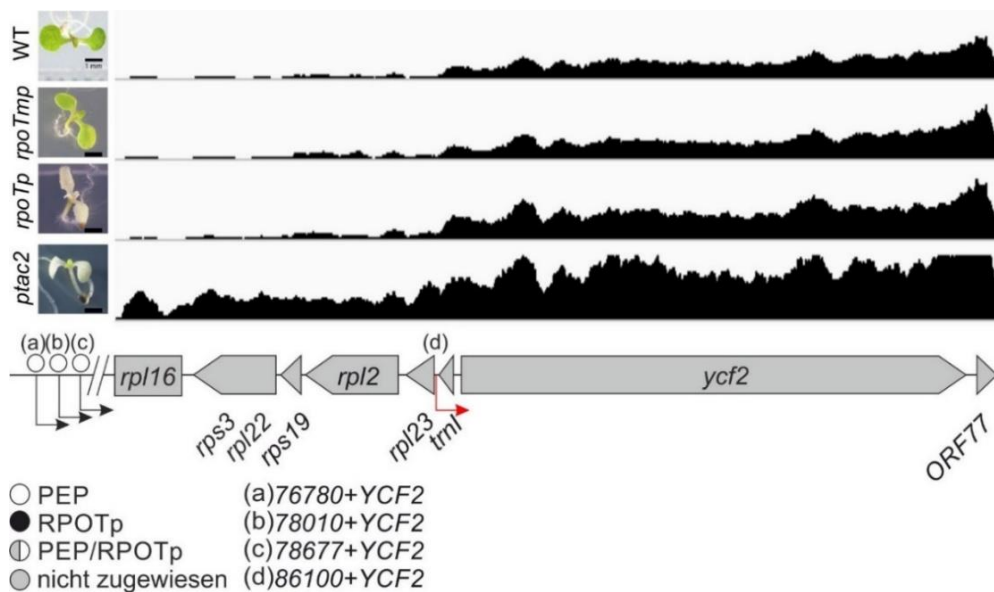


Abbildung III.32: Genomabdeckung aus der RNA-Seq-Analyse für Antisense *rpl16*, *rps3*, *rpl22*, *rps19*, *rpl2*, *rpl23*, *trnI* und die Gene *ycf2* sowie *ORF77* ohne Normalisierung für Wildtyp und Mutanten im späten Stadium. Die schwarzen Pfeile zeigen die drei PEP-abhängigen TSS an. Der rote Pfeil zeigt eine nicht identifizierte TSS an. (a), (b), (c): Identifizierte TSS mit Genomposition, (d): nicht identifizierte TSS mit abgeschätzter Genomposition.

Mittels der in dieser Arbeit optimierten dRNA-Seq konnte keine Genomabdeckung in diesem Bereich festgestellt werden. Insgesamt deutet die erhöhte Genomabdeckung in *ptac2* darauf hin, dass die PEP-abhängigen TSS entweder wenig zur Transkription von *ycf2* beitragen oder falsch zu dieser Polymeraseaktivität zugeordnet wurden.

### III.7.2 PEP transkribiert tRNAs und rRNAs während der Samenkeimung

Neben TSS für Photosynthesegene wurden auch TSS für tRNAs und rRNAs als PEP-abhängig identifiziert (Abb. III.31). Für *trnK* konnte die TSS 4451-TRNK identifiziert werden, welche mit dem statistischen Modell der PEP-Aktivität zugeschrieben werden konnte. Das der Transkription von dieser TSS entsprechende Transkript konnte im Northern Blot beobachtet werden (Abb. III.33 A). Die reduzierte Transkriptmenge in *ptac2* bestätigte weiterhin, dass diese tRNA von PEP transkribiert wird. Die RNA-Seq-Analyse (Abb. III.33 B) unterstützte diese Annahme und zeigte eine signifikante Reduktion in der PEP-Mutante *clb19*.

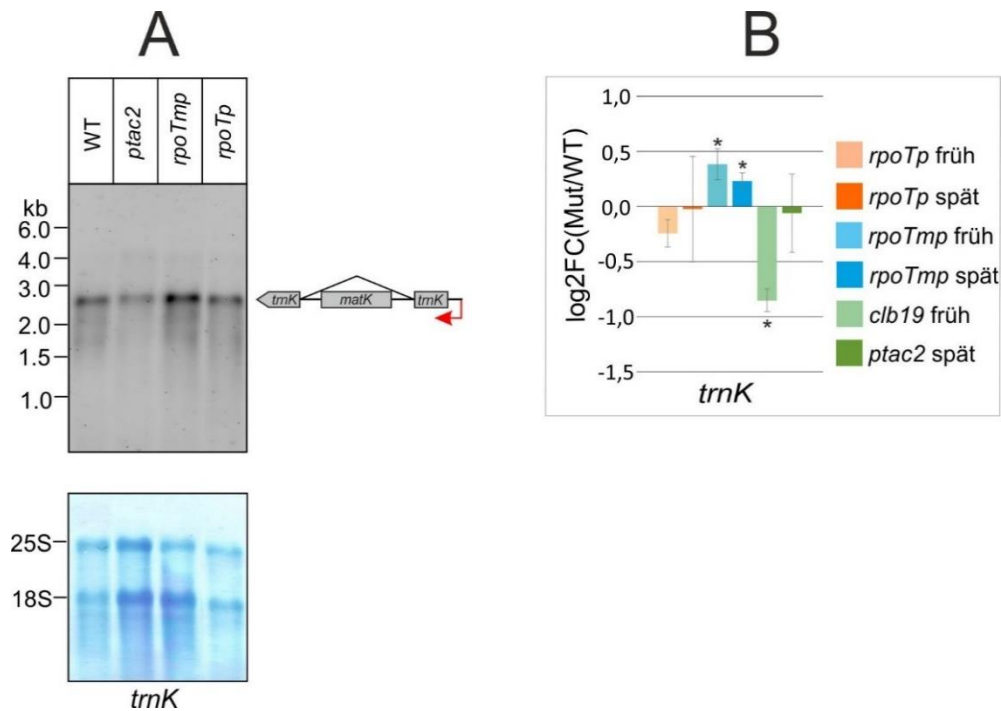


Abbildung III.33: Analyse der Transkriptmengen von *trnK*. A: Northern Blot-Analyse für das späte Stadium von Wildtyp und Mutanten. Die Beladung ist durch eine Methylenblaufärbung gezeigt. Die detektierten Signale können den verschiedenen prozessierten Transkripten zugeordnet werden (Genkarten, rechts). Der Pfeil zeigt die TSS an. Das Dach deutet das Intron von *trnK* an, in welchem *matK* liegt. Auf der linken Seite sind die Markergrößen angezeigt. kb: Kilobasen. B: RNA-Seq-Analyse in den Mutanten für das frühe und späte Stadium dargestellt als logarithmierte relative Transkriptlevel bezogen auf WT mit DESeq2-Normalisierung. Der Fehlerbalken zeigt den Standardfehler an. \* Wald-Test mit p-Wert  $\leq 0,05$ . Die Expressionsanalyse wurde von Jose Muino durchgeführt.

Eine mögliche Funktion von PEP in nicht-grünen Stadien der Pflanzenentwicklung wurde während der Untersuchung der Genomabdeckungen festgestellt (Abb. III.34). Das Beispiel zeigt die als PEP-abhängig identifizierten TSS 4451-TRNK und 104475+RRN23S.1 (Anhang VI.11 und VI.12). Die

Genomabdeckung der 5'-Enden war im frühen Stadium (WT früh) mindestens doppelt so hoch wie in grünen Keimlingen (WT spät), was die stärkere Nutzung dieser TSS in keimenden Samen anzeigt.

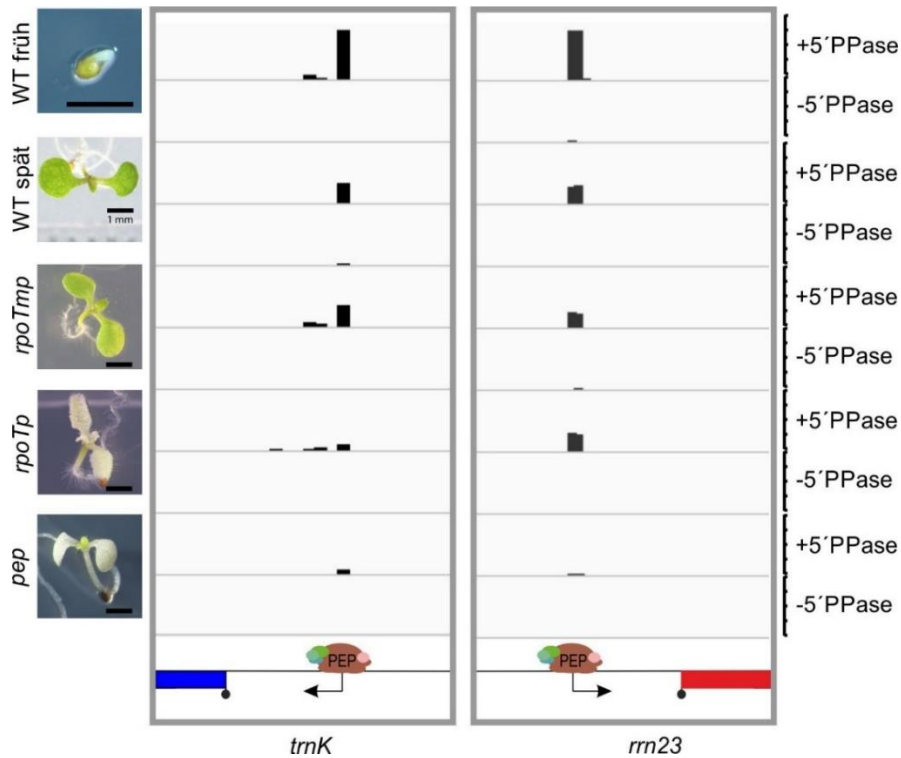


Abbildung III.34: Identifizierte TSS für *trnK*. Angezeigt sind die Genomabdeckungen der +5'PPase- sowie -5'PPase-cDNA-Bibliotheken mit RPM-Normalisierung für die 5'-Enden der Reads für Wildtyp und die RNA-Polymerase-Mutanten. In dunkelblau ist schematisch ein Teil von *trnK* mit angezeigt (links), in rot ein Teil von *rrn23* (rechts). Der Pfeil deutet die TSS an. Die Höhe der Skala ist für alle Genotypen gleichgesetzt. Der Größenbalken zeigt 1 mm an.

Nach der Identifizierung der PEP-abhängigen TSS wurde die Genomabdeckung für diese TSS im frühen und späten Stadium von Wildtyp analysiert, um zu untersuchen, in welchem Stadium der Pflanzenentwicklung diese TSS hauptsächlich als Transkriptionsstart dient (Abb. III.35). Die Analyse zeigte, dass im frühen Stadium vor allem PEP-abhängige TSS für tRNAs und rRNAs (in dunkelblau und rot) eine höhere Genomabdeckung haben, während im grünen Gewebe der Keimlinge eine höhere Nutzung der TSS für Gene des Photosyntheseapparats (in grün) festgestellt wurde. Einige PEP-abhängige TSS für tRNAs schienen die gleiche Nutzung in früher und später Entwicklung zu haben (*trnP*, *trnV*, *trnL*, *trnM*). Somit konnte PEP auch in frühen Entwicklungsstadien der Pflanze eine wichtige Rolle in der Transkription von tRNAs und rRNAs zugewiesen werden.

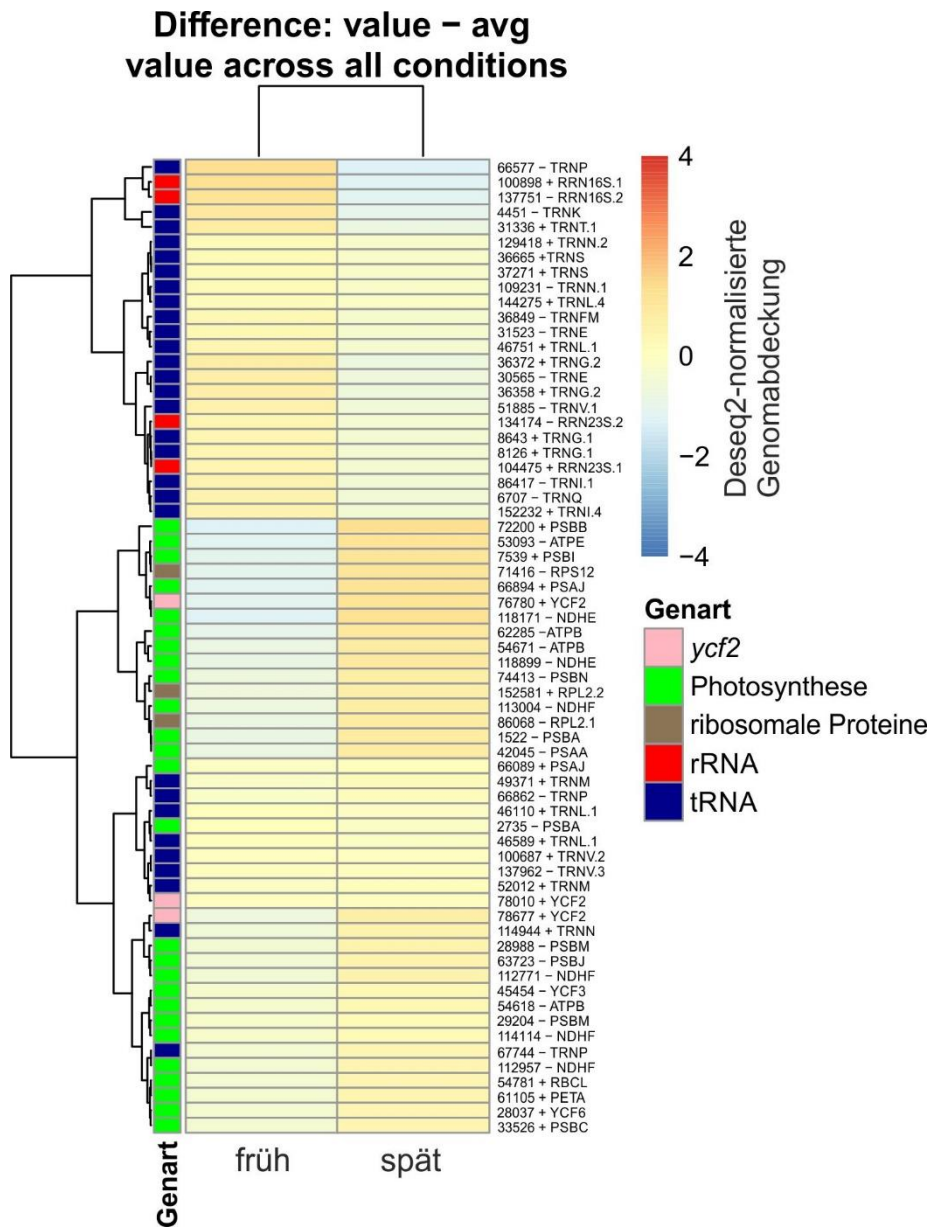


Abbildung III.35: Nutzung der PEP-abhängigen TSS in Wildtyp. Mengen der Reads in den +5'PPase-Proben mit DESeq2-Normalisierung für das frühe und späte Stadium von Wildtyp für die 65 PEP-abhängigen TSS. Jeweils drei Replikate wurden vereint. Die Farbkodierung gibt die Zugehörigkeit der TSS zu den einzelnen Genklassen wieder. Die Analyse wurde von Frederic Ehlers unter Anleitung von Julia Rott und Dermot Harnett durchgeführt.

### III.8 TSS-Nutzung in frühen Phasen der Samenkeimung

Es wurde eine erweiterte Untersuchung der Transkription durchgeführt, um die Funktion von der plastidären RNA-Polymerasen und besonders PEP vor und während der Samenkeimung zu identifizieren. Neben trockenen Samen wurden imbibierte, etiolierte und gekeimte Samen von Wildtyp analysiert (Abb. III.36). Eine tabellarische Übersicht über die Sequenzierungsstatistiken der cDNA-Bibliotheken ist im Anhang Tab. VI.9 hinterlegt.



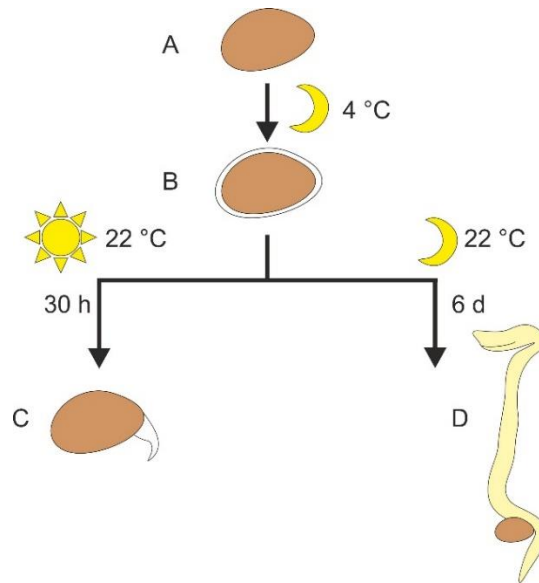


Abbildung III.36: Analyse der Transkription in sehr früher Samenentwicklung. Schematische Darstellung des experimentellen Versuchsaufbaus und Probenaufbereitung. A: trockene Samen, B: stratifizierte und imbibierte Samen, C: gekeimte Samen, D: 6 Tage alte etiolierte Keimlinge. Trockene Samen wurden auf Platten ausgebracht und bei 4 °C im Dunkeln stratifiziert. Anschließend wurden die imbibierten Samen entweder im Dauerlicht oder im Dunkeln bei 22 °C angezogen. Etiolierte Keimlinge wurden nach 6 Tagen geerntet, gekeimte Samen nach 30 Stunden. h: Stunden, d: Tage.

In dieser Analyse wurden insgesamt 57 TSS identifiziert, von denen 7 doppelte TSS aus den *inverted repeats* darstellen. Eine Liste mit der Übersicht befindet sich im Anhang Tab. VI.15. Die Aufteilung der 50 TSS nach Zugehörigkeit zu einer Genklasse (Abb. III.37 Mitte) zeigte, dass auch in nicht-grünen, sehr frühen Stadien der Samenkeimung neben *housekeeping*-Genen wie tRNAs und rRNAs auch einige Photosynthesegene transkribiert werden. Identifiziert wurden 27 TSS, die zu tRNAs gehören, sowie 15 TSS, die zu Photosynthesegenen gehören.

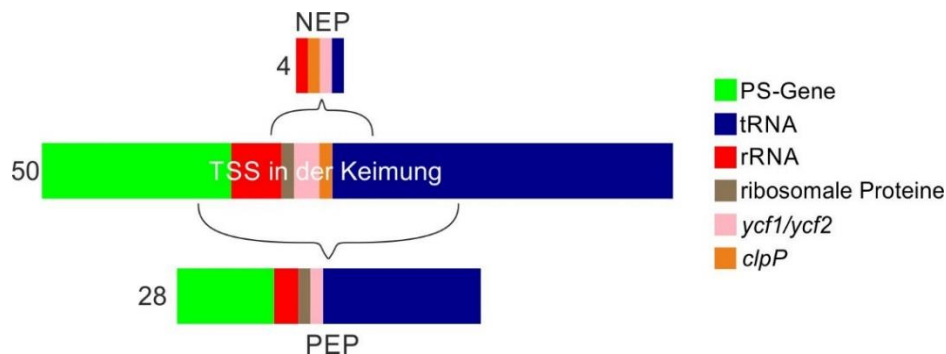


Abbildung III.37: Identifizierte TSS in der Samenentwicklung. Aufteilung der während der Samenentwicklung 50 identifizierten TSS nach Genklassen. 28 von diesen TSS konnten in vorigen Analysen als PEP-abhängig identifiziert werden (unten), 4 der TSS wurden den NEP-Polymeraseaktivitäten zugeordnet (oben). Die Identifizierung der TSS wurde von Frederic Ehlers unter Anleitung von Julia Rott und Dermot Harnett durchgeführt.

Ein Vergleich zu den TSS, welche im Rahmen der Mutantanalyse einer Polymeraseaktivität zugeordnet werden konnten (Anhang Tab. VI.14), zeigte, dass von den in der frühen Samenentwicklung identifizierten TSS 4 NEP-abhängig (RPOTp, RPOTmp oder überlappende Funktion) und 28 PEP-abhängig waren (Abb. III.37 oben und unten, Anhang Tab. VI.15). Es war bekannt, dass in trockenen Samen vor allem gespeicherte Transkripte vorhanden sind, welche in der Phase der Embryoentwicklung

und Samenbildung produziert werden (Allorent *et al.*, 2013; Demarsy *et al.*, 2012). Während der Imbibierung startet die Neusynthese der mRNAs, sodass der Vergleich der Transkriptomte dieser Entwicklungsphasen nötig ist, um eine Aussage über die Aktivität der Polymerasen treffen zu können. Insgesamt war die Menge der gelagerten Transkripte in trockenen Samen gering im Vergleich zu den neusynthetisierten Transkriptmengen (Anhang Abb. VI.4), was auf die Aktivität der plastidären RNA-Polymerasen bereits in der Phase der Imbibierung hinweist. Die Analyse der TSS-Nutzung im Samen konnte zeigen, dass 19 PEP-abhängige TSS, die zu tRNAs und rRNAs gehören, während der Imbibierung und Keimung aktiv genutzt werden. Weiterhin wurden auch 8 PEP-abhängige TSS für Photosynthesegene identifiziert, was darauf hindeutete, dass auch in Abwesenheit von Licht TSS für Photosynthesegene von PEP zur Transkription genutzt werden. Die 4 NEP-TSS, welche während der Keimung aktiv waren, sind 109366+TIC214.1, 7872-TRNQ, 71940-CLPP und 100873+RRN16S.1. Somit konnte gezeigt werden, dass beide NEPs schon während der Phasen der Imbibierung und Keimung wichtige Funktionen erfüllen.

Interessant ist, dass beide TSS für *rrn16* 137776-RRN16S.2 (Pc-Promotor, RPOTmp-abhängig) und 137751-RRN16S.2 (*Prrn16*-112, PEP-abhängig) auch in diesem Entwicklungsstadium identifiziert worden sind (Anhang Tab. VI.15). Das zeigte, dass die rRNA sowohl von PEP als auch von RPOTmp auch in sehr früher Samenentwicklung transkribiert wird. In trockenen Samen sind nur Transkripte gespeichert, welche durch PEP generiert worden sind (Abb. III.38). Die Nutzung der beiden TSS im Laufe der Samenentwicklung zeigte jedoch, dass in imbibierten Samen vor allem RPOTmp aktiv war und PEP zur Transkription des Gens weniger beitrug.

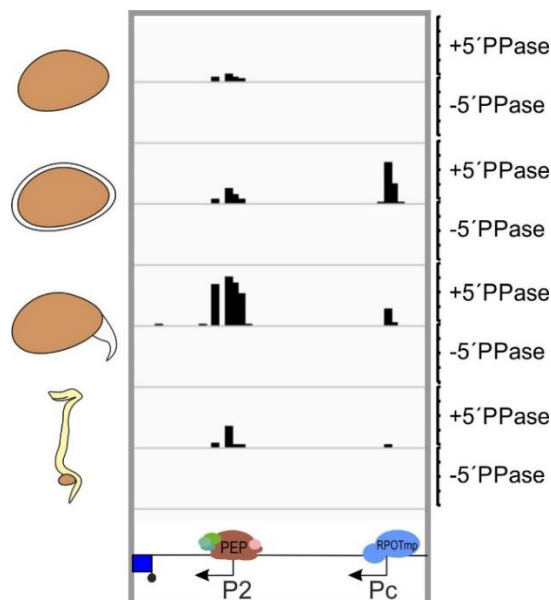


Abbildung III.38: Identifizierte TSS für *rrn16* während Stratifikation und Keimung. Angezeigt sind die Genomabdeckungen der +5' PPase- sowie -5' PPase-cDNA-Bibliotheken mit RPM-Normalisierung für die 5'-Enden der Reads für Wildtyp in den Stadien des trockenen, stratifizierten und keimenden Samens sowie im etiolierten Gewebe. In dunkelblau ist schematisch ein Teil des *rrn16*-Gens mit angezeigt. Der gebogene Pfeil deutet die TSS an. Die RNA-Polymerasen sind in braun (PEP) bzw. in blau (RPOTmp) dargestellt. Die Höhe der Skala ist für alle Proben gleichgesetzt.

Die Nutzung der beiden TSS in späteren Entwicklungsstadien der Pflanze (Abb. III.11) offenbarte, dass der PEP-abhängige Promotor *Prrn16-112* sowohl in keimenden Samen als auch in Keimlingen hauptsächlich als Transkriptionsstart diente, während der Pc-Promotor einen geringen Beitrag leistete, was mit den publizierten Ergebnissen übereinstimmte (Courtois *et al.*, 2007; Zoschke *et al.*, 2007). Die Transkription des *rrn16*-Operons in etiolierten Keimlingen wurde ebenfalls hauptsächlich von PEP ermöglicht (Abb. III.38), während für RPO*Tmp* eine sehr geringe Beteiligung beobachtet wurde. Die detektierten Transkriptmengen waren jedoch insgesamt viel niedriger, als in keimenden Samen.

### III.9 Untersuchung einer möglichen *rpoTp rpoTmp* Doppelmutante

Neben den Einzelmutanten *rpoTp* und *rpoTmp* sollte eine Doppelmutante untersucht werden, die keine Phagentyp-RNA-Polymerasen exprimiert. In 2006 wurde eine *rpoTp rpoTmp* Mutante vorgestellt, die nur bis zum Keimlingsstadium auf saccharosehaltigem Medium überlebt (Hricová *et al.*, 2006). Im Laufe dieser Arbeit wurden für zwei unabhängige Linien die Einzelmutanten *rpoTp-2* und *rpoTmp-2* sowie *rpoTp-3* und *rpoTmp-1* (Tab. II.1) miteinander gekreuzt. Es wurden Analysen zur Segregation und Untersuchungen der Samen und Embryos der selektierten unabhängigen *rpoTmp rpoTmp* RPO*Tp rpoTp* Mutantenlinien durchgeführt, um festzustellen, ob eine *rpoTp rpoTmp* Doppelmutante identifiziert und deren plastidäre Transkription analysiert werden kann.

#### III.9.1 Segregationsanalyse der *rpoTmp rpoTmp* RPO*Tp rpoTp* Mutante

Die Segregationsanalyse ergab keine sichtbare Doppelmutante, von welcher erwartet wurde, dass sie einen chlorotischen Phänotyp zeigt, da bereits die Einzelmutante *rpoTp* chlorotisch und stark wachstumsgehemmt (Abb. III.3) war. Bei der Annahme der Embryoletalität der Doppelmutante würde man erwarten, dass 25 % der Samen nicht keimen. Jedoch betrug die Keimungsrate mindestens 80 % Prozent (Tab. III.1). Die Genotypisierung einer Platte mit 100 Samen einer *rpoTmp rpoTmp* RPO*Tp rpoTp* Mutante (Tab. III.1) zeigte keine mögliche Doppelmutante, etwa 75 % der Samen trugen den Genotyp *rpoTmp rpoTmp* RPO*Tp rpoTp*.

Tabelle III.1: Segregationsanalyse einer *rpoTmp rpoTmp* RPO*Tp rpoTp* Mutante. Es wurden die zwei unabhängig generierte Linien *rpoTmp-2 rpoTmp-2* RPO*Tp-2 rpoTp-2* und *rpoTmp-1 rpoTmp-1* RPO*Tp-3 rpoTp-3* untersucht.

| Linie  | Keimungsrate % | RPO <i>Tp rpoTp</i> % | RPO <i>Tp</i> RPO <i>Tp</i> % |
|--|----------------|-----------------------|-------------------------------|
| <i>rpoTmp-2 rpoTmp-2</i> RPO <i>Tp-2 rpoTp-2</i> | 80             | 71,5                  | 28,5                          |
| <i>rpoTmp-1 rpoTmp-1</i> RPO <i>Tp-3 rpoTp-3</i> | 89             | 74                    | 26                            |

#### III.9.2 Phänotypische Untersuchung der Samen der *rpoTmp rpoTmp* RPO*Tp rpoTp* Mutante

Die Analyse der Schoten einer *rpoTmp rpoTmp* RPO*Tp rpoTp* Mutante (Linie *rpoTmp-2 rpoTmp-2* RPO*Tp-2 rpoTp-2*) zeigte zwei Arten von Samen (Abb. III.39). In grünen Schoten (A) erschien der

Samen mit dem Genotyp *rpoTmp rpoTmp RPOTp rpoTp* bzw. *rpoTmp rpoTmp RPOTp RPOTp* grün (n). Die Samen, die den Genotyp *rpoTmp rpoTmp rpoTp rpoTp* trugen, hatten eine weiß-durchsichtige Farbe (w). In der weiteren Reifung (B) zeigten die weißen Samen als erstes eine braune Färbung (w), während die grünen länger grün blieben (n). Beide Samenarten konnten in der reifen Schote voneinander unterschieden werden (C). Die Samen mit der Doppelmutation erschienen dunkelbraun und verkümmert (w), während die Samen mit dem Genotyp *rpoTmp rpoTmp RPOTp rpoTp* bzw. *rpoTmp rpoTmp RPOTp RPOTp* zu normal aussehenden, Wildtyp-ähnlichen vollen braunen Samen heranreifen (n). Die Schoten der zweiten Mutantenlinien *rpoTmp-1 rpoTmp-1 RPOTp-3 rpoTp-3* wiesen das gleiche phänotypische Bild auf.

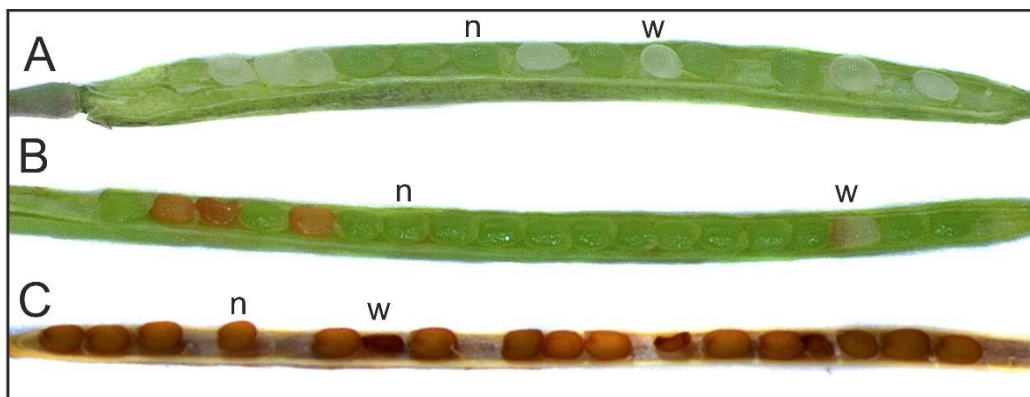


Abbildung III.39: Schoten von *rpoTmp-2 rpoTmp-2 RPOTp-2 rpoTp-2* Mutanten. A: Samen in grünen Schoten B: Samen in reifenden Schoten. C: Samen in reifen Schoten. Die Samen mit dem Genotyp *rpoTmp rpoTmp rpoTp rpoTp* sind mit w markiert, Samen mit dem Genotyp *rpoTmp rpoTmp RPOTp rpoTp* bzw. *rpoTmp rpoTmp RPOTp RPOTp* sind mit einem n markiert.

Für die Auswertung wurden die Samen in den Schoten ausgezählt. Die Auszählung ergab, dass etwa 25 % der Samen in den Schoten weiß oder verkümmert sind (Tab. III.2). Diese Analyse zeigte, dass eine Befruchtung und die anfängliche Entwicklung des Embryos stattfinden, sich der Embryo jedoch nicht zu einem normalen Samen entwickelt.

Tabelle III.2: Auszählung der Samen in Schoten von *rpoTmp rpoTmp RPOTp rpoTp* Mutanten.

| Linie                                    | ausgezählte Samen | normal grün/braun | weiß/ verkümmert | % normal grün/braun | % weiß/ verkümmert |
|--|-------------------|-------------------|------------------|---------------------|--------------------|
| <i>rpoTmp-2 rpoTmp-2 RPOTp-2 rpoTp-2</i> | 133               | 96                | 37               | 72,18               | 27,82              |
| <i>rpoTmp-1 rpoTmp-1 RPOTp-3 rpoTp-3</i> | 260               | 196               | 64               | 75,38               | 24,62              |

### III.9.3 Untersuchung der Embryonen der *rpoTmp rpoTmp RPOTp rpoTp* Mutante

Um den Zeitpunkt zu ermitteln, bei dem die Entwicklung in weißen Samen abbricht, wurden die Samen der Linie *rpoTmp-2 rpoTmp-2 RPOTp-2 rpoTp-2* entfärbt und mithilfe eines Lichtmikroskops mit Nomarski Optik untersucht. Abb. III.40 zeigt den Vergleich der Embryonen von einem grünen (*rpoTmp rpoTmp RPOTp rpoTp* bzw. *rpoTmp rpoTmp RPOTp RPOTp*) und einem weißen (*rpoTmp rpoTmp*

*rpoTp rpoTp*) Samen. In grünen Samen (B) wurde ein mittleres Torpedo-Stadium beobachtet, welches auf eine normale Embryoentwicklung hindeutete. Ein weißer Samen in der gleichen Schote (A) enthielt einen Embryo im Globularstadium. In einer reifenden Schote, in der sich der weiße Samen bereits bräunlicher verfärbt, wurde in diesem Samen weiterhin einen Embryo im Globularstadium gefunden (C), welcher sich nicht weiterentwickelt. Der grüne Samen in der gleichen Schote trug bereits einen reifen torpedoförmigen Embryo (D). Somit konnten sich Samen mit einer *rpoTmp rpoTmp rpoTp rpoTp* Doppelmutation nicht über das Globularstadium hinaus entwickeln und der Embryo starb ab.

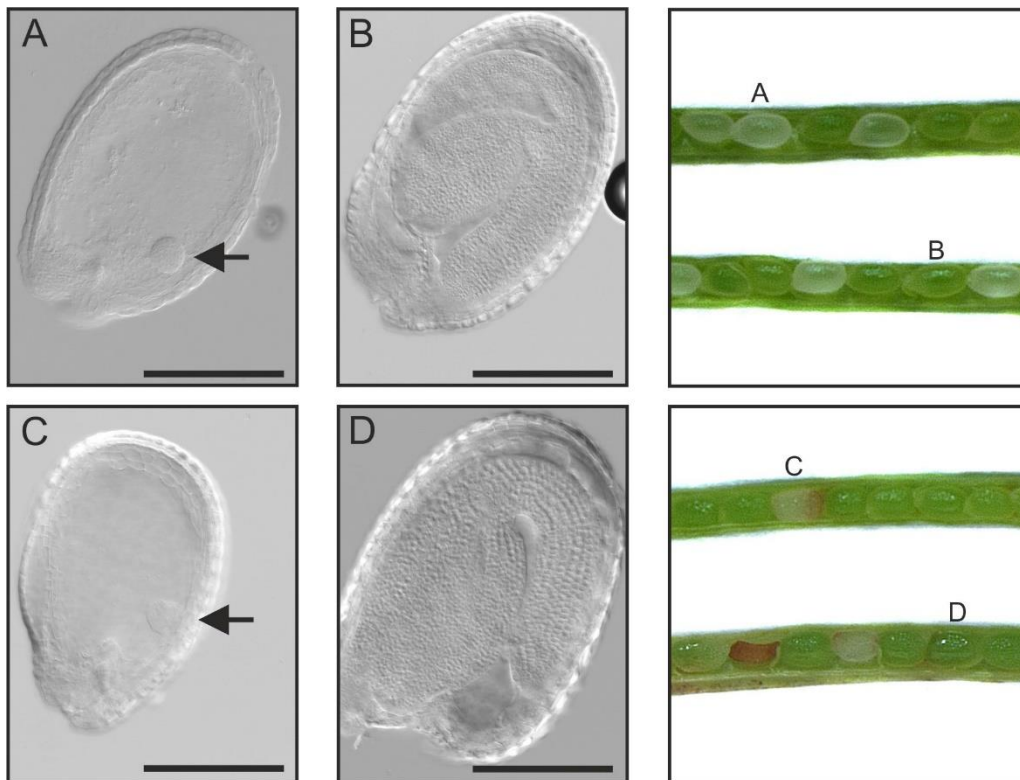


Abbildung III.40: Embryonalentwicklung in Samen von *rpoTmp-2 rpoTmp-2 RPOTp-2 rpoTp-2* Mutanten. A, B: Embryonen in grünen Schoten, C, D: Embryonen in reifenden Schoten. Die Buchstaben der Embryos entsprechen den Buchstaben der Samen in den Schoten rechts. Der Größenbalken zeigt 0,25 mm an. Die Pfeile deuten auf den Embryo im Globularstadium, welcher sich nicht weiterentwickelt.

### III.10 Elektronenmikroskopische Untersuchung der Plastiden in weißen Mutanten

Für die elektronenmikroskopische Analyse der Plastiden wurden die Keimblätter und die echten Blätter der Keimlinge betrachtet. Verglichen wurden der Wildtyp und die weißen Mutanten *rpoTp* und *ptac2*. Die Analyse der Keimblätter und der echten Blätter der Mutanten *rpoTp* und *ptac2* ergab sichtbare Unterschiede zu dem Aufbau der Plastiden in Wildtyp (Abb. III.41). Die *rpoTp*-Mutante zeigte in früher Entwicklung zwei verschiedene Arten von Plastiden, die innerhalb derselben Zelle vorkommen (Abb. III.41 Mitte). Es fanden sich mit Wildtyp vergleichbare Chloroplasten, die Thylakoidmembranen ausbilden und Stärke akkumulieren (n). Jedoch wurden in *rpoTp* weniger Granathylakoide ausgebildet. Die Hälfte der Plastiden der *rpoTp*-Mutante war im Inneren leer (a). Vereinzelt wurden bei diesen Plastiden Vesikelstrukturen, Plastoglobuli und kurze Stromathylakoide beobachtet. Die Keimblätter der

Mutante *ptac2* zeigte plastidäre Strukturen, die phänotypisch ebenfalls von den wildtypischen Plastiden abwichen (Abb. III.41 unten). Es waren vermehrt Vesikelstrukturen sichtbar, welche bei *ptac*-Mutanten bereits gezeigt worden sind (Pfalz *et al.*, 2006). In den Blättern wurden dagegen große Stapel von Granathylakoiden detektiert und nur wenige Stromathylakoide, wohingegen in Wildtyp das Gegenteil der Fall ist.

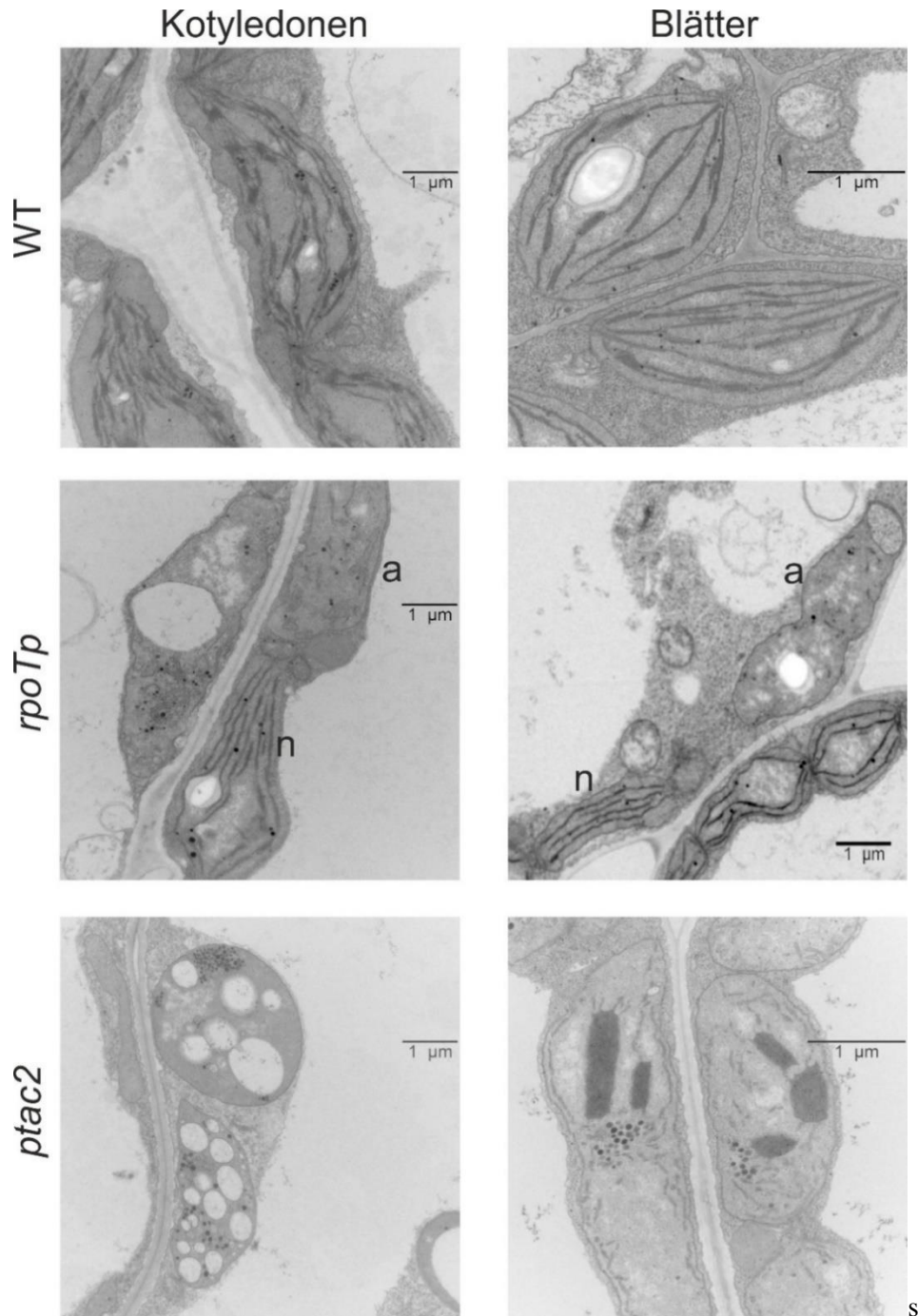


Abbildung III.41: Elektronenmikroskopische Aufnahme von Plastiden aus Wildtyp und Mutanten in Kotyledonen und echten Blättern. Die Pflanzen wurden zu den in Klammern angegebenen Zeitpunkten untersucht: Wildtyp (7 Tage); *rpoTp* (13 Tage); *ptac2* (18 Tage). n: normaler Plastid, a: abnormaler Plastid. Der Größenbalken zeigt 1 µm an. Die EM-Aufnahmen wurden von Gerd Hause gemacht.

Das Vorkommen von zwei verschiedenen Plastidenarten in der gleichen Zelle bei *rpoTp* wurde nicht bei der ursprünglichen Charakterisierung der Mutante gezeigt (Hricová *et al.*, 2006). So wurden zusätzlich ältere Blätter elektronenmikroskopisch untersucht (Abb. III.42). In den Blättern von 24 Tage alten Pflanzen der *rpoTp*-Mutante fanden sich nur noch normal aussehende Plastiden, was darauf schließen ließ, dass die abnormalen Plastiden entweder innerhalb von 1-2 Wochen abgebaut werden und sich nicht mehr teilen oder dass sie sich zu normalen Chloroplasten ausdifferenzieren. Insgesamt zeigten die Plastiden jedoch kleinere Granastapel, wie man sie in ersten jungen Blättern in Wildtyp gesehen hat (vergleichend zu Abb. III.41). Im Wildtyp würde man komplexere Strukturen in den Chloroplasten von 24 Tage alten Pflanzen erwarten (siehe Pogson *et al.*, 2015).

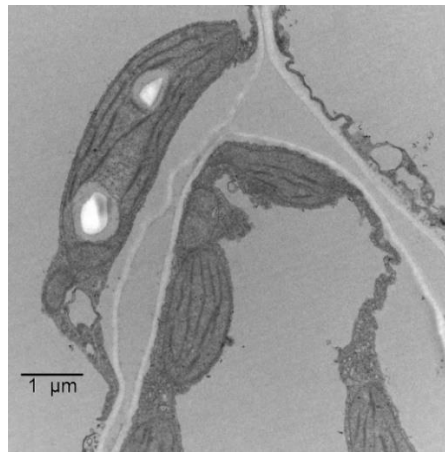


Abbildung III.42: Elektronenmikroskopische Aufnahme von Plastiden aus 24 Tage alten *rpoTp*-Mutanten. Die EM-Aufnahmen wurden von Gerd Hause gemacht.

## IV DISKUSSION

### IV.1 Erfolgreiche dRNA-Seq-Optimierung für quantitative Analyse der TSS-Nutzung in Plastiden

dRNA-Seq ist als eine Methode zur Kartierung von Transkriptionsstartpunkten in Bakterien entwickelt worden (Bischler *et al.*, 2015; Mitschke *et al.*, 2011; Sharma *et al.*, 2010; Sharma and Vogel, 2014). Da 5' Enden von organellären RNAs zu denen von bakteriellen RNAs als identisch beschrieben werden (Stern *et al.*, 2010), kann die Methode auch für die Identifizierung von TSS in Plastiden genutzt werden (Zhelyazkova *et al.*, 2012b). Im Rahmen dieser Arbeit wurde die Methode optimiert. Da nur ab einer bestimmten Datentiefe die Extraktion von TSS möglich ist, war eine Optimierung, die eine ausreichende Genomabdeckung des plastidären Genoms möglich macht, notwendig. Weiterhin wurden Optimierungen vorgenommen, welche die Quantifizierung der TSS verbessern und den Vergleich der TSS-Nutzung in den RNA-Polymerase-Mutanten erleichtern. Hier wurden zwei parallele Strategien genutzt, welche zum einen den Hintergrund reduzieren und zum anderen die Genomabdeckung erhöhen sollten. Für die Erhöhung der Genomabdeckung für NEP-abhängige TSS wurde ein sehr frühes Entwicklungsstadium mit Proplastiden genutzt, da im grünen Stadium einige NEP-abhängige Transkripte nur wenig abundant sind und somit die Detektion erschwert ist (Zoschke *et al.*, 2007). Die in frühen Entwicklungsstadien gering genutzten TSS für Photosynthesegene sollten durch die Nutzung eines Keimlingsstadiums mit ausdifferenzierten Chloroplasten kartiert werden können. Eine Modifizierung des Protokolls analog zu Kruesi *et al.* (2013) konnte mehr prozessierte Transkripte entfernen und die rRNA-Menge reduzieren, was zu der Reduktion des Hintergrunds geführt hat. Zusätzlich wurden durch die Nutzung des Enzyms 5'PPase Kerntranskripte mit 5'-Caps abgereichert. Die Modifizierung hat die Nutzung von Gesamtgewebe möglich gemacht, ohne dass eine vorherige Isolation von Chloroplasten erfolgen musste. So konnten in dieser Arbeit die Proben aus dem frühen und späten Stadiums gemeinsam betrachtet werden. Das verbesserte Protokoll ermöglichte die eindeutige Identifizierung von 121 Transkriptionsstartpunkten in Plastiden und erlaubte die Erstellung einer plastomweiten Karte von TSS in *A.thaliana*. Der quantitative Vergleich der TSS konnte dazu benutzt werden, um die Funktionen der einzelnen plastidären Polymerasen festzustellen und erlaubte einen tieferen Einblick in die komplexe plastidäre Transkription. Das gewonnene Wissen trägt zum Verständnis der Zusammenarbeit der unterschiedlichen Polymerasearten im Laufe der Entwicklung der Pflanze bei.

#### IV.1.1 Reduktion des Hintergrunds verbessert die eindeutige Festlegung der TSS

Im Vergleich zu ähnlichen Methoden zur Identifizierung von TSS (Castandet *et al.*, 2019; Sharma *et al.*, 2010; Zhelyazkova *et al.*, 2012b) liegen die Vorteile der in dieser Arbeit vorgestellten optimierten Anwendung in der Möglichkeit der tieferen Sequenzierung des Plastoms sowie der verbesserten



Identifizierung der TSS von wenig exprimierten Transkripten. Das optimierte Protokoll ermöglicht die Abreicherung von rRNA sowie von mRNA mit prozessierten Enden, sodass eine quantitative Untersuchung der TSS durchgeführt werden kann. Kürzlich wurde eine neue Methode, die Terminome-Sequenzierung, vorgestellt, welche sich zur Identifikation von TSS eignet (Castandet *et al.*, 2019). Ein Vergleich dieser Methode zu der modifizierten dRNA-Seq, welche im Laufe dieser Arbeit optimiert wurde, zeigt distinkte Transkriptionsstarts für dRNA-Seq (Anhang Abb. VI.6 A), während bei der Terminome-Seq teilweise Cluster von nebeneinanderliegenden TSS identifiziert wurden (Abb. VI.6 B, rote Spitzen). So wird mit dRNA-Seq nur eine TSS im ersten Intron von *clpP* identifiziert, während mit Terminome-Seq 13 nebeneinanderliegende TSS detektiert werden. Dies hat zur Folge, dass die Gesamtzahl der mit Terminome-Seq identifizierten TSS viel höher ist, als die Zahl der TSS, die mit der optimierten dRNA-Seq gefunden wurden (Kap. III.3.2). Während mit dem optimierten Protokoll 121 identifiziert wurden, wurden mittels Terminome-Seq 215 TSS gefunden (Castandet *et al.*, 2019). Diese Methode nutzt keine Strategien zur Reduktion von prozessierten Transkripten. Die Vermutung ist, dass größere Mengen an prozessierter mRNA und/oder primärer mRNA mit 5' Monophosphaten oder 5'-Hydroxygruppen zu viel Hintergrund generieren, sodass die TSS weniger definiert kartiert werden können und prozessierte Transkripte als Primärtranskripte identifiziert werden. Bei Nutzung des dRNA-Seq-Protokolls in Gerste wurden ebenfalls viel mehr TSS detektiert (Zhelyazkova *et al.*, 2012b). Hier könnte neben der Methode zur TSS-Identifikation auch die Normalisierung der Daten zu einer erhöhten Menge an falsch-positiven TSS beitragen. Die Festlegung der TSS wurde für Gerste manuell durchgeführt und eine TSS wurde mit einer Genomabdeckung von bereits zwei Reads, welche im Vergleich zum Hintergrund angereichert waren, festgelegt. Beide cDNA-Bibliotheken wurden auf RPM normalisiert, was dann ein Problem darstellt, wenn die Bibliotheken in der Readanzahl stark variieren. Da die nicht mit TEX behandelte Hintergrundbibliothek eine im Vergleich hohe Anzahl von rRNA enthält, ist die Gesamtanzahl der Reads hier viel größer und der RPM-Faktor für die Bibliotheken, welche die für Primärtranskripte angereichert sind, wird entsprechend höher. Die wenig abundanten Transkripte in der angereicherten Bibliothek werden auf diese Weise artifiziell erhöht und der Unterschied zu dem Hintergrund wird künstlich vergrößert, was zur Detektion von falsch-positiven TSS führen könnte. In der vorliegenden Arbeit wurde für die TSS-Identifizierung eine DESeq2-Normalisierung verwendet, die für Verzerrungen in der RNA-Zusammensetzung ausgleicht (Kap. III.3.2). Für die Berechnung der Genomabdeckungen wurde in dieser Arbeit eine RPM-Normalisierung durchgeführt. Für die in dieser Arbeit vorgestellte Modifizierung des Protokolls ist diese Normalisierung berechtigt, da beide cDNA-Bibliotheken +5'PPase und -5'PPase mit TEX behandelt wurden und somit in etwa die gleiche rRNA-Menge aufweisen.

Die Menge der prozessierten RNA in den Hintergrundbibliotheken konnte durch die Modifizierung des Protokolls reduziert werden. Im Protokoll von Zhelyazkova *et al.* (2012b) wird als Hintergrund eine Bibliothek verwendet, welche nicht mit TEX behandelt ist und folglich prozessierte RNA nicht entfernt

wird. Das erschwert die Detektion der echten Transkriptionsstarts. Für eine Verbesserung der TSS-Identifikation wurde aus Kruesi *et al.* (2013) eine Modifikation des Protokolls übernommen, bei der die komplette Probe mit TEX behandelt und ein großer Teil der prozessierten RNA entfernt wird. Hier wird als Hintergrund eine Bibliothek benutzt, welche die 5'-Enden der Primärtranskripte nicht konvertiert (-5'PPase) und schlussendlich nur die TEX-resistenten Transkripte enthält (-TEX). Der Vergleich der Genomabdeckungen für das *psbB*-Operon (Anhang Abb. VI.7) verdeutlicht den Unterschied zwischen den Hintergrundbibliotheken. In der Analyse des Transkriptoms von Gerste (Zhelyazkova *et al.*, 2012b) werden in der Hintergrundbibliothek (-TEX) deutliche Readmengen detektiert, welche von prozessierten Transkripten stammen (A). Die Nutzung einer -5'PPase-Hintergrundbibliothek reduziert den Hintergrund erheblich (B). Die Entfernung der prozessierten Transkripte kann auch ein Nachteil der modifizierten dRNA-Seq-Methode sein. Das Protokoll von (Zhelyazkova *et al.*, 2012b) kann für die Kartierung von prozessierten Transkripten genutzt werden und so können zum Beispiel Bindestellen von PPR-Proteinen identifiziert werden (Zhelyazkova *et al.*, 2012a). Die in dieser Arbeit optimierte Methode eignet sich nicht dazu, da prozessierte Transkripte gezielt abgereichert werden.

### IV.1.2 Nutzung eines UMIs korrigiert die Genomabdeckung für PCR-Duplikate

Der Prozess der PCR-Amplifikation kann zu einer Verzerrung der Molekülmengen führen, da präferentiell kürzere DNA-Moleküle amplifiziert werden. Auch der GC-Gehalt der DNA-Moleküle und die Temperaturregulierung während der PCR hat einen Einfluss auf die Amplifikationseffizienz verschiedener DNA-Moleküle (Aird *et al.*, 2011). Die eingeführte Verzerrung ist vor allem dann bedeutsam, wenn die PCR zur Herstellung von cDNA-Bibliotheken dient. So kommt für die Generierung von RNA-Seq-Datensätzen ein UMI (*unique molecular identifier*) zum Einsatz (Fu *et al.*, 2018; Hong and Gresham, 2017; Kivioja *et al.*, 2012), welcher die Verzerrung auflösen kann und eine quantitative Analyse der Datensätze sogar für einzelne Zellen ermöglicht (Egorov *et al.*, 2015; Islam *et al.*, 2014). Der UMI stellt eine randomisierte Nukleotidsequenz dar, die als Teil des Adapters an ein RNA- oder DNA-Molekül ligiert wird; es entsteht ein einzigartiges Molekül. Werden demnach im Datensatz identische Sequenzen identifiziert, stammen diese von PCR-Duplikaten (Fu *et al.*, 2018). Einige Kits für kleine RNAs, wie zum Beispiel der NEXTFLEX® Small RNA-Seq Kit v3 (PerkinElmer), bieten bereits Adaptoren mit randomisierten Nukleotiden an, welche als UMI genutzt werden könnten. Hier spielt jedoch die Länge des UMI eine entscheidende Rolle. Ein UMI mit 4 Nukleotiden generiert  $4^4$  mögliche Sequenz-Varianten, was bei einer hohen Anzahl an Ausgangsmolekülen zu einer Über-Deduplizierung führen würde, da zufällig RNAs mit gleicher Sequenz an einen UMI mit gleicher Sequenz ligiert werden könnten. Je länger der UMI ist, desto eindeutiger lässt sich ein Read zu einem spezifischen PCR-Produkt zurückverfolgen. Dieses ist vor allem bei Datensätzen mit niedriger Komplexität wie zum Beispiel kleinen RNAs oder organellären RNAs wichtig, die viele Moleküle mit gleicher Sequenz enthalten (Saunders *et al.*, 2020).

Für das optimierte dRNA-Seq-Protokoll wurde ein UMI mit 8 Nukleotiden gewählt, was ein Kompromiss aus Länge des Adapters und einer möglichst großen Anzahl an individuellen Sequenz-Varianten mit  $8^4$  Varianten darstellt. Kürzlich wurde berichtet, dass für die Sequenzierung von kleinen RNAs sogar ein UMI mit einer Länge von 8 Nukleotiden nicht ausreichend ist (Saunders *et al.*, 2020). Für organelläre Proben wird jedoch eine kleinere Anzahl von längeren Molekülen beobachtet, welche auch eine Verdünnung der eingesetzten Adapter erfordert, sodass keine Über-Deduplizierung stattfinden kann. Im Laufe der Optimierung des Protokolls ist die Bedeutung des UMI deutlich geworden, da die Herstellung der cDNA-Bibliotheken viele aufeinanderfolgenden Schritte beinhaltet, von denen jeder zu einem Verlust von RNA führt (Kap. II.2.9). Dies resultiert in einer kleineren cDNA-Menge, sodass ein cDNA-Molekül in der PCR mehrmals amplifiziert wird. Es ist essentiell, dass die PCR-Zykluszahl passend gewählt wird (Kap. III.2.3) und PCR-Duplikate eliminiert werden. Die Menge der PCR-Duplikate in der vorliegenden Analyse (Kap. III.2.2) lässt darauf schließen, dass ein Verlust von RNA während der TEX-Behandlung und durch die Aufreinigungen stattfindet. Die Alternativmethoden der Terminome-Seq (Castandet *et al.*, 2019) und der dRNA-Seq (Sharma *et al.*, 2010; Zhelyazkova *et al.*, 2012a) nutzen bei der Herstellung der cDNA-Bibliotheken keine UMIs, sodass nicht entfernte PCR-Duplikate eine zusätzliche Genomabdeckung ermöglichen und die Identifikation der TSS und besonders den quantitativen Vergleich der Nutzung einzelner TSS verzerren könnten.

### IV.1.3 Weitere mögliche Verbesserungen der Strategie

Einige zuvor beschriebene TSS, zum Beispiel für *rps4*, *rpoB*, *accD*, die als NEP-abhängig identifiziert wurden (Hajdukiewicz *et al.*, 1997; Hess *et al.*, 1993; Liere *et al.*, 2004; Swiatecka-Hagenbruch *et al.*, 2008, 2007), konnten in dieser Arbeit nicht detektiert werden (Kap. III.6.1). Da die cDNA-Bibliotheken eine hohe Menge an PCR-Duplikaten aufwiesen (Kap. III.2.2), könnte die RNA-Menge erhöht werden, mit welcher die TEX-Behandlung durchgeführt wird, um die Sequenziertiefe zu erhöhen. Dies ist der kritische Schritt, bei welchem die meiste RNA abgereichert wird. Nach den Berechnungen werden nur etwa 20 % der eingesetzten RNA-Menge nach der TEX-Behandlung zurückgewonnen. Als Lösung dieses Problems könnte man die vorherige Isolierung von Chloroplasten mit anschließender RNA-Extraktion durchführen, sodass Kernttranskripte weiter abgereichert werden. Jedoch würde diese Optimierung nur für das späte Stadium mit ausdifferenzierten Chloroplasten möglich sein, da im frühen Stadium in den gekeimten Samen keine reifen Chloroplasten vorhanden sind. Die unterschiedliche Behandlung der Proben würde wiederum den Vergleich der TSS-Nutzung in den beiden Stadien erschweren.

Ein möglicher Nachteil des Protokolls ist die Nutzung von *steady-state* Gesamt-RNA, sodass hier das Transkriptom leicht verzerrt dargestellt werden könnte, da RNAs mit verschiedener Geschwindigkeit prozessiert werden und die Probe sowohl kürzlich transkribierte wie auch länger im Plastiden

verbliebene Transkripte enthält. Da manche Transkripte durch RNA-bindende Proteine oder Sekundärstrukturen stabilisiert werden (Manavski *et al.*, 2018), könnten Unterschiede in der Genomabdeckung auftreten, die auf die Stabilisierung und nicht direkt auf die Transkription zurückzuführen sind. Jedoch wird angenommen, dass eher eine Stabilisation der prozessierten Transkripte, also des Hintergrunds, auftritt. Es gibt Hinweise darauf, dass in Mitochondrien Primärtranskripte durch PPR-Proteine stabilisiert werden könnten (Ruwe *et al.*, 2016). Ob Primärtranskripte in Chloroplasten stabilisiert werden, ist soweit nicht bekannt. Plastidäre monocistronische Transkripte wie *rbcL* und *psbA*, die nicht prozessiert werden (Eibl *et al.*, 1999), können jedoch Sekundärstrukturen am 5'-Ende ausbilden, welche zur Stabilität des Transkripts beitragen (Suay *et al.*, 2005; Zou *et al.*, 2003). Gleichzeitig werden diese Gene im Plastom am häufigsten transkribiert, sodass auch eine Stabilisierung dieser Transkripte keinen Einfluss auf die Analyse bzw. auf die Identifizierung der TSS haben sollte.

Basierend auf dem optimierten dRNA-Seq Protokoll wurde eine Methode zur Identifizierung mitochondrialer TSS in naszierender mRNA (Saleh *et al.*, in Druck) entwickelt, welche sich zur TSS-Kartierung in Organellen eignet und das Problem der stabilisierten Transkripte umgeht. Bei dieser Methode werden *nuclear Run-On assays* mit markierten Nukleotiden durchgeführt und die markierten Transkripte anschließend mittels Affinitätsaufreinigung isoliert. So können die zum Zeitpunkt des Experiments transkribierten Gene identifiziert und die TSS bestimmt werden. Auch für Chloroplasten wurde diese Methode bereits erfolgreich angewandt (Saleh *et al.*, nicht publiziert).

### IV.1.4 Durch Sekundäreffekte erschwerte Klassifikation der TSS

Die Klassifikation der TSS und die eindeutige Zuweisung zu einer Polymerase wurde durch mögliche Sekundäreffekte erschwert (Kap III.4.3). Da es bisher nicht möglich war, eine PEP-*Knockout*-Mutante in *Arabidopsis* zu generieren, wurden Mutanten verwendet, die eine sekundäre Mutation tragen und nur eine stark verminderte PEP-Aktivität zeigen (Kap. III.1). So ist in diesen Mutanten eine PEP-Restaktivität vorhanden. Gleichzeitig wurde eine sekundäre Reduktion der PEP-abhängigen TSS in der *rpoTp*-Mutante erwartet, da einige Publikationen vermuteten, dass RPOTp für die Transkription der Gene der PEP-Untereinheiten verantwortlich ist (Courtois *et al.*, 2007; Hricová *et al.*, 2006; Swiatecka-Hagenbruch *et al.*, 2008). Weiterhin wurde im Laufe dieser Arbeit die Reduktion der Sigma-Faktoren in dem frühen Stadium von *rpoTp* identifiziert (Anhang Abb. VI.8). So könnte das Fehlen von den für die Promotorerkennung wichtigen Sigma-Faktoren eine zusätzliche Reduktion der Nutzung von PEP-abhängigen TSS zur Folge haben. Die Reduktion von PEP-abhängigen TSS wurde auch in Sigma-Faktor-Mutanten beobachtet (Ji *et al.*, 2020; Loschelder *et al.*, 2006; Schweer *et al.*, 2010). So wird vermutet, dass im frühen Stadium nicht alle PEP-abhängigen TSS als solche identifiziert werden konnten. In dieser Arbeit wurde bei der Klassifizierung der TSS im frühen Stadium mehrere TSS als PROTp- und gleichzeitig PEP-abhängig identifiziert (Fall 4, Anhang Tab. VI.11). Der Großteil dieser

TSS konnte im späten Stadium der PEP-Aktivität zugeordnet werden. Aufgrund der sekundären Reduktion der TSS-Nutzung durch weniger PEP-Aktivität in *rpoTp* konnten diese TSS nicht korrekt eingeordnet werden, sodass die potentielle Rolle von PEP in frühen Entwicklungsstadien größer sein könnte, als angenommen. Da im späten Stadium in *rpoTp* wieder normale Transkriptmengen der Sigma-Faktoren detektiert werden (Anhang VI.8), konnten hier die PEP-abhängigen TSS eindeutiger zugewiesen werden. Zu Fall 4 (RPOTp/PEP) wurden im späten Stadium nur 4 TSS zugeordnet, während im frühen Stadium 68 TSS in Fall 4 zu finden sind. Da die meisten dieser Fall 4-TSS im späten Stadium als exklusiv PEP-abhängig identifiziert wurden, ist die Annahme berechtigt, dass im frühen Stadium Sekundäreffekte die Zuordnung erschweren. Für eine bessere Zuweisung der TSS zu den einzelnen Polymerasen in früher Pflanzenentwicklung wäre eine PEP-Mutante von Vorteil, die keine PEP-Aktivität besitzt. So könnten die Unterschiede zu der reduzierten Nutzung der PEP-Promotoren in *rpoTp* besser identifiziert werden. Da vor Kurzem eine Methode vorgestellt wurde, mit welcher man gezielt Gene im Plastom von *Arabidopsis* ausschalten kann (Ruf *et al.*, 2019), könnte eine *Knockout*-Linie mit einer ausgeschalteten PEP-Untereinheit generiert und analysiert werden, sofern diese überlebensfähig ist. Diese Methode ermöglicht die Generierung von fertilen Pflanzen, sodass auch Stadien mit keimenden Samen analysiert werden und dort die von RPOTp und RPOTmp genutzten TSS präziser bestimmt werden könnten. Der Nachteil dieser Methode ist, dass eine spezielle Mutantenlinie mit ausgeschaltetem ACC2 benötigt wird. Das könnte unvorhergesehene Sekundäreffekte zur Folge haben. In Tabak sind Mutanten, in denen eine PEP-Untereinheit ausgeschaltet ist, seit langem bekannt (Allison *et al.*, 1996). Für *Arabidopsis* ist das bis jetzt nicht möglich gewesen.

Es wurden insgesamt 9 TSS identifiziert, welche im statistischen Modell weder im frühen noch im späten Stadium zu einem Fall zugeordnet werden konnten (Anhang Tab. VI.14, bezeichnet als „potentiell NEP“). Die nahgelegene Erklärung wäre die Nutzung dieser TSS durch mehrere Polymerasen, welche sich ersetzen können und folglich keine Änderung in der Nutzung in den Mutanten sichtbar ist. Hier würde man vermuten, dass bei diesen TSS eher die beiden NEPs RPOTp und RPOTmp eine überlappende Funktion haben könnten, da sie nach jetzigem Wissen die gleichen Promotormotive erkennen (Kap. I.5.1). Für diese Erklärung spricht auch die Tatsache, dass im statistischen Modell keine TSS zu Fall 7 eingeordnet werden konnten, welcher die Abhängigkeit einer TSS von sowohl RPOTp als auch RPOTmp aufzeigt (Kap. III.4.1). Die Schwierigkeit ist, dass das statistische Modell annimmt, dass die Transkription von der TSS additiv von den zwei Polymerasen abhängt. Folglich, dass die Aktivität beider Polymerasen nötig ist, um die Transkriptlevel wie im Wildtyp zu erreichen. Diese Situation würde für die Transkription in der Wildtyppflanze erwartet werden. In der jeweiligen Mutante könnte jedoch eine komplette Übernahme der Transkription an der TSS von der anderen noch vorhandenen Phagentyp-Polymerase stattfinden und man würde folglich keine quantitativen Unterschiede zur Transkription in Wildtyp sehen. In diesem speziellen Fall könnten solche TSS nur schwer eingeordnet werden. Beim Hinzuziehen der pamK-Analyse (Anhang Tab.

VI.13) kann diese Schlussfolgerung weiter unterstützt werden. Hier wurden im frühen und späten Stadium Cluster identifiziert, welche TSS enthalten, die keine reduzierte Nutzung in den NEP-Mutanten ausweisen und gleichzeitig eine erhöhte Nutzung in den PEP-Mutanten zeigen. Tatsächlich finden sich 6 der 9 nicht zugewiesenen TSS im frühen Stadium einem möglichen Überlapp-Cluster (Anhang Abb. VI.2 und Tab. VI.13, Cluster 4). Diese Annahme unterstützen auch die weiterführende Untersuchung der Transkriptmengen im Northern Blot sowie die Motivanalyse (Kap. III.6.5). Nicht auszuschließen ist, dass die Übernahme der Transkription durch eine Phagentyp-Polymerase ebenfalls ein Sekundäreffekt sein könnte, welcher nur bei Abwesenheit der anderen Polymerase auftritt und die TSS im Wildtyp nur von einer Polymerase genutzt wird.

## IV.2 Funktion der plastidären Phagentyp-RNA-Polymerasen

### IV.2.1 RPOTp transkribiert *ycf1* (*tic214*)

In der vorliegenden Arbeit wurden einige wenige TSS beschrieben, die speziell von RPOTp oder RPOTmp abgelesen werden. Es wurde gezeigt, dass RPOTp das Gen *ycf1* transkribiert (Kap. III.6.2.1). Die alleinige Transkription von *ycf1* von RPOTp wird auch durch *in-vitro*-Analysen unterstützt, in welchen heterolog exprimiertes RPOTp den *ycf1*-Promotor erkennt, jedoch nicht RPOTmp (Kühn *et al.*, 2007). Für *Chlamydomonas* und Tabak wurde publiziert, dass *ycf1* essenziell für das Überleben der Pflanze ist (Boudreau *et al.*, 1997; Drescher *et al.*, 2000). In *A. thaliana* wurde diesem Protein eine Funktion in der Translokation über die innere Plastidenmembran zugesprochen (Kikuchi *et al.*, 2013). Das Gen mit bis dahin unbekannter Funktion soll für das Protein Tic214 kodieren, welches ein Teil des 1-MDa-TIC-Komplexes ist und dessen konservierte N-terminale Membrandomäne einen Teil des Kanals über die innere Plastidenmembran bildet (Nakai, 2015a). Das Protein wäre das einzige Protein des ganzen Komplexes, welches plastidenkodiert ist, während die anderen Komponenten kernkodiert sind. *A. thaliana*-Mutanten für die kernkodierten TIC-Proteine haben wie *rpoTp* einen chlorotischen Phänotyp, sind aber im Gegensatz zu *rpoTp* keimlingslethal (Kikuchi *et al.*, 2013; Köhler *et al.*, 2015). Das in Nakai (2015a, 2015b) vorgeschlagene Modell sieht den 1-MDa-TIC-Komplex als den Haupt-Importweg von überwiegend für die Photosynthese benötigten Proteine vor. Unterstützend für dieses Modell werden der chlorotischen Phänotyp der TIC-Mutanten sowie die reduzierte Akkumulation der Photosynthesepoteine in Plastiden genannt (Kikuchi *et al.*, 2013). Bei fehlender Transkription von *ycf1* (*tic214*) ist in *rpoTp* möglicherweise der Import in die Plastiden gestört, sodass wichtige Proteine, wie zum Beispiel die kernkodierten Bestandteile des Photosyntheseapparats, nicht oder ungenügend importiert werden. Jedoch werden auch andere Proteinarten mittels des 1-MDa-TIC-Komplexes importiert (Köhler *et al.*, 2016). Die Glaubwürdigkeit des Modells, bei welchem Tic214 als Teil eines Komplexes, welches den Hauptteil der Plastidenproteine importieren soll, fungiert, ist fraglich. Da *ycf1* evolutionär nicht konserviert ist und in vielen dikotylen und allen monokotylen Pflanzen fehlt (de Vries *et al.*, 2015), kann das Protein keine wichtige Komponente des Translocons sein. Gegen das Modell

sprechen Untersuchungen von Spectinomycin-behandelten Pflanzen, die kein *ycf1* translatieren können (Parker *et al.*, 2016). In diesen Pflanzen wurde der Import in Plastiden gezeigt und die wichtige Rolle von Ycf1 im plastidären Proteinimport angezweifelt (Bölter and Soll, 2017; Köhler *et al.*, 2016). Falls Ycf1 eine Rolle beim Proteinimport spielt, könnte dies die Ultrastruktur der Plastiden in *rpoTp* erklären (Kap. III.10). Mutanten für andere Komponenten des TIC-Komplexes zeigen ebenfalls leere Plastiden (Köhler *et al.*, 2015), was auf einen teilweise gestörten Import von kernkodierten Photosyntheseproteinen hinweisen könnte. Jedoch wurde auch bei diesen Mutanten Import nachgewiesen, was auf das Vorkommen eines alternativen Importwegs schließen ließ (Köhler *et al.*, 2015).

Für Ycf1 wurde eine weitere Funktion in der Biogenese der Photosynthese-Komplexe PSI, NDH und *Cytb<sub>6</sub>f* an Thylakoidmembranen vorgeschlagen (Yang *et al.*, 2016). Es wurde ein kernkodiertes RNA-Bindeprotein PBR1 identifiziert, welches speziell an *ycf1*-mRNA bindet und so die Translation reguliert (Yang *et al.*, 2016). In Mutanten mit fehlendem PBR1 ist die Translation von *ycf1* beeinträchtigt und die Abundanz mehrerer Proteine der Photosystemkomplexe PSI, NDH und *Cytb<sub>6</sub>f* reduziert. Die Detektion von Ycf1 in den Thylakoidmembranen mittels Western Blot deutete auf eine Interaktion mit den PSI und *Cytb<sub>6</sub>f* Komplexen hin, welche in der PBR1-Mutante fehlt und so die Biogenese der Komplexe beeinträchtigt. Auch die Interaktion von Ycf1 mit Ycf4 konnte gezeigt werden (Yang *et al.*, 2016). Weiterhin wurde gezeigt, dass die allotopische Expression von *ycf1* mit plastidärem Transitpeptid im Kern die Biogenesedefekte der oben genannten Komplexe in der PBR1-Mutante wieder hergestellt hat (Yang *et al.*, 2016). Das ist ein direkter Hinweis auf die Beteiligung von Ycf1 in diesen Prozess. In *rpoTp* könnte das Fehlen von Ycf1 folglich zu einer Beeinträchtigung der Assemblierung und/oder Stabilisierung der Photosynthese-Komplexe führen. Eine Untersuchung der Photosynthese-Komplexe der Thylakoidmembranen in *rpoTp* in einem Blaunativen Polyacrylamidgel und eine Western Blot-Analyse der Komplexe auf das Vorkommen von Ycf1 in Wildtyp und *rpoTp* könnte diese Frage beantworten. Wenn Ycf1 in die Biogenese der PS-Komplexe involviert ist, könnte das ebenfalls die fehlenden Thylakoiden in der Hälfte der Plastiden erklären, jedoch nicht die weniger ausgebildeten Grana-Thylakoiden, die in den mehr oder weniger wildtypisch aussehenden Plastiden vorkommen. Da PSI und *Cytb<sub>6</sub>f* in Stromathylakoiden lokalisiert sind (Austin and Staehelin, 2011), würde man eher die Beeinträchtigung in Stromathylakoiden erwarten.

Die differentielle Expressionsanalyse zeigt die mögliche Übernahme der *ycf1*-Transkription in *rpoTp* von RPOTmp (Abb. III.19). Es ist jedoch unklar, ob RPOTmp auch in Wildtyp *ycf1* transkribiert oder dieses nur ein sekundärer Effekt in *rpoTp* ist. Da *rpoTp* im Keimlingsstadium zwei Arten von Plastiden besitzt (Abb. III.41), normal aussehende und thylakoidlose, könnte man annehmen, dass hier eine basale Transkription von *ycf1* durch RPOTmp stattfindet, sodass die abnormalen Plastiden in weiterer Entwicklung nicht mehr auftreten. In *rpoTp*-Keimlingen wurde eine mögliche alternative TSS beobachtet (Abb. III.20), die zuvor als *Pycf1*-34 detektiert wurde und als eine PEP-abhängige TSS

dargestellt wird (Swiatecka-Hagenbruch *et al.*, 2008, 2007). Es ist jedoch wahrscheinlicher, dass diese TSS von RPOTmp genutzt werden könnte, da in *rpoTp* auch PEP-abhängige TSS reduziert genutzt werden sollten. Im frühen Stadium wird diese TSS nicht beobachtet, sodass die Transkription von *ycfI* durch RPOTmp in *rpoTp* möglicherweise erst im Keimlingsstadium einsetzt. Diese Vermutung korreliert auch mit der differentiellen Expressionsanalyse, die in *rpoTp* in frühem Stadium reduzierte *ycfI*-Transkriptmengen zeigt, während im späten Stadium erhöhte Mengen detektiert werden (Abb. III.19).

### IV.2.2 RPOTp transkribiert Komponenten des Translationsapparats

In Plastiden findet die Proteinbiosynthese an bakterienähnlichen 70S-Ribosomen statt, die aus einer kleinen (30S) und einer großen (50S) ribosomalen Untereinheit bestehen (Bieri *et al.*, 2017; Schippers and Mueller-Roeber, 2010; Tiller and Bock, 2014). Die Untereinheiten bestehen sowohl aus kernkodierten wie auch plastidkodierten Proteinen, wobei 21 von den insgesamt 55 Proteinen im Plastom zu finden sind (Yamaguchi *et al.*, 2000; Yamaguchi and Subramanian, 2000). Die im Laufe dieser Arbeit gewonnenen Erkenntnisse weisen auf die Funktion von RPOTp in der Transkription wichtiger Komponenten des Translationsapparats hin. Es wurden mehrere TSS für Gene für ribosomale Proteine wie *rpl32* identifiziert, welche ursächlich für die Transkription der Gene/Operons sind und gezeigt, dass das *rpl23*-Operon reduzierte Transkriptmengen in *rpoTp* aufweist (Abb. III.21 und III.22). Andere Publikationen stellten ebenfalls fest, dass RPOTp ribosomale Proteine während der Stratifizierung transkribiert (Demarsy *et al.*, 2012). Da sowohl *rpl23* wie auch *rpl32* für Pflanzen essentiell sind (Fleischmann *et al.*, 2011) könnte die starke Reduktion der Transkription dieser Gene zu einer gestörten Translation führen. Im *rpl23*-Operon sind Gene für die 30S-Untereinheit und die 50S-Untereinheit enthalten, was zu einer Beeinträchtigung der Assemblierung von sowohl den Untereinheiten wie auch des Ribosoms selbst führen könnte. Die Störung der Translation in *rpoTp* könnte das Fehlen von Thylakoid-Membranen in der Hälfte der Plastiden (Abb. III.41) und den Phänotyp (Abb. III.3) erklären, wenn wichtige Komponenten der Photosysteme, die plastomkodiert sind, nicht oder unzureichend translatiert werden. Auch hier könnte man das Verschwinden der thylakoidlosen Plastiden mit einer geringen Aktivität von RPOTmp in einen Zusammenhang bringen, welche geringe Transkriptlevel der ribosomalen Gene ermöglicht, die im Northern Blot in *rpoTp* zu beobachten sind (Abb. III.22 und III.23). Betrachtet man jedoch den chlorotischen Phänotyp und die Wachstumshemmung der RPOTp-Mutante, erscheint es naheliegend, dass die eventuelle Übernahme der Transkription an diesen TSS von RPOTmp nicht ausreichend für eine normale Plastidenentwicklung ist. Mutanten mit gestörter Ribosomenbiogenese zeigten einen ähnlichen Blattphänotyp wie die *rpoTp*-Mutante (Tiller and Bock, 2014). Eine Untersuchung der Translationseffizienz in *rpoTp* könnte potentiell Hinweise auf unzureichende Translation in der Mutante geben. Hier wäre eine Polysomenanalyse mit anschließendem Northern Blot (Barkan, 1998) oder für eine plastomweite Analyse eine Polysomenextraktion in Kombination mit einem Microarray (Kahlau and Bock, 2008)



möglich. Das Design des Experiments gestaltet sich jedoch als schwierig, da für einen quantitativen Vergleich Transkripte gewählt werden sollten, die vergleichbare Mengen in Wildtyp aufweisen. So kommen hier nur sehr wenige Transkripte in Frage, da in *rpoTp* sowohl RPOTp-abhängige wie auch PEP-abhängige Gene reduzierte Transkriptmengen aufweisen. Ein möglicher Kandidat könnte *clpP* sein, für welches eine TSS identifiziert wurde, die in überlappender Funktion von RPOTp und RPOTmp transkribiert wird und in *rpoTp* keine Reduktion zeigt (Kap. III.6.5).

### IV.2.3 RPOTmp transkribiert das *rrn16*-Operon vom Pc-Promotor während der Imbibition

Das *rrn16*-Operon besitzt zwei TSS, eine ist PEP-spezifisch (*Prrn16*-112, P2-Promotor), eine ist RPOTmp-spezifisch (*Prrn16*-139, Pc-Promotor) (Courtois *et al.*, 2007; Demarsy *et al.*, 2012; Swiatecka-Hagenbruch *et al.*, 2007). In der vorliegenden Arbeit konnte für RPOTmp außer der Transkription des *rrn16*-Operons vom Pc-Promotor (100873+RRN16S.1 bzw. 137776-RRN16S.2) keine andere individuelle Funktion gezeigt werden. Auch in früheren Untersuchungen des plastidären Transkriptoms mittels Microarray oder RT-qPCR konnten in *rpoTmp* keine großen Änderungen in der Transkriptabundanz feststellen werden (Kühn *et al.*, 2009; Tarasenko *et al.*, 2016) Interessant ist, dass der Pc-Promotor ausschließlich von RPOTmp genutzt wird, obwohl RPOTp und RPOTmp einige TSS in überlappender Funktion zur Transkription nutzen können (Kap. III.6.5) und folglich dieselbe Promotorart erkennen. Der Pc-Promotor wurde den Typ II-Nicht-Konsensus-Promotoren zugewiesen, da er kein YRTa-Motiv beinhaltet (Liere *et al.*, 2011). Jedoch ist RPOTmp nach den in dieser Arbeit erhaltenen Ergebnissen in der Lage, sowohl Typ I- wie auch Typ II-Promotoren mit YRTa-Motiv zu erkennen (Kap. III.6.5). Eine Microarray-Analyse des Transkriptoms in der frühen Samenentwicklung (Demarsy *et al.*, 2012) zog den Schluss, dass während der Imbibition die Transkription des *rrn16*-Operons sowohl von PEP (Transkription vom P2-Promotor) als auch von RPOTmp (Transkription vom Pc-Promotor) abhängig ist. Die Aktivierung des Pc-Promotors im Übergang vom trockenen zum imbibierten Samen wurde ebenfalls mittels *Primer Extension*-Analysen gezeigt (Courtois *et al.*, 2007), doch auch hier wurde kein Unterschied in der Nutzung der beiden Promotoren festgestellt. Mittels dRNA-Seq konnte ein quantitativer Vergleich der Nutzung dieser TSS gemacht werden. Dieser zeigt, dass während der Imbibition hauptsächlich RPOTmp *rrn16* transkribiert (Abb. III.38). Bereits nach der Illumination im keimenden Samen geht die Aktivität von RPOTmp zurück und das Operon wird im Licht vor allem von PEP transkribiert (Abb. III.38). Wenn ein Defizit des PEP-Enzyms besteht, wie es in *rpoTp* und *clb19* bzw. *ptac2* der Fall ist, wird in keimenden Samen hauptsächlich der Pc-Promotor als Transkriptionsstart verwendet (Abb. III.11 links), sodass hier RPOTmp für die reduzierte Aktivität von PEP kompensieren kann. Im Keimlingsstadium wird keine verstärkte Transkription vom Pc-Promotor detektiert, was auf eine Beschränkung der Aktivität von RPOTmp an diesem Promotor auf das frühe Entwicklungsstadium hindeuten könnte (Abb. III.11 rechts). In älteren Pflanzenstadien wird das Operon ebenfalls hauptsächlich von PEP transkribiert, eine Aktivität von RPOTmp ist jedoch gering vorhanden (Zoschke *et al.*, 2007).

### IV.2.4 PEP-Aktivität ist von beiden Phagentyp-Polymerasen abhängig

Bisher wurde die Meinung geäußert, dass nur RPO<sub>TP</sub> für die Transkription der PEP-Untereinheiten verantwortlich ist (Hricová *et al.*, 2006). Die Ergebnisse dieser Arbeit zeigen, dass beide Phagentyp-Polymerasen zur Transkription der PEP-Untereinheiten beitragen, da beide NEP-Mutanten reduzierte Transkriptmengen des *rpoB*-Operons aufweisen (Abb. III.30). Die Reduktion ist nur im frühen Stadium signifikant, während sich im späten Stadium die Transkriptmengen kaum von Wildtyp unterscheiden. Ausgehend von diesem Ergebnis ist eine entwicklungspezifische Regulation nicht auszuschließen, bei der in frühen, nicht-grünen Stadien der Pflanzenentwicklung die Aktivität beider NEPs erforderlich ist, während in späteren Stadien jeweils eine Phagentyp-Polymerase zur Transkription der *rpo*-Gene ausreicht. Es stellt sich jedoch die Frage, warum nur in *rpoTp* eine sekundäre Reduktion der PEP-abhängigen TSS zu beobachten ist. Es bleibt offen, ob die Reduktion ausschließlich auf der niedrigen Expression der Sigma-Faktoren in *rpoTp* zurückzuführen ist (Abb. VI.8), oder ob RPO<sub>TP</sub> einen größeren Beitrag zur Transkription dieser Gene leistet als RPO<sub>TMp</sub>. Bis jetzt wurde eine reduzierte PEP-Aktivität nur in *rpoTp* postuliert (Courtois *et al.*, 2007; Swiatecka-Hagenbruch *et al.*, 2008). In *rpoT<sub>mp</sub>* dagegen wurde keine Änderung der bekannten PEP-abhängigen Transkripte beobachtet (Kühn *et al.*, 2009; Tarasenko *et al.*, 2016). Weiterhin hat der Vergleich der Transkription des *rrn16*-Operons in *rpoTp* die reduzierte Nutzung des PEP-abhängigen P2-Promotors gezeigt, während in *rpoT<sub>mp</sub>* keine Transkriptionsreduktion von diesem Promotor nachgewiesen werden konnte (Courtois *et al.*, 2007). In der vorliegenden Arbeit wurden zwei TSS für Gene der PEP-Untereinheiten identifiziert (20245-RPOC2 und 78986-RPOA), welche mit dem statistischen Modell nicht eindeutig zu einer Polymerase-Aktivität zugewiesen werden konnten. Das Clustern mittels pamK konnte 20245-RPOC2 zu einem NEP-Cluster zuordnen (Anhang Tab. VI.13), was die überragende Funktion der NEPs bei der Transkription der PEP-Untereinheiten unterstützt. Die Spezifität der Bindung an die identifizierten Promotoren durch RPO<sub>TP</sub> und RPO<sub>TMp</sub> könnte mit *in-vitro*-Transkriptionsassays analysiert werden.

### IV.2.5 RPO<sub>TP</sub> und RPO<sub>TMp</sub> transkribieren *clpP*

Es wurden mehrere TSS identifiziert, welche keine oder geringe Reduktion in *rpoTp* und *rpoT<sub>mp</sub>* zeigen, jedoch eine erhöhte Nutzung in den PEP-Mutanten aufweisen (Anhang Tab. VI.14, bezeichnet als „potentiell NEP“). In mehreren Publikationen wurde beschrieben, dass in PEP-defizitären Mutanten und Mutanten mit verminderter PEP-Aktivität eine erhöhte Transkription von NEP-Promotoren beobachtet wurde (Allison *et al.*, 1996; Hajdukiewicz *et al.*, 1997; Loschelder *et al.*, 2006; Schweer *et al.*, 2010), sodass davon ausgegangen wird, dass diese TSS NEP-abhängig sind. Eine dieser TSS ist die zuvor identifizierte 71940-CLPP (*PclpP*-58). Da die Nutzung dieser TSS in keiner Mutante reduziert ist, konnte sie keiner Polymerase-Aktivität zugeordnet werden. Frühere Publikationen konnten feststellen, dass *clpP* durch NEP von dem Typ II-Nicht-Konsensus-Promotor *PclpP*-58 transkribiert wird (Hajdukiewicz *et al.*, 1997) und die Nutzung dieser TSS in der PEP-Mutante *ptac2* unverändert

ist (Pfalz *et al.*, 2006). Überexpression von RPOTp in Tabak und fehlende RPOTp-Aktivität in *Arabidopsis* zeigte keinen Einfluss auf die Nutzung des *clpP*-Promotors (Liere *et al.*, 2004; Swiatecka-Hagenbruch *et al.*, 2008) und es wurde zunächst angenommen, dass dieser Promotor RPOTmp-abhängig ist. Da auch der Pc-Promotor als Typ II-Nicht-Konsensus-Promotor definiert wurde, wurde vermutet, dass RPOTmp spezifisch diesen Promotortyp erkennt (Swiatecka-Hagenbruch *et al.*, 2008). Andere Publikationen konnten keine Reduktion der *clpP*-Transkriptmengen in *rpoTmp*-Mutanten feststellen (Baba *et al.*, 2004; Kühn *et al.*, 2009; Tarasenko *et al.*, 2016). In der vorliegenden Arbeit konnte durch Northern Blot und eine differentielle Expressionsanalyse gezeigt werden, dass *clpP* in überlappender Funktion von RPOTp und RPOMP transkribiert wird und bei Abwesenheit einer der Phagentyp-Polymerasen die andere Polymerase die komplette Transkription übernehmen kann (Abb. III.28). Die Ergebnisse deuten insgesamt darauf hin, dass beide NEPs die gleichen Promotorstrukturen *in-vivo* erkennen und beide Polymerasen sowohl Typ I- als auch Typ II-Promotoren nutzen können.

### IV.2.6 In Abwesenheit von RPOTp nutzt RPOTmp die gleichen TSS

Die Ergebnisse dieser Arbeit gezeigt, dass nur wenige TSS spezifisch von einer Phagentyp-RNA-Polymerase genutzt werden. Auch die Gene, für welche RPOTp-abhängige TSS identifiziert wurden, werden in Abwesenheit von RPOTp transkribiert. Da die RPOTmp-Mutante keine Photosynthesedefekte aufweist, kann vermutet werden, dass diese Gene in Wildtyp und *rpoTmp* ausschließlich von RPOTp transkribiert werden. Für einige dieser Transkripte wurde mittels Northern Blot eine starke Reduktion, aber keine Abwesenheit der Transkripte in *rpoTp* gezeigt (*rpl32*, *rpl23*, Abb. III.22 und III.23). Gleichzeitig wurde in *rpoTmp* keine Reduktion der Transkriptlevel beobachtet. Das weist eher auf eine sekundäre Aktivität von RPOTmp in der *rpoTp*-Mutante hin und zeigt, dass RPOTmp in der Lage ist, die gleichen TSS zu erkennen und zur Transkription zu nutzen, jedoch nicht in vollem Maße, wie es RPOTp tut.

### IV.2.7 Variable Motive in den NEP-Promotoren

Die Einteilung der NEP-Promotoren in die Typen Ia, Ib und II (Kap. I.5.1) erfolgte anhand weniger identifizierter TSS (Weihe and Börner, 1999). Typ Ia-Promotoren haben ein YRTa-Motiv (Y: C/T, R: A/G) gemeinsam, während Typ Ib-Promotoren zusätzlich ein GAA-Motiv stromaufwärts des YRTa-Motivs zeigten. Unter Typ II-Promotoren („Nicht-Konsensus“-Promotoren) wurden alle Promotoren zusammengefasst, welche kein YRTa-Motiv aufweisen. Zwischen diesen Typ II-Promotoren konnten jedoch bis jetzt keine Gemeinsamkeiten gefunden werden, die diese Einteilung in dieselbe Gruppe rechtfertigen (Weihe and Börner, 1999).

In der vorliegenden Arbeit zeigten nicht alle als RPOTp-, RPOTmp- oder RPOTp/RPOTmp-abhängig identifizierten TSS das YRTa-Motiv (Kap. III.6.4 und III.6.5). Neben den schon bekannten TSS ohne YRTa-Motiv wie 71940-CLPP1 (*PclpP*-58) (Sriraman *et al.*, 1998) und 100873+RRN16S.1 (*Prrn16*-139, Pc-Promotor) (Courtois *et al.*, 2007; Demarsy *et al.*, 2012; Swiatecka-Hagenbruch *et al.*, 2007)

wurden die RPOTp-abhängigen TSS 32644+PSBD, 7872-TRNQ und 56771-ATPB und NEP-abhängigen TSS 20245-RPOC2 und 112979+RPL32 identifiziert, in welchen das YRTa-Motiv nicht gefunden werden konnte. Es wurden nur zwei NEP-Promotoren mit einem möglichen GAA-Motiv identifiziert (Kap. III.6.4). Die Rolle der GAA-Box ist unklar, da bis jetzt nur für einen Promotor (*PatpB*-289) eine Funktionalität der GAA-Box gezeigt wurde (Xie and Allison, 2002). Auch in der genomweiten Analyse der TSS in Gerste wurden keine GAA-Motive identifiziert (Zhelyazkova *et al.*, 2012b). Mit der Einteilung der NEP-Promotoren nach Vorkommen von bestimmten Motiven wurde versucht, die Spezifitäten der NEPs zu erklären. Vorangehende Untersuchungen von NEP-Promotoren vermuteten, dass Typ I-Promotoren von RPOTp erkannt werden, während die Promotoren des Typs II ausschließlich von RPOTmp genutzt werden (Swiatecka-Hagenbruch *et al.*, 2008). Die Ergebnisse der vorliegenden Arbeit widersprechen dieser Annahme und zeigen, dass RPOTp und RPOTmp die gleichen Promotoren erkennen können, welche variable Motivsequenzen aufweisen (Kap. III.6.5). Ausgehend von den Ergebnissen der vorliegenden Arbeit kann spekuliert werden, dass die von den NEPs genutzten Promotormotive variabler sind, als bis jetzt angenommen und die vorgeschlagene Einteilung in die drei Gruppen unzureichend ist, um plastidäre Promotoren zu charakterisieren.

### IV.2.8 Modell für die Funktionen der Phagentyp-Polymerasen

Ausgehend von den Ergebnissen, welche die vorliegende Arbeit geliefert hat, wird ein Modell vorgeschlagen, welches die Funktionen der Phagentyp-Polymerasen beschreibt (Abb. IV.1). Die Ergebnisse zeigen sowohl individuelle als auch überlappende Rollen für RPOTp und RPOTmp in der Transkription. Im Wildtyp (Abb. IV.1 A) hat RPOTmp eine kleinere Rolle in der Transkription vom Pc-Promotor und der Co-Transkription von *clpP* zusammen mit RPOTp. RPOTp transkribiert in Wildtyp *ycfI* und einige ribosomale Gene wie *rpl32* und das *rpl23*-Operon. Zusammen mit RPOTmp wird *clpP* transkribiert. In Abwesenheit von RPOTmp (Abb. IV.1 B) kann RPOTp die Transkriptlevel von *clpP* alleine aufrechterhalten, es findet jedoch keine Transkription vom Pc-Promotor statt. Bei fehlender RPOTp-Aktivität (Abb. IV.1 C) kann RPOTmp wiederum die komplette Transkription von *clpP* übernehmen. Gleichfalls kann RPOTmp im Keimlingsstadium, aber nicht in keimenden Samen, zu einem geringen Level die Transkription von *ycfI* und den ribosomalen Genen übernehmen. Während der Imbibition übernehmen beide Polymerasen ihre individuellen Rollen in der Transkription von *ycfI* und *rrn16*, gleichzeitig transkribieren sie in überlappender Funktion *clpP* (Anhang Abb. VI.4).

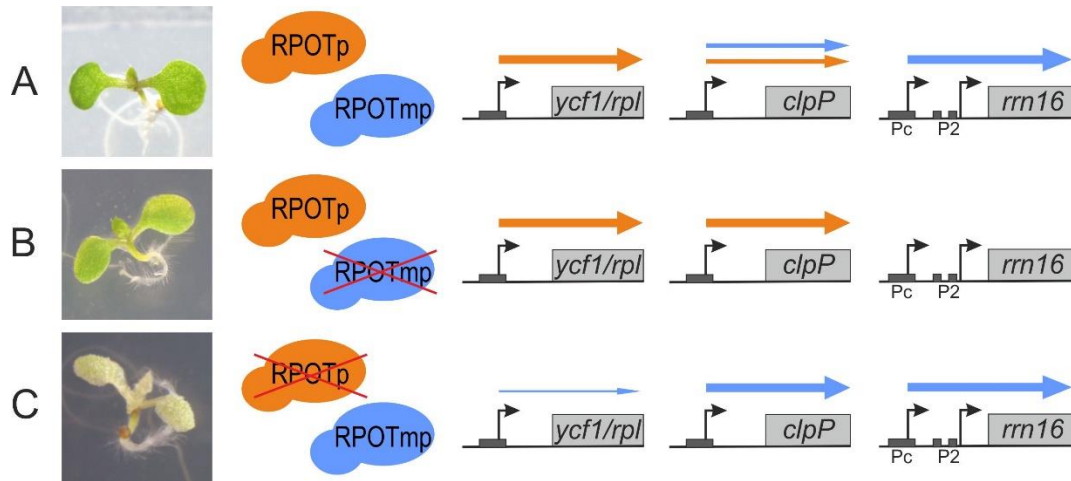


Abbildung IV.1: Funktion von RPOTp und RPOTmp in der Transkription der Plastidene gene *ycf1*, *rpl*, *clpP* und *rrn16*. A: Transkription in Wildtyp, B: Transkription in der *rpoTnp*-Mutante, C: Transkription in der *rpoTp*-Mutante. Die gebogenen Pfeile stellen die TSS dar. Farbige Pfeile zeigen die Transkriptionsaktivitäten der Polymerasen. Ein dicker Pfeil zeigt auf eine hohe, ein dünner Pfeil auf eine niedrige Transkriptionsaktivität. Pc: RPOTmp-abhängige TSS, P2: PEP-abhängige TSS.

### IV.3 Der embryolethalen NEP-Doppelmutante fehlt die Transkription essentieller Gene

Während der Samenreifeung durchläuft der Embryo mehrere Stadien, die sich unter anderem durch die Entwicklung von voll funktionsfähigen Chloroplasten auszeichnen (Le *et al.*, 2010). Dazu müssen die plastidären Transkriptionsmaschinerien aufgebaut werden, welche die Expression der plastidären Gene ermöglichen (Allorent *et al.*, 2013). Chloroplasten werden bereits in Embryonen im Globularstadium detektiert (Tejos *et al.*, 2010). Eine Doppelmutante, welcher RPOTp und RPOTmp fehlen, wurde im Laufe dieser Arbeit als embryolethal identifiziert (Kap. III.9). Den Ergebnissen zufolge fehlt in der Doppelmutante die Transkription essentieller Gene. Dazu gehören Gene für ribosomale Proteine (Fleischmann *et al.*, 2011) und *ycf1* (Boudreau *et al.*, 1997; Drescher *et al.*, 2000), was zu Embryolethalität führen könnte. Da die Entwicklung des Embryos im Globularstadium endet, scheint der kritische Punkt während der frühen Embryogenese zu sein, bei der speziell die Proteinmengen von RPOTp und RPOTmp steigen sollten (Allorent *et al.*, 2013) und die Expressionslevel der beiden Gene am höchsten sind (Kremnev and Strand, 2014). Gleichzeitig sollten in dieser Entwicklungsphase die Transkriptmengen von ribosomalen Proteinen steigen (Allorent *et al.*, 2013), sodass das Fehlen der NEPs in der Doppelmutante mit der fehlenden Transkription der ribosomalen Gene verknüpft werden könnte. Ein weiterer möglicher Grund der angehaltenen Weiterentwicklung des Embryos könnte die fehlende PEP-Aktivität sein. Da in der vorliegenden Arbeit bestätigt wurde, dass die Gene der PEP-Untereinheiten sowohl von RPOTp als auch von RPOTmp transkribiert werden, würde das Fehlen dieser Polymerasen auch bedeuten, dass kein PEP-Enzym vorhanden ist. Kremnev and Strand (2014) stellten fest, dass die PEP-Aktivität für die Entwicklung des Embryos essentiell ist und die Expression der plastidkodierte Photosynthesegene eine wichtige Rolle in der Embryoentwicklung spielt. Da der Embryo einen vollständigen Photosyntheseapparat während der Samenentwicklung exprimiert

(Allorent *et al.*, 2015), ist die PEP-Aktivität in diesen Stadien wichtig, um die Expression der plastidkodierte Photosynthesegene zu ermöglichen. Die Transkription der Gene der PEP-Untereinheiten startet im Globularstadium (Kremnev and Strand, 2014), was eine weitere Erklärung dafür sein könnte, warum der Embryo der Doppelmutante in diesem Stadium seine Entwicklung abbricht.

#### IV.4 Zusammenspiel von NEPs und PEP in der Transkription der Plastidengene

Die frühere Annahme, dass PEP hauptsächlich an der Transkription photosyntheserelevanter Gene/Operons beteiligt ist, während die NEP in *housekeeping*-Gene transkribiert (Börner *et al.*, 2015), konnte in dieser Arbeit nicht bestätigt werden. Es scheint in Plastiden eine viel größere Transkriptionsflexibilität möglich zu sein als bisher angenommen. Für die meisten Gene konnten mittels der optimierten dRNA-Seq mehrere TSS identifiziert werden. Da auch multiple Promotoren für mitochondriale Gene beschrieben wurden (Kühn *et al.*, 2005), scheint das für die Organellen allgemein zuzutreffen. Die meisten Gene/Operons haben mehrere Promotoren und diese werden von mehreren RNA-Polymerasen erkannt. Eine überlappende Funktion von PEP und NEP wurde bereits in früheren Publikationen vermutet, da trotz fehlender PEP-Aktivität die meisten Gene transkribiert worden sind (Legen *et al.*, 2002; Zhelyazkova *et al.*, 2012b). Vor allem tRNAs scheinen mehrere, oft mehr als zwei, Promotoren zu haben. Dies könnte für eine flexiblere Regulation des Translationsvorgangs während der Pflanzenentwicklung von Nutzen sein. Es wurden sowohl PEP- als auch NEP-abhängige TSS für dieselben tRNAs identifiziert, was auf eine vielseitige Regulation der tRNA-Expression hindeutet. Zuvor wurden bereits einige tRNAs als PEP-abhängig beschrieben (Williams-Carrier *et al.*, 2014) und auch die Beteiligung von dem Sigma-Faktor SIG2 (Kap. I.5.2) bei der Erkennung von tRNA-Promotoren gezeigt (Hanaoka *et al.*, 2003). Die Beteiligung von RPOTp an der Transkription von tRNAs wurde bis jetzt nicht festgestellt.

PEP und RPOTmp zeigen eine überlappende Funktion in der Transkription des *rrn16*-Operons. Beide Promotoren werden im Laufe der ganzen Entwicklung aktiv genutzt. Die Nutzung einer der TSS während der Pflanzenentwicklung variiert jedoch. In dieser Arbeit konnte gezeigt werden, dass im Laufe der Imbibition RPOTmp die Hauptaktivität in der Transkription des *rrn16*-Operons vom Pc-Promotor hat. Die Hauptaktivität von PEP am *Prrn16-112* setzt während der Keimung ein und die Aktivität von RPOTmp geht zurück (Abb. III.38), kann aber in geringem Maße detektiert werden. Ausgehend aus diesen Ergebnissen kann die Schlussfolgerung gezogen werden, dass das *rrn16*-Operon in *Arabidopsis* entwicklungspezifisch transkribiert wird.

Es konnten mindestens zwei Photosynthesegene *psbD* und *atpB* identifiziert werden, welche sowohl PEP- als auch RPOTp-abhängige TSS haben. Für *atpB* wurden insgesamt 5 TSS identifiziert, für *psbD* wurden 3 TSS identifiziert, von denen jeweils eine der RPOTp-Aktivität zugeordnet werden konnte. Dies deutet auf eine Beteiligung von RPOTp in der Transkription von Photosynthesegenen hin.

Ausgehend von den Ergebnissen der Northern Blot-Analyse scheint RPOTp einen geringen Beitrag zur Transkription dieser Gene zu leisten, während die hauptsächliche Transkription von den PEP-TSS ausgeht (Abb. III.24 und III.25). Der RPOTp-spezifische *atpB*-Promotor 56771-ATPB wurde zuvor als „rettende“ TSS beschrieben, die bei Abwesenheit von PEP aktiviert und im Wildtyp nicht zur Transkription verwendet wird (Schweer *et al.*, 2006). Für diese Annahme sprechen auch die Ergebnisse der vorliegenden Arbeit. In der Northern Blot-Analyse wird das für diese TSS 56771-ATPB spezifische Transkript in Wildtyp nicht detektiert (Abb. III.25). Gleichzeitig zeigt die dRNA-Seq-Analyse, dass diese RPOTp-abhängige TSS nur in den PEP-Mutanten signifikant angereichert ist, aber nicht in den anderen Genotypen. Die Reduktion der Transkriptmengen von *atpB* in der *rpoTp*-Mutante könnte ein Sekundäreffekt sein, der durch die reduzierte Menge des PEP-Enzyms auftritt. Zum anderen könnte eine weitere, RPOTp-abhängige TSS vorhanden sein, welche in dieser Arbeit nicht identifiziert, aber in anderen Analysen gefunden wurde (*PatpB*-318) (Schweer *et al.*, 2006; Swiatecka-Hagenbruch *et al.*, 2008, 2007). Durch die Nutzung dieser alternativen TSS in der PEP-Mutante wird in einer Northern Blot-Analyse ein langes Transkript detektiert (Abb. III.25). Schon zuvor wurde für PEP-Mutanten in Tabak das Vorkommen von längeren Transkripten beschrieben (Krause *et al.*, 2000; Legen *et al.*, 2002), welche auf alternativ genutzte TSS hindeuten. Das Vorkommen dieser verlängerten Transkripte kann auf den Ausfall bestimmter PEP-TSS und der daraus resultierenden Nutzung einer anderen TSS zurückzuführen sein, welche von RPOTp und/oder RPOTmp erkannt wird (Schweer *et al.*, 2006). Von den drei für *psbD* identifizierten TSS 31614+PSBD (PEP/RPOTp-abhängig), 32644+PSBD (RPOTp-abhängig) und 32672+PSBD (PEP/RPOTp-abhängig) ließ sich nur eine TSS eindeutig zuordnen. Ausgehend aus unseren Ergebnissen lässt sich jedoch nicht sagen, ob die Nutzung der RPOTp-abhängigen TSS 32644+PSBD einen wichtigen oder eher geringen Beitrag zur Transkription des Operons leistet. Da diese TSS sehr nahe zu 32672+PSBD liegt, können die Transkripte im Northern Blot nicht voneinander unterschieden werden (Abb. III.24). Die TSS 31614+PSBD (*PpsbD*-948) ist bereits zuvor identifiziert worden (Shimmura *et al.*, 2017) und hier kann der Schluss gezogen werden, dass diese TSS PEP-abhängig ist und in der *rpoTp*-Mutante eine sekundäre Reduktion zu beobachten ist. Diese TSS 31614+PSBD zeigt die typischen Konsensusmotive von PEP-Promotoren mit -10- und -35-Regionen (Shimmura *et al.*, 2017) und wird durch SIG5 aktiviert (Nagashima *et al.*, 2004; Tsunoyama *et al.*, 2004), sodass die TSS eindeutig der PEP-Aktivität zugeordnet werden kann.

### IV.5 Rolle von PEP in der frühen Samenkeimung

In trockenen Samen werden sowohl Transkripte als auch Proteine der PEP-Untereinheit detektiert (Allorent *et al.*, 2013; Demarsy *et al.*, 2006), was auf eine Funktion von PEP in nicht-grünen Entwicklungsstadien hinweist. In Plastiden von stratifizierten Samen wurde die PEP-Aktivität gezeigt und die Transkription von PEP-abhängigen Genen wie *rbcL* und *psbA* im keimenden Samen nachgewiesen (Demarsy *et al.*, 2012). Obwohl in dieser Publikation dargestellt wird, dass PEP erst nach der Illumination des Samens Photosynthesegene transkribiert, zeigen die vorliegenden Ergebnisse,

dass auch im Dunkeln bereits TSS für Photosynthesegene wie *psbE* genutzt werden (Anhang Tab. VI.15). Mit der sensitiveren Methode des dRNA-Seq konnte der Promotor 64445-PSBE (*PpsbE*-125) bereits in stratifizierten Samen identifiziert werden. Mit *Primer Extension*-Analysen wurde die aktive Transkription von diesem Promotor erst im keimenden Samen, also erst in der Lichtphase, gezeigt (Demarsy *et al.*, 2012). Die identifizierte Transkription von *psbE* im Dunkeln ist im Einklang mit der Beobachtung, dass Teile des PSII bereits in Etioplasten zu finden sind und unabhängig von Licht synthetisiert werden (Baena-González *et al.*, 2002). Zu dem Schluss, dass PEP bereits in nicht-grünen Pflanzenstadien und im Dunkeln Photosynthesegene transkribiert, kommt auch eine kürzlich publizierte Studie, die Transkription in Proplastiden analysiert (Ji *et al.*, 2020).

Die dRNA-Seq-Analyse zeigt ebenfalls, dass PEP in früher Samenkeimung auch eine wichtige Rolle in der Transkription von tRNAs und rRNAs spielt. Die PEP-abhängige Transkription von einigen tRNAs war bereits bekannt (Williams-Carrier *et al.*, 2014), jedoch wurde die Rolle von PEP in der Transkription der tRNAs in nicht-grünen Pflanzenstadien bis jetzt nicht untersucht. Gleichzeitig wurde die Beteiligung von den früh exprimierten Sigma-Faktoren SIG2 und SIG6 bei der Erkennung von Promotoren von tRNAs festgestellt (Hanaoka *et al.*, 2003; Ishizaki *et al.*, 2005; Kanamaru *et al.*, 2001). Das ist ein weiterer Hinweis dafür, dass PEP in der frühen Pflanzenentwicklung tRNAs transkribiert, was über die Bindung von SIG2 und SIG6 an die entsprechenden Promotorstrukturen vermittelt wird. Mit Microarray-Analysen konnte die PEP-Aktivität während der Imbibition ausschließlich für die Transkription des *rrn16*-Operons gezeigt werden (Demarsy *et al.*, 2012). In der vorliegenden Arbeit wurden neben dieser bekannten TSS für *rrn16* (P2-Promotor, 100898+RRN16S.1) eine weitere TSS 104475+RRN23S.1 identifiziert, welche vor *rrn23* liegt und von der PEP-Aktivität abhängig ist. Für beide diese TSS (100898+RRN16S.1 und 104475+RRN23S.1) wurde die Nutzung in stratifizierten Samen gezeigt (Anhang Abb. VI.4). Somit erfüllt PEP eine wichtige Funktion bereits in der Dunkelheit während der Stratifikation.

### IV.5.1 Regulation der PEP-Aktivität während der Pflanzenentwicklung

Es wurde gezeigt, dass in Etioplasten von Senf nur ein minimaler PEP-Komplex, bestehend aus den PEP-Untereinheiten (und einem Sigma-Faktor), vorhanden ist (Pfannschmidt and Link, 1994), was zu der Vermutung führte, dass im Licht und im Dunkeln die Zusammensetzung des PEP-Komplexes unterschiedlich ist und die Frage aufwarf, ob diese unterschiedlichen Komplexe unterschiedliche Rollen während der Lichtphase und der Dunkelphase in der Transkription erfüllen. Ji *et al.* (2020) konnten jedoch den vollen PEP-Komplex in Etioplasten von *A. thaliana* detektieren, sodass die Regulation der PEP-Aktivität im Licht und im Dunkeln vermutlich anders verläuft.

Da im PEP-Transkriptionskomplex die Bindung der Promotorregionen durch Sigma-Faktoren initiiert wird, kann vermutet werden, dass in frühen und späten Entwicklungsstadien der Pflanze verschiedene Sigma-Faktoren die Promotorbindung steuern und so die Funktion der PEP in verschiedenen



Entwicklungsstadien reguliert wird. SIG2 und SIG6 werden bereits einen Tag nach Imbibierung in *Arabidopsis* detektiert, während die anderen Sigma-Faktoren erst nach zwei Tagen oder später nachgewiesen werden können (Shiina *et al.*, 2009). Die Expression von *SIG2* und *SIG6* wurde in *A. thaliana* Zellkultur gezeigt, welche im Dunkeln angezogen wurde (Ji *et al.*, 2020). Dies unterstützt die Annahme, dass diese Sigma-Faktoren eine Funktion in sehr früher Pflanzenentwicklung bei der Promotorerkennung erfüllen. SIG2 wurde außerdem als essentiell für die Chloroplastenentwicklung beschrieben (Kanamaru and Tanaka, 2004). Mutanten beider Sigma-Faktoren zeigen einen chlorotischen Phänotyp in sehr frühen Entwicklungsstadien (Woodson *et al.*, 2015). In den Kotyledonen der Mutanten *sig2* und *sig6* wurden reduzierte Mengen der PEP-abhängigen Transkripte nachgewiesen. Während in *sig6* vor allem Transkripte der Photosynthesegene reduziert waren, zeigte *sig2* die Reduktion von einigen tRNAs, wobei die meisten Photosynthese-Transkripte nicht reduziert waren (Kanamaru *et al.*, 2001). So wird der Schluss gezogen, dass beide Sigma-Faktoren die Transkriptionsaktivität der PEP in früher Entwicklung der Pflanze steuern, wobei SIG2 die Transkription einiger tRNAs fördert, während SIG6 für die frühe Transkription von Photosynthesegenen verantwortlich ist. Anhand einer *sig2/sig6*-Doppelmutante konnte jedoch auch eine teilweise redundante Rolle gezeigt werden, da diese eine stärkere Reduktion von plastidären Transkripten aufwies, als die Einzelmutanten (Woodson *et al.*, 2015). *In-vivo*-Analysen bestätigten, dass SIG6 die Transkription vom *rrn16*-Operon vom PEP-abhängigen P2-Promotor initiiert (Ishizaki *et al.*, 2005). Studien *in-vitro* zeigen die mögliche Beteiligung von SIG2, aber nicht SIG1 oder SIG3, an der Transkription dieses Operons (Privat *et al.*, 2003). Jedoch ist der Mechanismus dieser möglichen Regulation der PEP-Transkriptionsaktivität durch Sigma-Faktoren nicht beschrieben. Alle Sigma-Faktoren scheinen die Standard-PEP-Promotorstrukturen mit -35 und -10-Konsensusmotiven zu erkennen und es wurde festgestellt, dass die Promotoren für *psbA* und *rbcL* von mehreren Sigma-Faktoren gebunden werden können (Shiina *et al.*, 2009). Die Effizienz der Promotorbindung der einzelnen Sigma-Faktoren scheint jedoch zu variieren (Schweer, 2010), sodass hier zusätzliche regulatorische Mechanismen eine Rolle spielen könnten. Die Phosphorylierung des Sigma-Faktors durch eine plastidäre Transkriptionskinase (PTK) ist ein möglicher Regulationsweg, wie die Transkription plastidärer Gene aktiviert oder gehemmt werden kann (Schweer *et al.*, 2010). Die Regulation könnte ebenfalls über Interaktionspartner der Sigma-Faktoren erfolgen. Ein PPR-Protein DG1 wurde identifiziert, das in die Regulation der frühen Chloroplastentwicklung involviert ist (Chi *et al.*, 2008) Dieses PPR-Protein interagiert mit SIG6 und gemeinsam scheinen sie die PEP-abhängige Genexpression zu steuern (Chi *et al.*, 2010).

## IV.6 Schlusswort

Im Laufe dieser Arbeit konnte eine Methode zur Kartierung von Transkriptionsstarts in Plastiden der dikotylen Pflanze *A.thaliana* etabliert werden. Mit Hilfe dieser Methode des *differential* RNA-Seq konnten Transkriptionsstarts erfolgreich identifiziert und quantitativ untersucht werden. Eine

Klassifizierung der TSS nach der Nutzung in Wildtyp sowie Mutanten der plastidären RNA-Polymerasen trug zum Verständnis der plastidären Transkription bei und ermöglichte die Zuweisung von individuellen sowie überlappenden Rollen der Phagentyp-Polymerasen. In der vorliegenden Arbeit wurden sowohl für die NEPs als auch für PEP wichtige Funktionen im Laufe der Pflanzenentwicklung gezeigt (Abb. IV.2). RPOTp und RPOTmp transkribieren in nicht-grünen und in grünen Stadien *housekeeping*-Gene wie *clpP* und Gene für die Untereinheiten der PEP. Die individuelle Funktion von RPOTp wurde bei der Transkription der Gene für ribosomale Proteine sowie *ycf1* nachgewiesen, während für RPOTmp die ausschließliche Transkription von *rrn16* vom Pc-Promotor bestätigt wurde. Die Hauptfunktion von PEP wechselt im Laufe der Entwicklung. In nicht-grünem Gewebe (Abb. IV.2 links) werden von PEP hauptsächlich tRNAs und rRNAs transkribiert, in grünen Pflanzenstadien transkribiert PEP zumeist Photosynthesegene (rechts). Bei der entwicklungspezifischen Funktion der PEP regulieren offenbar Sigma-Faktoren die PEP-Aktivität. So sind in nicht-grünen Stadien vor allem die Sigma-Faktoren *SIG2* und *SIG6* exprimiert, während in grünem Gewebe alle sechs Sigma-Faktoren *SIG1* bis *SIG6* an der Promotorerkennung beteiligt sind.

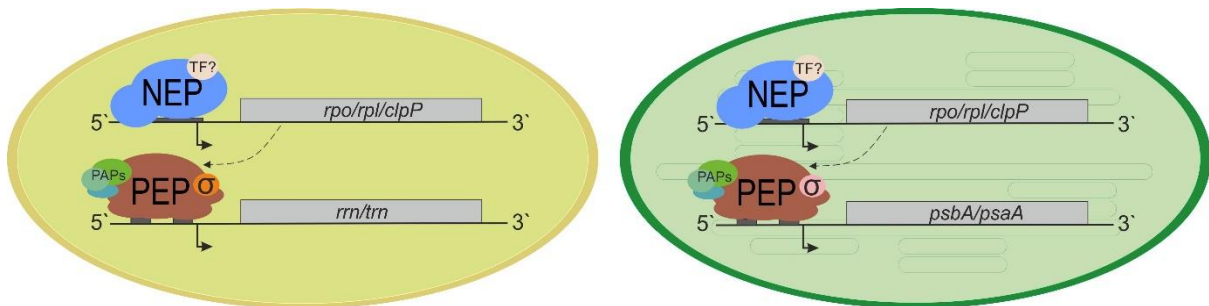


Abbildung IV.2: Funktionen der plastidären RNA-Polymerasen. Während die NEPs sowohl in nicht-grünen (links) als auch in grünen (rechts) Entwicklungsstadien hauptsächlich *housekeeping*-Gene transkribieren, transkribiert PEP im Dunkeln und während der frühen Samenkeimung tRNAs und rRNAs, was über *SIG2/SIG6* gesteuert wird. In Pflanzenstadien mit reifen Chloroplasten werden von PEP vor allem Photosynthesegene transkribiert. Dies wird über die Sigma-Faktoren *SIG1* bis *SIG6* vermittelt. PAP: PEP-assoziiertes Protein, TF: Transkriptionsfaktor.

## V LITERATURVERZEICHNIS

- Aird, D., Ross, M., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D., Nusbaum, C., Gnirke, A., 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* 12, 1–14.
- Allison, L.A., Simon, L.D., Maliga, P., 1996. Deletion of *rpoB* reveals a second distinct transcription system in plastids of higher plants. *EMBO J.* 15, 2802–2809.
- Allorent, G., Courtois, F., Chevalier, F., Lerbs-Mache, S., 2013. Plastid gene expression during chloroplast differentiation and dedifferentiation into non-photosynthetic plastids during seed formation. *Plant Mol. Biol.* 82, 59–70.
- Allorent, G., Osorio, S., Ly Vu, J., Falconet, D., Jouhet, J., Kuntz, M., Fernie, A.R., Lerbs-Mache, S., Macherel, D., Courtois, F., Finazzi, G., 2015. Adjustments of embryonic photosynthetic activity modulate seed fitness in *Arabidopsis thaliana*. *New Phytol.* 205, 707–719.
- Anders, S., Huber, W., 2010. Differential expression analysis for sequence count data. *Genome Biol.* 11, R106.
- Austin, J.R., Staehelin, A.L., 2011. Three-dimensional architecture of grana and stroma thylakoids of higher plants as determined by electron tomography. *Plant Physiol.* 155, 1601–1611.
- Azevedo, J., Courtois, F., Hakimi, M.-A., Demarsy, E., Lagrange, T., Alcaraz, J.-P., Jaiswal, P., Maréchal-Drouard, L., Lerbs-Mache, S., 2008. Intraplastidial trafficking of a phage-type RNA polymerase is mediated by a thylakoid RING-H2 protein. *Proc. Natl. Acad. Sci. U. S. A.* 105, 9123–9128.
- Azevedo, J., Courtois, F., Lerbs-Mache, S., 2006. Sub-plastidial localization of two different phage-type RNA polymerases in spinach chloroplasts. *Nucleic Acids Res.* 34, 436–44.
- Baba, K., Schmidt, J., Espinosa-Ruiz, A., Villarejo, A., Shiina, T., Gardeström, P., Sane, A.P., Bhallerao, R.P., 2004. Organellar gene transcription and early seedling development are affected in the *rpoT*;2 mutant of *Arabidopsis*. *Plant J.* 38, 38–48.
- Baena-González, E., Aro, E.M., Styring, S., Andersson, B., Hundal, T., Diner, B., Wikstrom, M., Fromme, P., Junge, W., 2002. Biogenesis, assembly and turnover of photosystem II units. *Philos. Trans. R. Soc. B Biol. Sci.* 357, 1451–1460.
- Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., Noble, W.S., 2009. MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res.* 37, 202–208.
- 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., 1998. Approaches to investigating nuclear genes that function in chloroplast biogenesis in land plants. *Methods Enzymol.* 297, 38–57.
- Barkan, A., 2011. Expression of Plastid Genes: Organelle-Specific Elaborations on a Prokaryotic Scaffold. *Plant Physiol.* 155, 1520–1532.
- Bensing, B. a, Meyer, B.J., Dunny, G.M., 1996. Sensitive detection of bacterial transcription initiation sites and differentiation from RNA processing sites in the pheromone-induced plasmid transfer system of *Enterococcus faecalis*. *Proc. Natl. Acad. Sci. U. S. A.* 93, 7794–7799.
- Berendzen, K., Searle, I., Ravenscroft, D., Koncz, C., Batschauer, A., Coupland, G., Somssich, I.E., Ülker, B., 2005. A rapid and versatile combined DNA/RNA extraction protocol and its application to the analysis of a novel DNA marker set polymorphic between *Arabidopsis thaliana* ecotypes Col-0 and Landsberg erecta. *Plant Methods* 1.
- Bieri, P., Leibundgut, M., Saurer, M., Boehringer, D., Ban, N., 2017. The complete structure of the chloroplast 70S ribosome in complex with translation factor pY. *EMBO J.* 36, 475–486.
- Bischler, T., Tan, H.S., Nieselt, K., Sharma, C.M., 2015. Differential RNA-seq (dRNA-seq) for annotation of transcriptional start sites and small RNAs in *Helicobacter pylori*. *Methods* 86, 89–101.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
- Bölter, B., Soll, J., 2017. Ycf1/Tic214 Is Not Essential for the Accumulation of Plastid Proteins. *Mol. Plant* 10, 219–221.

- Boorstein, W.R., Craig, E.A., 1989. Primer Extension Analysis of RNA. *Methods Enzymol.* 180, 347–369.
- Börner, T., Aleynikova, A.Y., Zubo, Y.O., Kusnetsov, V. V., 2015. Chloroplast RNA polymerases: Role in chloroplast biogenesis. *Biochim. Biophys. Acta - Bioenerg.* 1847, 761–769.
- Börner, T., Zhelyazkova, P., Legen, J., Schmitz-Linneweber, C., 2014. Chloroplast Gene Expression—RNA Synthesis and Processing, *Advances in Plant Biology.*
- Boudreau, E., Turmel, M., Goldschmidt-Clermont, M., Rochaix, J.D., Sivan, S., Michaels, A., Leu, S., 1997. A large open reading frame (orf1995) in the chloroplast DNA of *Chlamydomonas reinhardtii* encodes an essential protein. *Mol. Gen. Genet.* 253, 649–653.
- Boyes, D.C., Zayed, A.M., Ascenzi, R., McCaskill, A.J., Hoffman, N.E., Davis, K.R., Görlach, J., 2001. Growth stage-based phenotypic analysis of *Arabidopsis*: a model for high throughput functional genomics in plants. *Plant Cell* 13, 1499–510.
- Carey, M.F., Peterson, C.L., Smale, S.T., 2013. The primer extension assay. *Cold Spring Harb. Protoc.* 8, 164–173.
- Castandet, B., Germain, A., Hotto, A.M., Stern, D.B., 2019. Systematic sequencing of chloroplast transcript termini from *Arabidopsis thaliana* reveals >200 transcription initiation sites and the extensive imprints of RNA-binding proteins and secondary structures. *Nucleic Acids Res.* 47, 11889–11905.
- Castandet, B., Hotto, A.M., Fei, Z., Stern, D.B., 2013. Strand-specific RNA sequencing uncovers chloroplast ribonuclease functions. *FEBS Lett.* 587, 3096–3101.
- Chateigner-Boutin, A.-L., Ramos-Vega, M., Guevara-García, A., Andr s, C., de la Luz Guti rrez-Nava, M., Cantero, A., Delannoy, E., Jim nez, L.F., Lurin, C., Small, I., Le n, P., 2008. CLB19, a pentatricopeptide repeat protein required for editing of *rpoA* and *clpP* chloroplast transcripts. *Plant J.* 56, 590–602.
- Cheetham, G.M., Steitz, T.A., 2000. Insights into transcription: Structure and function of single-subunit DNA-dependent RNA polymerases. *Curr. Opin. Struct. Biol.* 10, 117–123.
- Chen, D., Patton, J.T., 2001. Reverse Transcriptase Adds Nontemplated Nucleotides to cDNAs During 5'-RACE and Primer Extension. *Biotechniques* 30, 574–582.
- Chi, W., He, B., Mao, J., Jiang, J., Zhang, L., 2015. Plastid sigma factors: Their individual functions and regulation in transcription. *Biochim. Biophys. Acta - Bioenerg.* 1847, 770–778.
- Chi, W., Ma, J., Zhang, D., Guo, J., Chen, F., Lu, C., Zhang, L., 2008. The pentatricopeptide repeat protein Delayed Greening1 is involved in the regulation of early chloroplast development and chloroplast gene expression in *Arabidopsis*. *Plant Physiol.* 147, 573–584.
- Chi, W., Mao, J., Li, Q., Ji, D., Zou, M., Lu, C., Zhang, L., 2010. Interaction of the pentatricopeptide-repeat protein DELAYED GREENING 1 with sigma factor SIG6 in the regulation of chloroplast gene expression in *Arabidopsis* cotyledons. *Plant J.* 64, 14–25.
- Churchman, L.S., Weissman, J.S., 2011. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 469, 368–373.
- Core, L.J., Waterfall, J.J., Lis, J.T., 2008. Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters 322.
- Courtois, F., Merendino, L., Demarsy, E., Mache, R., Lerbs-Mache, S., 2007. Phage-type RNA polymerase RPO<sub>tmp</sub> transcribes the *rrn* operon from the PC promoter at early developmental stages in *Arabidopsis*. *Plant Physiol.* 145, 712–721.
- Davis, M.C., Kesthely, C.A., Franklin, E.A., MacLellan, S.R., 2017. The essential activities of the bacterial sigma factor. *Can. J. Microbiol.* 63, 89–99.
- Day, D.A., 2004. Mitochondrial Structure and Function in Plants. *Plant Mitochondria From Genome to Funct.* 1–12.
- de Vries, J., Sousa, F.L., B lter, B., Soll, J., Goulida, S.B., 2015. YCF1: A green TIC? *Plant Cell* 27, 1827–1833.
- Demarsy, E., Buhr, F., Lambert, E., Lerbs-Mache, S., 2012. Characterization of the plastid-specific germination and seedling establishment transcriptional programme. *J. Exp. Bot.* 63, 925–939.
- Demarsy, E., Courtois, F., Azevedo, J., Buhot, L., Lerbs-Mache, S., 2006. Building up of the plastid transcriptional machinery during germination and early plant development. *Plant Physiol.* 142, 993–1003.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M.,

- Gingeras, T.R., 2013. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- Dobrogojski, J., Adamiec, M., Luciński, R., 2020. The chloroplast genome: a review. *Acta Physiol. Plant.* 42, 1–13.
- Drescher, A., Stephanie, R., Calsa, T., Carrer, H., Bock, R., 2000. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *Plant J.* 22, 97–104.
- Edwards, K., Johnstone, C., Thompson, C., 1991. A simple and rapid method for preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Res.* 19, 1349.
- Egorov, E.S., Merzlyak, E.M., Shelenkov, A.A., Britanova, O. V., Sharonov, G. V., Staroverov, D.B., Bolotin, D.A., Davydov, A.N., Barsova, E., Lebedev, Y.B., Shugay, M., Chudakov, D.M., 2015. Quantitative Profiling of Immune Repertoires for Minor Lymphocyte Counts Using Unique Molecular Identifiers. *J. Immunol.* 194, 6155–6163.
- Eibl, C., Zou, Z., Beck, A., Kim, M., Mullet, J., Koop, H.U., 1999. In vivo analysis of plastid psbA, rbcL and rpl32 UTR elements by chloroplast transformation: Tobacco plastid gene expression is controlled by modulation of transcript levels and translation efficiency. *Plant J.* 19, 333–345.
- Emanuel, C., Von Groll, U., Müller, M., Börner, T., Weihe, A., 2006. Development- and tissue-specific expression of the RpoT gene family of Arabidopsis encoding mitochondrial and plastid RNA polymerases. *Planta* 223, 998–1009.
- Fey, V., Wagner, R., Bräutigam, K., Wirtz, M., Hell, R., Dietzmann, A., Leister, D., Oelmüller, R., Pfanschmidt, T., 2005. Retrograde plastid redox signals in the expression of nuclear genes for chloroplast proteins of Arabidopsis thaliana. *J. Biol. Chem.* 280, 5318–5328.
- Fleischmann, T.T., Scharff, L.B., Alkatib, S., Hasdorf, S., Schöttler, M.A., Bock, R., 2011. Nonessential plastid-encoded ribosomal proteins in tobacco: A developmental role for plastid translation and implications for reductive genome evolution. *Plant Cell* 23, 3137–3155.
- Frank, J., Happeck, R., Meier, B., Hoang, M.T.T., Stribny, J., Hause, G., Ding, H., Morsomme, P., Baginsky, S., Peiter, E., 2019. Chloroplast-localized BICAT proteins shape stromal calcium signals and are required for efficient photosynthesis. *New Phytol.* 221, 866–880.
- Fu, Y., Wu, P.H., Beane, T., Zamore, P.D., Weng, Z., 2018. Elimination of PCR duplicates in RNA-seq and small RNA-seq using unique molecular identifiers. *BMC Genomics* 1–14.
- Gentle, J.E., Kaufman, L., Rousseuw, P.J., 1991. Finding Groups in Data : An Introduction to Cluster Analysis. *Biometrics* 47, 788.
- Ghulam, M.M., Courtois, F., Lerbs-Mache, S., Merendino, L., 2013. Complex processing patterns of mRNAs of the large ATP synthase operon in Arabidopsis chloroplasts. *PLoS One* 8, 3–10.
- Ghulam, M.M., Zghidi-Abouzid, O., Lambert, E., Lerbs-Mache, S., Merendino, L., 2012. Transcriptional organization of the large and the small ATP synthase operons, atpI/H/F/A and atpB/E, in Arabidopsis thaliana chloroplasts. *Plant Mol. Biol.* 79, 259–72.
- Girardot, C., Scholtalbers, J., Sauer, S., Su, S.Y., Furlong, E.E.M., 2016. Je, a versatile suite to handle multiplexed NGS libraries with unique molecular identifiers. *BMC Bioinformatics* 17, 4–9.
- Gray, M.W., 1999. Evolution of organellar genomes. *Curr. Opin. Genet. Dev.* 9, 678–687.
- Greiner, S., Lehwark, P., Bock, R., 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47, W59–W64.
- Hajdukiewicz, P.T.J., Allison, L. a., Maliga, P., 1997. The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. *EMBO J.* 16, 4041–4048.
- Hanaoka, M., Kanamaru, K., Fujiwara, M., Takahashi, H., Tanaka, K., 2005. Glutamyl-tRNA mediates a switch in RNA polymerase use during chloroplast biogenesis. *EMBO Rep.* 6, 545–550.
- Hanaoka, M., Kanamaru, K., Takahashi, H., Tanaka, K., 2003. Molecular genetic analysis of chloroplast gene promoters dependent on SIG2, a nucleus-encoded sigma factor for the plastid-encoded RNA polymerase, in Arabidopsis thaliana. *Nucleic Acids Res.* 31, 7090–7098.
- Hedtke, B., Börner, T., Weihe, a, 2000. One RNA polymerase serving two genomes. *EMBO Rep.* 1, 435–440.
- Hedtke, B., Börner, T., Weihe, A., 1997. Mitochondrial and chloroplast phage-type RNA polymerases in Arabidopsis. *Science* 277, 809–811.
- Hess, W.R., Prombona, A., Fieder, B., Subramanian, a R., Börner, T., 1993. Chloroplast rps15 and

- the rpoBIC1/C2 gene cluster are strongly transcribed in ribosome-deficient plastids: evidence for a functioning non-chloroplast-encoded RNA polymerase. *EMBO J.* 12, 563–571.
- Hetzl, J., Duttke, S.H., Benner, C., Chory, J., 2016. Nascent RNA sequencing reveals distinct features in plant transcription. *Proc. Natl. Acad. Sci.* 201603217.
- Hong, J., Gresham, D., 2017. Incorporation of unique molecular identifiers in TruSeq adapters improves the accuracy of quantitative sequencing. *Biotechniques* 63, 221–226.
- Hotto, A.M., Germain, A., Stern, D.B., 2012. Plastid non-coding RNAs: Emerging candidates for gene regulation. *Trends Plant Sci.* 17, 737–744.
- Hotto, A.M., Schmitz, R.J., Fei, Z., Ecker, J.R., Stern, D.B., Kim, J.K., 2011. Unexpected Diversity of Chloroplast Noncoding RNAs as Revealed by Deep Sequencing of the Arabidopsis Transcriptome. *Genes|Genomes|Genetics* 1, 559–570.
- Hricová, A., Quesada, V., Micol, J.L., 2006. The SCABRA3 nuclear gene encodes the plastid RpoTp RNA polymerase, which is required for chloroplast biogenesis and mesophyll cell proliferation in Arabidopsis. *Plant Physiol.* 141, 942–56.
- Igloi, G.L., Kössel, H., 1992. The Transcriptional Apparatus of Chloroplasts. *CRC. Crit. Rev. Plant Sci.* 10, 525–558.
- Ishizaki, Y., Tsunoyama, Y., Hatano, K., Ando, K., Kato, K., Shinmyo, A., Kobori, M., Takeba, G., Nakahira, Y., Shiina, T., 2005. A nuclear-encoded sigma factor, Arabidopsis SIG6, recognizes sigma-70 type chloroplast promoters and regulates early chloroplast development in cotyledons. *Plant J.* 42, 133–144.
- Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnerberg, P., Linnarsson, S., 2014. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* 11, 163–166.
- Jarvis, P., López-Juez, E., 2013. Biogenesis and homeostasis of chloroplasts and other plastids. *Nat. Rev. Mol. Cell Biol.* 14, 787–802.
- Ji, Y., Lehotai, N., Zan, Y., Dubreuil, C., Guinea Díaz, M., Strand, Å., 2020. A fully assembled PEP complex detected in etioplasts and proplastids in Arabidopsis. *Physiol. Plant.* 1–12.
- Kahlau, S., Bock, R., 2008. Plastid transcriptomics and translatomics of tomato fruit development and chloroplast-to-chromoplast differentiation: Chromoplast gene expression largely serves the production of a single protein. *Plant Cell* 20, 856–874.
- Kanamaru, K., Nagashima, A., Fujiwara, M., Shimada, H., Shirano, Y., Nakabayashi, K., Shibata, D., Tanaka, K., Takahashi, H., 2001. An arabidopsis sigma factor (SIG2)-dependent expression of plastid-encoded tRNAs in chloroplasts. *Plant Cell Physiol.* 42, 1034–1043.
- Kanamaru, K., Tanaka, K., 2004. Roles of chloroplast RNA polymerase sigma factors in chloroplast development and stress response in higher plants. *Biosci. Biotechnol. Biochem.* 68, 2215–2223.
- Kapoor, S., Sugiura, M., 1999. Identification of two essential sequence elements in the nonconsensus type II PatpB-290 plastid promoter by using plastid transcription extracts from cultured tobacco BY-2 cells. *Plant Cell* 11, 1799–1810.
- Kapoor, S., Suzuki, J.Y., Sugiura, M., 1997. Identification and functional significance of a new class of non-consensus-type plastid promoters. *Plant J.* 11, 327–337.
- Kapoor, S., Wakasugi, T., Deno, H., Sugiura, M., 1994. An atpE-specific promoter within the coding region of the atpB gene in tobacco chloroplast DNA. *Curr. Genet.* 26, 263–268.
- Kikuchi, S., Asakura, Y., Imai, M., Nakahira, Y., Kotani, Y., Hashiguchi, Y., Nakai, Y., Takafuji, K., Bédard, J., Hirabayashi-Ishioka, Y., Mori, H., Shiina, T., Nakai, M., 2018. A Ycf2-FtsHi heteromeric AAA-ATPase complex is required for chloroplast protein import. *Plant Cell* 30, 2677–2703.
- Kikuchi, S., Bédard, J., Hirano, M., Hirabayashi, Y., Oishi, M., Imai, M., Takase, M., Ide, T., Nakai, M., 2013. Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science* (80-. ). 339, 571–574.
- Kivioja, T., Vähärautio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S., Taipale, J., 2012. Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* 9, 72–74.
- Köhler, D., Helm, S., Agne, B., Baginsky, S., 2016. Importance of translocon subunit Tic56 for rRNA processing and chloroplast ribosome assembly. *Plant Physiol.* 172, 2429–2444.
- Köhler, D., Montandon, C., Hause, G., Majovsky, P., Kessler, F., Baginsky, S., Agne, B., 2015.

- Characterization of chloroplast protein import without Tic56, a component of the 1-megadalton translocon at the inner envelope membrane of chloroplasts. *Plant Physiol.* 167, 972–990.
- Krause, K., Maier, R.M., Kofer, W., Krupinska, K., Herrmann, R.G., 2000. Disruption of plastid-encoded RNA polymerase genes in tobacco: expression of only a distinct set of genes is not based on selective transcription of the plastid chromosome. *Mol. Gen. Genet.* 263, 1022–30.
- Kremnev, D., Strand, Å., 2014. Plastid encoded RNA polymerase activity and expression of photosynthesis genes required for embryo and seed development in Arabidopsis. *Front. Plant Sci.* 5, 1–12.
- Kruesi, W.S., Core, L.J., Waters, C.T., Lis, J.T., Meyer, B.J., 2013. Condensin controls recruitment of RNA polymerase ii to achieve nematode X-chromosome dosage compensation. *Elife* 2013, 1–31.
- Krupinska, K., Melonek, J., Krause, K., 2013. New insights into plastid nucleoid structure and functionality. *Planta* 237, 653–664.
- Kühn, K., Bohne, A.-V., Liere, K., Weihe, A., Börner, T., 2007. Arabidopsis Phage-Type RNA Polymerases: Accurate in Vitro Transcription of Organellar Genes. *Plant Cell Online* 19, 959–971.
- Kühn, K., Richter, U., Meyer, E.H., Delannoy, E., Falcon de Longevialle, A., O’Toole, N., Börner, T., Millar, A.H., Small, I.D., Whelan, J., 2009. Phage-Type RNA Polymerase RPOTmp Performs Gene-Specific Transcription in Mitochondria of Arabidopsis thaliana. *Plant Cell Online* 21, 2762–2779.
- Kwak, H., Fuda, N.J., Core, L.J., Lis, J.T., 2012. Precise Maps of RNA Polymerase Reveal How Promoters Direct Initiation and Pausing. *Science* (80-. ). 339, 950–953.
- Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10.
- Le, B.H., Cheng, C., Bui, A.Q., Wagmaister, J.A., Henry, K.F., Pelletier, J., Kwong, L., Belmonte, M., Kirkbride, R., Horvath, S., Drews, G.N., Fischer, R.L., Okamuro, J.K., Harada, J.J., Goldberg, R.B., 2010. Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* 107, 8063–8070.
- Legen, J., Kemp, S., Krause, K., Profanter, B., Herrmann, R.G., Maier, R.M., 2002. Comparative analysis of plastid transcription profiles of entire plastid chromosomes from tobacco attributed to wild-type and PEP-deficient transcription machineries. *Plant J.* 31, 171–88.
- Lerbs-Mache, S., 1993. The 110-kDa polypeptide of spinach plastid DNA-dependent RNA polymerase: Single-subunit enzyme or catalytic core of multimeric enzyme complexes? 90, 5509–5513.
- Lerbs-Mache, S., 2011. Function of plastid sigma factors in higher plants: Regulation of gene expression or just preservation of constitutive transcription? *Plant Mol. Biol.* 76, 235–249.
- Li, H., 2011a. Improving SNP discovery by base alignment quality. *Bioinformatics* 27, 1157–1158.
- Li, H., 2011b. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Liere, K., Börner, T., 2007. Transcription and transcriptional regulation in plastids. *Top. Curr. Genet.* 19, 121–174.
- Liere, K., Kaden, D., Maliga, P., Börner, T., 2004. Overexpression of phage-type RNA polymerase RpoTp in tobacco demonstrates its role in chloroplast transcription by recognizing a distinct promoter type. *Nucleic Acids Res.* 32, 1159–1165.
- Liere, K., Kestermann, M., Müller, U., Link, G., 1995. Identification and characterization of the Arabidopsis thaliana chloroplast DNA region containing the genes psbA, trnH and rps19’. *Curr. Genet.* 28, 128–130.
- Liere, K., Maliga, P., 1999. In vitro characterization of the tobacco rpoB promoter reveals a core sequence motif conserved between phage-type plastid and plant mitochondrial promoters. *EMBO J.* 18, 249–257.
- Liere, K., Maliga, P., 2001. Plastid RNA Polymerases in Higher Plants. *Regul. Photosynth.* 29–49.

- Liere, K., Weihe, A., Börner, T., 2011. The transcription machineries of plant mitochondria and chloroplasts: Composition, function, and regulation. *J. Plant Physiol.* 168, 1345–1360.
- Loschelder, H., Schweer, J., Link, B., Link, G., 2006. Dual temporal role of plastid sigma factor 6 in *Arabidopsis* development. *Plant Physiol.* 142, 642–650.
- Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 1–21.
- Lysenko, E. a, Kuznetsov, V. V., 2005. [Plastid RNA polymerases]. *Mol. Biol. (Mosk).* 39, 762–75.
- Manavski, N., Schmid, L.M., Meurer, J., 2018. RNA-stabilization factors in chloroplasts of vascular plants. *Essays Biochem.* 62, 51–64.
- Matteau, D., Rodrigue, S., 2015. Precise Identification of Genome-Wide Transcription Start Sites in Bacteria by 5'-Rapid Amplification of cDNA Ends (5'-RACE), *DNA-Protein Interactions: Principles and Protocols: Fourth Edition.*
- McFadden, G.I., 2014. Origin and evolution of plastids and photosynthesis in eukaryotes. *Cold Spring Harb. Perspect. Biol.* 6, 1–9.
- Mitschke, J., Georg, J., Scholz, I., Sharma, C.M., Dienst, D., Bantscheff, J., Voß, B., Steglich, C., Wilde, A., Vogel, J., Hess, W.R., 2011. An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. *Proc. Natl. Acad. Sci. U. S. A.* 108, 2124–2129.
- Nagashima, A., Hanaoka, M., Shikanai, T., Fujiwara, M., Kanamaru, K., Takahashi, H., Tanaka, K., 2004. The multiple-stress responsive plastid sigma factor, SIG5, directs activation of the psbD Blue Light-Responsive Promoter (BLRP) in *Arabidopsis thaliana*. *Plant Cell Physiol.* 45, 357–368.
- Nakai, M., 2015a. The TIC complex uncovered: The alternative view on the molecular mechanism of protein translocation across the inner envelope membrane of chloroplasts. *Biochim. Biophys. Acta - Bioenerg.* 1847, 957–967.
- Nakai, M., 2015b. Ycf1: A green TIC: Response to the de Vries et al. Commentary. *Plant Cell* 27, 1834–1838.
- Narsai, R., Law, S.R., Carrie, C., Xu, L., Whelan, J., 2011. In-depth temporal transcriptome profiling reveals a crucial developmental switch with roles for RNA processing and organelle metabolism that are essential for germination in *Arabidopsis*. *Plant Physiol.* 157, 1342–1362.
- Oñate-Sánchez, L., Vicente-Carbajosa, J., 2008. DNA-free RNA isolation protocols for *Arabidopsis thaliana*, including seeds and siliques. *BMC Res. Notes* 1, 1–7.
- Ortelt, J., Link, G., 2014. Plastid gene transcription: Promoters and RNA polymerases. *Methods Mol. Biol.* 1132, 47–72.
- Parker, N., Wang, Y., Meinke, D., 2016. Analysis of *Arabidopsis* accessions hypersensitive to a loss of chloroplast translation. *Plant Physiol.* 172, 1862–1875.
- Pfalz, J., Liere, K., Kandlbinder, A., Dietz, K.-J., Oelmüller, R., 2006. pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. *Plant Cell* 18, 176–197.
- Pfalz, J., Pfannschmidt, T., 2013. Essential nucleoid proteins in early chloroplast development. *Trends Plant Sci.* 18, 186–194.
- Pfannschmidt, Link, G., 1997. The A and B forms of plastid DNA-dependent RNA polymerase from mustard (*Sinapis alba* L.) transcribe the same genes in a different developmental context. *Mol. Gen. Genet.* 257, 35–44.
- Pfannschmidt, T., Blanvillain, R., Merendino, L., Courtois, F., Chevalier, F., Liebers, M., Grübler, B., Hommel, E., Lerbs-Mache, S., 2015. Plastid RNA polymerases: Orchestration of enzymes with different evolutionary origins controls chloroplast biogenesis during the plant life cycle. *J. Exp. Bot.* 66, 6957–6973.
- Pfannschmidt, T., Link, G., 1994. Separation of two classes of plastid DNA-dependent RNA polymerases that are differentially expressed in mustard (*Sinapis alba* L.) seedlings. *Plant Mol. Biol.* 25, 69–81.
- Pogson, B.J., Ganguly, D., Albrecht-Borth, V., 2015. Insights into chloroplast biogenesis and development. *Biochim. Biophys. Acta - Bioenerg.* 1847, 1017–1024.
- Powikrowska, M., Oetke, S., Jensen, P.E., Krupinska, K., 2014. Dynamic composition, shaping and organization of plastid nucleoids. *Front. Plant Sci.* 5, 1–13.



- Privat, I., Hakimi, M.A., Buhot, L., Favory, J.J., Lerbs-Mache, S., 2003. Characterization of Arabidopsis plastid sigma-like transcription factors SIG1, SIG2 and SIG3. *Plant Mol. Biol.* 51, 385–399.
- Quinlan, A.R., Hall, I.M., 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., Manke, T., 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 44, W160–W165.
- Rolland, N., Bouchnak, I., Moyet, L., Salvi, D., Kuntz, M., 2018. The main functions of plastids. *Methods Mol. Biol.* 1829, 73–85.
- Rousseau-Gueutin, M., Lloyd, A.H., Sheppard, A.E., Timmis, J.N., 2012. Organelle genetics: Evolution of organelle genomes and gene expression, Bullerwell C. (eds) *Organelle Genetics*.
- Ruf, S., Forner, J., Hasse, C., Kroop, X., Seeger, S., Schollbach, L., Schadach, A., Bock, R., 2019. High-efficiency generation of fertile transplastomic Arabidopsis plants. *Nat. Plants* 5, 282–289.
- 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.
- Ruwe, H., Wang, G., Gusewski, S., Schmitz-Linneweber, C., 2016. Systematic analysis of plant mitochondrial and chloroplast small RNAs suggests organelle-specific mRNA stabilization mechanisms. *Nucleic Acids Res.* 44, 7406–7417.
- Sambrook, J., Russell, D.W., 2001. *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Saunders, K., Bert, A.G., Dredge, B.K., Toubia, J., Gregory, P.A., Pillman, K.A., Goodall, G.J., Bracken, C.P., 2020. Insufficiently complex unique-molecular identifiers (UMIs) distort small RNA sequencing. *Sci. Rep.* 10, 1–9.
- Schippers, J.H.M., Mueller-Roeber, B., 2010. Ribosomal composition and control of leaf development. *Plant Sci.* 179, 307–315.
- Schweer, J., 2010. Plant sigma factors come of age : flexible transcription factor network for regulated plastid gene expression. *J. Endocytobiosis Cell Res.* 20, 1–12.
- Schweer, J., Loschelder, H., Link, G., 2006. A promoter switch that can rescue a plant sigma factor mutant. *FEBS Lett.* 580, 6617–6622.
- Schweer, J., Türkeri, H., Kolpack, A., Link, G., 2010. Role and regulation of plastid sigma factors and their functional interactors during chloroplast transcription - Recent lessons from Arabidopsis thaliana. *Eur. J. Cell Biol.* 89, 940–946.
- Scotto-Lavino, E., Du, G., Frohman, M.A., 2007. 5' end cDNA amplification using classic RACE. *Nat. Protoc.* 1, 2555–2562.
- Sharma, C.M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiß, S., Sittka, A., Chabas, S., Reiche, K., Hackermüller, J., Reinhardt, R., Stadler, P.F., Vogel, J., 2010. The primary transcriptome of the major human pathogen Helicobacter pylori. *Nature* 464, 250–255.
- Sharma, C.M., Vogel, J., 2014. Differential RNA-seq: The approach behind and the biological insight gained. *Curr. Opin. Microbiol.* 19, 97–105.
- Shen, Y., Danon, A., Christopher, D.A., 2001. RNA binding-proteins interact specifically with the Arabidopsis chloroplast psbA mRNA 5' untranslated region in a redox-dependent manner. *Plant Cell Physiol.* 42, 1071–1078.
- Shiina, T., Tsunoyama, Y., Nakahira, Y., Khan, M.S., 2005. Plastid RNA polymerases, promoters, and transcription regulators in higher plants. *Int. Rev. Cytol.* 244, 1–68.
- Shimmura, S., Nozoe, M., Kitora, S., Kin, S., Matsutani, S., Ishizaki, Y., Nakahira, Y., Shiina, T., 2017. Comparative analysis of chloroplast psbD promoters in terrestrial plants. *Front. Plant Sci.* 8, 1–13.
- Sriraman, P., Silhavy, D., Maliga, P., 1998. The phage-type PcpP-53 plastid promoter comprises sequences downstream of the transcription initiation site. *Nucleic Acids Res.* 26, 4874–4879.
- Stern, D.B., Goldschmidt-Clermont, M., Hanson, M.R., 2010. Chloroplast RNA metabolism. *Annu. Rev. Plant Biol.* 61, 125–155.
- Suay, L., Salvador, M.L., Abesha, E., Klein, U., 2005. Specific roles of 5' RNA secondary structures in stabilizing transcripts in chloroplasts. *Nucleic Acids Res.* 33, 4754–4761.
- Sugiura, M., 1992. *The Chloroplast Genome*. *Plant Mol. Biol.* 19, 149–168.

- Swiatecka-Hagenbruch, M., Emanuel, C., Hedtke, B., Liere, K., Börner, T., 2008. Impaired function of the phage-type RNA polymerase RpoTp in transcription of chloroplast genes is compensated by a second phage-type RNA polymerase. *Nucleic Acids Res.* 36, 785–92.
- Swiatecka-Hagenbruch, M., Liere, K., Börner, T., 2007. High diversity of plastidial promoters in *Arabidopsis thaliana*. *Mol. Genet. Genomics* 277, 725–734.
- Tan, X.Y., Liu, X.L., Wang, W., Jia, D.J., Chen, L.Q., Zhang, X.Q., Ye, D., 2010. Mutations in the *Arabidopsis* nuclear-encoded mitochondrial phage-type rna polymerase gene *rpotm* led to defects in pollen tube growth, female gametogenesis and embryogenesis. *Plant Cell Physiol.* 51, 635–649.
- Tarasenko, V.I., Katyshev, A.I., Yakovleva, T. V., Garnik, E.Y., Chernikova, V. V., Konstantinov, Y.M., Koulintchenko, M. V., 2016. RPOTmp, an *Arabidopsis* RNA polymerase with dual targeting, plays an important role in mitochondria, but not in chloroplasts. *J. Exp. Bot.* 67, 5657–5669.
- Tejos, R.I., Mercado, A. V., Meisel, L.A., 2010. Analysis of chlorophyll fluorescence reveals stage specific patterns of chloroplast-containing cells during *Arabidopsis* embryogenesis. *Biol. Res.* 43, 99–111.
- Thomas, F., Massenet, O., Dorne, A.-M., Briat, J.-F., Mache, R., 1988. Expression of the *rpl23*, *rpl2* and *rps19* genes in *spinach* chloroplasts. *Nucleic Acids Res.* 16, 349.
- Tiller, N., Bock, R., 2014. The translational apparatus of plastids and its role in plant development. *Mol. Plant* 7, 1105–1120.
- Triezenberg, S.J., 1992. Primer Extension. *Curr. Protoc. Mol. Biol.* 20, 1–5.
- Tsunoyama, Y., Ishizaki, Y., Morikawa, K., Kobori, M., Nakahira, Y., Takeba, G., Toyoshima, Y., Shiina, T., 2004. Blue light-induced transcription of plastid-encoded *psbD* gene is mediated by a nuclear-encoded transcription initiation factor, *AtSig5*. *Proc. Natl. Acad. Sci. U. S. A.* 101, 3304–3309.
- Vera, A., Sugiura, M., 1992. Combination of *in vitro* capping and ribonuclease protection improves the detection of transcription start sites in chloroplasts. *Plant Mol. Biol.* 19, 309–311.
- Weihe, A., Börner, T., 1999. Transcription and the architecture of promoters in chloroplasts. *Trends Plant Sci.* 4, 169–170.
- Weihe, A., Liere, K., Thomas, B., 2012. Transcription and Transcription Regulation in Chloroplasts and Mitochondria of Higher Plants.
- Weinholdt, C., Wichmann, H., Kotrba, J., Ardell, D.H., Kappler, M., Eckert, A.W., Vordermark, D., Grosse, I., 2019. Prediction of regulatory targets of alternative isoforms of the epidermal growth factor receptor in a glioblastoma cell line. *BMC Bioinformatics* 20, 1–14.
- Westhoff, P., Hermann, R.G., 1988. Complex RNA maturation in chloroplasts: The *psbB* operon from *spinach*. *Eur. J. Biochem.* 171, 551–564.
- Wicke, S., Schneeweiss, G.M., dePamphilis, C.W., Müller, K.F., Quandt, D., 2011. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol. Biol.* 76, 273–297.
- Williams-Carrier, R., Zoschke, R., Belcher, S., Pfalz, J., Barkan, A., 2014. A Major Role for the Plastid-Encoded RNA Polymerase Complex in the Expression of Plastid Transfer RNAs. *Plant Physiol.* 164, 239–248.
- Woodson, J.D., Perez-Ruiz, J.M., Schmitz, R.J., Ecker, J.R., Chory, J., 2015. Sigma factor-mediated plastid retrograde signals control nuclear gene expression 4, 327–333.
- Xie, G., Allison, L.A., 2002. Sequences upstream of the YRTA core region are essential for transcription of the tobacco *atpB* NEP promoter in chloroplasts *in vivo*. *Curr. Genet.* 41, 176–182.
- Yagi, Y., Shiina, T., 2014. Recent advances in the study of chloroplast gene expression and its evolution. *Front. Plant Sci.* 5, 61.
- Yamaguchi, K., Subramanian, A.R., 2000. The plastid ribosomal proteins. Identification of all the proteins in the 50 S subunit of an organelle ribosome (chloroplast). *J. Biol. Chem.* 275, 28466–28482.
- Yamaguchi, K., Von Knoblauch, K., Subramanian, A.R., 2000. The plastid ribosomal proteins. Identification of all the proteins in the 30 S subunit of an organelle ribosome (chloroplast). *J. Biol. Chem.* 275, 28455–28465.

- Yang, X.-F., Wang, Y.-T., Chen, S.-T., Li, J.-K., Shen, H.-T., Guo, F.-Q., 2016. PBR1 selectively controls biogenesis of photosynthetic complexes by modulating translation of the large chloroplast gene *Ycf1* in *Arabidopsis*. *Cell Discov.* 2, 1–19.
- Yin, C., Richter, U., Börner, T., Weihe, A., 2010. Evolution of plant phage-type RNA polymerases: the genome of the basal angiosperm *Nuphar advena* encodes two mitochondrial and one plastid phage-type RNA polymerases. *BMC Evol. Biol.* 10, 379.
- Yuan Meng, B., Tanaka, M., Wakasugi, T., Ohme, M., Shinozaki, K., Sugiura, M., 1988. Cotranscription of the genes encoding two P700 chlorophyll a apoproteins with the gene for ribosomal protein CS14: determination of the transcriptional initiation site by *in vitro* capping. *Curr. Genet.* 14, 395–400.
- Zghidi, W., Merendino, L., Cottet, A., Mache, R., Lerbs-Mache, S., 2007. Nucleus-encoded plastid sigma factor SIG3 transcribes specifically the *psbN* gene in plastids. *Nucleic Acids Res.* 35, 455–464.
- Zhelyazkova, P., Hammani, K., Rojas, M., Voelker, R., Vargas-Suárez, M., Börner, 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, C.M., Forstner, K.U., 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.
- Zimorski, V., Ku, C., Martin, W.F., Gould, S.B., 2014. Endosymbiotic theory for organelle origins. *Curr. Opin. Microbiol.* 22, 38–48.
- Zoschke, R., Liere, K., Börner, T., 2007. From seedling to mature plant: *Arabidopsis* plastidial genome copy number, RNA accumulation and transcription are differentially regulated during leaf development. *Plant J.* 50, 710–722.
- Zoschke, R., Watkins, K.P., Miranda, R.G., Barkan, A., 2016. The PPR-SMR protein PPR53 enhances the stability and translation of specific chloroplast RNAs in maize. *Plant J.* 85, 594–606.
- Zou, Z., Eibl, C., Koop, H.U., 2003. The stem-loop region of the tobacco *psbA* 5'UTR is an important determinant of mRNA stability and translation efficiency. *Mol. Genet. Genomics* 269, 340–349.

VI ANHANG

VI.1 Tabellen

Tabelle VI.1: Liste der Hersteller und Anbieter

| Hersteller                   | Sitz                         |
|------------------------------|------------------------------|
| Biologio B. V.               | Nijmegen, Niederlande        |
| Carl Roth GmbH+Co. KG        | Karlsruhe, Deutschland       |
| HP Labortechnik GmbH         | Oberschleißheim, Deutschland |
| GE Healthcare                | Chicago, IL, USA             |
| Lucigen Corporation          | Middleton, WI, USA           |
| Merck Chemicals GmbH         | Darmstadt, Deutschland       |
| New England Biolabs GmbH     | Ipswich, USA                 |
| Novogene Co. LTD             | Peking, China                |
| PeqLab                       | Erlangen, Deutschland        |
| PerkinElmer, Inc.            | Waltham, MA, USA             |
| Qiagen                       | Venlo, Niederlande           |
| Roche Molecular Biochemicals | Mannheim, Germany            |
| Thermo Fisher Scientific™    | Waltham, MA, USA             |
| Sigma-Aldrich                | St. Louis, MO, USA           |
| Vilber Lourmat Deutschland   | Eberhardzell, Deutschland    |
| VWR International            | Radnor, PA, USA              |
| Carl Zeiss                   | Jena, Deutschland            |

Tabelle VI.2: Liste der verwendeten Oligonukleotide und deren Anwendung

| Anwendung       | Primernamen    | Sequenz 5` nach 3`             | Mutante         | Organismus/<br>T-DNA |
|-----------------|----------------|--------------------------------|-----------------|----------------------|
| Genotypisierung | 067191-RP-2    | GGTATGATTCTTGGAACTTTGGA        | <i>rpoTp-2</i>  | <i>A. thaliana</i>   |
| Genotypisierung | 067191-LP-2    | TGTGCAACCCTATTGCAGAA           |                 | <i>A. thaliana</i>   |
| Genotypisierung | 323-086115-RP2 | GTTGATAAAGAAACAGAAGTTGTGCAGCAG | <i>rpoTmp-2</i> | <i>A. thaliana</i>   |
| Genotypisierung | 323-132842-RP  | AGCGACATTCACATTTCCAAC          |                 | <i>A. thaliana</i>   |
| Genotypisierung | 123752-RP      | CTCTTAACGCTTGACAAACC           | <i>clb19-3</i>  | <i>A. thaliana</i>   |
| Genotypisierung | 123752-LP      | GAGGAAGCGACAATGGTTATG          |                 | <i>A. thaliana</i>   |
| Genotypisierung | 075736-LP      | GGCTTGTTGATGAGTGCAGA           | <i>ptac2</i>    | <i>A. thaliana</i>   |
| Genotypisierung | 075736-RP      | TCCTGCAAGAAGGAGAAAGC           |                 | <i>A. thaliana</i>   |
| Genotypisierung | 093884-LP      | ATCAGCCCAGAGCCTTTCTAC          | <i>rpoTp-3</i>  | <i>A. thaliana</i>   |
| Genotypisierung | 093884-RP      | CGAGTTTTCGACCTGCATTC           |                 | <i>A. thaliana</i>   |
| Genotypisierung | 323-286E07-RP  | AGCAGCTCTTTCCATTCTTC           | <i>rpoTmp-1</i> | <i>A. thaliana</i>   |
| Genotypisierung | 286E07-LP-2    | TAAACAAGCAATTTTCGAGGTCAG       |                 | <i>A. thaliana</i>   |
| Genotypisierung | LB_6313R       | TCAAACAGGATTTTCGCCTGCT         |                 | Salk T-DNA           |
| Genotypisierung | GABI 8409 LB   | ATATTGACCATCATACTCATTGC        |                 | Gabi T-DNA           |
| Anwendung       | Primernamen    | Sequenz 5` nach 3`             | Gen             | Organismus           |
| Northern Blot   | rpL32-N-F3     | CTCTCATACGGATTTCAAACG          | <i>rpl32</i>    | <i>A. thaliana</i>   |
| Northern Blot   | rpL32-N-R1     | CGATGTCCAATATCCCTTCC           |                 | <i>A. thaliana</i>   |
| Northern Blot   | atpB-N-F       | TGCTAAAGCTCATGGTGGTG           | <i>atpB</i>     | <i>A. thaliana</i>   |
| Northern Blot   | atpB-N-R       | CGGTACTAAGGGTGGGTTGA           |                 | <i>A. thaliana</i>   |
| Northern Blot   | clpP-N-F1      | AGGCCAAGAGGTTGATACCG           | <i>clpP</i>     | <i>A. thaliana</i>   |
| Northern Blot   | clpP-N-R1      | TGGTAATTGCTCCTCCGACT           |                 | <i>A. thaliana</i>   |
| Northern Blot   | psbD-N-F       | CTGGTCCCATTGCTGTTTTT           | <i>psbD</i>     | <i>A. thaliana</i>   |
| Northern Blot   | psbD-N-R       | GTTTCTTCGGCTTGAGTTGG           |                 | <i>A. thaliana</i>   |
| Northern Blot   | rpl23-N-F      | CGGTTATTGGGGAAAAATCA           | <i>rpl23</i>    | <i>A. thaliana</i>   |
| Northern Blot   | rpl23-N-R      | AATAACCCGGTTGAAGCGTA           |                 | <i>A. thaliana</i>   |
| Northern Blot   | matK-N-F       | TTTGGGGCATACCAGTCTT            | <i>matK</i>     | <i>A. thaliana</i>   |

## Anhang

|                                |                             |   |  |                    |
|--------------------------------|-----------------------------|---|--|--------------------|
| Northern Blot                  | matK-N-t7-R                 | TAATACGACTCACTATAGGGATTCATCAG<br>AAGCGGCGTA |  | <i>A. thaliana</i> |
| cDNA-Bibliothek<br>Herstellung | Modifizierter<br>5' Adapter | GUUCAGAGUUCUACAGUCCGACGAUCGAT<br>CNNNNNNNN  |  |                    |

Tabelle VI.3: Abkürzungsverzeichnis

| Abkürzung                | Vollform  |
|--------------------------|---|
| 5' PPase                 | 5' Polyphosphatase  |
| Abb.                     | Abbildung   |
| APS                      | Ammoniumperoxodisulfat  |
| A, T, G, C               | Bausteine der DNA: Adenin, Thymin, Guanin, Cytosin  |
| b                        | Base(n)   |
| bp                       | Basenpaar   |
| bzw.                     | beziehungsweise   |
| ca.                      | circa   |
| cDNA                     | komplementäre DNA   |
| Col-0                    | Columbia-0  |
| CSPD                     | Dinatrium 3-(4-methoxy Spiro {1,2-dioxetan-3,2'-(5'-chloro)tricyclo [3.3.1.1 <sup>3,7</sup> ]decan}-4-yl)phenylphosphat |
| <i>Cytb<sub>6f</sub></i> | Cytochrom-b <sub>6f</sub> -Komplex  |
| ddH <sub>2</sub> O       | doppelt destilliertes Wasser  |
| DIC                      | <i>differential interference contrast</i> (Differentialinterferenzkontrast)   |
| DNA                      | Desoxyribonukleinsäure  |
| DNase                    | Desoxyribonuklease  |
| dNTP                     | Desoxyribonukleosidtriphosphat(e) dATP, dCTP, dGTP, dTTP  |
| dRNA-Seq                 | <i>differential RNA-Seq</i> (Differenzielle RNA-Sequenzierung)  |
| EDTA                     | Ethylendiamintetraessigsäure  |
| <i>et al.</i>            | <i>et altera</i> (und andere)   |
| EtBr                     | Ethidiumbromid  |
| EtOH                     | Ethanol   |
| <i>g</i>                 | Erdbeschleunigung   |
| µg                       | Mikrogramm  |
| g                        | Gramm   |
| h                        | Stunde  |
| IGV                      | Integrated Genome Viewer  |
| KaOAc                    | Kaliumacetat  |
| kb                       | Kilobase (= 1000 Basenpaare)  |
| l                        | Liter   |
| mA                       | Milliamper  |
| <i>MEME</i>              | <i>Multiple Expectation-Maximization for Motif Elicitation</i>  |
| min                      | Minute  |
| µL                       | Mikroliter  |
| mL                       | Milliliter  |
| mm                       | Millimeter  |
| MES                      | 2-(N-Morpholino)ethansulfonsäure  |
| MOPS                     | 3-(N-Morpholino)propansulfonsäure   |
| MS                       | Murashige&Skoog Medium  |
| NaOAc                    | Natriumacetat   |
| ng                       | Nanogramm   |
| nt                       | Nukleotid   |
| NTP                      | Ribonukleosidtriphosphat(e) ATP, CTP, GTP, TTP  |
| o.g.                     | oben genannt  |
| ORF                      | <i>open reading frame</i> (offenes Leseraster)  |
| PAGE                     | Polyacrylamidgelelektrophorese  |
| PCR                      | <i>polymerase chain reaction</i> (Polymerasekettenreaktion)   |

## Anhang

|                |  |
|----------------|--|
| Primer         | Oligonukleotid   |
| PS I           | Photosystem I  |
| PS II          | Photosystem II   |
| PVP            | Polyvinylpyrrolidon  |
| RNA            | Ribonukleinsäure   |
| RNase          | Ribonuklease   |
| rRNA           | ribosomale RNA   |
| rpm            | <i>rounds per minute</i> (Umdrehungen pro Minute)                            |
| RT             | Raumtemperatur   |
| SDS            | Natriumdodecylsulfat   |
| s              | Sekunde  |
| Tab.           | Tabelle  |
| TAE            | Tris-Acetat-EDTA-Puffer  |
| TAP            | <i>Tobacco Acid Pyrophosphatase</i>  |
| Taq Polymerase | <i>Thermus aquaticus</i> DNA-Polymerase                                      |
| TBE            | Tris-Borat-EDTA-Puffer   |
| TEMED          | NNN'N' Tetramethyl-ethylendiamin   |
| TEX            | Terminator <sup>TM</sup> 5'-Phosphat-abhängige Exonuclease                   |
| Tm             | Schmelztemperatur  |
| Tris           | Tris(hydroxymethyl)-aminomethan  |
| tRNA           | Transfer-RNA   |
| TSS            | Transkriptionsstartseite   |
| U              | <i>Unit</i> , Einheit(en)  |
| ü.N.           | über Nacht   |
| UMI            | <i>unique molecular identifier</i> (einzigartiger molekularer Identifikator) |
| UV             | Ultraviolett   |
| V              | Volt   |
| WT             | Wildtyp  |

Tabelle VI.4: Liste der Genbezeichnungen

| Gen                 | Funktion   |
|---------------------|--|
| <i>accD</i>         | Gen für die $\beta$ -Untereinheit der Acetyl-CoA-Carboxylase                     |
| <i>atpA</i>         | Gen für die CF <sub>1</sub> ATPase-Komplex $\alpha$ -Untereinheit                |
| <i>atpB</i>         | Gen für die CF <sub>1</sub> ATPase-Komplex $\beta$ -Untereinheit                 |
| <i>atpF</i>         | Gen für die CF <sub>1</sub> ATPase-Komplex F-Untereinheit                        |
| <i>atpI</i>         | Gen für die CF <sub>0</sub> ATPase-Komplex $\alpha$ -Untereinheit                |
| <i>ccsA (ycf5)</i>  | Gen für das Protein für den Cytochrom C Zusammenbau                              |
| <i>cemA (ycf10)</i> | Gen für das CemA-ähnliches Protonextrusionsprotein                               |
| <i>clpP</i>         | Gen für die proteolytische Untereinheit der ATP-abhängigen Clp-Protease          |
| <i>matK</i>         | Gen für die Maturase K   |
| <i>ndhC</i>         | Gen für die Untereinheit III des NADH-Dehydrogenase                              |
| <i>ndhK (psbG)</i>  | Gen für Untereinheit von NADH-Dehydrogenase                                      |
| <i>petB</i>         | Gen für die $\beta$ -Untereinheit des Cytochrom <i>b<sub>6</sub>f</i> -Komplexes |
| <i>petD</i>         | Gen für die Untereinheit IV des Cytochrom <i>b<sub>6</sub>f</i> -Komplexes       |
| <i>petN (ycf6)</i>  | Gen für die Untereinheit VIII des Cytochrom <i>b<sub>6</sub>f</i> -Komplexes     |
| <i>psaA</i>         | Gen für das P700 Chlorophyll a A2 Apoprotein des PS I                            |
| <i>psaB</i>         | Gen für das P700 Chlorophyll a A1 Apoprotein des PS I                            |
| <i>psaJ</i>         | Gen für die J-Untereinheit des PS I  |
| <i>psbA</i>         | Gen für das D1-Protein des PS II   |
| <i>psbB</i>         | Gen für das P680 Chlorophyll A Apoprotein des PS II                              |
| <i>psbD</i>         | Gen für das D2-Protein des PS II   |
| <i>psbF</i>         | Gen für das Protein VI des PS II   |
| <i>psbI</i>         | Gen für das Protein I des PS II  |
| <i>psbT</i>         | Gen für das Protein T des PS II  |
| <i>rbcL</i>         | Gen für die große Untereinheit der RuBisCO                                       |
| <i>rpl23</i>        | Gen für das ribosomale Protein L23   |

## Anhang

|                      |   |
|----------------------|---|
| <i>rpl33</i>         | Gen für das ribosomale Protein L33  |
| <i>rpoA</i>          | Gen für die $\alpha$ -Untereinheit der RNA-Polymerase   |
| <i>rpoB</i>          | Gen für die $\beta$ -Untereinheit der RNA-Polymerase  |
| <i>rpoC1</i>         | Gen für die $\beta'$ -Untereinheit der RNA-Polymerase   |
| <i>rpoC2</i>         | Gen für die $\beta''$ -Untereinheit der RNA-Polymerase  |
| <i>rps2</i>          | Gen für das ribosomale Protein S2   |
| <i>rps4</i>          | Gen für das ribosomale Protein S4   |
| <i>rps8</i>          | Gen für das ribosomale Protein S8   |
| <i>rps11</i>         | Gen für das ribosomale Protein S11  |
| <i>rps15</i>         | Gen für das ribosomale Protein S15  |
| <i>rps16</i>         | Gen für das ribosomale Protein S16  |
| <i>rps18</i>         | Gen für das ribosomale Protein S18  |
| <i>rrn16</i>         | Gen für die ribosomale RNA 16S  |
| <i>rrn23</i>         | Gen für die ribosomale RNA 23S  |
| <i>rrn4,5</i>        | Gen für die ribosomale RNA 4,5S   |
| <i>rrn5</i>          | Gen für die ribosomale RNA 5S   |
| <i>tic214 (ycf1)</i> | Gen für das Protein des Transloconkomplexes der inneren Chloroplastenmembran  |
| <i>tmA</i>           | Gen für die tRNA <sup>Ala</sup>   |
| <i>tmC</i>           | Gen für die tRNA <sup>Cys</sup>   |
| <i>tmD</i>           | Gen für die tRNA <sup>Asp</sup>   |
| <i>tmE</i>           | Gen für die tRNA <sup>Glu</sup>   |
| <i>tmF</i>           | Gen für die tRNA <sup>Phe</sup>   |
| <i>tmfM</i>          | Gen für die tRNA <sup>fMet</sup>  |
| <i>tmG</i>           | Gen für die tRNA <sup>Gly</sup>   |
| <i>tmH</i>           | Gen für die tRNA <sup>His</sup>   |
| <i>tmI</i>           | Gen für die tRNA <sup>Ile</sup>   |
| <i>tmK</i>           | Gen für die tRNA <sup>Lys</sup>   |
| <i>tmL</i>           | Gen für die tRNA <sup>Leu</sup>   |
| <i>tmM</i>           | Gen für die tRNA <sup>Met</sup>   |
| <i>tmN</i>           | Gen für die tRNA <sup>Asn</sup>   |
| <i>tmP</i>           | Gen für die tRNA <sup>Pro</sup>   |
| <i>tmQ</i>           | Gen für die tRNA <sup>Gln</sup>   |
| <i>tmR</i>           | Gen für die tRNA <sup>Arg</sup>   |
| <i>tmS</i>           | Gen für die tRNA <sup>Ser</sup>   |
| <i>tmT</i>           | Gen für die tRNA <sup>Thr</sup>   |
| <i>tmV</i>           | Gen für die tRNA <sup>Val</sup>   |
| <i>tmW</i>           | Gen für die tRNA <sup>Trp</sup>   |
| <i>tmY</i>           | Gen für die tRNA <sup>Tyr</sup>   |
| <i>ycf2</i>          | Gen für Protein, welches Teil des AAA-ATPase Komplexes ist  |
| <i>ycf3</i>          | Gen für Protein, welches in Assemblierung und Stabilität des PSI involviert ist   |
| <i>ycf4</i>          | Gen für Protein, welches in Assemblierung und Stabilität des PSI involviert ist   |
| <i>ycf15 (ORF77)</i> | plastidäres offenes Leseraster 15 ( <i>hypothetical chloroplast open reading frame 15</i> ), putativer Proteinvorläufer |

Tabelle VI.5: Statistische Daten zu RNA-Seq für die Analyse von Wildtyp und Mutanten. Rep.: Replikat. Die Analyse wurde von Jose Muino durchgeführt.

| Probe                  | Reads sequenziert | Reads nach Trimmen | Reads gemappt zum Gesamtgenom | Reads gemappt zum Plastidengenom |
|------------------------|-------------------|--------------------|-------------------------------|----------------------------------|
| <i>clb19</i> früh Rep1 | 44744378          | 44339754           | 45683680                      | 3207896                          |
| <i>clb19</i> früh Rep2 | 41148740          | 40853070           | 40779618                      | 2490815                          |
| <i>clb19</i> früh Rep3 | 45983332          | 45599944           | 46320566                      | 2598404                          |
| <i>clb19</i> spät Rep1 | 54537776          | 54141592           | 56879912                      | 18688397                         |
| <i>clb19</i> spät Rep2 | 45308944          | 44918184           | 50121058                      | 12791591                         |
| <i>clb19</i> spät Rep3 | 51091390          | 50693628           | 52477766                      | 15486963                         |

## Anhang

|                         |          |          |          |          |
|-------------------------|----------|----------|----------|----------|
| <i>ptac2</i> spät Rep1  | 50468964 | 49941678 | 51225360 | 14096709 |
| <i>ptac2</i> spät Rep2  | 53232474 | 52653176 | 56267036 | 12686166 |
| <i>ptac2</i> spät Rep3  | 44685484 | 44118788 | 44575050 | 8363745  |
| <i>rpoTmp</i> früh Rep1 | 44962724 | 44658196 | 45369960 | 8303056  |
| <i>rpoTmp</i> früh Rep2 | 53620852 | 53068932 | 46897812 | 9511731  |
| <i>rpoTmp</i> früh Rep3 | 41719076 | 41458404 | 42852452 | 7451915  |
| <i>rpoTmp</i> spät Rep1 | 43547398 | 43276146 | 45213430 | 24358959 |
| <i>rpoTmp</i> spät Rep2 | 45134008 | 44819596 | 46538530 | 22959028 |
| <i>rpoTmp</i> spät Rep3 | 48881446 | 48366072 | 49771106 | 24442347 |
| <i>rpoTp</i> früh Rep1  | 48867006 | 48401102 | 49717536 | 1798958  |
| <i>rpoTp</i> früh Rep2  | 44887412 | 44469640 | 45717214 | 1604015  |
| <i>rpoTp</i> früh Rep3  | 42354560 | 41903068 | 42901614 | 1589076  |
| <i>rpoTp</i> spät Rep1  | 41304790 | 40972894 | 42800386 | 16279630 |
| <i>rpoTp</i> spät Rep2  | 50949808 | 50636692 | 50161844 | 13090981 |
| <i>rpoTp</i> spät Rep3  | 47648344 | 47160566 | 49074086 | 10977786 |
| WT früh Rep1            | 47126538 | 46623038 | 48824912 | 6012072  |
| WT früh Rep2            | 44042726 | 43577698 | 45904556 | 6648056  |
| WT früh Rep3            | 48773674 | 48265218 | 50755000 | 6247310  |
| WT spät Rep1            | 46589076 | 46180186 | 50432236 | 26173927 |
| WT spät Rep2            | 42208874 | 41746902 | 41812246 | 24362360 |
| WT spät Rep3            | 44336966 | 43899526 | 44076418 | 25078878 |

Tabelle VI.6: Differentielle Expressionsanalyse in Mutanten im frühen Stadium als logarithmierte relative Transkriptlevel bezogen auf WT mit Angabe des p-Werts und des Standardfehlers. Die Analyse wurde von Jose Muino durchgeführt.

| Gene          | log2FC(Mut/WT)       |                       |                      | p-Wert               |                       |                      | Standardfehler       |                       |                      |
|---------------|----------------------|-----------------------|----------------------|----------------------|-----------------------|----------------------|----------------------|-----------------------|----------------------|
|               | <i>rpoTp</i><br>früh | <i>rpoTmp</i><br>früh | <i>clb19</i><br>früh | <i>rpoTp</i><br>früh | <i>rpoTmp</i><br>früh | <i>clb19</i><br>früh | <i>rpoTp</i><br>früh | <i>rpoTmp</i><br>früh | <i>clb19</i><br>früh |
| <i>trnH</i>   | -0,04                | 0,45                  | -1,11                | 0,89                 | 0,06                  | 0,00                 | 0,31                 | 0,24                  | 0,32                 |
| <i>psbA</i>   | -0,32                | 0,34                  | -2,68                | 0,04                 | 0,08                  | 0,00                 | 0,15                 | 0,19                  | 0,14                 |
| <i>trnK</i>   | -0,24                | 0,38                  | -0,85                | 0,05                 | 0,01                  | 0,00                 | 0,12                 | 0,14                  | 0,11                 |
| <i>rps16</i>  | -0,70                | -0,40                 | -0,22                | 0,00                 | 0,04                  | 0,13                 | 0,17                 | 0,20                  | 0,15                 |
| <i>trnQ</i>   | -0,97                | 0,38                  | 0,16                 | 0,00                 | 0,12                  | 0,47                 | 0,33                 | 0,25                  | 0,22                 |
| <i>psbK</i>   | 1,23                 | 0,09                  | -0,43                | 0,00                 | 0,49                  | 0,00                 | 0,16                 | 0,13                  | 0,13                 |
| <i>psbI</i>   | 1,06                 | 0,13                  | -0,33                | 0,00                 | 0,31                  | 0,02                 | 0,15                 | 0,13                  | 0,14                 |
| <i>trnS.1</i> | 1,22                 | 0,30                  | -0,38                | 0,00                 | 0,04                  | 0,01                 | 0,15                 | 0,14                  | 0,14                 |
| <i>trnG.1</i> | -0,47                | 0,25                  | -1,47                | 0,00                 | 0,27                  | 0,00                 | 0,15                 | 0,22                  | 0,14                 |
| <i>trnR.1</i> | -0,99                | 0,08                  | -0,48                | 0,01                 | 0,75                  | 0,09                 | 0,37                 | 0,27                  | 0,28                 |
| <i>atpA</i>   | 0,64                 | -0,17                 | 0,77                 | 0,00                 | 0,39                  | 0,00                 | 0,14                 | 0,20                  | 0,15                 |
| <i>atpF</i>   | 0,13                 | -0,09                 | 0,25                 | 0,30                 | 0,43                  | 0,03                 | 0,13                 | 0,11                  | 0,12                 |
| <i>atpH</i>   | -0,36                | -0,29                 | 0,15                 | 0,01                 | 0,03                  | 0,16                 | 0,13                 | 0,13                  | 0,11                 |
| <i>atpI</i>   | 0,38                 | -0,33                 | 0,83                 | 0,01                 | 0,03                  | 0,00                 | 0,14                 | 0,15                  | 0,15                 |
| <i>rps2</i>   | 0,18                 | -0,32                 | 1,35                 | 0,26                 | 0,10                  | 0,00                 | 0,16                 | 0,20                  | 0,17                 |
| <i>rpoC2</i>  | -0,34                | -0,27                 | 1,35                 | 0,02                 | 0,11                  | 0,00                 | 0,15                 | 0,17                  | 0,17                 |
| <i>rpoC1</i>  | -0,33                | -0,24                 | 1,19                 | 0,01                 | 0,02                  | 0,00                 | 0,12                 | 0,10                  | 0,12                 |
| <i>rpoB</i>   | -0,87                | -0,40                 | 1,37                 | 0,00                 | 0,01                  | 0,00                 | 0,14                 | 0,16                  | 0,15                 |
| <i>trnC</i>   | -0,95                | 0,20                  | -2,37                | 0,00                 | 0,32                  | 0,00                 | 0,16                 | 0,21                  | 0,15                 |
| <i>petN</i>   | 0,50                 | 0,34                  | -2,49                | 0,00                 | 0,12                  | 0,00                 | 0,17                 | 0,22                  | 0,15                 |
| <i>psbM</i>   | -0,16                | 0,30                  | -0,91                | 0,34                 | 0,18                  | 0,00                 | 0,17                 | 0,22                  | 0,17                 |
| <i>trnD</i>   | -0,61                | 0,06                  | -1,53                | 0,00                 | 0,76                  | 0,00                 | 0,20                 | 0,19                  | 0,18                 |
| <i>trnY</i>   | -1,03                | 0,08                  | -2,14                | 0,00                 | 0,67                  | 0,00                 | 0,21                 | 0,20                  | 0,14                 |



## Anhang

|        |       |       |       |      |      |      |      |      |      |
|--------|-------|-------|-------|------|------|------|------|------|------|
| trnE   | -1,41 | 0,24  | -0,07 | 0,00 | 0,23 | 0,78 | 0,33 | 0,19 | 0,24 |
| trnT.1 | -0,10 | 0,29  | -1,56 | 0,82 | 0,42 | 0,00 | 0,47 | 0,36 | 0,50 |
| psbD   | 0,53  | 0,06  | -0,64 | 0,00 | 0,76 | 0,00 | 0,15 | 0,19 | 0,16 |
| psbC   | 0,60  | 0,24  | -1,23 | 0,00 | 0,11 | 0,00 | 0,13 | 0,15 | 0,14 |
| trnS.2 | 0,28  | -0,05 | -1,14 | 0,07 | 0,70 | 0,00 | 0,15 | 0,14 | 0,15 |
| psbZ   | 0,26  | -0,13 | -1,25 | 0,07 | 0,42 | 0,00 | 0,14 | 0,16 | 0,14 |
| trnG.2 | -0,75 | 1,11  | -1,41 | 0,02 | 0,00 | 0,00 | 0,32 | 0,21 | 0,29 |
| trnFM  | -0,10 | 0,53  | -2,10 | 0,68 | 0,01 | 0,00 | 0,25 | 0,20 | 0,27 |
| rps14  | 0,73  | -0,12 | -0,17 | 0,00 | 0,36 | 0,14 | 0,14 | 0,13 | 0,12 |
| psaB   | 0,12  | -0,46 | 0,05  | 0,43 | 0,03 | 0,74 | 0,15 | 0,21 | 0,16 |
| psaA   | -0,15 | -0,51 | -0,07 | 0,33 | 0,01 | 0,66 | 0,15 | 0,19 | 0,16 |
| ycf3   | 0,98  | -0,11 | 0,82  | 0,00 | 0,35 | 0,00 | 0,13 | 0,11 | 0,12 |
| trnS.3 | -0,61 | 0,00  | 0,44  | 0,01 | 0,98 | 0,01 | 0,22 | 0,16 | 0,17 |
| rps4   | -0,29 | -0,18 | 0,67  | 0,03 | 0,23 | 0,00 | 0,14 | 0,15 | 0,14 |
| trnT.2 | -1,04 | 0,41  | -1,10 | 0,00 | 0,10 | 0,00 | 0,32 | 0,25 | 0,25 |
| trnL.1 | -0,62 | 1,02  | -2,34 | 0,00 | 0,00 | 0,00 | 0,14 | 0,22 | 0,14 |
| trnF   | -0,07 | 0,41  | -3,06 | 0,58 | 0,05 | 0,00 | 0,13 | 0,21 | 0,12 |
| ndhJ   | 0,75  | -0,11 | 0,28  | 0,00 | 0,46 | 0,02 | 0,16 | 0,15 | 0,13 |
| ndhK   | 0,87  | -0,27 | -0,35 | 0,00 | 0,02 | 0,00 | 0,14 | 0,12 | 0,12 |
| ndhC   | 0,65  | -0,19 | -0,93 | 0,00 | 0,15 | 0,00 | 0,14 | 0,13 | 0,12 |
| trnV.1 | -0,87 | 0,66  | -2,70 | 0,00 | 0,01 | 0,00 | 0,14 | 0,24 | 0,12 |
| trnM   | 0,15  | 0,06  | 0,69  | 0,80 | 0,89 | 0,12 | 0,58 | 0,43 | 0,45 |
| atpE   | -0,24 | -0,29 | 0,34  | 0,05 | 0,01 | 0,00 | 0,12 | 0,11 | 0,11 |
| atpB   | -0,15 | -0,37 | 0,62  | 0,22 | 0,00 | 0,00 | 0,12 | 0,11 | 0,12 |
| rbcL   | -0,51 | 0,70  | -2,99 | 0,00 | 0,00 | 0,00 | 0,14 | 0,17 | 0,15 |
| accD   | 1,38  | -0,44 | 0,19  | 0,00 | 0,00 | 0,10 | 0,12 | 0,12 | 0,11 |
| psaI   | 0,35  | 0,09  | 0,93  | 0,02 | 0,65 | 0,00 | 0,15 | 0,20 | 0,12 |
| ycf4   | 0,63  | -0,17 | 0,76  | 0,00 | 0,27 | 0,00 | 0,12 | 0,16 | 0,13 |
| ycf10  | 0,77  | -0,13 | 0,67  | 0,00 | 0,29 | 0,00 | 0,12 | 0,12 | 0,13 |
| petA   | 0,70  | -0,12 | 0,58  | 0,00 | 0,22 | 0,00 | 0,12 | 0,10 | 0,11 |
| psbJ   | -0,41 | 0,06  | -0,78 | 0,00 | 0,66 | 0,00 | 0,14 | 0,14 | 0,15 |
| psbL   | -0,29 | -0,13 | -1,46 | 0,04 | 0,35 | 0,00 | 0,14 | 0,14 | 0,13 |
| psbF   | -0,19 | -0,21 | -1,57 | 0,20 | 0,14 | 0,00 | 0,15 | 0,14 | 0,14 |
| psbE   | -0,28 | -0,26 | -1,36 | 0,03 | 0,04 | 0,00 | 0,13 | 0,13 | 0,12 |
| petL   | 0,93  | 0,29  | -0,11 | 0,00 | 0,10 | 0,52 | 0,15 | 0,17 | 0,17 |
| petG   | 1,14  | 0,26  | -0,32 | 0,00 | 0,06 | 0,02 | 0,14 | 0,14 | 0,13 |
| trnW   | -0,41 | 0,44  | -1,33 | 0,12 | 0,08 | 0,00 | 0,26 | 0,25 | 0,25 |
| trnP   | -0,87 | 0,29  | -1,04 | 0,00 | 0,22 | 0,00 | 0,23 | 0,23 | 0,20 |
| psaJ   | 0,68  | 0,56  | -0,78 | 0,00 | 0,00 | 0,00 | 0,14 | 0,15 | 0,13 |
| rpl33  | 0,69  | 0,42  | -0,62 | 0,00 | 0,01 | 0,00 | 0,13 | 0,16 | 0,13 |
| rpS18  | -0,39 | 0,43  | -0,43 | 0,00 | 0,02 | 0,00 | 0,13 | 0,18 | 0,12 |
| rpl20  | 0,67  | -0,23 | 1,21  | 0,00 | 0,11 | 0,00 | 0,13 | 0,15 | 0,14 |
| rps12  | 0,37  | -0,25 | 0,97  | 0,01 | 0,19 | 0,00 | 0,14 | 0,19 | 0,13 |
| clpP   | 0,15  | -0,23 | 0,57  | 0,32 | 0,14 | 0,00 | 0,15 | 0,16 | 0,14 |
| psbB   | 0,23  | 0,10  | -1,81 | 0,15 | 0,60 | 0,00 | 0,16 | 0,20 | 0,17 |
| psbT   | 0,47  | 0,07  | -1,52 | 0,00 | 0,78 | 0,00 | 0,17 | 0,23 | 0,19 |
| psbN   | 0,13  | -0,47 | 1,25  | 0,49 | 0,00 | 0,00 | 0,18 | 0,15 | 0,14 |
| psbH   | 0,44  | 0,22  | -1,48 | 0,01 | 0,29 | 0,00 | 0,18 | 0,21 | 0,18 |
| petB   | 0,52  | 0,31  | -1,32 | 0,00 | 0,05 | 0,00 | 0,14 | 0,16 | 0,15 |
| petD   | 0,41  | 0,32  | -1,19 | 0,00 | 0,07 | 0,00 | 0,14 | 0,18 | 0,14 |
| rpoA   | 0,17  | -0,16 | 1,21  | 0,19 | 0,36 | 0,00 | 0,13 | 0,18 | 0,14 |
| rps11  | -0,03 | -0,07 | 1,16  | 0,86 | 0,73 | 0,00 | 0,15 | 0,21 | 0,16 |
| rpl36  | -0,07 | -0,10 | 1,13  | 0,63 | 0,58 | 0,00 | 0,15 | 0,18 | 0,15 |
| rps8   | -0,15 | -0,02 | 1,09  | 0,33 | 0,91 | 0,00 | 0,16 | 0,18 | 0,15 |
| rpl14  | -0,30 | 0,00  | 0,85  | 0,03 | 0,97 | 0,00 | 0,14 | 0,12 | 0,13 |
| rpl16  | -0,24 | 0,05  | 0,77  | 0,12 | 0,72 | 0,00 | 0,15 | 0,14 | 0,14 |
| rps3   | 0,37  | 0,02  | 0,67  | 0,01 | 0,89 | 0,00 | 0,14 | 0,17 | 0,15 |
| rpl22  | 0,42  | -0,02 | 0,61  | 0,01 | 0,91 | 0,00 | 0,15 | 0,18 | 0,15 |

## Anhang

|          |       |       |       |      |      |      |      |      |      |
|----------|-------|-------|-------|------|------|------|------|------|------|
| rps19    | 0,72  | -0,13 | 0,65  | 0,00 | 0,51 | 0,00 | 0,15 | 0,20 | 0,16 |
| rpl2.1   | 0,46  | -0,28 | 0,60  | 0,00 | 0,04 | 0,00 | 0,12 | 0,13 | 0,12 |
| rpl23.1  | -0,85 | -0,19 | 0,34  | 0,00 | 0,11 | 0,01 | 0,14 | 0,12 | 0,13 |
| trnI.1   | -1,95 | 0,21  | 0,26  | 0,00 | 0,31 | 0,22 | 0,36 | 0,21 | 0,21 |
| ycf2.1   | 0,59  | -0,48 | 0,79  | 0,00 | 0,00 | 0,00 | 0,13 | 0,12 | 0,12 |
| ORF 77.1 | 0,77  | -0,54 | 0,89  | 0,00 | 0,00 | 0,00 | 0,14 | 0,14 | 0,12 |
| trnL.2   | -1,32 | 0,20  | -0,38 | 0,00 | 0,32 | 0,07 | 0,32 | 0,20 | 0,21 |
| ndhB.1   | -0,64 | -0,18 | 0,96  | 0,00 | 0,12 | 0,00 | 0,13 | 0,12 | 0,13 |
| rps7.1   | -0,72 | -0,12 | 1,19  | 0,00 | 0,41 | 0,00 | 0,12 | 0,15 | 0,12 |
| rps12C   | -1,01 | -0,40 | 1,18  | 0,00 | 0,00 | 0,00 | 0,13 | 0,13 | 0,12 |
| trnV.2   | -0,46 | 0,34  | -0,40 | 0,05 | 0,08 | 0,08 | 0,24 | 0,19 | 0,23 |
| trnI.2   | -0,59 | 0,08  | -2,04 | 0,00 | 0,71 | 0,00 | 0,14 | 0,22 | 0,15 |
| trnA.1   | -0,83 | -0,87 | -2,66 | 0,00 | 0,00 | 0,00 | 0,13 | 0,29 | 0,13 |
| trnR.2   | -0,76 | -0,94 | -2,59 | 0,00 | 0,00 | 0,00 | 0,16 | 0,21 | 0,15 |
| trnN.1   | -0,18 | 0,60  | -2,43 | 0,22 | 0,00 | 0,00 | 0,15 | 0,14 | 0,15 |
| ycf1.1   | -0,87 | -0,23 | 1,07  | 0,00 | 0,08 | 0,00 | 0,13 | 0,13 | 0,12 |
| ndhF     | 0,19  | -0,05 | 0,06  | 0,14 | 0,74 | 0,60 | 0,13 | 0,14 | 0,12 |
| rpl32    | -0,08 | 0,17  | 0,84  | 0,61 | 0,50 | 0,00 | 0,16 | 0,26 | 0,16 |
| trnL.3   | -0,85 | 0,33  | 0,54  | 0,01 | 0,30 | 0,02 | 0,32 | 0,32 | 0,23 |
| ycf5     | -0,44 | -0,29 | -0,06 | 0,00 | 0,05 | 0,68 | 0,15 | 0,15 | 0,14 |
| ndhD     | 0,58  | -0,09 | 1,00  | 0,00 | 0,47 | 0,00 | 0,12 | 0,12 | 0,13 |
| psaC     | -0,06 | 0,11  | 0,75  | 0,68 | 0,46 | 0,00 | 0,14 | 0,15 | 0,13 |
| ndhE     | 0,60  | 0,24  | 0,27  | 0,00 | 0,03 | 0,03 | 0,12 | 0,11 | 0,12 |
| ndhG     | 0,73  | 0,33  | 0,62  | 0,00 | 0,00 | 0,00 | 0,12 | 0,11 | 0,12 |
| ndhI     | 1,04  | 0,36  | 1,00  | 0,00 | 0,02 | 0,00 | 0,15 | 0,15 | 0,15 |
| ndhA     | 0,64  | 0,20  | 0,77  | 0,00 | 0,17 | 0,00 | 0,12 | 0,14 | 0,13 |
| ndhH     | 0,73  | -0,14 | 1,30  | 0,00 | 0,42 | 0,00 | 0,15 | 0,17 | 0,16 |
| rps15    | 0,45  | 0,27  | 1,11  | 0,00 | 0,05 | 0,00 | 0,15 | 0,14 | 0,15 |
| ycf1.2   | -0,18 | -0,30 | 1,36  | 0,26 | 0,04 | 0,00 | 0,16 | 0,15 | 0,16 |
| trnN.2   | -0,18 | 0,60  | -2,43 | 0,23 | 0,00 | 0,00 | 0,15 | 0,14 | 0,15 |
| trnR.3   | -0,76 | -0,94 | -2,60 | 0,00 | 0,00 | 0,00 | 0,16 | 0,21 | 0,15 |
| trnA.2   | -0,83 | -0,87 | -2,66 | 0,00 | 0,00 | 0,00 | 0,13 | 0,29 | 0,13 |
| trnI.3   | -0,59 | 0,08  | -2,04 | 0,00 | 0,71 | 0,00 | 0,14 | 0,22 | 0,15 |
| trnV.3   | -0,45 | 0,34  | -0,41 | 0,06 | 0,08 | 0,07 | 0,23 | 0,19 | 0,23 |
| rps12B   | -1,01 | -0,40 | 1,18  | 0,00 | 0,00 | 0,00 | 0,13 | 0,13 | 0,12 |
| rps7.2   | -0,72 | -0,12 | 1,19  | 0,00 | 0,40 | 0,00 | 0,12 | 0,15 | 0,12 |
| ndhB.2   | -0,64 | -0,18 | 0,97  | 0,00 | 0,12 | 0,00 | 0,13 | 0,12 | 0,13 |
| trnL.4   | -1,32 | 0,20  | -0,39 | 0,00 | 0,32 | 0,07 | 0,32 | 0,20 | 0,21 |
| ORF 77.2 | 0,77  | -0,53 | 0,90  | 0,00 | 0,00 | 0,00 | 0,14 | 0,14 | 0,12 |
| ycf2.2   | 0,59  | -0,48 | 0,79  | 0,00 | 0,00 | 0,00 | 0,13 | 0,12 | 0,12 |
| trnI.4   | -2,03 | 0,23  | 0,27  | 0,00 | 0,29 | 0,20 | 0,37 | 0,21 | 0,21 |
| rpl23.2  | -0,85 | -0,19 | 0,34  | 0,00 | 0,11 | 0,01 | 0,14 | 0,12 | 0,13 |
| rpl2.2   | 0,48  | -0,28 | 0,60  | 0,00 | 0,03 | 0,00 | 0,12 | 0,13 | 0,12 |

Tabelle VI.7: Differentielle Expressionsanalyse in Mutanten im späten Stadium als logarithmierte relative Transkriptlevel bezogen auf WT mit Angabe des p-Werts und des Standardfehlers. NA: Not available. Die Signifikante differentielle Expression wurde mit dem Wald-Test bestimmt. Die Analyse wurde von Jose Muino durchgeführt.

| Gene  | log2FC(Mut/WT)    |                    |                   | p-Wert            |                    |                   | Standardfehler    |                    |                   |
|-------|-------------------|--------------------|-------------------|-------------------|--------------------|-------------------|-------------------|--------------------|-------------------|
|       | <i>rpoTp</i> spät | <i>rpoTmp</i> spät | <i>ptac2</i> spät | <i>rpoTp</i> spät | <i>rpoTmp</i> spät | <i>ptac2</i> spät | <i>rpoTp</i> spät | <i>rpoTmp</i> spät | <i>ptac2</i> spät |
| trnH  | -1,31             | -0,06              | -1,38             | 0,15              | 0,82               | 0,02              | 0,91              | 0,28               | 0,60              |
| psbA  | -0,92             | -0,89              | -2,99             | 0,03              | 0,00               | 0,00              | 0,43              | 0,20               | 0,28              |
| trnK  | -0,02             | 0,23               | -0,06             | 0,96              | 0,00               | 0,86              | 0,48              | 0,07               | 0,35              |
| rps16 | 0,39              | 0,57               | 1,31              | 0,34              | 0,00               | 0,00              | 0,42              | 0,14               | 0,35              |
| trnQ  | 0,53              | 0,41               | 1,22              | 0,29              | 0,04               | 0,03              | 0,51              | 0,20               | 0,57              |

## Anhang

|        |       |       |       |      |      |      |      |      |      |
|--------|-------|-------|-------|------|------|------|------|------|------|
| psbK   | -1,15 | -0,57 | -1,02 | 0,01 | 0,00 | 0,00 | 0,46 | 0,17 | 0,35 |
| psbI   | -0,91 | -0,60 | -0,95 | 0,06 | 0,00 | 0,00 | 0,49 | 0,16 | 0,31 |
| trnS.1 | -0,58 | -0,40 | -1,13 | 0,09 | 0,12 | 0,00 | 0,34 | 0,26 | 0,33 |
| trnG.1 | 0,31  | 0,18  | 0,24  | 0,33 | 0,01 | 0,42 | 0,32 | 0,07 | 0,29 |
| trnR.1 | -0,23 | 0,54  | -0,02 | 0,82 | 0,00 | 0,96 | 0,99 | 0,17 | 0,52 |
| atpA   | -0,38 | -0,57 | 0,08  | 0,39 | 0,00 | 0,81 | 0,44 | 0,14 | 0,34 |
| atpF   | -0,98 | -0,43 | 0,29  | 0,00 | 0,00 | 0,38 | 0,31 | 0,14 | 0,33 |
| atpH   | -0,89 | -0,27 | -0,21 | 0,00 | 0,14 | 0,53 | 0,30 | 0,18 | 0,34 |
| atpI   | -1,12 | -0,45 | -0,49 | 0,00 | 0,00 | 0,19 | 0,31 | 0,09 | 0,37 |
| rps2   | -0,19 | -0,10 | 1,52  | 0,44 | 0,40 | 0,00 | 0,24 | 0,12 | 0,29 |
| rpoC2  | -0,09 | 0,04  | 2,06  | 0,73 | 0,84 | 0,00 | 0,27 | 0,21 | 0,36 |
| rpoC1  | 0,08  | -0,05 | 2,48  | 0,71 | 0,69 | 0,00 | 0,20 | 0,11 | 0,37 |
| rpoB   | -0,27 | 0,10  | 1,97  | 0,41 | 0,52 | 0,00 | 0,33 | 0,16 | 0,36 |
| trnC   | -0,52 | 0,21  | -2,22 | 0,20 | 0,07 | 0,00 | 0,40 | 0,12 | 0,61 |
| petN   | -1,04 | -0,68 | -2,82 | 0,08 | 0,01 | 0,00 | 0,59 | 0,25 | 0,37 |
| psbM   | -1,26 | -0,41 | -0,55 | 0,13 | 0,06 | 0,14 | 0,83 | 0,22 | 0,37 |
| trnD   | 0,66  | 0,77  | 0,11  | 0,02 | 0,00 | 0,85 | 0,28 | 0,24 | 0,62 |
| trnY   | -0,14 | 0,40  | -1,43 | 0,55 | 0,00 | 0,00 | 0,23 | 0,13 | 0,51 |
| trnE   | -0,30 | 0,07  | 0,02  | 0,35 | 0,74 | 0,96 | 0,32 | 0,22 | 0,50 |
| trnT.1 | -0,44 | 0,05  | -0,82 | 0,73 | 0,89 | 0,05 | 1,28 | 0,37 | 0,41 |
| psbD   | -0,68 | -0,74 | -1,61 | 0,07 | 0,00 | 0,00 | 0,38 | 0,11 | 0,32 |
| psbC   | -0,60 | -0,68 | -1,41 | 0,15 | 0,00 | 0,00 | 0,41 | 0,16 | 0,32 |
| trnS.2 | -1,43 | -0,57 | -3,03 | 0,00 | 0,00 | 0,00 | 0,45 | 0,07 | 0,29 |
| psbZ   | -1,19 | -0,56 | -3,04 | 0,00 | 0,00 | 0,00 | 0,28 | 0,11 | 0,34 |
| trnG.2 | -0,31 | 0,34  | -1,23 | 0,43 | 0,05 | 0,02 | 0,39 | 0,17 | 0,52 |
| trnFM  | 0,83  | 0,67  | -1,62 | 0,02 | 0,04 | 0,00 | 0,37 | 0,32 | 0,46 |
| rps14  | -0,95 | -0,30 | -1,54 | 0,00 | 0,00 | 0,00 | 0,32 | 0,07 | 0,36 |
| psaB   | -1,12 | -0,44 | -1,77 | 0,02 | 0,00 | 0,00 | 0,47 | 0,06 | 0,29 |
| psaA   | -1,05 | -0,37 | -1,89 | 0,03 | 0,00 | 0,00 | 0,49 | 0,05 | 0,28 |
| ycf3   | 1,15  | 0,07  | 2,12  | 0,00 | 0,03 | 0,00 | 0,31 | 0,03 | 0,35 |
| trnS.3 | 1,29  | 0,72  | 1,53  | 0,00 | 0,00 | 0,02 | 0,27 | 0,14 | 0,65 |
| rps4   | 0,57  | 0,12  | 1,65  | 0,01 | 0,08 | 0,00 | 0,22 | 0,07 | 0,34 |
| trnT.2 | 0,17  | 0,54  | -0,44 | 0,72 | 0,03 | 0,38 | 0,49 | 0,25 | 0,50 |
| trnL.1 | -0,17 | 0,07  | -2,79 | 0,68 | 0,62 | 0,00 | 0,40 | 0,14 | 0,37 |
| trnF   | 0,79  | 0,23  | -1,51 | 0,02 | 0,00 | 0,00 | 0,35 | 0,06 | 0,37 |
| ndhJ   | -0,12 | -0,48 | 0,62  | 0,72 | 0,00 | 0,06 | 0,34 | 0,14 | 0,33 |
| ndhK   | -0,22 | -0,66 | 0,36  | 0,55 | 0,00 | 0,30 | 0,37 | 0,16 | 0,35 |
| ndhC   | -0,37 | -0,72 | 0,41  | 0,19 | 0,00 | 0,21 | 0,28 | 0,15 | 0,32 |
| trnV.1 | 0,59  | 0,17  | -1,24 | 0,01 | 0,21 | 0,00 | 0,22 | 0,14 | 0,35 |
| trnM   | 0,45  | 0,50  | 1,55  | 0,42 | 0,13 | 0,00 | 0,55 | 0,33 | 0,46 |
| atpE   | -0,81 | -0,40 | 0,27  | 0,02 | 0,03 | 0,49 | 0,35 | 0,18 | 0,39 |
| atpB   | -0,51 | -0,45 | 0,31  | 0,32 | 0,01 | 0,37 | 0,51 | 0,18 | 0,35 |
| rbcL   | -0,35 | -0,19 | -4,85 | 0,53 | 0,24 | 0,00 | 0,57 | 0,16 | 0,28 |
| accD   | 0,31  | -0,27 | 0,50  | 0,08 | 0,00 | 0,12 | 0,18 | 0,06 | 0,32 |
| psaI   | -1,40 | -0,55 | 0,10  | 0,08 | 0,01 | 0,79 | 0,79 | 0,22 | 0,37 |
| ycf4   | -0,59 | -0,24 | 0,42  | 0,03 | 0,00 | 0,24 | 0,27 | 0,07 | 0,36 |
| ycf10  | -0,55 | -0,26 | 0,20  | 0,13 | 0,00 | 0,57 | 0,36 | 0,07 | 0,35 |
| petA   | 0,05  | -0,20 | 0,39  | 0,78 | 0,00 | 0,25 | 0,18 | 0,06 | 0,34 |
| psbJ   | -0,81 | -0,36 | -1,47 | 0,03 | 0,13 | 0,00 | 0,38 | 0,24 | 0,37 |
| psbL   | -1,60 | -0,53 | -2,11 | 0,02 | 0,02 | 0,00 | 0,71 | 0,22 | 0,34 |
| psbF   | -1,03 | -0,61 | -2,05 | 0,01 | 0,00 | 0,00 | 0,41 | 0,21 | 0,33 |
| psbE   | -0,67 | -0,70 | -1,99 | 0,23 | 0,00 | 0,00 | 0,56 | 0,23 | 0,38 |
| petL   | -0,55 | -0,73 | -0,49 | 0,31 | 0,01 | 0,13 | 0,55 | 0,27 | 0,32 |
| petG   | -0,20 | -0,80 | -0,56 | 0,54 | 0,00 | 0,08 | 0,33 | 0,23 | 0,32 |
| trnW   | 1,56  | 0,96  | -0,08 | 0,00 | 0,00 | 0,88 | 0,51 | 0,20 | 0,53 |
| trnP   | 0,72  | 0,43  | -0,58 | 0,15 | 0,03 | 0,34 | 0,50 | 0,19 | 0,61 |
| psaJ   | -0,40 | -0,60 | -1,54 | 0,32 | 0,06 | 0,00 | 0,40 | 0,32 | 0,40 |
| rpl33  | 0,07  | -0,20 | -0,29 | 0,77 | 0,00 | 0,36 | 0,23 | 0,06 | 0,32 |

## Anhang

|          |       |       |       |      |      |      |      |      |      |
|----------|-------|-------|-------|------|------|------|------|------|------|
| rpS18    | -0,54 | 0,02  | 0,46  | 0,38 | 0,55 | 0,14 | 0,62 | 0,04 | 0,31 |
| rpl20    | -0,14 | -0,12 | 2,81  | 0,56 | 0,08 | 0,00 | 0,24 | 0,07 | 0,37 |
| rps12    | 0,11  | -0,08 | 2,56  | 0,62 | 0,29 | 0,00 | 0,22 | 0,07 | 0,30 |
| clpP     | 0,18  | -0,06 | 2,37  | 0,48 | 0,59 | 0,00 | 0,26 | 0,12 | 0,32 |
| psbB     | -0,69 | -0,66 | -4,11 | 0,15 | 0,00 | 0,00 | 0,48 | 0,11 | 0,28 |
| psbT     | -1,05 | -0,33 | -3,71 | 0,02 | 0,02 | 0,00 | 0,46 | 0,14 | 0,29 |
| psbN     | 0,76  | 0,08  | 3,22  | 0,09 | 0,71 | 0,00 | 0,44 | 0,21 | 0,39 |
| psbH     | -0,33 | -0,34 | -3,34 | 0,40 | 0,00 | 0,00 | 0,40 | 0,12 | 0,32 |
| petB     | -0,50 | -0,46 | -2,71 | 0,08 | 0,00 | 0,00 | 0,29 | 0,09 | 0,32 |
| petD     | -0,55 | -0,31 | -2,47 | 0,01 | 0,01 | 0,00 | 0,22 | 0,11 | 0,33 |
| rpoA     | -0,14 | 0,04  | 1,53  | 0,54 | 0,57 | 0,00 | 0,22 | 0,06 | 0,34 |
| rps11    | 0,40  | 0,07  | 1,54  | 0,33 | 0,24 | 0,00 | 0,42 | 0,06 | 0,29 |
| rpl36    | -0,48 | 0,05  | 1,54  | 0,22 | 0,47 | 0,00 | 0,40 | 0,08 | 0,27 |
| rps8     | -0,44 | 0,04  | 1,45  | 0,18 | 0,57 | 0,00 | 0,33 | 0,07 | 0,31 |
| rpl14    | 0,24  | 0,07  | 1,06  | 0,34 | 0,23 | 0,00 | 0,25 | 0,06 | 0,31 |
| rpl16    | -0,06 | 0,06  | 1,14  | 0,77 | 0,01 | 0,00 | 0,19 | 0,02 | 0,36 |
| rps3     | 0,27  | -0,08 | 0,94  | 0,24 | 0,11 | 0,00 | 0,23 | 0,05 | 0,33 |
| rpl22    | 0,14  | -0,15 | 0,76  | 0,60 | 0,00 | 0,01 | 0,27 | 0,04 | 0,29 |
| rps19    | 0,06  | -0,08 | 0,68  | 0,88 | 0,36 | 0,03 | 0,41 | 0,09 | 0,31 |
| rpl2.1   | 2,13  | 1,47  | 1,52  | 0,04 | 0,19 | 0,17 | 1,05 | 1,11 | 1,11 |
| rpl23.1  | 1,54  | 1,72  | 1,32  | 0,27 | 0,25 | NA   | 1,40 | 1,49 | 1,83 |
| trnI.1   | 1,87  | 2,27  | 1,20  | 0,23 | 0,14 | 0,50 | 1,56 | 1,54 | 1,78 |
| ycf2.1   | 2,65  | 1,65  | 2,63  | 0,06 | 0,28 | 0,14 | 1,40 | 1,51 | 1,80 |
| ORF 77.1 | 2,35  | 1,73  | 2,30  | 0,03 | 0,13 | NA   | 1,06 | 1,14 | 1,27 |
| trnL.2   | 1,40  | 1,53  | 0,33  | 0,06 | 0,01 | 0,54 | 0,75 | 0,61 | 0,54 |
| ndhB.1   | 0,99  | 1,33  | 2,50  | 0,28 | 0,17 | 0,03 | 0,91 | 0,97 | 1,14 |
| rps7.1   | 1,42  | 1,36  | 3,07  | 0,17 | 0,22 | 0,02 | 1,04 | 1,10 | 1,27 |
| rps12C   | 1,19  | 1,07  | 3,50  | 0,09 | 0,12 | 0,00 | 0,71 | 0,69 | 0,87 |
| trnV.2   | 1,75  | 1,49  | -0,44 | 0,03 | 0,04 | 0,47 | 0,79 | 0,74 | 0,61 |
| trnI.2   | 1,79  | 1,24  | -1,53 | 0,01 | 0,07 | 0,00 | 0,72 | 0,68 | 0,49 |
| trnA.1   | 2,71  | 1,83  | -1,43 | NA   | 0,21 | NA   | 1,40 | 1,47 | 1,73 |
| trnR.2   | 1,76  | 1,50  | -1,27 | 0,02 | 0,04 | 0,09 | 0,73 | 0,74 | 0,74 |
| trnN.1   | 2,23  | 2,00  | -0,25 | NA   | 0,08 | 0,81 | 1,12 | 1,14 | 1,02 |
| ycf1.1   | 1,44  | 1,44  | 3,60  | 0,03 | 0,04 | 0,00 | 0,67 | 0,71 | 0,74 |
| ndhF     | -1,44 | 0,07  | -1,30 | 0,00 | 0,05 | 0,00 | 0,45 | 0,04 | 0,31 |
| rpl32    | -0,84 | -0,30 | 2,22  | 0,30 | 0,19 | 0,00 | 0,81 | 0,23 | 0,37 |
| trnL.3   | 0,34  | 1,32  | 1,61  | 0,73 | 0,00 | 0,00 | 0,97 | 0,34 | 0,45 |
| ycf5     | -0,82 | -0,13 | -0,60 | 0,06 | 0,17 | 0,06 | 0,43 | 0,09 | 0,32 |
| ndhD     | -0,31 | -0,28 | 1,77  | 0,43 | 0,00 | 0,00 | 0,39 | 0,08 | 0,32 |
| psaC     | -0,41 | -0,23 | 0,61  | 0,14 | 0,04 | 0,09 | 0,28 | 0,11 | 0,36 |
| ndhE     | -0,49 | -0,13 | 0,15  | 0,15 | 0,02 | 0,64 | 0,34 | 0,05 | 0,31 |
| ndhG     | -0,51 | -0,04 | 0,37  | 0,20 | 0,55 | 0,28 | 0,40 | 0,06 | 0,34 |
| ndhI     | -0,27 | 0,03  | 0,38  | 0,35 | 0,63 | 0,17 | 0,29 | 0,06 | 0,27 |
| ndhA     | -0,39 | -0,01 | 0,23  | 0,17 | 0,89 | 0,49 | 0,29 | 0,07 | 0,33 |
| ndhH     | 0,24  | 0,13  | 1,46  | 0,33 | 0,14 | 0,00 | 0,25 | 0,09 | 0,31 |
| rps15    | 0,33  | 0,22  | 2,15  | 0,53 | 0,02 | 0,00 | 0,51 | 0,10 | 0,30 |
| ycf1.2   | 0,40  | 0,55  | 3,52  | 0,41 | 0,00 | 0,00 | 0,48 | 0,13 | 0,30 |
| trnN.2   | 2,23  | 2,00  | -0,25 | NA   | 0,08 | 0,81 | 1,12 | 1,15 | 1,03 |
| trnR.3   | 1,76  | 1,51  | -1,26 | 0,02 | 0,04 | 0,09 | 0,73 | 0,74 | 0,74 |
| trnA.2   | 2,71  | 1,83  | -1,43 | NA   | 0,21 | NA   | 1,40 | 1,46 | 1,72 |
| trnI.3   | 1,79  | 1,24  | -1,53 | 0,01 | 0,07 | 0,00 | 0,72 | 0,68 | 0,49 |
| trnV.3   | 1,75  | 1,48  | -0,46 | 0,03 | 0,04 | 0,46 | 0,78 | 0,73 | 0,61 |
| rps12B   | 1,19  | 1,07  | 3,50  | 0,09 | 0,12 | 0,00 | 0,71 | 0,69 | 0,88 |
| rps7.2   | 1,42  | 1,36  | 3,07  | 0,17 | 0,22 | 0,02 | 1,04 | 1,10 | 1,27 |
| ndhB.2   | 0,99  | 1,33  | 2,50  | 0,28 | 0,17 | 0,03 | 0,91 | 0,97 | 1,14 |
| trnL.4   | 1,40  | 1,53  | 0,32  | 0,06 | 0,01 | 0,55 | 0,75 | 0,62 | 0,54 |
| ORF 77.2 | 2,35  | 1,73  | 2,29  | 0,03 | 0,13 | NA   | 1,06 | 1,14 | 1,27 |
| ycf2.2   | 2,65  | 1,65  | 2,63  | 0,06 | 0,27 | 0,14 | 1,40 | 1,51 | 1,79 |

## Anhang

|         |      |      |      |      |      |      |      |      |      |
|---------|------|------|------|------|------|------|------|------|------|
| trnI.4  | 1,87 | 2,26 | 1,19 | 0,23 | 0,14 | 0,50 | 1,56 | 1,54 | 1,78 |
| rpl23.2 | 1,54 | 1,73 | 1,33 | 0,28 | 0,26 | NA   | 1,43 | 1,53 | 1,88 |
| rpl2.2  | 1,95 | 1,33 | 1,43 | 0,02 | 0,15 | 0,14 | 0,87 | 0,91 | 0,97 |

Tabelle VI.8: Statistische Daten zu dRNA-Seq für die Analyse von WT und Mutanten, Rep: Replik, +PP: +PPase-cDNA-Bibliothek, -PP: -PPase-cDNA-Bibliothek. Dedup.: Deduplikation

| Probe                       | Reads sequenziert | Reads nach Trimmen | Kern-Reads gemappt nach Dedup. | Organellen-Reads gemappt nach Dedup. | Reads gemappt zum Plastidengenom | PCR-Duplikate | % PCR-Duplikate |
|-----------------------------|-------------------|--------------------|--------------------------------|--------------------------------------|----------------------------------|---------------|-----------------|
| WT früh Rep1 +PP            | 4851267           | 4543774            | 1058442                        | 91023                                | 69422                            | 159194        | 63,6            |
| WT früh Rep1 -PP            | 3995698           | 3796423            | 1244419                        | 131439                               | 104104                           | 130959        | 49,9            |
| WT früh Rep2 +PP            | 5346148           | 4991336            | 1627571                        | 122873                               | 95259                            | 81335         | 39,8            |
| WT früh Rep2 -PP            | 4594905           | 4377593            | 1040107                        | 80869                                | 60609                            | 155013        | 65,7            |
| WT früh Rep3 +PP            | 8814658           | 6532114            | 1053750                        | 77621                                | 54174                            | 410922        | 84,1            |
| WT früh Rep3 -PP            | 8097013           | 7410734            | 934418                         | 80211                                | 55431                            | 475481        | 85,6            |
| <i>rpoTmp</i> früh Rep1 +PP | 6068199           | 4840209            | 1132036                        | 168078                               | 137553                           | 177314        | 51,3            |
| <i>rpoTmp</i> früh Rep1 -PP | 5478809           | 4459641            | 1341979                        | 196796                               | 160077                           | 216028        | 52,3            |
| <i>rpoTmp</i> früh Rep2 +PP | 5643939           | 4675899            | 980187                         | 156281                               | 131458                           | 187325        | 54,5            |
| <i>rpoTmp</i> früh Rep2 -PP | 5351242           | 4312913            | 1237842                        | 167429                               | 138507                           | 272147        | 61,9            |
| <i>rpoTmp</i> früh Rep3 +PP | 6641494           | 5283324            | 1352523                        | 189074                               | 156901                           | 222777        | 54,1            |
| <i>rpoTmp</i> früh Rep3 -PP | 5101517           | 3429561            | 952996                         | 117292                               | 95220                            | 192364        | 62,1            |
| <i>rpoTp</i> früh Rep1 +PP  | 4186061           | 3769785            | 1696824                        | 64771                                | 35588                            | 21945         | 25,3            |
| <i>rpoTp</i> früh Rep1 -PP  | 3877178           | 3596588            | 1496904                        | 56548                                | 30619                            | 41309         | 42,2            |
| <i>rpoTp</i> früh Rep2 +PP  | 7389195           | 6709453            | 973838                         | 58414                                | 34561                            | 307875        | 84,1            |
| <i>rpoTp</i> früh Rep2 -PP  | 6608630           | 5785629            | 1108956                        | 61881                                | 36031                            | 180124        | 74,4            |
| <i>rpoTp</i> früh Rep3 +PP  | 9170751           | 8660571            | 841157                         | 43044                                | 25003                            | 291461        | 87,1            |
| <i>rpoTp</i> früh Rep3 -PP  | 7364602           | 6811567            | 958939                         | 47879                                | 24936                            | 227304        | 82,6            |
| <i>clb19</i> früh Rep1 +PP  | 3195438           | 2977925            | 1084023                        | 42053                                | 22953                            | 38075         | 47,5            |
| <i>clb19</i> früh Rep1 -PP  | 3648451           | 3206825            | 1014765                        | 49771                                | 30084                            | 32703         | 39,7            |
| <i>clb19</i> früh Rep2 +PP  | 10269716          | 9016005            | 1872498                        | 69450                                | 35966                            | 215027        | 75,6            |
| <i>clb19</i> früh Rep2 -PP  | 6594679           | 5499943            | 683423                         | 23086                                | 11804                            | 163680        | 87,6            |
| <i>clb19</i> früh Rep3 +PP  | 5673652           | 4671997            | 1633749                        | 64162                                | 32853                            | 69149         | 51,9            |
| <i>clb19</i> früh Rep3 -PP  | 5825851           | 4535630            | 1585468                        | 65310                                | 34800                            | 72213         | 52,5            |
| WT spät Rep1 +PP            | 4993028           | 4709646            | 509245                         | 361745                               | 349127                           | 1552568       | 81,1            |
| WT spät Rep1 -PP            | 4057433           | 3652255            | 618606                         | 407630                               | 388650                           | 973887        | 70,5            |
| WT spät Rep2 +PP            | 5857847           | 4578534            | 1189893                        | 724599                               | 691925                           | 682571        | 48,5            |
| WT spät Rep2 -PP            | 4803400           | 4552695            | 412350                         | 304078                               | 289879                           | 1599392       | 84,0            |
| WT spät Rep3 +PP            | 5252716           | 4648238            | 356315                         | 179113                               | 164346                           | 1328582       | 88,1            |
| WT spät Rep3 -PP            | 5435164           | 5247103            | 733672                         | 330241                               | 312554                           | 1110630       | 77,1            |
| <i>rpoTmp</i> spät Rep1 +PP | 5083603           | 4925330            | 342748                         | 223250                               | 211286                           | 2010887       | 90,0            |
| <i>rpoTmp</i> spät Rep1 -PP | 3819093           | 3634199            | 500205                         | 309569                               | 292400                           | 968767        | 75,8            |
| <i>rpoTmp</i> spät Rep2 +PP | 8243336           | 5744202            | 1418458                        | 665314                               | 616424                           | 672373        | 50,3            |
| <i>rpoTmp</i> spät Rep2 -PP | 4816171           | 4193330            | 804536                         | 381097                               | 352521                           | 910055        | 70,5            |
| <i>rpoTmp</i> spät Rep3 +PP | 4940505           | 4765650            | 310966                         | 140617                               | 130036                           | 1321318       | 90,4            |

## Anhang

|                             |         |         |         |        |        |         |      |
|-----------------------------|---------|---------|---------|--------|--------|---------|------|
| <i>rpoTmp</i> spät Rep3 -PP | 1535115 | 1393775 | 170197  | 55010  | 49639  | 235349  | 81,1 |
| <i>rpoTp</i> spät Rep1 +PP  | 838996  | 733768  | 136718  | 54169  | 50147  | 185616  | 77,4 |
| <i>rpoTp</i> spät Rep1 -PP  | 3876173 | 3651038 | 257945  | 97919  | 90458  | 976043  | 90,9 |
| <i>rpoTp</i> spät Rep2 +PP  | 5421189 | 4679422 | 990142  | 380708 | 360615 | 627390  | 62,2 |
| <i>rpoTp</i> spät Rep2 -PP  | 5540808 | 5137159 | 634362  | 139733 | 129555 | 583612  | 80,7 |
| <i>rpoTp</i> spät Rep3 +PP  | 4615991 | 2287268 | 733912  | 191064 | 166413 | 113034  | 37,2 |
| <i>rpoTp</i> spät Rep3 -PP  | 3830901 | 808718  | 289373  | 70751  | 60780  | 19846   | 21,9 |
| <i>ptac2</i> spät Rep1 +PP  | 6670351 | 5783832 | 1113877 | 219273 | 198801 | 992935  | 81,9 |
| <i>ptac2</i> spät Rep1 -PP  | 729931  | 6199354 | 869002  | 185711 | 172051 | 1650934 | 89,9 |
| <i>ptac2</i> spät Rep2 +PP  | 3418314 | 1535557 | 465559  | 83338  | 63906  | 21375   | 20,4 |
| <i>ptac2</i> spät Rep2 -PP  | 4091627 | 2316485 | 1086721 | 55574  | 38287  | 35442   | 38,9 |
| <i>ptac2</i> spät Rep3 +PP  | 6210264 | 4543280 | 747988  | 46864  | 34447  | 320717  | 87,3 |
| <i>ptac2</i> spät Rep3 -PP  | 4765170 | 3700246 | 515705  | 114390 | 84890  | 68075   | 37,3 |
| <i>ptac2</i> spät Rep4 +PP  | 4376413 | 3139477 | 966476  | 170560 | 156249 | 220760  | 56,4 |
| <i>ptac2</i> spät Rep4 -PP  | 8952183 | 2940903 | 893258  | 140010 | 130005 | 309826  | 68,9 |

Tabelle VI.9: Statistische Daten zu dRNA-Seq für die Analyse der frühen Samenentwicklung in WT. 0+: imbibierte Samen, 0.5: gekeimte Samen, dry: trockene Samen, etio: etiolierte Keimlinge, Rep: Replikat, +PP: +PPase-cDNA-Bibliothek, -PP: -PPase-cDNA-Bibliothek. Dedup.: Deduplikation

| Probe           | Reads sequenziert | Reads nach Trimmen | Reads gemappt zu Organellen | Reads gemappt nach Dedup. | Reads gemappt zum Plastidengenom | PCR-Duplikate | % PCR-Duplikate |
|-----------------|-------------------|--------------------|-----------------------------|---------------------------|----------------------------------|---------------|-----------------|
| WT 0+ Rep1 +PP  | 4441520           | 1845853            | 141893                      | 78476                     | 49644                            | 63417         | 44,7            |
| WT 0+ Rep1 -PP  | 3845935           | 2671399            | 124117                      | 99113                     | 74137                            | 25004         | 20,1            |
| WT 0+ Rep2 +PP  | 4380586           | 3396268            | 217854                      | 52076                     | 30692                            | 165778        | 76,1            |
| WT 0+ Rep2 -PP  | 4740638           | 4052627            | 141608                      | 62241                     | 46928                            | 79367         | 56,0            |
| WT 0+ Rep3 +PP  | 4535897           | 3539305            | 184280                      | 57347                     | 31716                            | 126933        | 68,9            |
| WT 0+ Rep3 -PP  | 4364043           | 3977272            | 124481                      | 82933                     | 52263                            | 41548         | 33,4            |
| WT 0+ Rep4 +PP  | 3846224           | 2154444            | 140515                      | 66025                     | 41314                            | 74490         | 53,0            |
| WT 0+ Rep4 -PP  | 3592958           | 3210108            | 142823                      | 40361                     | 28338                            | 102462        | 71,7            |
| WT 0.5 Rep1 +PP | 3950775           | 1487753            | 170906                      | 102334                    | 67829                            | 68572         | 40,1            |
| WT 0.5 Rep1 -PP | 4484240           | 2694876            | 198621                      | 101492                    | 71204                            | 97129         | 48,9            |
| WT 0.5 Rep2 +PP | 5540278           | 2923054            | 364746                      | 174840                    | 122373                           | 189906        | 52,1            |
| WT 0.5 Rep2 -PP | 692862            | 587281             | 41323                       | 28349                     | 19478                            | 12974         | 31,4            |
| WT 0.5 Rep3 +PP | 4196292           | 2771437            | 690316                      | 62002                     | 45805                            | 628314        | 91,0            |
| WT 0.5 Rep3 -PP | 3178161           | 1200530            | 296037                      | 71078                     | 57220                            | 224959        | 76,0            |
| WT 0.5 Rep4 +PP | 4647703           | 4085206            | 325525                      | 30565                     | 20508                            | 294960        | 90,6            |
| WT 0.5 Rep4 -PP | 3561627           | 3080317            | 192072                      | 32695                     | 25112                            | 159377        | 83,0            |
| WT dry Rep1 +PP | 2929053           | 1657919            | 89590                       | 44887                     | 26646                            | 44703         | 49,9            |
| WT dry Rep1 -PP | 4150821           | 3333772            | 91707                       | 59182                     | 43943                            | 32525         | 35,5            |
| WT dry Rep2 +PP | 5142006           | 2005189            | 116535                      | 58710                     | 36323                            | 57825         | 49,6            |
| WT dry Rep2 -PP | 3159281           | 2633672            | 83851                       | 63069                     | 44171                            | 20782         | 24,8            |
| WT dry Rep3 +PP | 3649127           | 1706311            | 90092                       | 55842                     | 32746                            | 34250         | 38,0            |
| WT dry Rep3 -PP | 4272687           | 3158810            | 103387                      | 77918                     | 56613                            | 25469         | 24,6            |

## Anhang

|                  |         |         |        |        |        |        |      |
|------------------|---------|---------|--------|--------|--------|--------|------|
| WT dry Rep4 +PP  | 2010053 | 1251000 | 60598  | 39780  | 22011  | 20818  | 34,4 |
| WT dry Rep4 -PP  | 3470482 | 2677626 | 68105  | 56481  | 39569  | 11624  | 17,1 |
| WT etio Rep1 +PP | 2462698 | 1790508 | 450470 | 62519  | 45453  | 387951 | 86,1 |
| WT etio Rep1 -PP | 4145789 | 3115325 | 328225 | 185703 | 152294 | 142522 | 43,4 |
| WT etio Rep2 +PP | 5860245 | 3537445 | 612239 | 75771  | 57916  | 536468 | 87,6 |
| WT etio Rep2 -PP | 3470897 | 1625844 | 301867 | 75771  | 55973  | 226096 | 74,9 |
| WT etio Rep3 +PP | 4265235 | 3821007 | 401013 | 67222  | 30276  | 333791 | 83,2 |
| WT etio Rep3 -PP | 1846935 | 1295204 | 109517 | 40442  | 30674  | 69075  | 63,1 |

Tabelle VI.10: Liste der identifizierten TSS in Mutantanalyse mit Referenzen zu beschriebenen TSS.

| TSS          | Alternative Schreibweise | Referenz   |
|--------------|--------------------------|--|
| 1522-PSBA    | <i>PpsbA</i> -77         | Liere <i>et al.</i> , 1995; Shen <i>et al.</i> , 2001; Swiatecka-Hagenbruch <i>et al.</i> , 2007; Zhelyazkova <i>et al.</i> 2012b; Castandet <i>et al.</i> 2019  |
| 2735-PSBA    |                          |  |
| 4451-TRNK    | beschriebener Promotor   | Castandet <i>et al.</i> , 2019   |
| 6253-RPS16   |                          |  |
| 6708-TRNQ    |                          |  |
| 7539+PSBI    | beschriebener Promotor   | Castandet <i>et al.</i> 2019   |
| 7872-TRNQ    |                          |  |
| 8126+TRNG.1  |                          |  |
| 8213+TRNG.1  | beschriebener Promotor   | Castandet <i>et al.</i> 2019   |
| 8230+TRNG.1  |                          |  |
| 8643+TRNG.1  | <i>PtrnG</i> -3          | Castandet <i>et al.</i> 2019   |
| 11392-TRNQ   |                          | Castandet <i>et al.</i> 2019   |
| 13921-ATPH   | <i>PatpH</i> -413        | Zghidi <i>et al.</i> , 2007; Castandet <i>et al.</i> 2019  |
| 20245-RPOC2  |                          |  |
| 28037+PETN   | <i>PpetN</i>             | Zhelyazkova <i>et al.</i> , 2012b; Castandet <i>et al.</i> 2019  |
| 28988-PSBM   |                          |  |
| 29204-PSBM   |                          |  |
| 29912-TRND   |                          |  |
| 30565-TRNE   | <i>PtrnE_UUC</i> -26     | Hanaoka <i>et al.</i> , 2003; Castandet <i>et al.</i> 2019   |
| 31319+TRNT.1 |                          |  |
| 31336+TRNT.1 |                          |  |
| 31523-TRNE   |                          |  |
| 31614+PSBD   | <i>PpsbD</i> -948        | Nagashima <i>et al.</i> , 2004; Shimmura <i>et al.</i> , 2017; Tsunoyama <i>et al.</i> , 2004  |
| 32644+PSBD   | beschriebener Promotor   | Castandet <i>et al.</i> 2019   |
| 32672+PSBD   | beschriebener Promotor   | Castandet <i>et al.</i> 2019   |
| 33526+PSBC   | <i>PpsbC</i> -194        | Zhelyazkova <i>et al.</i> 2012b; Castandet <i>et al.</i> 2019  |
| 35410-TRNE   |                          |  |
| 36324+TRNG.2 |                          |  |
| 36358+TRNG.2 |                          |  |
| 36372+TRNG.2 |                          |  |
| 36665+TRNS   |                          |  |
| 36849-TRNFM  |                          |  |
| 37271+TRNS   |                          |  |
| 42045-PSAA   | <i>PpsaA</i> -188        | Fey <i>et al.</i> , 2005; Yuan Meng <i>et al.</i> , 1988; Swiatecka-Hagenbruch <i>et al.</i> 2007; Zhelyazkova <i>et al.</i> 2012b; Castandet <i>et al.</i> 2013 |
| 44234+TRNS.3 |                          |  |
| 44755-YCF3   |                          |  |
| 44824+TRNS.3 |                          |  |

## Anhang

|                 |                                 |   |
|-----------------|---------------------------------|---|
| 45454-YCF3      |                                 |   |
| 46110+TRNL.1    | beschriebener Promotor          | Castandet <i>et al.</i> 2019  |
| 46339-TRNT.2    |                                 |   |
| 46499+TRNL.1    | beschriebener Promotor          | Castandet <i>et al.</i> 2019  |
| 46585-TRNT.2    |                                 |   |
| 46589+TRNL.1    | beschriebener Promotor          | Castandet <i>et al.</i> 2019  |
| 46751+TRNL.1    | beschriebener Promotor          | Zhelyazkova <i>et al.</i> 2012b; Castandet <i>et al.</i> 2019   |
| 46928+TRNF      |                                 |   |
| 47545+TRNF      | beschriebener Promotor          | Castandet <i>et al.</i> 2013; Castandet <i>et al.</i> 2019  |
| 47703+TRNF      | beschriebener Promotor          | Castandet <i>et al.</i> 2013; Castandet <i>et al.</i> 2019  |
| 49371+TRNM      |                                 |   |
| 51885-TRNV.1    |                                 |   |
| 52012+TRNM      |                                 |   |
| 53093-ATPE      | <i>PatpE</i> -430               | Kapoor <i>et al.</i> , 1994; Ghulam <i>et al.</i> , 2012  |
| 54618-ATPB      | <i>PatpB</i> -467               | Ghulam <i>et al.</i> , 2012   |
| 54671-ATPB      |                                 |   |
| 54781+RBCL      | <i>PrbcL</i> -179               | Hanaoka <i>et al.</i> 2003; Castandet <i>et al.</i> 2019  |
| 56771-ATPB      | beschriebener Promotor          | Schweer <i>et al.</i> , 2006  |
| 59146+PSAI      |                                 |   |
| 61105+PETA      |                                 |   |
| 62285-ATPB      | beschriebener Promotor          | Castandet <i>et al.</i> 2019  |
| 63131-ATPB      |                                 |   |
| 63723-PSBJ      |                                 |   |
| 64448-PSBE      | <i>PpsbE</i> -126               | Zhelyazkova <i>et al.</i> 2012b; Castandet <i>et al.</i> 2019; Demarsy <i>et al.</i> , 2012; Castandet <i>et al.</i> , 2013 |
| 66089+PSAJ      |                                 | Castandet <i>et al.</i> 2019  |
| 66382-TRNW      |                                 |   |
| 66577-TRNP      |                                 |   |
| 66862-TRNP      |                                 |   |
| 66894+PSAJ      | <i>PpsaJ</i> -37                | Nagashima <i>et al.</i> , 2004; Zhelyazkova <i>et al.</i> 2012b; Castandet <i>et al.</i> 2019                               |
| 67744-TRNP      | beschriebener Promotor          | Castandet <i>et al.</i> 2019  |
| 67995-TRNP      |                                 |   |
| 71416-RPS12     | beschriebener Promotor          | Castandet <i>et al.</i> 2019  |
| 71940-CLPP1     | <i>PclpP</i> -58                | Swiatecka-Hagenbruch <i>et al.</i> 2007; Castandet <i>et al.</i> 2019   |
| 72200+PSBB      | <i>PpsbB</i> -176               | Westhoff & Hermann, 1988; Zhelyazkova <i>et al.</i> 2012b; Castandet <i>et al.</i> 2019                                     |
| 74413-PSBN      | <i>PpsbN</i> -32                | Zghidi <i>et al.</i> 2007   |
| 76780+YCF2.1    | beschriebener Promotor          | Castandet <i>et al.</i> 2019  |
| 78010+YCF2.1    |                                 |   |
| 78677+YCF2.1    |                                 |   |
| 78986-RPOA      |                                 |   |
| 80625-RPS8      |                                 |   |
| 82650-RPL14     | beschriebener Promotor          | Castandet <i>et al.</i> 2019  |
| 85045-RPS19     |                                 |   |
| 86068-RPL2.1    | beschriebener Promotor          | Castandet <i>et al.</i> 2019; Thomas <i>et al.</i> , 1988   |
| 86211-RPL23     | beschriebener Promotor          | Thomas <i>et al.</i> 1988; Zhelyazkova <i>et al.</i> 2012b; Castandet <i>et al.</i> 2019                                    |
| 86417-TRNI.1    |                                 | Castandet <i>et al.</i> 2019  |
| 92598-TRNI      |                                 |   |
| 94519-TRNL.2    |                                 |   |
| 100687+TRNV.2   | beschriebener Promotor          | Castandet <i>et al.</i> 2019  |
| 100873+RRN16S.1 | Pc-Promotor, <i>Prrn16</i> -139 | Swiatecka-Hagenbruch <i>et al.</i> 2007; Courtois <i>et al.</i> , 2007  |
| 100898+RRN16S.1 | P2-Promotor, <i>Prrn16</i> -112 | Swiatecka-Hagenbruch <i>et al.</i> 2007; Castandet <i>et al.</i> 2019   |
| 104475+RRN23S.1 |                                 |   |
| 109231-TRNN.1   | beschriebener Promotor          | Castandet <i>et al.</i> 2013; Castandet <i>et al.</i> 2019  |
| 109366+TIC214   | <i>PycfI</i> -39                | Swiatecka-Hagenbruch <i>et al.</i> , 2008   |
| 110032-TRNN.1   |                                 |   |



## Anhang

|                 |                                 |   |
|-----------------|---------------------------------|---|
| 110341-TRNN.1   |                                 |   |
| 112358+RPL32    | beschriebener Promotor          | Castandet <i>et al.</i> 2019  |
| 112771-NDHF     |                                 |   |
| 112953+RPL32    | beschriebener Promotor          | Castandet <i>et al.</i> 2019  |
| 112957-NDHF     |                                 |   |
| 112979+RPL32    | beschriebener Promotor          | Castandet <i>et al.</i> 2019  |
| 113004-NDHF     |                                 |   |
| 114114-NDHF     |                                 |   |
| 114261+TRNL.3   |                                 |   |
| 114944+TRNN     |                                 |   |
| 116996-NDHF     | beschriebener Promotor          | Castandet <i>et al.</i> 2019  |
| 118082-PSAC     |                                 |   |
| 118171-NDHE     |                                 |   |
| 118899-NDHE     | beschriebener Promotor          | Castandet <i>et al.</i> 2019  |
| 122076-NDHA     | beschriebener Promotor          | Castandet <i>et al.</i> 2019, in Mais als prozessiertes 5`Ende beschrieben Zoschke <i>et al.</i> , 2016 |
| 128308+TRNN.2   |                                 |   |
| 128616+TRNN.2   |                                 |   |
| 129283-TIC214   | <i>Pycfl-39</i>                 | Swiatecka-Hagenbruch <i>et al.</i> , 2008   |
| 129418+TRNN.2   |                                 |   |
| 134174-RRN23S.2 |                                 |   |
| 137751-RRN16S.2 | P2-Promotor, <i>Prrn16</i> -112 | Swiatecka-Hagenbruch <i>et al.</i> 2007; Castandet <i>et al.</i> 2019                                   |
| 137776-RRN16S.2 | Pc-Promotor, <i>Prrn16</i> -139 | Swiatecka-Hagenbruch <i>et al.</i> 2007; Courtois <i>et al.</i> , 2007                                  |
| 137962-TRNV.3   |                                 |   |
| 144130+TRNL.4   |                                 |   |
| 144275+TRNL.4   |                                 |   |
| 146051+TRNI     |                                 |   |
| 152232+TRNI.4   |                                 |   |
| 152438+RPL23.2  |                                 |   |
| 152581+RPL2.2   |                                 |   |
| 153604+PSBK     |                                 |   |

Tabelle VI.11: Klassifizierung mit dem statistischen Modell im frühen Stadium. Die Zugehörigkeit der TSS zu einem Fall ist in Prozent angegeben. Rottöne: hohe Wahrscheinlichkeit für diesen Fall (60-100 %), Blautöne: niedrige Wahrscheinlichkeit für diesen Fall (0-59 %). Pol.: Polymerase. . Die Analyse wurde von Frederic Ehlers unter Anleitung von Julia Rott und Ivo Große durchgeführt.

| TSS abhängig von | keine Pol | PEP    | <i>rpoTp</i> | PEP/<br><i>rpoTp</i> | <i>rpoTmp</i> | PEP/<br><i>rpoTmp</i> | <i>rpoTp</i> /<br><i>rpoTmp</i> | alle Pols |
|------------------|-----------|--------|--------------|----------------------|---------------|-----------------------|---------------------------------|-----------|
| TSS              | Fall 1    | Fall 2 | Fall 3       | Fall 4               | Fall 5        | Fall 6                | Fall 7                          | Fall 8    |
| 114114-NDHF      | 0,0       | 99,9   | 0,0          | 0,0                  | 0,0           | 0,0                   | 0,0                             | 0,0       |
| 36665+TRNS       | 0,1       | 99,7   | 0,0          | 0,2                  | 0,0           | 0,1                   | 0,0                             | 0,0       |
| 112771-NDHF      | 0,6       | 97,0   | 0,2          | 1,7                  | 0,1           | 0,2                   | 0,0                             | 0,2       |
| 46110+TRNL.1     | 0,1       | 95,3   | 0,0          | 4,5                  | 0,0           | 0,0                   | 0,0                             | 0,0       |
| 36372+TRNG.2     | 0,1       | 93,8   | 0,0          | 6,1                  | 0,0           | 0,0                   | 0,0                             | 0,0       |
| 109231-TRNN.1    | 0,1       | 93,5   | 0,0          | 6,4                  | 0,0           | 0,0                   | 0,0                             | 0,0       |
| 129418+TRNN.2    | 0,1       | 83,7   | 0,0          | 16,1                 | 0,0           | 0,0                   | 0,0                             | 0,0       |
| 29204-PSBM       | 1,0       | 82,8   | 0,3          | 15,1                 | 0,1           | 0,3                   | 0,1                             | 0,3       |
| 30565-TRNE       | 0,1       | 81,0   | 0,0          | 18,8                 | 0,0           | 0,0                   | 0,0                             | 0,1       |
| 113004-NDHF      | 4,1       | 80,9   | 1,2          | 10,4                 | 0,6           | 1,3                   | 0,2                             | 1,2       |
| 33526+PSBC       | 0,1       | 80,5   | 0,0          | 19,3                 | 0,0           | 0,0                   | 0,0                             | 0,0       |
| 66577-TRNP       | 0,1       | 79,1   | 0,0          | 20,7                 | 0,0           | 0,0                   | 0,0                             | 0,1       |
| 152581+RPL2.2    | 0,1       | 78,0   | 0,0          | 21,8                 | 0,0           | 0,0                   | 0,0                             | 0,1       |
| 54781+RBCL       | 0,2       | 76,9   | 0,1          | 22,6                 | 0,0           | 0,1                   | 0,0                             | 0,2       |
| 36358+TRNG.2     | 0,8       | 75,9   | 0,2          | 22,2                 | 0,1           | 0,3                   | 0,1                             | 0,4       |
| 72200+PSBB       | 0,3       | 64,0   | 0,1          | 35,3                 | 0,0           | 0,1                   | 0,0                             | 0,2       |
| 86068-RPL2.1     | 0,1       | 60,9   | 0,0          | 38,8                 | 0,0           | 0,0                   | 0,0                             | 0,1       |
| 37271+TRNS       | 0,5       | 60,2   | 0,2          | 38,6                 | 0,1           | 0,2                   | 0,0                             | 0,2       |
| 32644+PSBD       | 0,0       | 0,0    | 100,0        | 0,0                  | 0,0           | 0,0                   | 0,0                             | 0,0       |

## Anhang

|                 |     |      |      |       |     |     |      |     |
|-----------------|-----|------|------|-------|-----|-----|------|-----|
| 153604+PSBK     | 0,0 | 0,0  | 99,7 | 0,1   | 0,0 | 0,0 | 0,0  | 0,0 |
| 109366+TIC214   | 0,0 | 0,0  | 99,7 | 0,1   | 0,0 | 0,0 | 0,1  | 0,0 |
| 92598-TRNI      | 0,0 | 0,0  | 99,5 | 0,4   | 0,0 | 0,0 | 0,0  | 0,0 |
| 129283-TIC214   | 0,1 | 0,0  | 99,4 | 0,2   | 0,0 | 0,0 | 0,2  | 0,1 |
| 85045-RPS19     | 0,2 | 0,1  | 99,0 | 0,6   | 0,0 | 0,0 | 0,1  | 0,1 |
| 146051+TRNI     | 0,0 | 0,0  | 98,9 | 1,0   | 0,0 | 0,0 | 0,0  | 0,0 |
| 112358+RPL32    | 0,8 | 0,1  | 98,0 | 0,3   | 0,2 | 0,0 | 0,5  | 0,1 |
| 7872-TRNQ       | 3,2 | 0,8  | 90,9 | 2,3   | 0,6 | 0,1 | 1,3  | 0,9 |
| 56771-ATPB      | 2,1 | 0,3  | 66,7 | 0,7   | 0,6 | 0,2 | 28,7 | 0,8 |
| 86417-TRNI.1    | 0,0 | 0,0  | 0,0  | 100,0 | 0,0 | 0,0 | 0,0  | 0,0 |
| 152232+TRNI.4   | 0,0 | 0,0  | 0,0  | 100,0 | 0,0 | 0,0 | 0,0  | 0,0 |
| 63723-PSBJ      | 0,0 | 0,1  | 0,0  | 99,9  | 0,0 | 0,0 | 0,0  | 0,0 |
| 78986-RPOA      | 0,0 | 0,1  | 0,0  | 99,9  | 0,0 | 0,0 | 0,0  | 0,0 |
| 35410-TRNE      | 0,0 | 0,0  | 0,1  | 99,8  | 0,0 | 0,0 | 0,0  | 0,0 |
| 6708-TRNQ       | 0,0 | 0,2  | 0,0  | 99,8  | 0,0 | 0,0 | 0,0  | 0,0 |
| 128308+TRNN.2   | 0,0 | 0,1  | 0,0  | 99,8  | 0,0 | 0,0 | 0,0  | 0,0 |
| 49371+TRNM      | 0,0 | 0,1  | 0,0  | 99,7  | 0,0 | 0,0 | 0,0  | 0,0 |
| 46585-TRNT.2    | 0,0 | 0,2  | 0,0  | 99,7  | 0,0 | 0,0 | 0,0  | 0,0 |
| 74413-PSBN      | 0,0 | 0,2  | 0,0  | 99,7  | 0,0 | 0,0 | 0,0  | 0,0 |
| 144130+TRNL.4   | 0,1 | 0,2  | 0,1  | 99,6  | 0,0 | 0,0 | 0,0  | 0,0 |
| 66382-TRNW      | 0,0 | 0,0  | 0,3  | 99,6  | 0,0 | 0,0 | 0,0  | 0,0 |
| 94519-TRNL.2    | 0,1 | 0,1  | 0,1  | 99,6  | 0,0 | 0,0 | 0,0  | 0,0 |
| 144275+TRNL.4   | 0,1 | 0,5  | 0,0  | 99,4  | 0,0 | 0,0 | 0,0  | 0,0 |
| 36849-TRNFM     | 0,0 | 0,8  | 0,0  | 99,1  | 0,0 | 0,0 | 0,0  | 0,0 |
| 100687+TRNV.2   | 0,0 | 0,9  | 0,0  | 99,1  | 0,0 | 0,0 | 0,0  | 0,0 |
| 110032-TRNN.1   | 0,2 | 0,3  | 0,3  | 99,0  | 0,0 | 0,0 | 0,0  | 0,1 |
| 116996-NDHF     | 0,1 | 0,5  | 0,1  | 99,0  | 0,0 | 0,0 | 0,0  | 0,1 |
| 67995-TRNP      | 0,1 | 0,9  | 0,0  | 98,9  | 0,0 | 0,0 | 0,0  | 0,0 |
| 52012+TRNM      | 0,1 | 1,1  | 0,0  | 98,8  | 0,0 | 0,0 | 0,0  | 0,0 |
| 31336+TRNT.1    | 0,0 | 1,2  | 0,0  | 98,7  | 0,0 | 0,0 | 0,0  | 0,0 |
| 128616+TRNN.2   | 0,3 | 0,4  | 0,4  | 98,6  | 0,0 | 0,1 | 0,1  | 0,1 |
| 66862-TRNP      | 0,1 | 1,5  | 0,0  | 98,3  | 0,0 | 0,0 | 0,0  | 0,0 |
| 122076-NDHA     | 0,3 | 0,9  | 0,3  | 98,3  | 0,0 | 0,1 | 0,0  | 0,1 |
| 66894+PSAJ      | 0,3 | 1,1  | 0,2  | 98,2  | 0,0 | 0,1 | 0,0  | 0,1 |
| 7539+PSBI       | 0,2 | 1,4  | 0,1  | 98,1  | 0,0 | 0,1 | 0,0  | 0,1 |
| 44234+TRNS.3    | 0,4 | 0,6  | 0,5  | 98,1  | 0,1 | 0,1 | 0,1  | 0,2 |
| 44824+TRNS.3    | 0,4 | 0,7  | 0,4  | 98,0  | 0,0 | 0,1 | 0,1  | 0,2 |
| 137962-TRNV.3   | 0,1 | 2,3  | 0,0  | 97,6  | 0,0 | 0,0 | 0,0  | 0,0 |
| 112957-NDHF     | 0,3 | 2,3  | 0,2  | 96,8  | 0,0 | 0,1 | 0,1  | 0,3 |
| 67744-TRNP      | 0,5 | 1,9  | 0,3  | 96,7  | 0,1 | 0,1 | 0,1  | 0,2 |
| 31523-TRNE      | 0,0 | 3,3  | 0,0  | 96,6  | 0,0 | 0,0 | 0,0  | 0,0 |
| 28988-PSBM      | 0,0 | 3,4  | 0,0  | 96,5  | 0,0 | 0,0 | 0,0  | 0,0 |
| 78677+YCF2.1    | 0,2 | 3,7  | 0,1  | 95,6  | 0,0 | 0,1 | 0,0  | 0,3 |
| 6253-RPS16      | 0,4 | 3,5  | 0,2  | 95,3  | 0,1 | 0,1 | 0,1  | 0,2 |
| 54618-ATPB      | 1,1 | 2,7  | 0,7  | 94,7  | 0,1 | 0,2 | 0,1  | 0,3 |
| 64448-PSBE      | 0,6 | 4,6  | 0,3  | 94,0  | 0,1 | 0,2 | 0,1  | 0,2 |
| 46928+TRNF      | 0,3 | 0,1  | 6,4  | 92,9  | 0,0 | 0,0 | 0,1  | 0,1 |
| 45454-YCF3      | 1,2 | 4,7  | 0,6  | 92,8  | 0,1 | 0,3 | 0,1  | 0,3 |
| 114261+TRNL.3   | 0,1 | 7,3  | 0,1  | 92,4  | 0,0 | 0,0 | 0,0  | 0,1 |
| 46589+TRNL.1    | 2,2 | 3,5  | 1,5  | 91,3  | 0,2 | 0,4 | 0,3  | 0,6 |
| 110341-TRNN.1   | 1,6 | 6,1  | 0,8  | 89,9  | 0,2 | 0,5 | 0,2  | 0,6 |
| 134174-RRN23S.2 | 0,1 | 10,8 | 0,0  | 89,0  | 0,0 | 0,0 | 0,0  | 0,1 |
| 66089+PSAJ      | 1,3 | 9,4  | 0,5  | 87,6  | 0,1 | 0,4 | 0,2  | 0,5 |
| 82650-RPL14     | 2,4 | 1,3  | 6,4  | 87,6  | 0,3 | 0,4 | 0,7  | 0,9 |
| 76780+YCF2.1    | 3,7 | 3,3  | 3,3  | 87,2  | 0,3 | 0,6 | 0,6  | 1,1 |
| 118899-NDHE     | 3,7 | 3,4  | 3,4  | 85,9  | 0,5 | 0,9 | 0,9  | 1,4 |
| 46339-TRNT.2    | 0,1 | 14,2 | 0,0  | 85,5  | 0,0 | 0,0 | 0,0  | 0,1 |

## Anhang

|                 |      |      |      |      |      |      |      |      |
|-----------------|------|------|------|------|------|------|------|------|
| 42045-PSAA      | 0,3  | 0,1  | 13,9 | 85,5 | 0,0  | 0,0  | 0,1  | 0,2  |
| 152438+RPL23.2  | 4,8  | 4,7  | 3,7  | 83,1 | 0,5  | 0,9  | 0,8  | 1,5  |
| 1522-PSBA       | 0,3  | 16,3 | 0,1  | 82,9 | 0,0  | 0,1  | 0,0  | 0,2  |
| 63131-ATPB      | 4,8  | 5,5  | 3,1  | 82,7 | 0,5  | 1,0  | 0,8  | 1,5  |
| 8230+TRNG.1     | 1,3  | 17,3 | 0,5  | 79,6 | 0,2  | 0,4  | 0,2  | 0,6  |
| 61105+PETA      | 3,3  | 5,4  | 1,9  | 79,4 | 0,8  | 1,1  | 0,9  | 7,1  |
| 104475+RRN23S.1 | 0,1  | 22,1 | 0,0  | 77,7 | 0,0  | 0,0  | 0,0  | 0,1  |
| 44755-YCF3      | 6,1  | 8,6  | 3,2  | 77,0 | 0,7  | 1,5  | 1,0  | 1,9  |
| 31319+TRNT.1    | 0,3  | 22,8 | 0,1  | 76,6 | 0,0  | 0,1  | 0,0  | 0,1  |
| 8643+TRNG.1     | 0,1  | 24,4 | 0,0  | 75,4 | 0,0  | 0,0  | 0,0  | 0,1  |
| 31614+PSBD      | 0,6  | 24,2 | 0,2  | 74,3 | 0,1  | 0,2  | 0,1  | 0,3  |
| 29912-TRND      | 9,5  | 6,8  | 6,8  | 73,6 | 0,6  | 0,7  | 0,7  | 1,4  |
| 100898+RRN16S.1 | 0,1  | 27,4 | 0,0  | 72,3 | 0,0  | 0,0  | 0,0  | 0,1  |
| 78010+YCF2.1    | 0,4  | 29,6 | 0,1  | 69,5 | 0,0  | 0,1  | 0,0  | 0,2  |
| 54671-ATPB      | 2,3  | 26,7 | 0,8  | 68,1 | 0,3  | 0,7  | 0,3  | 0,8  |
| 47545+TRNF      | 0,6  | 30,5 | 0,2  | 67,9 | 0,1  | 0,2  | 0,1  | 0,4  |
| 53093-ATPE      | 10,3 | 12,1 | 4,8  | 66,0 | 1,0  | 1,9  | 1,1  | 2,9  |
| 62285-ATPB      | 8,2  | 14,5 | 3,4  | 65,4 | 1,4  | 2,4  | 1,6  | 3,2  |
| 137751-RRN16S.2 | 0,3  | 34,5 | 0,1  | 64,8 | 0,0  | 0,1  | 0,0  | 0,2  |
| 86211-RPL23     | 11,0 | 9,7  | 5,8  | 63,8 | 1,5  | 2,6  | 2,1  | 3,4  |
| 51885-TRNV.1    | 0,1  | 56,9 | 0,0  | 42,7 | 0,0  | 0,1  | 0,0  | 0,1  |
| 8213+TRNG.1     | 8,7  | 56,4 | 2,5  | 23,7 | 1,6  | 2,9  | 1,0  | 3,1  |
| 47703+TRNF      | 0,3  | 55,1 | 0,1  | 44,1 | 0,0  | 0,1  | 0,0  | 0,2  |
| 114944+TRNN     | 5,7  | 52,5 | 1,7  | 35,7 | 0,7  | 1,6  | 0,4  | 1,7  |
| 28037+YCF6      | 6,6  | 52,3 | 1,9  | 32,4 | 1,2  | 2,1  | 0,9  | 2,6  |
| 2735-PSBA       | 0,2  | 49,3 | 0,1  | 50,2 | 0,0  | 0,1  | 0,0  | 0,1  |
| 112953+RPL32    | 3,2  | 47,8 | 1,0  | 44,4 | 0,6  | 1,1  | 0,5  | 1,5  |
| 4451-TRNK       | 0,2  | 43,3 | 0,1  | 56,3 | 0,0  | 0,1  | 0,0  | 0,1  |
| 46751+TRNL.1    | 0,5  | 42,1 | 0,2  | 56,7 | 0,1  | 0,2  | 0,0  | 0,2  |
| 137776-RRN16S.2 | 0,6  | 0,2  | 0,1  | 0,0  | 98,0 | 0,8  | 0,2  | 0,2  |
| 100873+RRN16S.1 | 0,4  | 0,1  | 0,1  | 0,0  | 98,4 | 0,8  | 0,2  | 0,1  |
| 8126+TRNG.1     | 0,9  | 0,9  | 0,2  | 0,3  | 1,1  | 80,7 | 0,4  | 15,5 |
| 32672+PSBD      | 13,4 | 28,7 | 4,3  | 41,0 | 2,3  | 3,9  | 2,1  | 4,4  |
| 118171-NDHE     | 32,1 | 24,3 | 9,3  | 21,2 | 3,6  | 5,8  | 1,5  | 2,2  |
| 71416-RPS12     | 12,8 | 23,8 | 4,4  | 49,0 | 1,6  | 3,4  | 1,3  | 3,7  |
| 36324+TRNG.2    | 5,7  | 22,5 | 1,7  | 18,9 | 1,7  | 8,7  | 1,7  | 39,1 |
| 118082-PSAC     | 30,8 | 20,3 | 10,0 | 30,9 | 2,4  | 3,2  | 1,2  | 1,2  |
| 13921-ATPH      | 27,2 | 8,9  | 8,0  | 9,4  | 8,9  | 11,5 | 9,4  | 16,7 |
| 20245-RPOC2     | 30,0 | 8,8  | 15,5 | 16,0 | 6,7  | 5,8  | 8,5  | 8,7  |
| 46499+TRNL.1    | 28,3 | 8,2  | 13,6 | 13,2 | 8,0  | 8,0  | 9,6  | 11,1 |
| 71940-CLPP1     | 16,2 | 7,1  | 11,0 | 37,8 | 4,2  | 4,7  | 5,1  | 13,8 |
| 11392-TRNQ      | 18,2 | 6,2  | 21,6 | 41,8 | 2,1  | 1,6  | 4,0  | 4,6  |
| 80625-RPS8      | 22,9 | 6,1  | 19,6 | 8,1  | 6,8  | 6,1  | 19,9 | 10,6 |
| 59146+PSAI      | 9,6  | 5,2  | 7,3  | 45,5 | 2,6  | 3,1  | 3,5  | 23,1 |
| 112979+RPL32    | 17,7 | 4,1  | 40,6 | 6,1  | 5,1  | 2,7  | 17,7 | 5,9  |

Tabelle VI.12: Klassifizierung mit dem statistischen Modell im späten Stadium. Die Zugehörigkeit der TSS zu einem Fall ist in Prozent angegeben. Rottöne: hohe Wahrscheinlichkeit für diesen Fall (60-100 %), Blautöne: niedrige Wahrscheinlichkeit für diesen Fall (0-59 %). Pol.: Polymerase. . Die Analyse wurde von Frederic Ehlers unter Anleitung von Julia Rott und Ivo Große durchgeführt.

| TSS abhängig von | keine Pol | PEP    | <i>rpoTp</i> | PEP/<br><i>rpoTp</i> | <i>rpoTmp</i> | PEP/<br><i>rpoTmp</i> | <i>rpoTp</i> /<br><i>rpoTmp</i> | alle Pols |
|------------------|-----------|--------|--------------|----------------------|---------------|-----------------------|---------------------------------|-----------|
| TSS              | Fall 1    | Fall 2 | Fall 3       | Fall 4               | Fall 5        | Fall 6                | Fall 7                          | Fall 8    |
| 54781+RBCL       | 0,0       | 100,0  | 0,0          | 0,0                  | 0,0           | 0,0                   | 0,0                             | 0,0       |
| 78010+YCF2.1     | 0,0       | 100,0  | 0,0          | 0,0                  | 0,0           | 0,0                   | 0,0                             | 0,0       |
| 31523-TRNE       | 0,0       | 100,0  | 0,0          | 0,0                  | 0,0           | 0,0                   | 0,0                             | 0,0       |

## Anhang

|                 |     |       |     |      |     |     |     |     |
|-----------------|-----|-------|-----|------|-----|-----|-----|-----|
| 33526+PSBC      | 0,0 | 100,0 | 0,0 | 0,0  | 0,0 | 0,0 | 0,0 | 0,0 |
| 53093-ATPE      | 0,0 | 99,9  | 0,0 | 0,0  | 0,0 | 0,0 | 0,0 | 0,0 |
| 112771-NDHF     | 0,0 | 99,9  | 0,0 | 0,0  | 0,0 | 0,0 | 0,0 | 0,0 |
| 36849-TRNFM     | 0,0 | 99,9  | 0,0 | 0,0  | 0,0 | 0,0 | 0,0 | 0,0 |
| 46751+TRNL.1    | 0,0 | 99,9  | 0,0 | 0,0  | 0,0 | 0,0 | 0,0 | 0,0 |
| 61105+PETA      | 0,0 | 99,9  | 0,0 | 0,0  | 0,0 | 0,0 | 0,0 | 0,0 |
| 4451-TRNK       | 0,0 | 99,9  | 0,0 | 0,0  | 0,0 | 0,0 | 0,0 | 0,0 |
| 2735-PSBA       | 0,0 | 99,8  | 0,0 | 0,1  | 0,0 | 0,0 | 0,0 | 0,0 |
| 45454-YCF3      | 0,0 | 99,8  | 0,0 | 0,2  | 0,0 | 0,0 | 0,0 | 0,0 |
| 114944+TRNN     | 0,0 | 99,8  | 0,0 | 0,1  | 0,0 | 0,1 | 0,0 | 0,0 |
| 7539+PSBI       | 0,0 | 99,7  | 0,0 | 0,1  | 0,0 | 0,1 | 0,0 | 0,0 |
| 74413-PSBN      | 0,0 | 99,7  | 0,0 | 0,2  | 0,0 | 0,1 | 0,0 | 0,0 |
| 72200+PSBB      | 0,0 | 99,7  | 0,0 | 0,2  | 0,0 | 0,0 | 0,0 | 0,0 |
| 8643+TRNG.1     | 0,0 | 99,7  | 0,0 | 0,2  | 0,0 | 0,0 | 0,0 | 0,0 |
| 31336+TRNT.1    | 0,1 | 99,7  | 0,0 | 0,2  | 0,0 | 0,0 | 0,0 | 0,0 |
| 100687+TRNV.2   | 0,0 | 99,6  | 0,0 | 0,2  | 0,0 | 0,1 | 0,0 | 0,0 |
| 54618-ATPB      | 0,0 | 99,5  | 0,0 | 0,5  | 0,0 | 0,0 | 0,0 | 0,0 |
| 137962-TRNV.3   | 0,1 | 99,4  | 0,0 | 0,4  | 0,0 | 0,1 | 0,0 | 0,0 |
| 109231-TRNN.1   | 0,1 | 99,3  | 0,0 | 0,5  | 0,0 | 0,0 | 0,0 | 0,0 |
| 29204-PSBM      | 0,0 | 99,3  | 0,0 | 0,5  | 0,0 | 0,1 | 0,0 | 0,0 |
| 118171-NDHE     | 0,1 | 99,3  | 0,0 | 0,1  | 0,0 | 0,3 | 0,0 | 0,0 |
| 6708-TRNQ       | 0,2 | 98,9  | 0,0 | 0,3  | 0,0 | 0,4 | 0,0 | 0,1 |
| 51885-TRNV.1    | 0,0 | 98,8  | 0,0 | 1,1  | 0,0 | 0,0 | 0,0 | 0,0 |
| 134174-RRN23S.2 | 0,0 | 98,7  | 0,0 | 1,1  | 0,0 | 0,0 | 0,0 | 0,0 |
| 129418+TRNN.2   | 0,2 | 98,6  | 0,0 | 1,0  | 0,0 | 0,1 | 0,0 | 0,1 |
| 104475+RRN23S.1 | 0,0 | 98,4  | 0,0 | 1,5  | 0,0 | 0,0 | 0,0 | 0,0 |
| 49371+TRNM      | 0,0 | 98,2  | 0,0 | 1,7  | 0,0 | 0,0 | 0,0 | 0,0 |
| 66862-TRNP      | 0,1 | 98,1  | 0,0 | 1,8  | 0,0 | 0,0 | 0,0 | 0,0 |
| 8126+TRNG.1     | 0,5 | 97,9  | 0,1 | 0,3  | 0,1 | 1,0 | 0,0 | 0,1 |
| 28037+YCF6      | 0,3 | 97,8  | 0,1 | 0,2  | 0,1 | 1,4 | 0,0 | 0,1 |
| 76780+YCF2.1    | 0,3 | 97,8  | 0,1 | 0,9  | 0,1 | 0,6 | 0,0 | 0,2 |
| 63723-PSBJ      | 0,3 | 96,9  | 0,1 | 1,9  | 0,1 | 0,5 | 0,0 | 0,3 |
| 1522-PSBA       | 0,1 | 95,1  | 0,0 | 4,6  | 0,0 | 0,1 | 0,0 | 0,1 |
| 112957-NDHF     | 0,7 | 95,1  | 0,2 | 2,0  | 0,2 | 1,4 | 0,1 | 0,6 |
| 37271+TRNS      | 1,2 | 94,9  | 0,3 | 2,2  | 0,2 | 0,6 | 0,1 | 0,4 |
| 66894+PSAJ      | 0,6 | 94,6  | 0,1 | 1,3  | 0,2 | 2,4 | 0,1 | 0,6 |
| 42045-PSAA      | 0,6 | 94,6  | 0,2 | 4,0  | 0,1 | 0,3 | 0,0 | 0,2 |
| 46589+TRNL.1    | 1,7 | 93,2  | 0,4 | 2,6  | 0,3 | 1,1 | 0,1 | 0,6 |
| 54671-ATPB      | 0,9 | 91,8  | 0,2 | 1,9  | 0,2 | 3,7 | 0,1 | 1,1 |
| 62285-ATPB      | 0,0 | 90,1  | 0,0 | 9,7  | 0,0 | 0,0 | 0,0 | 0,1 |
| 113004-NDHF     | 2,3 | 89,6  | 0,5 | 2,6  | 0,6 | 3,1 | 0,2 | 1,0 |
| 67744-TRNP      | 1,8 | 89,1  | 0,5 | 6,0  | 0,4 | 1,3 | 0,2 | 0,9 |
| 52012+TRNM      | 1,5 | 88,8  | 0,4 | 3,4  | 0,4 | 4,0 | 0,2 | 1,3 |
| 100898+RRN16S.1 | 3,3 | 88,8  | 0,7 | 2,7  | 0,7 | 2,6 | 0,2 | 1,0 |
| 144275+TRNL.4   | 3,9 | 85,8  | 1,0 | 6,2  | 0,6 | 1,3 | 0,2 | 0,9 |
| 86417-TRNL.1    | 1,3 | 84,0  | 0,4 | 11,5 | 0,3 | 1,1 | 0,2 | 1,1 |
| 46110+TRNL.1    | 3,1 | 83,0  | 0,8 | 6,8  | 0,7 | 3,2 | 0,4 | 1,9 |
| 86068-RPL2.1    | 0,8 | 82,5  | 0,2 | 14,9 | 0,2 | 0,6 | 0,1 | 0,7 |
| 78677+YCF2.1    | 0,6 | 82,3  | 0,2 | 15,5 | 0,1 | 0,5 | 0,1 | 0,8 |
| 137751-RRN16S.2 | 5,4 | 82,1  | 1,3 | 4,2  | 1,2 | 3,7 | 0,4 | 1,6 |
| 152581+RPL2.2   | 0,5 | 81,9  | 0,1 | 16,4 | 0,1 | 0,3 | 0,1 | 0,5 |
| 30565-TRNE      | 0,5 | 80,8  | 0,1 | 17,9 | 0,1 | 0,2 | 0,0 | 0,2 |
| 36665+TRNS      | 7,1 | 78,7  | 1,6 | 4,3  | 1,6 | 4,3 | 0,4 | 1,9 |
| 152232+TRNL.4   | 1,3 | 76,9  | 0,4 | 19,4 | 0,3 | 0,7 | 0,2 | 0,8 |
| 71416-RPS12     | 3,7 | 76,6  | 1,0 | 9,2  | 0,9 | 4,7 | 0,6 | 3,3 |
| 28988-PSBM      | 2,5 | 75,7  | 0,7 | 6,3  | 0,7 | 8,4 | 0,5 | 5,1 |
| 118899-NDHE     | 4,2 | 68,6  | 1,2 | 17,4 | 1,0 | 3,4 | 0,7 | 3,5 |

## Anhang

|                 |      |      |       |      |      |      |      |      |
|-----------------|------|------|-------|------|------|------|------|------|
| 36372+TRNG.2    | 8,5  | 68,3 | 2,2   | 10,1 | 1,9  | 4,8  | 1,0  | 3,1  |
| 66089+PSAJ      | 5,6  | 64,2 | 1,6   | 19,6 | 1,3  | 3,4  | 0,9  | 3,5  |
| 67995-TRNP      | 4,0  | 57,1 | 1,1   | 25,1 | 1,0  | 4,2  | 0,9  | 6,5  |
| 118082-PSAC     | 1,2  | 54,8 | 0,3   | 40,9 | 0,2  | 0,8  | 0,2  | 1,5  |
| 44234+TRNS.3    | 1,8  | 48,3 | 0,5   | 47,4 | 0,3  | 0,6  | 0,2  | 0,9  |
| 8230+TRNG.1     | 11,9 | 46,4 | 3,1   | 8,6  | 3,3  | 16,5 | 2,5  | 7,8  |
| 152438+RPL23.2  | 20,2 | 42,5 | 5,1   | 10,4 | 4,9  | 8,7  | 2,4  | 5,8  |
| 6253-RPS16      | 17,2 | 42,4 | 4,3   | 8,6  | 4,6  | 13,7 | 2,9  | 6,4  |
| 144130+TRNL.4   | 3,6  | 41,6 | 1,0   | 50,8 | 0,5  | 1,0  | 0,3  | 1,1  |
| 47703+TRNF      | 23,1 | 40,1 | 5,2   | 7,9  | 5,7  | 9,9  | 1,9  | 6,0  |
| 86211-RPL23     | 21,1 | 38,7 | 5,6   | 12,8 | 5,0  | 8,0  | 2,8  | 6,1  |
| 128308+TRNN.2   | 17,0 | 31,3 | 4,8   | 30,9 | 3,2  | 5,0  | 2,6  | 5,3  |
| 153604+PSBK     | 0,0  | 0,0  | 100,0 | 0,0  | 0,0  | 0,0  | 0,0  | 0,0  |
| 85045-RPS19     | 0,0  | 0,0  | 99,9  | 0,0  | 0,0  | 0,0  | 0,0  | 0,0  |
| 146051+TRNI     | 0,7  | 0,1  | 97,1  | 0,3  | 0,2  | 0,0  | 1,4  | 0,1  |
| 92598-TRNI      | 1,8  | 0,4  | 94,1  | 1,2  | 0,4  | 0,1  | 1,5  | 0,5  |
| 116996-NDHF     | 0,5  | 5,7  | 0,2   | 88,0 | 0,1  | 0,4  | 0,1  | 4,9  |
| 32672+PSBD      | 0,5  | 22,2 | 0,1   | 74,6 | 0,1  | 0,4  | 0,1  | 2,0  |
| 44824+TRNS.3    | 9,1  | 7,6  | 4,9   | 63,6 | 2,0  | 2,7  | 2,5  | 7,6  |
| 110032-TRNN.1   | 9,9  | 10,6 | 4,2   | 61,2 | 2,1  | 2,9  | 2,6  | 6,5  |
| 94519-TRNL.2    | 3,7  | 33,8 | 1,1   | 58,5 | 0,5  | 1,0  | 0,3  | 1,1  |
| 82650-RPL14     | 6,0  | 25,3 | 1,9   | 57,7 | 1,2  | 2,2  | 1,3  | 4,3  |
| 128616+TRNN.2   | 9,8  | 5,3  | 7,4   | 57,1 | 2,4  | 2,8  | 2,9  | 12,4 |
| 78986-RPOA      | 12,7 | 26,7 | 3,8   | 43,6 | 2,4  | 3,8  | 2,3  | 4,8  |
| 35410-TRNE      | 17,5 | 26,9 | 5,0   | 34,2 | 3,7  | 5,3  | 3,4  | 4,1  |
| 11392-TRNQ      | 21,5 | 15,8 | 7,2   | 35,8 | 3,8  | 5,7  | 4,0  | 6,1  |
| 100873+RRN16S.1 | 0,6  | 0,1  | 0,1   | 0,0  | 98,6 | 0,2  | 0,2  | 0,1  |
| 137776-RRN16S.2 | 0,8  | 0,1  | 0,1   | 0,0  | 98,3 | 0,3  | 0,3  | 0,1  |
| 112979+RPL32    | 27,5 | 3,4  | 7,1   | 0,5  | 41,6 | 7,1  | 11,5 | 1,3  |
| 56771-ATPB      | 14,8 | 1,0  | 22,8  | 2,3  | 5,2  | 0,6  | 51,1 | 2,1  |
| 7872-TRNQ       | 15,6 | 2,1  | 29,6  | 4,2  | 4,8  | 1,7  | 39,4 | 2,5  |
| 46339-TRNT.2    | 42,8 | 11,0 | 10,8  | 5,3  | 11,5 | 7,4  | 7,5  | 3,7  |
| 44755-YCF3      | 41,9 | 11,2 | 8,3   | 2,8  | 12,6 | 11,6 | 8,0  | 3,5  |
| 129283-TIC214   | 39,4 | 3,4  | 15,7  | 3,3  | 11,8 | 1,8  | 19,0 | 5,7  |
| 66577-TRNP      | 38,1 | 10,5 | 9,9   | 8,3  | 10,1 | 8,8  | 7,1  | 7,1  |
| 114114-NDHF     | 37,8 | 10,5 | 10,9  | 10,9 | 9,9  | 9,0  | 10,1 | 0,9  |
| 109366+TIC214   | 37,7 | 3,7  | 15,3  | 3,8  | 10,9 | 1,9  | 17,2 | 9,5  |
| 31614+PSBD      | 37,6 | 9,6  | 10,4  | 7,9  | 10,4 | 8,8  | 10,4 | 5,0  |
| 46499+TRNL.1    | 36,8 | 9,8  | 9,4   | 5,6  | 12,2 | 10,2 | 10,1 | 5,9  |
| 63131-ATPB      | 35,9 | 9,2  | 10,1  | 9,1  | 9,9  | 7,5  | 10,0 | 8,4  |
| 29912-TRND      | 35,5 | 9,7  | 9,8   | 9,0  | 10,3 | 9,8  | 9,9  | 6,0  |
| 80625-RPS8      | 35,3 | 5,2  | 10,1  | 3,5  | 14,3 | 6,4  | 15,4 | 9,7  |
| 66382-TRNW      | 35,2 | 9,8  | 9,5   | 9,1  | 9,6  | 9,3  | 8,1  | 9,5  |
| 122076-NDHA     | 34,7 | 9,7  | 9,2   | 9,0  | 9,9  | 10,2 | 9,3  | 8,0  |
| 112358+RPL32    | 34,1 | 2,2  | 18,9  | 2,7  | 10,7 | 1,1  | 26,8 | 3,5  |
| 59146+PSAI      | 33,9 | 11,6 | 8,7   | 9,4  | 9,4  | 10,5 | 7,5  | 9,0  |
| 114261+TRNL.3   | 33,4 | 13,4 | 9,2   | 11,1 | 8,3  | 9,3  | 6,6  | 8,6  |
| 46585-TRNT.2    | 33,2 | 20,2 | 8,7   | 11,0 | 7,7  | 9,4  | 3,9  | 5,9  |
| 36358+TRNG.2    | 33,1 | 12,7 | 8,8   | 9,5  | 9,1  | 10,4 | 7,4  | 9,0  |
| 31319+TRNT.1    | 32,6 | 15,3 | 9,0   | 12,9 | 7,4  | 9,0  | 5,7  | 8,0  |
| 112953+RPL32    | 31,9 | 14,2 | 6,4   | 8,2  | 9,0  | 15,9 | 5,5  | 8,8  |
| 64448-PSBE      | 31,4 | 8,7  | 12,6  | 11,5 | 8,5  | 8,3  | 9,4  | 9,6  |
| 47545+TRNF      | 31,1 | 22,7 | 6,6   | 8,7  | 8,4  | 13,2 | 3,6  | 5,9  |
| 8213+TRNG.1     | 30,6 | 15,7 | 7,0   | 8,5  | 8,5  | 15,6 | 6,1  | 8,0  |
| 36324+TRNG.2    | 30,0 | 16,8 | 8,1   | 10,9 | 8,1  | 10,6 | 6,7  | 8,8  |
| 13921-ATPH      | 29,7 | 11,6 | 7,7   | 8,4  | 9,0  | 15,8 | 8,1  | 9,7  |
| 32644+PSBD      | 28,4 | 4,9  | 16,8  | 7,3  | 9,0  | 4,8  | 24,5 | 4,3  |

## Anhang

|               |      |      |      |      |      |     |      |     |
|---------------|------|------|------|------|------|-----|------|-----|
| 71940-CLPP1   | 28,0 | 6,0  | 11,9 | 7,5  | 10,0 | 7,1 | 21,1 | 8,3 |
| 20245-RPOC2   | 27,4 | 3,3  | 12,0 | 4,1  | 12,0 | 4,1 | 30,9 | 6,0 |
| 110341-TRNN.1 | 27,4 | 14,4 | 8,4  | 22,0 | 6,0  | 7,6 | 6,2  | 8,1 |
| 46928+TRNF    | 23,5 | 28,8 | 6,5  | 19,4 | 5,0  | 6,9 | 3,5  | 6,3 |

Tabelle VI.13: Klassifizierung der TSS mit pamK, angezeigt sind logarithmierte relative Transkriptlevel bezogen auf WT, normalisiert mit DESeq2. Die Analyse wurde von Frederic Ehlers unter Anleitung von Julia Rott und Dermot Harnett durchgeführt.

| TSS          | frühes Stadium |               |              |         | spätes Stadium |               |              |         |
|--------------|----------------|---------------|--------------|---------|----------------|---------------|--------------|---------|
|              | <i>rpoTp</i>   | <i>rpoTmp</i> | <i>clb19</i> | Cluster | <i>rpoTp</i>   | <i>rpoTmp</i> | <i>ptac2</i> | Cluster |
| 1522-PSBA    | -0,42          | -0,01         | -1,46        | 1       | -0,12          | 0,40          | -0,86        | 1       |
| 2735-PSBA    | -0,93          | -0,22         | -4,82        | 2       | 0,67           | 0,41          | -2,19        | 2       |
| 4451-TRNK    | -0,67          | 0,41          | -2,69        | 1       | 1,71           | 1,34          | -0,14        | 3       |
| 6253-RPS16   | -0,25          | -0,04         | -1,42        | 1       | 1,25           | 0,34          | 0,98         | 4       |
| 6708-TRNQ    | -1,97          | 0,75          | -4,08        | 2       | 1,50           | 0,53          | -1,48        | 2       |
| 7539+PSBI    | -0,17          | 0,54          | -1,07        | 3       | 1,69           | 0,79          | -2,19        | 2       |
| 7872-TRNQ    | -0,40          | 0,45          | 2,57         | 4       | -1,18          | -0,93         | 4,19         | 5       |
| 8126+TRNG.1  | 0,07           | -3,48         | -0,83        | 4       | 2,50           | 0,90          | 0,45         | 3       |
| 8213+TRNG.1  | 0,56           | -0,25         | 0,03         | 3       | 1,82           | 0,45          | 2,14         | 4       |
| 8230+TRNG.1  | -0,08          | -0,25         | -0,77        | 3       | 0,89           | -0,22         | 0,87         | 1       |
| 8643+TRNG.1  | -0,96          | -0,51         | -2,76        | 1       | 0,78           | 0,89          | -0,67        | 1       |
| 11392-TRNQ   | -0,51          | 0,78          | 0,59         | 3       | 0,59           | 1,56          | 1,61         | 4       |
| 13921-ATPH   | 0,06           | -2,41         | 0,14         | 4       | 2,36           | 0,25          | 2,11         | 4       |
| 20245-RPOC2  | 0,44           | -0,18         | 2,26         | 4       | 1,17           | -0,02         | 4,18         | 6       |
| 28037+YCF6   | 0,14           | -0,22         | -0,74        | 3       | 2,61           | 0,31          | -1,77        | 2       |
| 28988-PSBM   | -1,18          | -0,03         | -3,02        | 1       | 0,09           | -0,75         | -0,56        | 1       |
| 29204-PSBM   | 0,05           | 0,09          | -2,85        | 1       | -0,05          | 0,28          | -2,03        | 2       |
| 29912-TRND   | -0,33          | 3,45          | -0,01        | 3       | 1,45           | 0,42          | 3,20         | 4       |
| 30565-TRNE   | -0,48          | 0,27          | -2,66        | 1       | 0,92           | 1,38          | 0,60         | 3       |
| 31319+TRNT.1 | -0,44          | 1,03          | -2,20        | 1       | 1,30           | 1,41          | 2,10         | 4       |
| 31336+TRNT.1 | -1,39          | 0,66          | -3,26        | 2       | 1,68           | 1,75          | -1,33        | 3       |
| 31523-TRNE   | -1,59          | -0,28         | -5,11        | 2       | 1,04           | 0,62          | -4,44        | 2       |
| 31614+PSBD   | -0,43          | 0,48          | -2,49        | 1       | 2,29           | 1,35          | 5,22         | 6       |
| 32644+PSBD   | -4,81          | 0,01          | 2,12         | 5       | 0,77           | 0,27          | 3,25         | 4       |
| 32672+PSBD   | -0,24          | -0,09         | -2,02        | 1       | -0,50          | -0,18         | -0,65        | 1       |
| 33526+PSBC   | -0,32          | 0,37          | -2,57        | 1       | 0,70           | 0,41          | -1,39        | 2       |
| 35410-TRNE   | -1,94          | 0,93          | -1,24        | 6       | 0,64           | 1,49          | 1,52         | 4       |
| 36324+TRNG.2 | -0,72          | -2,26         | -1,37        | 1       | 1,73           | 1,07          | 2,01         | 4       |
| 36358+TRNG.2 | -0,37          | 0,53          | -2,18        | 1       | 1,83           | 0,88          | 2,59         | 4       |
| 36372+TRNG.2 | -0,38          | 0,36          | -3,49        | 1       | 1,14           | 0,75          | 0,97         | 4       |
| 36665+TRNS   | 1,07           | -0,34         | -1,84        | 1       | 2,24           | 1,57          | 1,58         | 4       |
| 36849-TRNFM  | -1,18          | 1,13          | -2,47        | 1       | 1,34           | 1,14          | -1,69        | 2       |
| 37271+TRNS   | 0,08           | 0,96          | -3,71        | 1       | 1,57           | 1,62          | -0,51        | 3       |
| 42045-PSAA   | -1,89          | -0,10         | 0,32         | 6       | 1,17           | 1,61          | 0,59         | 3       |
| 44234+TRNS.3 | -1,32          | -0,06         | -0,39        | 6       | 0,24           | 1,44          | -0,61        | 1       |
| 44755-YCF3   | -0,07          | 0,58          | 0,61         | 3       | 3,11           | 1,32          | 3,76         | 6       |
| 44824+TRNS.3 | -0,61          | 0,39          | -0,46        | 3       | -0,53          | 0,37          | 1,15         | 1       |
| 45454-YCF3   | -0,24          | 1,74          | -0,75        | 3       | 0,33           | 0,84          | -3,02        | 2       |
| 46110+TRNL.1 | -0,44          | -0,12         | -3,36        | 1       | 0,83           | 0,44          | 0,59         | 1       |
| 46339-TRNT.2 | -0,65          | 1,23          | -4,17        | 2       | 4,00           | 3,10          | 6,05         | 6       |
| 46499+TRNL.1 | -0,13          | -1,06         | 1,42         | 4       | 3,15           | 0,45          | 3,56         | 6       |
| 46585-TRNT.2 | -1,90          | 0,28          | -5,05        | 2       | 3,84           | 2,25          | 1,70         | 4       |
| 46589+TRNL.1 | 0,24           | 1,67          | -0,27        | 3       | 1,52           | 1,22          | -0,11        | 3       |
| 46751+TRNL.1 | -0,45          | 1,09          | -2,09        | 1       | 2,10           | 1,77          | -0,90        | 3       |
| 46928+TRNF   | -1,63          | 0,98          | 0,30         | 6       | 2,04           | 1,70          | 1,68         | 4       |

## Anhang

|                 |       |       |       |   |       |       |       |   |
|-----------------|-------|-------|-------|---|-------|-------|-------|---|
| 47545+TRNF      | -1,40 | -0,45 | -4,45 | 2 | 3,52  | 1,49  | 1,86  | 4 |
| 47703+TRNF      | -0,79 | -0,10 | -2,94 | 1 | 3,96  | 2,42  | 2,68  | 4 |
| 49371+TRNM      | -1,71 | 0,82  | -2,23 | 6 | 0,38  | 1,77  | -3,48 | 2 |
| 51885-TRNV.1    | -1,26 | -0,13 | -4,02 | 2 | 0,21  | 0,85  | -1,42 | 2 |
| 52012+TRNM      | -2,11 | 1,08  | -5,23 | 2 | 0,66  | -0,33 | -0,99 | 1 |
| 53093-ATPE      | -0,06 | 1,03  | -0,68 | 3 | 0,70  | 0,71  | -2,95 | 2 |
| 54618-ATPB      | -0,49 | 1,15  | -0,76 | 3 | -0,35 | 0,94  | -4,60 | 2 |
| 54671-ATPB      | 0,16  | 0,08  | -0,70 | 3 | 0,59  | -0,52 | -0,67 | 1 |
| 54781+RBCL      | -0,36 | -0,63 | -2,10 | 1 | 0,25  | 0,49  | -3,49 | 2 |
| 56771-ATPB      | -1,88 | -1,63 | 2,77  | 4 | 0,39  | 0,83  | 5,60  | 6 |
| 59146+PSAI      | -1,53 | -1,90 | -0,08 | 4 | 3,36  | 1,87  | 2,78  | 4 |
| 61105+PETA      | -1,16 | -1,74 | -2,32 | 1 | 1,49  | 0,64  | -2,35 | 2 |
| 62285-ATPB      | -0,78 | -0,46 | -1,94 | 1 | -0,80 | 0,42  | -3,17 | 2 |
| 63131-ATPB      | -0,08 | 0,48  | 0,41  | 3 | 1,82  | 1,31  | 4,08  | 6 |
| 63723-PSBJ      | -0,73 | -0,88 | -1,42 | 1 | 0,65  | 0,27  | -0,68 | 1 |
| 64448-PSBE      | -0,39 | 1,00  | -1,02 | 3 | 0,42  | 0,71  | 3,03  | 4 |
| 66089+PSAJ      | -0,02 | 0,48  | -0,30 | 3 | 0,69  | 0,54  | 0,06  | 1 |
| 66382-TRNW      | -3,26 | 1,46  | -1,08 | 6 | 2,61  | 1,53  | 3,94  | 6 |
| 66577-TRNP      | -0,83 | -0,07 | -4,08 | 2 | 3,31  | 2,96  | 4,76  | 6 |
| 66862-TRNP      | -1,11 | 1,16  | -3,00 | 1 | 0,08  | 1,14  | -2,19 | 2 |
| 66894+PSAJ      | -0,85 | 0,39  | -1,78 | 1 | 0,86  | 0,14  | -0,49 | 1 |
| 67744-TRNP      | -1,55 | 0,19  | -2,16 | 6 | 0,60  | 0,61  | -0,56 | 1 |
| 67995-TRNP      | -0,60 | 0,24  | -2,55 | 1 | 0,35  | 0,08  | 0,15  | 1 |
| 71416-RPS12     | 0,01  | 0,32  | -0,82 | 3 | 0,87  | 0,02  | 0,45  | 1 |
| 71940-CLPP1     | -0,20 | -1,23 | 0,69  | 4 | 0,51  | -0,36 | 3,29  | 4 |
| 72200+PSBB      | -0,51 | -0,01 | -2,39 | 1 | 0,00  | -0,31 | -2,78 | 2 |
| 74413-PSBN      | -1,15 | 0,97  | -1,50 | 6 | 0,48  | 0,14  | -1,81 | 2 |
| 76780+YCF2.1    | -1,05 | 1,01  | -0,73 | 6 | 0,55  | 0,09  | -1,00 | 1 |
| 78010+YCF2.1    | -0,68 | 1,37  | -2,70 | 1 | 1,84  | 0,94  | -1,74 | 2 |
| 78677+YCF2.1    | -1,13 | -1,01 | -3,16 | 2 | -0,47 | -0,04 | -2,03 | 2 |
| 78986-RPOA      | -1,69 | 0,31  | -1,70 | 6 | 0,60  | 1,22  | 1,53  | 4 |
| 80625-RPS8      | -0,91 | -1,39 | 1,39  | 4 | 1,71  | -0,03 | 4,68  | 6 |
| 82650-RPL14     | -1,20 | -0,01 | -0,36 | 6 | 0,24  | 0,26  | 0,40  | 1 |
| 85045-RPS19     | -2,36 | 0,40  | 1,98  | 4 | -2,38 | 1,58  | 4,24  | 5 |
| 86068-RPL2.1    | -0,35 | -0,06 | -2,92 | 1 | 0,18  | 0,51  | -0,42 | 1 |
| 86211-RPL23     | -0,01 | 0,42  | 0,72  | 3 | 1,44  | 1,17  | 1,55  | 4 |
| 86417-TRNI.1    | -2,18 | -0,38 | -3,73 | 2 | 0,38  | 0,40  | -1,15 | 1 |
| 92598-TRNI      | -6,19 | -0,10 | 0,69  | 5 | -2,88 | 1,32  | 3,30  | 5 |
| 94519-TRNL.2    | -0,97 | 0,85  | -0,61 | 6 | 0,70  | 1,92  | 0,91  | 3 |
| 100687+TRNV.2   | -1,38 | 1,00  | -3,24 | 1 | 1,36  | 1,26  | -0,48 | 3 |
| 100873+RRN16S.1 | 3,90  | -5,57 | 3,32  | 7 | 4,78  | -3,02 | 5,51  | 7 |
| 100898+RRN16S.1 | -0,81 | -0,03 | -2,53 | 1 | 2,53  | 1,45  | 0,37  | 3 |
| 104475+RRN23S.1 | -0,78 | -0,08 | -2,31 | 1 | 0,65  | 1,16  | -0,39 | 1 |
| 109231-TRNN.1   | -0,03 | 0,57  | -3,29 | 1 | 1,51  | 1,97  | -0,04 | 3 |
| 109366+TIC214   | -4,64 | -1,27 | 1,76  | 5 | 3,17  | 1,91  | 6,47  | 6 |
| 110032-TRNN.1   | -1,54 | 0,48  | -1,22 | 6 | -0,52 | 0,53  | 0,61  | 1 |
| 110341-TRNN.1   | 0,07  | 0,60  | -0,90 | 3 | 0,98  | 0,92  | 1,89  | 4 |
| 112358+RPL32    | -1,39 | 0,79  | 3,65  | 4 | 2,75  | 3,15  | 8,61  | 6 |
| 112771-NDHF     | 1,33  | 0,29  | -0,95 | 3 | 2,59  | 1,06  | -2,52 | 2 |
| 112953+RPL32    | -0,11 | -0,20 | -1,20 | 3 | 2,20  | 0,69  | 2,31  | 4 |
| 112957-NDHF     | -1,95 | -0,51 | -4,39 | 2 | 1,22  | -0,18 | -3,30 | 2 |
| 112979+RPL32    | -0,71 | -1,05 | 3,01  | 4 | 4,17  | -0,91 | 6,87  | 7 |
| 113004-NDHF     | 1,11  | -0,33 | 0,26  | 3 | 0,91  | 0,15  | -0,54 | 1 |
| 114114-NDHF     | 1,17  | -0,51 | -1,97 | 1 | 0,65  | 0,92  | 2,16  | 4 |
| 114261+TRNL.3   | -0,93 | 0,96  | -3,16 | 1 | 1,72  | 1,88  | 3,16  | 4 |
| 114944+TRNN     | 0,85  | 0,56  | -0,55 | 3 | 1,52  | 0,65  | -2,01 | 2 |
| 116996-NDHF     | -1,50 | 0,44  | -1,50 | 6 | -1,33 | -0,46 | -1,13 | 1 |
| 118082-PSAC     | 1,89  | 2,59  | 1,42  | 3 | -0,36 | 0,02  | -2,00 | 2 |

## Anhang

|                 |       |       |       |   |       |       |       |   |
|-----------------|-------|-------|-------|---|-------|-------|-------|---|
| 118171-NDHE     | 2,21  | 1,54  | 1,38  | 3 | 2,20  | 0,98  | 0,09  | 3 |
| 118899-NDHE     | -1,87 | 0,22  | -1,55 | 6 | 0,69  | 0,28  | -0,60 | 1 |
| 122076-NDHA     | -2,20 | 3,10  | -2,59 | 6 | 3,06  | 0,92  | 4,56  | 6 |
| 128308+TRNN.2   | -1,42 | -0,01 | -1,42 | 6 | 1,02  | 1,17  | 1,54  | 4 |
| 128616+TRNN.2   | -1,54 | 0,06  | -1,22 | 6 | -1,02 | -0,07 | 0,77  | 1 |
| 129283-TIC214   | -4,53 | -0,98 | 2,04  | 5 | 3,94  | 2,79  | 7,44  | 6 |
| 129418+TRNN.2   | -0,16 | 0,40  | -2,81 | 1 | 1,50  | 1,96  | 0,07  | 3 |
| 134174-RRN23S.2 | -0,76 | -0,14 | -2,04 | 1 | 0,69  | 1,09  | -0,39 | 1 |
| 137751-RRN16S.2 | -0,84 | -0,07 | -2,30 | 1 | 2,41  | 1,39  | 0,55  | 3 |
| 137776-RRN16S.2 | 4,03  | -5,04 | 3,43  | 7 | 4,71  | -2,86 | 5,49  | 7 |
| 137962-TRNV.3   | -1,12 | 1,00  | -3,02 | 1 | 1,27  | 1,29  | -0,49 | 3 |
| 144130+TRNL.4   | -0,88 | 0,89  | -0,77 | 6 | 0,75  | 1,95  | 0,90  | 3 |
| 144275+TRNL.4   | -1,26 | 1,39  | -3,32 | 1 | 2,11  | 3,03  | 1,05  | 3 |
| 146051+TRNI     | -6,30 | -0,15 | 0,43  | 5 | -2,89 | 0,85  | 3,61  | 5 |
| 152232+TRNI.4   | -2,24 | 0,16  | -4,34 | 2 | -0,04 | 0,60  | -2,05 | 2 |
| 152438+RPL23.2  | -0,03 | 0,48  | 0,63  | 3 | 1,53  | 1,06  | 1,57  | 4 |
| 152581+RPL2.2   | -0,51 | -0,27 | -3,78 | 2 | 0,32  | 0,52  | -0,33 | 1 |
| 153604+PSBK     | -3,39 | 0,05  | 2,17  | 5 | -3,95 | 1,02  | 3,54  | 5 |

Tabelle VI.14: Liste der TSS nach zugewiesener Polymeraseabhängigkeit mit flankierender Sequenz, welche für die Motivsuche verwendet wurde. Die flankierenden Sequenzen gehen von der TSS, welche die Position +1 hat, 40 Nukleotide stromabwärts und 5 Nukleotide stromaufwärts (-40 bis +5).

| TSS          | Transkribiert von | Sequenz 5'-->3'                                 |
|--------------|-------------------|---|
| 1522-PSBA    | PEP               | TGGTTGACATGGCTATATAAGTCATGTTTACTGTTTCATAACAAG   |
| 2735-PSBA    | PEP               | GGGTTGCTCAAGGATCCTTTTCATACATTATGTTAGATATCACGGAA |
| 4451-TRNK    | PEP               | TCATATTGACAACAGTGTATGGACCAAATATAATTCTCTCATAGAT  |
| 6708-TRNQ    | PEP               | TGAATTGACGAACAACCAATAGGAATATTACTCTTTTAGTAGTCTT  |
| 7539+PSBI    | PEP               | TTTATTGGTATCAAAATTAGATATTTGATATAAAAAATAGAGAATCT |
| 8126+TRNG.1  | PEP               | TTATATAAATATATATATCCTATATTTAAATATATAATATATATAT  |
| 8643+TRNG.1  | PEP               | TTGTTTCGACAAAAGGTTAATTTATATGCAATAATTGCATTGTAGCG |
| 28037+YCF6   | PEP               | TTTTTTGACTCTGCACCAGTGATTTCACTATTATTAGTGAACAATA  |
| 28988-PSBM   | PEP               | GTGGATAAAAGATCTTCTATGTTTATACTATTAAATTTCTTGACGAT |
| 29204-PSBM   | PEP               | ATTGAACTTCGTAATTTTCTATATTGATTTCTATTGTTTTAT      |
| 30565-TRNE   | PEP               | ATTATATTGACAATTTAAAAAACTGATCATACTATGATCATAGTA   |
| 31336+TRNT.1 | PEP               | GTATATATATTTAATATATATATATATATAGATTTTATTTACTACC  |
| 31523-TRNE   | PEP               | TAACCTTTATTATTAGTTATTATATTATTAGAATGAAATTTATTAGA |
| 33526+PSBC   | PEP               | GGGTTGCTTTTTTCCAATAAACGTTGGTTACATTTCTTTATGTTATT |
| 36358+TRNG.2 | PEP               | ATATATATTATTAATATATATATTTATAATATATATATATATATTA  |
| 36372+TRNG.2 | PEP               | TATATATATTTATAATATATATATATATATTAATAGAATTGTTAAT  |
| 36665+TRNS   | PEP               | TTTCTCGCGTCCCAAAAGACACGTATTAATTTATGACTTGTTAGT   |
| 36849-TRNFM  | PEP               | GCCCTTTACCATTCTGTCTAAATAGGCTATTCTATTTGTACAGAT   |
| 37271+TRNS   | PEP               | TCTTTGCCATAATGCTTCAGTTCCTGTTATTATCAATGATAATGAT  |
| 42045-PSAA   | PEP               | GGGTCCGTTGAGCACCTATGGATATGTCATAATAGATCCGAACAC   |
| 45454-YCF3   | PEP               | ATCATGGACATATTTTAGTTAATGGTCGTATAGTTGATATACCAAG  |
| 46110+TRNL.1 | PEP               | TATATTCCTAATATGGAAGTTTATATGACATAATATAAATGGAGTG  |
| 46589+TRNL.1 | PEP               | TTCCATTGCACATTTGTAGAATCTAAGTTTCAATAATAATAAAT    |
| 46751+TRNL.1 | PEP               | CATATTGAATTGCAAAATACAAAATGATAGAATCTTTGTTGATTAG  |
| 49371+TRNM   | PEP               | ATGATGGTGGATAGAGTAGTCTTGATCATAATTTCCAATATGCCG   |
| 51885-TRNV.1 | PEP               | AATCTTGACAAGAAATATCTACATGATAAGATAAAATGTGTATCA   |
| 52012+TRNM   | PEP               | TGGTATTGCTTGAAAAATAATTTGGATTTATAATCCTATCGATGTG  |
| 53093-ATPE   | PEP               | ACATTTGCACATTTAGATGCGACTACCGTACTATCAAGAGGATTAG  |
| 54618-ATPB   | PEP               | CTCTTGACAGTAATATCTGTTGTATATGTAATCCTAGATATGACA   |
| 54671-ATPB   | PEP               | ACTTTTTATAAATCCAATATTTAATATTTAAATTTAAATATTTAAT  |
| 54781+RBCL   | PEP               | TAGGTTGCGCTATACATATGAAAGAATATACAATAATGATGTATTT  |
| 61105+PETA   | PEP               | CACTTCTCGACAAATCTAATATCTTTGTTATTCTAAGTGGTTATT   |
| 62285-ATPB   | PEP               | CCCCTTTCTTTTCGTAATAATTTTACTTATTATCCCTCCTGCCGTA  |



## Anhang

|                 |                  |  |
|-----------------|------------------|--|
| 63723-PSBJ      | PEP              | TAATTTTATTATCCCATTCGGATGGATTATATATAATAATTATCTC   |
| 66089+PSAJ      | PEP              | TTGGCTGGATTATTCGTAACCTGCTTATTTACAATACAGACGTGGTG  |
| 66577-TRNP      | PEP              | GGATGTGGAAAACAAGACAGGAATTCCTCTACAATGACACTGTAAAC  |
| 66862-TRNP      | PEP              | TTGTATTACATTATATATAGTAATATATATAAATTACAGATGTACAT  |
| 66894+PSAJ      | PEP              | TACATGTACATCTGTAATTATATATATATTACTATATATAATGTAATA |
| 67744-TRNP      | PEP              | TTATTTGTTTATATTTATATATATATGTAATAATATATAGATACTAGC |
| 71416-RPS12     | PEP              | CAAATCAAATACAAATAATATATAATATAAATTCAAAATTTGTTGAC  |
| 72200+PSBB      | PEP              | CCCATTGCATATTGGTACTTATCGGATATAGAATAGATCCGCTTCC   |
| 74413-PSBN      | PEP              | GACTTTGTATACCATTCCGTTGTAAATAAATGATCTTAGCATAGAT   |
| 76780+YCF2.1    | PEP              | AAAAAAGAAAAATAATTATTAATAGTAATAGTATGTAAATGCATTC   |
| 78010+YCF2.1    | PEP              | GTCTTCTTGACTGTATTCAAAGGTCCAATAATGTATGTATATTG     |
| 78677+YCF2.1    | PEP              | TGTACTGATTCTTGAATACCTGCTATGTTAGAATAGTCATGTGGTA   |
| 86068-RPL2.1    | PEP              | AAGTATTCGGTTATTTGGGGAAAAATCAATATACTTTTAATGTCGAA  |
| 86417-TRNI.1    | PEP              | GCATTGATTTATCCTAAAGATTTCAATTTCAATTTGGAATTTGGTTAT |
| 100687+TRNV.2   | PEP              | GTGGCCTTGAGTTTCTCGACCCCTTGACTTAGGATTAGTCAGTTCT   |
| 100898+RRN16S.1 | PEP              | GGATTGACGTGAGGGGGTAGGGGTAGCTATATTTCTGGGAGCGAAC   |
| 104475+RRN23S.1 | PEP              | AGTTCTTGTCTATCGTTGGCCTCTATGGTAGAATCAGTCAGGGGCC   |
| 109231-TRNN.1   | PEP              | AAATGAATGGAATTTGGGATATGGATGGAATATAATGAAATAGAGCC  |
| 112771-NDHF     | PEP              | ATATATTGCTTTTCAAAGATTTATGAAGGATTATATTAATTCGAAA   |
| 112957-NDHF     | PEP              | AAATCTTTGTTCTATTCTTAATATATGTATATAAATTAATTGTAATA  |
| 113004-NDHF     | PEP              | GTCAAGCTTTTTTATATAAATTAATATTAAGTTTTTTTTTGTCTGTG  |
| 114114-NDHF     | PEP              | TTTTGTCTTTTATATTTTTTTCGAGAAACAATGAAAAATGTAAC     |
| 114944+TRNN     | PEP              | AGGCTATGGCGCTCTGTTATGCGGATCATTATTAATCAATAGCTCTT  |
| 118171-NDHE     | PEP              | AATATTTTTTTTTATTCGACTTTTTTGTATATTCTAGTTGATGGA    |
| 118899-NDHE     | PEP              | GGATTTGATTTTTTCTTATTTTAATAATATAATGGATTTGCCTGG    |
| 129418+TRNN.2   | PEP              | AAATGAATGGAATTTGGGATATGGATGGAATATAATGAAATAGAGCC  |
| 134174-RRN23S.2 | PEP              | AGTTCTTGTCTATCGTTGGCCTCTATGGTAGAATCAGTCAGGGGCC   |
| 137751-RRN16S.2 | PEP              | GGATTGACGTGAGGGGGTAGGGGTAGCTATATTTCTGGGAGCGAAC   |
| 137962-TRNV.3   | PEP              | GTGGCCTTGAGTTTCTCGACCCCTTGACTTAGGATTAGTCAGTTCT   |
| 144275+TRNL.4   | PEP              | TATTCATACCTTACTAGATAGATATCATATTTATGGAATACGATT    |
| 152232+TRNI.4   | PEP              | GCATTGATTTATCCTAAAGATTTCAATTTCAATTTGGAATTTGGTTAT |
| 152581+RPL2.2   | PEP              | AAGTATTCGGTTATTTGGGGAAAAATCAATATACTTTTAATGTCGAA  |
| 7872-TRNQ       | <i>rpoTp</i>     | TTTTTTTTATGGAACCTAATTAATATCCCTTTTCGATTAGGAGAG    |
| 32644+PSBD      | <i>rpoTp</i>     | CAAATGATAGAAATTTGAGCGCCCTGAAAATGCCATAATATATAACA  |
| 85045-RPS19     | <i>rpoTp</i>     | ACCCATAATATGTGGAAGTATCGACGTAATTTTCATAGAGTCATTCCG |
| 92598-TRNI      | <i>rpoTp</i>     | CTGATCAACGATAGAACAAGATCCATCTTGCATCATATATAAGGGA   |
| 109366+TIC214   | <i>rpoTp</i>     | CTATTTCTATATATGGAAGTTGCAAAAATCATCATATAATAATCCA   |
| 112358+RPL32    | <i>rpoTp</i>     | AACCAAAATCCCAACAGTTGTAATTAATATTGACATAATAGAAGTA   |
| 129283-TIC214   | <i>rpoTp</i>     | CTATTTCTATATATGGAAGTTGCAAAAATCATCATATAATAATCCA   |
| 146051+TRNI     | <i>rpoTp</i>     | CTGATCAACGATAGAACAAGATCCATCTTGCATCATATATAAGGGA   |
| 153604+PSBK     | <i>rpoTp</i>     | ACCCATAATATGTGGAAGTATCGACGTAATTTTCATAGAGTCATTCCG |
| 56771-ATPB      | <i>rpoTp</i>     | TAAGAATATTAAGAACTTTTTTAGTAAAATCTCGTATACTAATATA   |
| 100873+RRN16S.1 | <i>rpoTmp</i>    | AATAAGAATGGATAAGAGGCTCGTGGGATTGACGTGAGGGGGTAGG   |
| 137776-RRN16S.2 | <i>rpoTmp</i>    | AATAAGAATGGATAAGAGGCTCGTGGGATTGACGTGAGGGGGTAGG   |
| 11392-TRNQ      | potentiell NEP   | TAATATTATCCGTGAACGTATTGAGCAATATAATAGAGAAGTAACG   |
| 13921-ATPH      | potentiell NEP   | TACATTATGTAAGAAACACTTGTATATGTGATATTTGATATTGCCT   |
| 20245-RPOC2     | potentiell NEP   | AAGGTTTTTCTCAAGCCTGTTTCATATGATACCTAATAATATGCTAT  |
| 36324+TRNG.2    | potentiell NEP   | TAATATTAATATATTAATATATAGAATAAATATAATATATATTAT    |
| 46499+TRNL.1    | potentiell NEP   | TAACCTTAATTTCTATATATATATCTATATAGTCATAGAATAGTATA  |
| 59146+PSAI      | potentiell NEP   | TATATATTTACTTAAAGATACTTAGTATAAATTAATAATAGATATA   |
| 71940-CLPP1     | potentiell NEP   | TAGGAAAAAAGGATAATAGTATTAATAAACGAAACCCCAATTTTAC   |
| 80625-RPS8      | potentiell NEP   | CGGATTAATAAACAGGTATTGAAATAAATTAATTAATAGATTCT     |
| 112979+RPL32    | potentiell NEP   | AATTAATAATCCATATTACAATAATTTATATACATATATTAAGAAT   |
| 6253-RPS16      | nicht zugewiesen | ATTCATTTCATAGAGAAAAGGATAGATTATATATACTGAC         |
| 8213+TRNG.1     | nicht zugewiesen | ATATATAATATATATAAATATATATATATATAAATATATATATTACT  |
| 8230+TRNG.1     | nicht zugewiesen | AATATATATTATATAAATATATATATATTACTATTATATAACTGTTTT |
| 29912-TRND      | nicht zugewiesen | AAAAATTTGAATTTAAAATGAAATGATTAATAATGAAATCATAAGA   |

## Anhang

|                |                  |   |
|----------------|------------------|---|
| 31319+TRNT.1   | nicht zugewiesen | ACCACTATGCGTCTATTGTATATATATTTAATATATATATATATAT    |
| 31614+PSBD     | nicht zugewiesen | TAAAAAATAAAAAATGAACATTTGTATATTATAATACAATTACAATG   |
| 32672+PSBD     | nicht zugewiesen | ATGCCATAATATATAACATCAAAATCTATAACATTAAGGTGTTCGG    |
| 35410-TRNE     | nicht zugewiesen | GCTATGTTCTATTTGGCGAAAAAATAAAAAAGAAAATTATCTTT      |
| 44234+TRNS.3   | nicht zugewiesen | GCTTCTTCCCTATGCCTCCTATTCGTATATTGTATTAGTGTAGTA     |
| 44755-YCF3     | nicht zugewiesen | AATAGCGAGTTATTCATCGTTTTTTAGATTATGGTTAATAGATACC    |
| 44824+TRNS.3   | nicht zugewiesen | TAACTCGCTATTCAACCCAGATACTCAGTCATAATCTTGATGTCGGA   |
| 46339-TRNT.2   | nicht zugewiesen | AATTCTAACTAGATAAGGAAAAATATTAATAGAATAGTTAATTGAT    |
| 46585-TRNT.2   | nicht zugewiesen | TTTTTTTTTTTACTTTTCATGAAAAGAAATTTATTATTATTATT      |
| 46928+TRNF     | nicht zugewiesen | AAAAAAGGGGATATGGCGGAATTGGTAGACGCTACGGACTTAATTG    |
| 47545+TRNF     | nicht zugewiesen | AAAGGTTGACACCTTACCTTTTTTTTCGTTATTATTTATTGAATTA    |
| 47703+TRNF     | nicht zugewiesen | TTCATTGACATAGACCTAAGTCATATATTAATAATGATACTGATACT   |
| 63131-ATPB     | nicht zugewiesen | AAAAAAAAGAAAACACAAAAAGTAACTACTATAATATATATATAT     |
| 64448-PSBE     | nicht zugewiesen | TTGCGTTGCTGTGTGTCAGAAGAAGGATAGCTATACTGATTCCGGTAGA |
| 66382-TRNW     | nicht zugewiesen | TTTAATTCTGAAATTTATGATATACTATGTATGATATAGTATATAC    |
| 67995-TRNP     | nicht zugewiesen | ATAAATCGACTAATTAAACTCATGTTTCTATAATCAATTCCGATCCC   |
| 78986-RPOA     | nicht zugewiesen | TTTCGTACGAGATGTAACCCCTATGCCACATAATGGTTGTAGACCC    |
| 82650-RPL14    | nicht zugewiesen | ATAAAAAATTTCCATCAAAACTCCTTTGATATAATTGCTATGCTTAG   |
| 86211-RPL23    | nicht zugewiesen | ATCATCTATACATAACGAATTGGTGTGGTATATTCATATCATAACA    |
| 94519-TRNL.2   | nicht zugewiesen | TTGTCCATCTTTTTGATATAATATATATTAATATATATATAAGGCA    |
| 110032-TRNN.1  | nicht zugewiesen | GAATTTCTCAATTCTGACACAAGGAACCTATACTTATTAGATCTAA    |
| 110341-TRNN.1  | nicht zugewiesen | TTTCCATATAGATTTTCAGAAACAGTTTTATAGTTATAGTATGTTTC   |
| 112953+RPL32   | nicht zugewiesen | GATTTTATGCAAAACAGAAAAAGTGAATTATAATTCCATATTACA     |
| 114261+TRNL.3  | nicht zugewiesen | TAATCTCGACGATTGAATCAAAACTTGTTAATATTGTTTTGAACAA    |
| 116996-NDHF    | nicht zugewiesen | TCTTCTAACGACCTACGCTTTTTTGTATAATTTTAAAAATGGACGAT   |
| 118082-PSAC    | nicht zugewiesen | AAGGAGTTGCTGAATGATACTCGAACATGTACTTGTTTTGAGTGCC    |
| 122076-NDHA    | nicht zugewiesen | TATTAATTTGCAAAATCTTCCGAACTAGTTAAAAGAATGAAATTG     |
| 128308+TRNN.2  | nicht zugewiesen | TTTCCATATAGATTTTCAGAAACAGTTTTATAGTTATAGTATGTTTC   |
| 128616+TRNN.2  | nicht zugewiesen | AGAATTTCTCAATTCTGACACAAGGAACCTATACTTATTAGATCTA    |
| 144130+TRNL.4  | nicht zugewiesen | TTGTCCATCTTTTTGATATAATATATATTAATATATATATAAGGCA    |
| 152438+RPL23.2 | nicht zugewiesen | ATCATCTATACATAACGAATTGGTGTGGTATATTCATATCATAACA    |

Tabelle VI.15: Liste der TSS in der sehr frühen Samenentwicklung, mit flankierender Sequenz. Die während der Analyse der RNA-Polymerasenmutanten identifizierten Abhängigkeiten zu Polymerasenaktivitäten sind aufgeführt. TSS, welche exklusiv in der Analyse der frühen Samenentwicklung detektiert wurden, sind mit „in früher Samenentw.“ gekennzeichnet.

| TSS             | Transkribiert von | Sequenz 5' -->3'                                |
|-----------------|-------------------|---|
| 7872-TRNQ       | <i>rpoTp</i>      | TTTTTTTTATGGAACCTAATTAATATCCCTTTTCGATTAGGAGAG   |
| 109366+TIC214.1 | <i>rpoTp</i>      | CTATTTCTATATATGGAAGTTGCAAAATCATCATATAATAATCCA   |
| 129283-TIC214.2 | <i>rpoTp</i>      | CTATTTCTATATATGGAAGTTGCAAAATCATCATATAATAATCCA   |
| 100873+RRN16S.1 | <i>rpoTmp</i>     | AATAAGAATGGATAAGAGGCTCGTGGGATTGACGTGAGGGGGTAGG  |
| 137776-RRN16S.2 | <i>rpoTmp</i>     | AATAAGAATGGATAAGAGGCTCGTGGGATTGACGTGAGGGGGTAGG  |
| 71940-CLPP      | NEP               | TAGGAAAAAAGGATAATAGTATTAATAAACGAAACCCCAATTTTAC  |
| 1522-PSBA       | PEP               | TGGTTGACATGGCTATATAAGTCATGTTATACTGTTTCATAACAAG  |
| 4451-TRNK       | PEP               | TCATATTGACAACAGTGTATGGACCAATATAATTCTCTCATAGAT   |
| 6708-TRNQ       | PEP               | TGAATTGACGAACAACCAATAGGAATATTACTCTTTTAGTAGTCTT  |
| 8643+TRNG.1     | PEP               | TTGTTTCGACAAAAGGTTAATTTATATGCAATAATGCATTGTAGCG  |
| 28988-PSBM      | PEP               | GTGGATAAAAGATCTTCTATGTTATACTATTAATCTTGACGAT     |
| 30565-TRNE      | PEP               | ATTATATTGACAATTTAAAAAACTGATCATACTATGATCATAGTA   |
| 31336+TRNT.1    | PEP               | GTATATATATTTAATATATATATATATATATAGATTTTATTACTACC |
| 33526+PSBC      | PEP               | GGGTTGCTTTTTCCAATAAACGTTGGTTACATTTCTTTATGTTATT  |
| 36849-TRNFM     | PEP               | GCCCTTTACCATCTGTCTAAATAGGCTATTCTATTTGTACAGAT    |
| 42045-PSAA      | PEP               | GGGTCCGTTGACACCCTATGGATATGTCATAATAGATCCGAACAC   |
| 46110+TRNL.1    | PEP               | TATATTCCTAATATGGAAGTTTATATGACATAATATAAATGGAGTG  |
| 46751+TRNL.1    | PEP               | CATATTGAATTGCAAAATCAAAAAATGATAGAATCTTTGTTGATTAG |

## Anhang

|                 |                      |  |
|-----------------|----------------------|--|
| 51885-TRNV.1    | PEP                  | AATCTTGACAAGAAATTATCTACATGATAAGATAAAATGTGTATCA   |
| 52012+TRNM      | PEP                  | TGGTATTGCTTAGAAATAATATTGGATTTATAATCCTATCGATGTG   |
| 54618-ATPB      | PEP                  | CTCTTGACAGTAATATCTGTTGTATATGTAAATCCTAGATATGACA   |
| 54781+RBCL      | PEP                  | TAGGTTGCGCTATACATATGAAAGAATATACAATAATGATGTATTT   |
| 66577-TRNP      | PEP                  | GGATGTGGAACAAGACAGGAATTCTCTACAATGACACTGTAAAC     |
| 66894+PSAJ      | PEP                  | TACATGTACATCTGTAATTATATATATTACTATATATAATGTAATA   |
| 72200+PSBB      | PEP                  | CCCATTGCATATTGGTACTTATCGGATATAGAATAGATCCGCTTCC   |
| 78010+YCF2.1    | PEP                  | GTCTTCTTGACTGTTATTCAAAAGGTCCAATAATGTATGTATATTG   |
| 86068-RPL2.1    | PEP                  | AAGTATTCGGTTATTGGGGAAAAATCAATATACTTTTAATGTGCGAA  |
| 100898+RRN16S.1 | PEP                  | GGATTGACGTGAGGGGGTAGGGGTAGCTATATTTCTGGGAGCGAAC   |
| 104475+RRN23S.1 | PEP                  | AGTTCTTGTCTATCGTTGGCCTCTATGGTAGAATCAGTCAGGGGCC   |
| 109231-TRNN.1   | PEP                  | AAATGAATGGAATGGGATATGGATGGAATATAATGAAATAGAGCC    |
| 129418+TRNN.2   | PEP                  | AAATGAATGGAATGGGATATGGATGGAATATAATGAAATAGAGCC    |
| 134174-RRN23S.2 | PEP                  | AGTTCTTGTCTATCGTTGGCCTCTATGGTAGAATCAGTCAGGGGCC   |
| 137751-RRN16S.2 | PEP                  | GGATTGACGTGAGGGGGTAGGGGTAGCTATATTTCTGGGAGCGAAC   |
| 152581+RPL2.2   | PEP                  | AAGTATTCGGTTATTGGGGAAAAATCAATATACTTTTAATGTGCGAA  |
| 8213+TRNG.1     | nicht zugewiesen     | ATATATAATATATATAATATATATATTATATAATATATATATTACT   |
| 8230+TRNG.1     | nicht zugewiesen     | AATATATATTATATAATATATATATATTACTATTATATAACTGTTTTT |
| 31319+TRNT.1    | nicht zugewiesen     | ACCACTATGCGTCTATTGTATATATATTTAATATATATATATATAT   |
| 46928+TRNF      | nicht zugewiesen     | AAAAAAGGGGATATGGCGGAATTGGTAGACGCTACGGACTTAATG    |
| 94519-TRNL.2    | nicht zugewiesen     | TTGTCCATCTTTTTTGATATAATATATATTAATATATATATAAGGCA  |
| 110032-TRNN.1   | nicht zugewiesen     | GAATTTCTCAATTCTGACACAAGGAAGCTTATACTTATTAGATCTAA  |
| 114261+TRNL.3   | nicht zugewiesen     | TAATCTCGACGATTGAATCAAACTTGTTAATATTGTTTTGAACAA    |
| 122076-NDHA     | nicht zugewiesen     | TATTAATTTGCAAAATCTTCCTGAAGTAAAGAATGAAATTG        |
| 144130+TRNL.4   | nicht zugewiesen     | TTGTCCATCTTTTTTGATATAATATATATTAATATATATATAAGGCA  |
| 31369+TRNT.1    | in früher Samenentw. | TTTATTTACTACCGGCTCAGTCAGGAATTAACCAAGACGGCCCTT    |
| 36000-TRNE      | in früher Samenentw. | TTTTTCTTTATTTCGATTTTATGGACTTATATTGAATTTGAATGT    |
| 59449+YCF4      | in früher Samenentw. | TTTCGATTCGATAAGACCCATTTGGTAGAATGTTGTATAACACATAG  |
| 64445-PSBE      | in früher Samenentw. | CGTTGCTGTGTGAGAAGAAGGATAGCTATACTGATTCGGTAGACTC   |
| 66302-PSBE      | in früher Samenentw. | TTCTAGCCTTTTTCGTTTTTTAGAAAAAAAAAAAAAAGAAAAACGCTC |
| 66778+PSAJ      | in früher Samenentw. | AGGCCTTTACGTTTTCAAATGGAATCGATAAGATCGTTCTAGTCGA   |
| 74411-PSBN      | in früher Samenentw. | CTTGTATACCATCCGTTGTAATAAATGATCTTAGCATAGATCC      |
| 83825+YCF2.1    | in früher Samenentw. | ATAACATCCTCGATAAGGCATAAGTTCTAATATCATAAGTGCTTCT   |
| 107825+RRN5S.1  | in früher Samenentw. | CATTGTTCAACTCTTTGACAACATGAAAAACCTAAAAATCTGCCC    |
| 109084-TRNN     | in früher Samenentw. | GAACGGGAATTGAACTCTATGAGATCTAATCTCCCGTTGTTCCCTCA  |
| 129565+TRNN.2   | in früher Samenentw. | GAACGGGAATTGAACTCTATGAGATCTAATCTCCCGTTGTTCCCTCA  |
| 130824-RRN5S.2  | in früher Samenentw. | CATTGTTCAACTCTTTGACAACATGAAAAACCTAAAAATCTGCCC    |
| 63121-ATPB      | in früher Samenentw. | AAACTACAAAAAGTAACTACTATAATATATATATATATATGTAAC    |
| 8197+TRNG.1     | in früher Samenentw. | ATAAATCCTCTATTATATATATAATATATATATAATATATATTATA   |

VI.2 Abbildungen

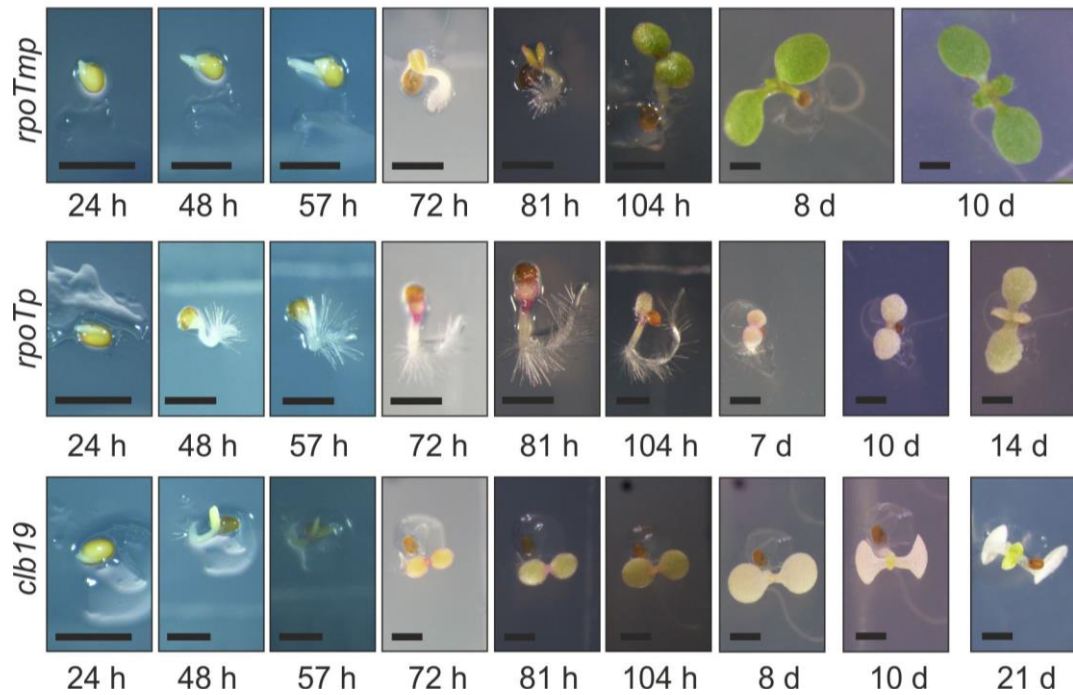


Abbildung VI.1: Wachstumsanalyse der Mutanten. Die Samen wurden wie in Kap. II.1.3 Methoden auf Agarplatten unter Langtagbedingungen angezogen und die Entwicklung zu festgelegten Zeitpunkten protokolliert. h: Stunden. Der Größenbalken zeigt 1 mm an.

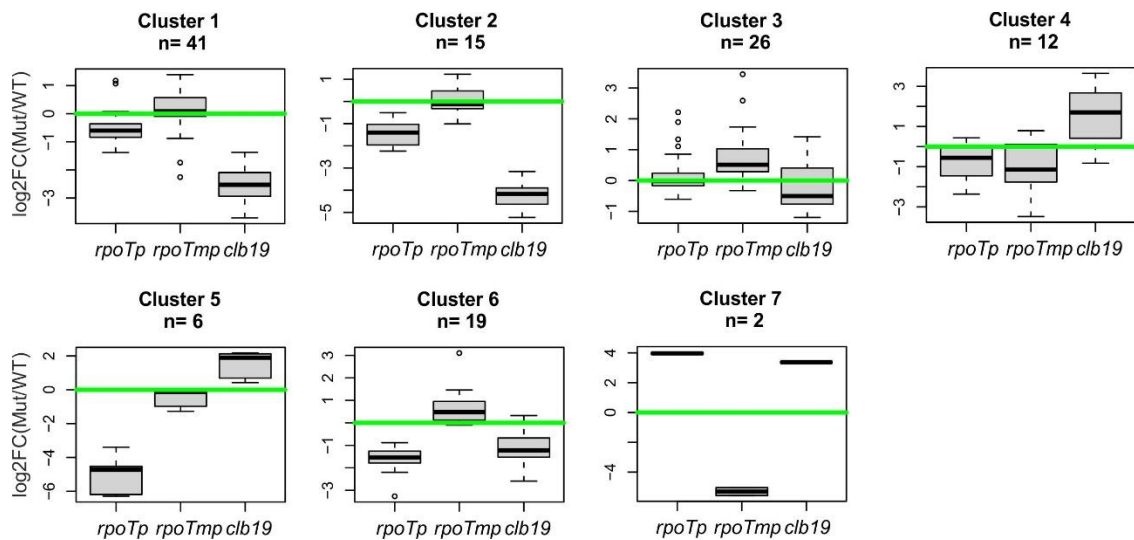


Abbildung VI.2: Clusteranalyse mittels pamK für das frühe Stadium. Die TSS wurden nach den logarithmierten relativen Transkriptlevel bezogen auf WT zu Clustern zusammengefasst. Die grüne Linie zeigt 0 an. Der schwarze Strich innerhalb der Boxen zeigt den Median an, die Begrenzungen der Box sind jeweils das untere (25 %) bzw. das obere (75 %) Quartil. Die unterbrochene Linie führt von der Box zum Minimum bzw. Maximum des Datensatzes. Die Ausreißer werden als einzelne Punkte außerhalb der Boxen dargestellt. n = Anzahl der TSS im Cluster. Die Analyse wurde von Frederic Ehlers unter Anleitung von Julia Rott und Dermot Harnett durchgeführt.

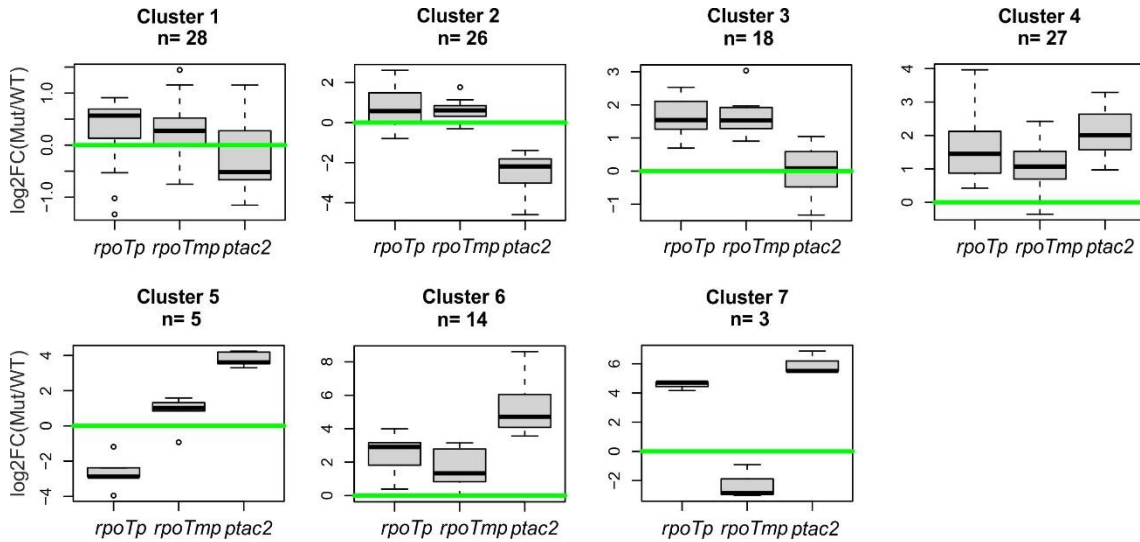


Abbildung VI.3: Clusteranalyse mittels pamK für das späte Stadium. Die TSS wurden nach den logarithmierten relativen Transkriptlevel bezogen auf WT zu Clustern zusammengefasst. Die grüne Linie zeigt 0 an. Der schwarze Strich innerhalb der Boxen zeigt den Median an, die Begrenzungen der Box sind jeweils das untere (25 %) bzw. das obere (75 %) Quartil. Die unterbrochene Linie führt von der Box zum Minimum bzw. Maximum des Datensatzes. Die Ausreißer werden als einzelne Punkte außerhalb der Boxen dargestellt. n = Anzahl der TSS im Cluster. Die Analyse wurde von Frederic Ehlers unter Anleitung von Julia Rott und Dermot Harnett durchgeführt.

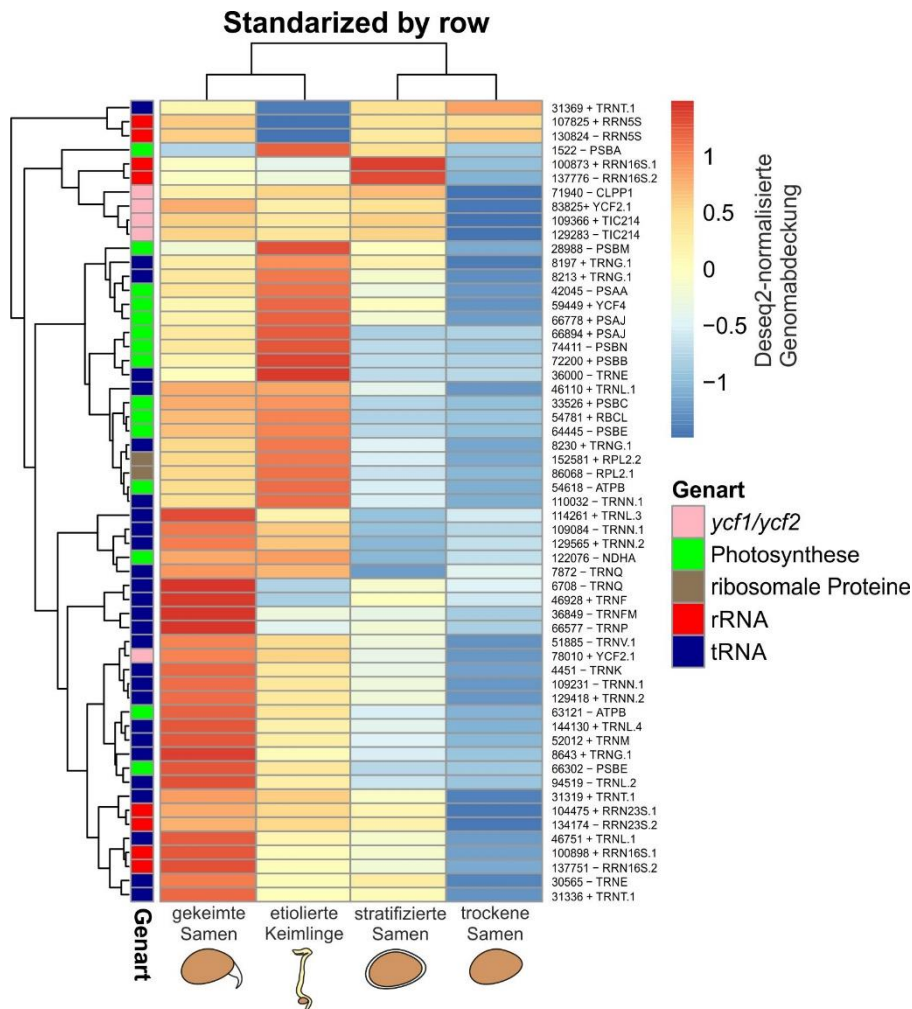


Abbildung VI.4: TSS-Nutzung in der sehr frühen Samenentwicklung. Es sind die DESeq2-normalisierten Genomabdeckungen in den +5'PPase-Proben für Wildtyp in den Entwicklungsstadien trockene Samen, stratifizierte Samen, gekeimte Samen und

etiolierte Keimlinge für die detektierten 57 TSS dargestellt. Jeweils vier Replikate wurden vereint. Die Farbkodierung gibt die Zugehörigkeit der TSS zu den einzelnen Genarten wieder. Die Analyse wurde von Frederic Ehlers unter Anleitung von Julia Rott und Dermot Harnett durchgeführt.

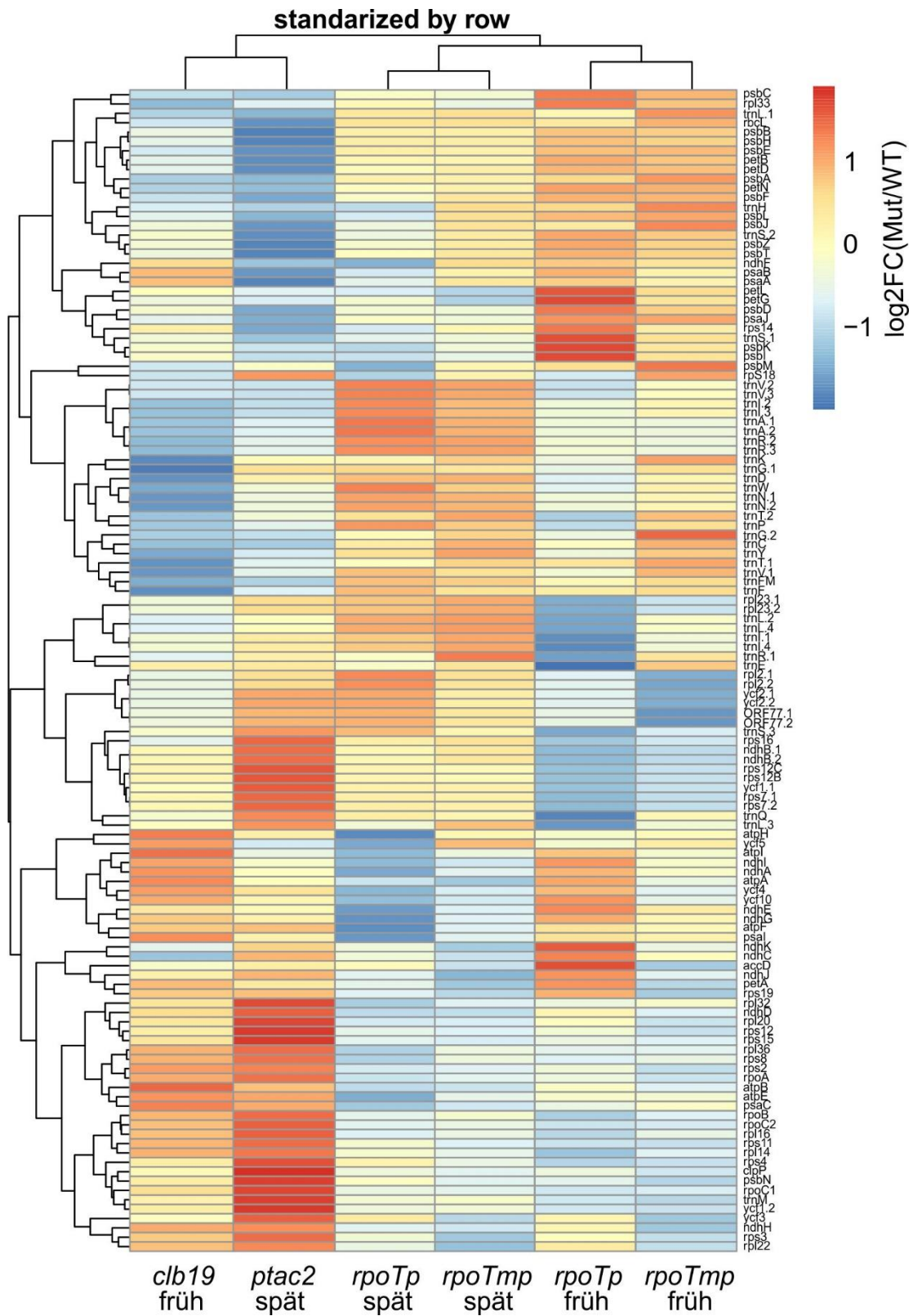


Abbildung VI.5: RNA-Seq-Analyse in den Mutanten für das frühe und späte Stadium dargestellt als logarithmierte relative Transkriptlevel bezogen auf WT. Die Expressionsanalyse wurde von Jose Muino durchgeführt.

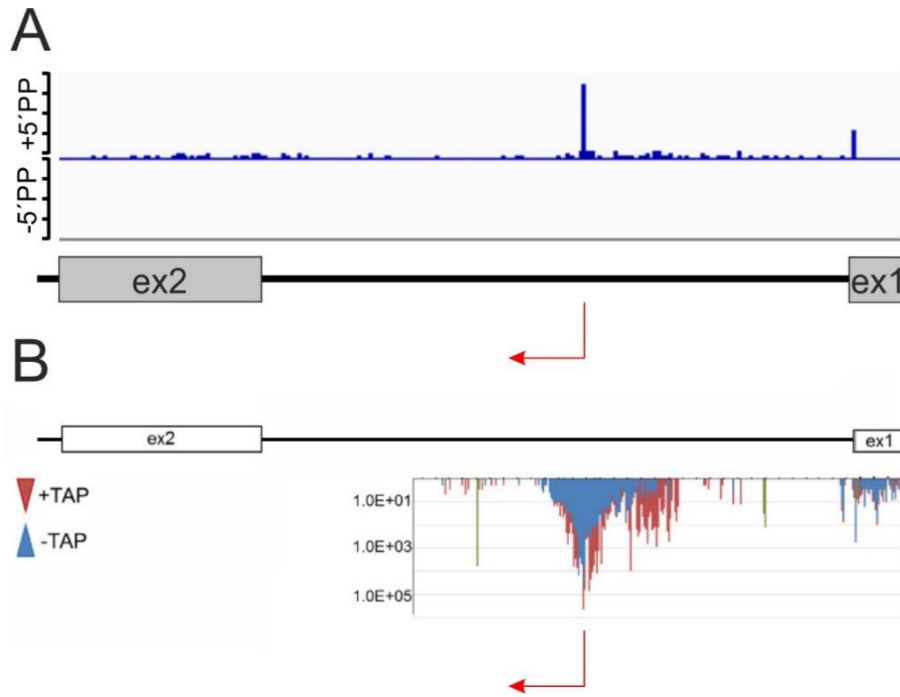


Abbildung VI.6: Gegenüberstellung der Genomabdeckung für die 5'-Enden des ersten Introns von *clpP*. A: Optimiertes Protokoll, B: Terminome-Seq, modifiziert nach Castandet *et al.*, 2019. Der gebogene Pfeil zeigt die TSS, welche mit beiden Methoden identifiziert wurde. +5'PP: +5'PPase, -5'PP: -5'PPase. Ex1: erstes Exon, ex2: zweites Exon, die schwarze Linie zwischen den Exons stellt das erste Intron dar. -TAP/-5'PP stellen Hintergrundbibliotheken dar, welche prozessierte Transkripte enthalten.

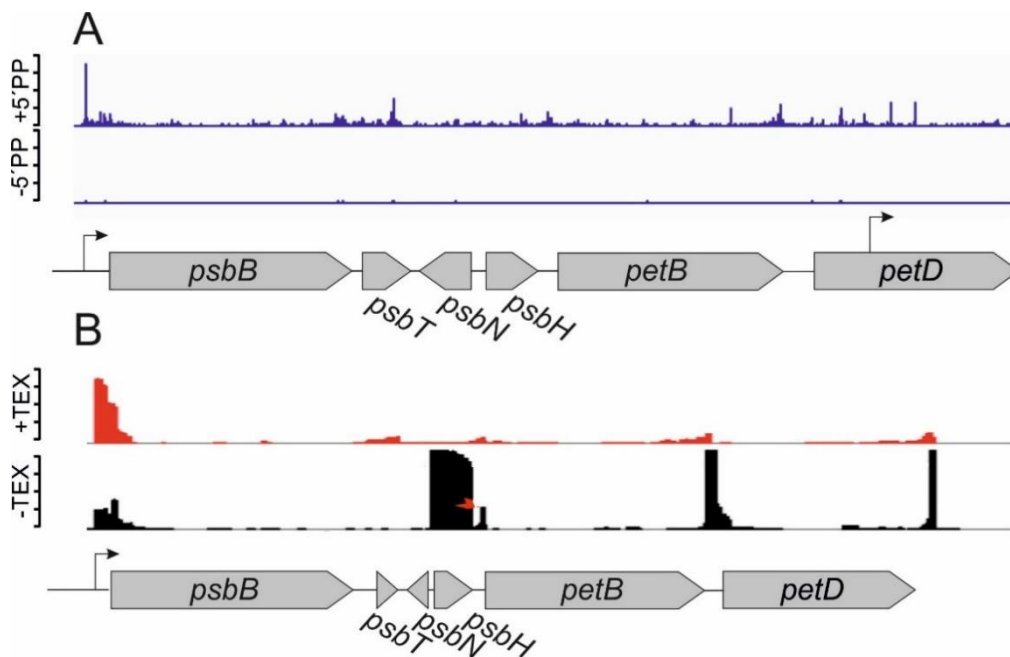


Abbildung VI.7: Genomabdeckung für das *psbB*-Operon. A: Optimiertes Protokoll, angewendet in *A. thaliana*, B: Originalprotokoll angewendet in Gerste, Bild modifiziert nach Zhelyazkova *et al.*, 2012. Der gebogene Pfeil zeigt die TSS an. +5'PP: +5'PPase, -5'PP: -5'PPase. Die Gene sind als graue Pfeile dargestellt. -TEX/-5'PP stellen Hintergrundbibliotheken dar, welche prozessierte Transkripte enthalten.

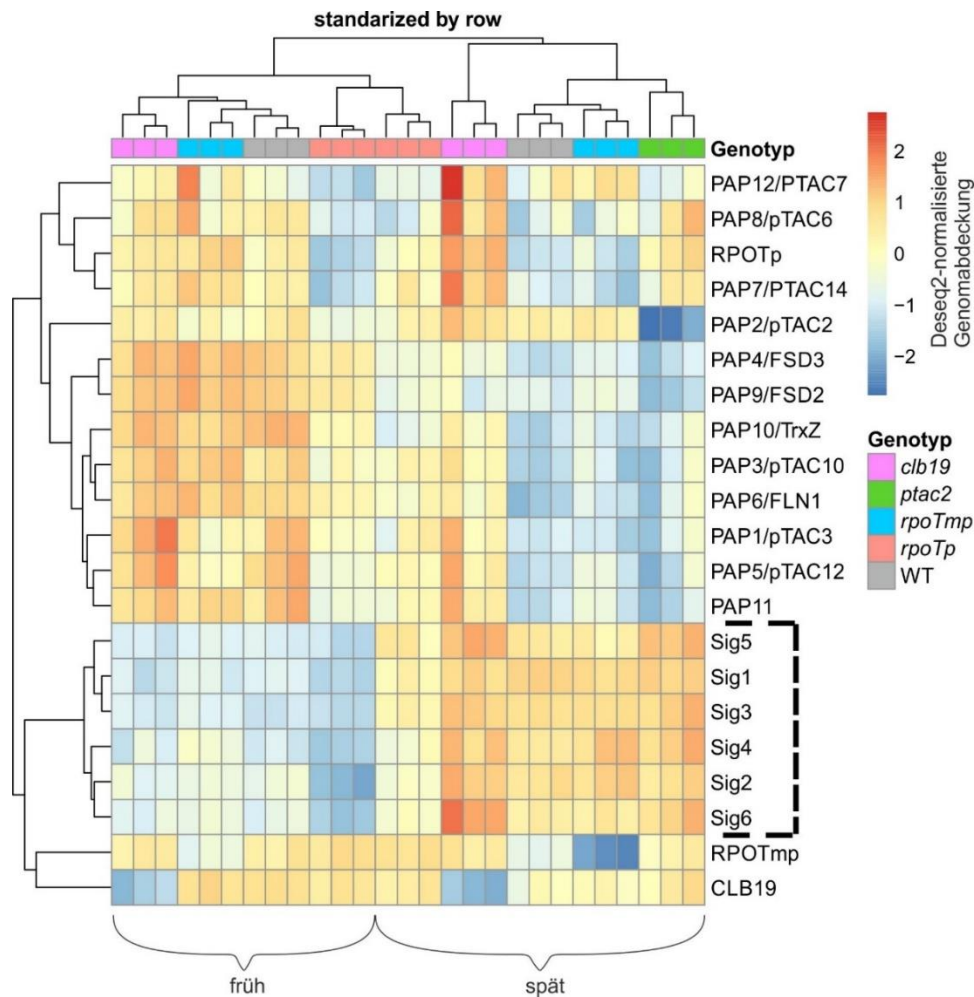


Abbildung VI.8: RNA-Seq-Analyse ausgewählter plastidenlokalisierter Kerngene in Wildtyp und Mutanten im frühen und späten Entwicklungsstadium. Angezeigt sind die Genomabdeckungen mit DESeq2-Normalisierung. Die unterbrochene Linie markiert die Expression der Sigma-Faktoren. PAP: PEP-assoziiertes Protein. Geno: Genotypen, die Farbkodierung ist auf der rechten Seite angegeben. Die Analyse wurde durch Jose Muino durchgeführt.

### VI.3 Bioinformatische Datenanalyse

#### TSS Identifizierung

```
run<-function(name, rep, n=-1, b=2, w=1L, nper=100) {

  bam<-paste(bamfolder, name, ".bam", sep="")
  stopifnot(file.exists(bam))

  bam <- scanBam(bam, param=param) [[1]]; names(bam) <- c("chr", "strand", "pos", "lengthRead", "mq");
  bam2<-as.data.frame(bam); bam2$Nhits=1;
  bam2$pos<-bam2$pos-1##to put it in 0-based coordinates (as genome browser)
  bam2$pos[bam2$strand=="-"]<-bam2$pos[bam2$strand=="-"]+bam2$lengthRead[bam2$strand=="-"]-1
  if(is.null(bam)) return(NA)
  ip=bam2

  name<-sub("sample", "control", name)
  controlbam <- paste(here("bamFiles/"), name, ".bam", sep="")
  bam <- scanBam(controlbam, param=param) [[1]]; names(bam) <-
  c("chr", "strand", "pos", "lengthRead", "mq");
  bam1<-as.data.frame(bam); bam1$Nhits=1;
  bam1$pos<-bam1$pos-1##to put it in 0-based coordinates (as genome browser)
  bam1$pos[bam1$strand=="-"]<-bam1$pos[bam1$strand=="-"]+bam1$lengthRead[bam1$strand=="-"]-1
  control<-bam1

  for(strand in c("Reverse", "Foward")){
```



## Anhang

```
chr=c("ChrC")#, "ChrM")
chrL=c(154478)#, 366924)
minlength=0
name<-sub("control","", name)
print(w)
rrna= -1;#if(strand=="Foward"){rrna<-
c(101012:102502,104691:107500,107599:107701,107949:108069)}else{rrna<-
c(130580:130700,130948:131050,131149:133958,136147:137637)}
nhitsS<- mappedReads2Nhits(
bam2[!is.na(bam2$pos) & bam2$lengthRead>minlength & !is.element(bam2$pos,rrna) ],,
"hitS", w = w, considerStrand = strand, uniquelyMapped = F, uniquePosition =
F, chr=chr, chrL=chrL
)
nhitsS$digits=0

score2wig(nhitsS, file=paste(name, "_sample", "Strand", strand, n, "b", b, "w", w, "minl", minlength, ".wig", sep=""), times=10000)
nhitsC<- mappedReads2Nhits(bam1[!is.na(bam1$pos) & bam1$lengthRead>minlength &
!is.element(bam1$pos,rrna) ],, "hitC", w = w, considerStrand = strand, uniquelyMapped = F,
uniquePosition = F, chr=chr, chrL=chrL)
nhitsC$digits=0

score2wig(nhitsC, file=paste(name, "_control", "Strand", strand, n, "b", b, "w", w, "minl", minlength, ".wig", sep=""), times=10000)
test3<-ChIPseqScore(control=nhitsC, sample=nhitsS, file="test3", norm=
n, backg=b, times=1000000, test="Ratio")
score2wig(test3, file=paste("wig/TestRatio-
", name, "Strand", strand, n, "b", b, "w", w, "minl", minlength, ".wig", sep=""), times=1000000, t=1)
win<-sigWin(test3, g=3)

if(!is.null(win)){
foreach(i=1:nper,
.packages='CSAR')%dopar%{permutatedWinScores(nn=i, sample=bam2[!is.na(bam2$pos) &
bam2$lengthRead>minlength & !is.element(bam2$pos,rrna) ],, control=bam1[!is.na(bam1$pos) &
bam1$lengthRead>minlength & !is.element(bam1$pos,rrna) ],, fileOutput="per",
backg=b, chr = chr, chrL = chrL, w = w, considerStrand = strand, uniquelyMapped = F,
uniquePosition = F, norm= n, g=3)}
nulldist<-getPermutatedWinScores(file="per", nn=1:nper)
unlink("*permutatedWin")
save(nulldist, win, file=paste("Perm-
", name, strand, n, "b", b, "w", w, "minl", minlength, ".RD", sep=""))
load(file=paste("Perm-", name, strand, n, "b", b, "w", w, "minl", minlength, ".RD", sep=""))
tr=getThreshold(winscores=values(win)$score, permutatedScores=nulldist, FDR=.05)
win1<-NULL;
win1<-as.data.frame(win)[values(win)$score>tr$threshold, c(1, 6, 6)]
print(head(win1))
if(dim(win1)[1]>0){
win1$posPeak<-win1$posPeak-5; win1$posPeak.1<-win1$posPeak.1+5; win1$posPeak[win1$posPeak<1]<-1
win1$name<-". "; win1$score<-". "
if(strand=="Foward"){win1$strand<-"+"}
if(strand=="Reverse"){win1$strand<-"-"}
write.table(win1, file=paste("beds-Step1/TestRatio-
", name, "Strand", strand, n, "b", b, "w", w, "minl", minlength, ".bed", sep=""), quote=F, sep="\t", col.name
s=F, row.names=F)
}
}
}
gc()
return("done")
}

ids<-list.files(pattern=".bam$", here("bamFiles"))
ids<-unique(sub(".bam","", ids))
ids<-ids[grep("sample", ids)]
cores=detectCores()
cl <- makeCluster(4) #not to overload your computer
registerDoParallel(cl)
what <- c("qwidth", "strand", "rname", "pos", "mapq")
param <- ScanBamParam( what = what)

for(nam in ids){print(nam); run(name=nam, rep=1)}
stopCluster(cl)
system(paste0('cat ', paste(list.files(outputfolder, full=T), collapse=' '), '>> merged.bed'))
system('sort -k1,1 -k2,2n -i merged.bed > mergeds.bed')
system('mergeBed -i mergeds.bed -iobuf 5G -s >merged-1b2w1minl0.bed')
```

## Anhang

```
q1<-read.table("merged-1b2w1min10.bed");q1$V1<-as.character(q1$V1)
gtf<-data.frame(q1$V1,"CSAR","gene",q1$V2,q1$V3,"1000",q1$V4,".",paste("gene_id \"",q1$V2,"\"";
transcript_id \"",q1$V2,"\"";",sep=""))
gtf[,9]<-as.character(gtf[,9]);gtf[,7]<-as.character(gtf[,7])
write.table(gtf,file="merged-1b2w1min10.gtf",quote=F,row.names=F,col.names=F,sep="\t")

q1<-read.table("merged-1b2w1min10.bed");q1<-q1[q1$V1=="ChrC",]
gff<-read.table(here("ext_data/Araport11_GFF3_genes_transposons.201606.gtf"),sep="\t")
gff1<-gff[gff$V1=="ChrC" & gff$V3=="gene",]
disF<-sapply(q1$V3[q1$V4=="+"],function(x){temp<-gff1$V4[gff1$V7=="+"]-
x;c(as.character(gff1$V9[gff1$V7=="+"]) [temp==min(temp[temp>0 & temp <1000])],NA) [1]})
disR<-sapply(q1$V2[q1$V4=="-"],function(x){temp<-gff1$V5[gff1$V7=="-"]-
x;c(as.character(gff1$V9[gff1$V7=="-"]) [temp==max(temp[temp<0 & temp > -1000])],NA) [1]})
fin<-data.frame(id=c(q1$V2[q1$V4=="+"],q1$V2[q1$V4=="-"]),gene=c(disF,disR))
fin$gene<- unlist(lapply(strsplit(as.character(fin$gene),split=" |"),function(x)x[2]))
write.csv(fin,file="annotationTSS1kb.csv")

SAMPLESCORETHRESHOLD <- 10
FDR_THRESHOLD <- 0.05
CONTROLSCORETHRESHOLD <- 5
ECTHRESHvalues <- c(1,1.5,2)

outdir <- paste0('run_',fdr_',FDR_THRESHOLD,'_pseudocount_',CONTROLSCORETHRESHOLD)
dir.create(showWarnings = F,outdir)
setwd(outdir)
library(tidyverse)
if('beds-Step2' %in% list.files(outdir)){
  list.files(outdir,full=T,rec=T)%>%file.remove
}

ids<-list.files(pattern=".counts$", "../countFiles")
res<-read.table(paste("../countFiles/",ids[1],sep=""),header=T)

for(id in ids[-1]){
  temp<-read.table(paste("../countFiles/",id,sep=""),header=T) [,7]
  res<-cbind(res,temp)
}

mid<-as.integer((res$Start+res$End)/2)
rrna<-sapply(1:length(res$Strand),function(x){if(res$Strand[x]=="+") {rrna<-
c(101012:102502,104691:107500,107599:107701,107949:108069)}
else{rrna<-c(130580:130700,130948:131050,131149:133958,136147:137637)};
is.element(mid[x],rrna)})
res<-res[!rrna,]
coor<-res[,1:6]
counts<-res[,7:length(res[,1])]
colnames(counts)<-ids;rownames(counts)<-res$Geneid
ww<-which(rowSums(counts[,grep("sample",colnames(counts))]>SAMPLESCORETHRESHOLD)>0)
write.csv(counts,file="counts.csv")

mdis=5000
gff<-read.table(here("ext_data/Araport11_GFF3_genes_transposons.201606.gtf"),sep="\t")
gff<-gff[gff$V3=="gene" & gff$V1=="ChrC",]
dis<-lapply(coor$Start[coor$Strand=="+"],function(x){temp<- x-
gff$V4[gff$V7=="+"];temp[abs(temp)<1000]})
disp<-lapply(coor$Start[coor$Strand=="+"],function(x){temp<- x-
gff$V4[gff$V7=="+"];names(temp)<-gff$V9[gff$V7=="+"];temp[temp==max(temp[(temp)< 0 & temp> -
mdis ])}))
names(disp)<-coor$Geneid[coor$Strand=="+"]
disn<-lapply(coor$Start[coor$Strand=="-"],function(x){temp<- gff$V5[gff$V7=="-"]-
x;names(temp)<-gff$V9[gff$V7=="-"];temp[temp==max(temp[(temp)< 0 & temp> -mdis ])}))
names(disn)<-coor$Geneid[coor$Strand=="-"]
temp<-sapply(as.character(coor$Geneid),function(x){temp<-
unlist(c(disp[x],disn[x]));c(names(temp)," transcript_id None;") [1]})
target<-unlist(lapply(strsplit(temp,"transcript_id "),function(x)x[2]))
coor$target<-sub(";",",",target)
coor$dis<-sapply(as.character(coor$Geneid),function(x){temp<-
unlist(c(disp[x],disn[x]));c(temp,mdis) [1]})

geno<-as.factor(unlist(lapply(strsplit(ids,"_"),function(x)x[1])))
dev<-as.factor(unlist(lapply(strsplit(ids,"_"),function(x)x[2])))
rep<-as.factor(unlist(lapply(strsplit(ids,"_"),function(x)x[3])))
control<-as.factor(unlist(lapply(strsplit(ids,"_"),function(x)x[4])))
cond<-as.factor(substr(unlist(lapply(strsplit(ids,"_"),function(x)x[3])),1,3))
```

```

colData<-
data.frame(cond=factor(cond), rep=as.factor(rep), geno=geno, dev=dev); rownames(colData)<-ids
save(counts, coor, file=file.path("Counts.RD"))

samp<-paste(geno, dev, sep="_")
for(i in unique(samp)){
  print(i)
  dds <- DESeqDataSetFromMatrix(countData = as.matrix(counts[, samp==i]), colData =
colData[samp==i,], design = ~ cond+rep)
  message(digest::digest(dds))

  dds1 <- DESeq(dds, fitType="local")
  res<-as.data.frame(results(dds1, contrast=c("cond", "sam", "con"), altHypothesis="greater"))
  res$pos<-coor$Start; res$pos2<-coor$End; res$strand<-coor$Strand; res$target<-
coor$target; res$dis<-coor$dis

  deseq_sig <- (res$padj<FDR_THRESHOLD & !is.na(res$padj) )
  sig<-res[deseq_sig, 7:10];
  sig$pos<-sig$pos
  sig$chr<- "ChrC"; sig$point<- "."; sig<-sig[, c(5, 1, 2, 4, 6, 3)]
  dir.create(file.path(("beds-Step2/")))

  write.table(sig, file=file.path(paste("beds-Step2/", i, "-
SigFDR.bed", sep="")), sep="\t", col.names=F, row.names=F, quote=F)
  write.csv(res, file=file.path(paste(i, "-Step2.csv", sep="")))
}

allbeds = Sys.glob(file.path(paste("beds-Step2/", "*", "-SigFDR.bed", sep=""))) %>%setNames(., .)
names(allbeds) %<>%basename%>%str_replace('\\-SigFD.*', '')
library(rtracklayer)
allbeddf<-allbeds%>%map(import) %>%GRangesList%>%unlist%>%unique
names(allbeddf)<-NULL
allbeddf$score = 0
allbeddf%>%export('beds-Step2/merge4julia.bed')

ids<-list.files(pattern="-Step2.csv")
res<-read.csv(ids[1], header=T) [, 8:11]
resf<-read.csv(ids[1], header=T) [, 8:11]
for(id in ids){
  print(id)
  temp<-read.csv(id, header=T) [, 7]
  res<-cbind(res, temp)
  tempf<-read.csv(id, header=T) [, 3]
  resf<-cbind(resf, tempf)
};
res[is.na(res)]<-1; colnames(res) [5:13]<-ids;

FCTHRESH=1
for(FCTHRESH in FCTHRESHvalues){
  deseq_sig <- (res$padj<FDR_THRESHOLD & !is.na(res$padj) )
  res1<-res[rowSums(res[, 5:13]<FDR_THRESHOLD & resf[, 5:13]>FCTHRESH )>0,];
  rownames(res1)<-paste(res1$pos, res1$target, sep="-")
  write.csv(res1, file.path(paste0("Targets-FDR-lFC", FCTHRESH, ".csv")))
  stopifnot(nrow(res1) <= length(allbeddf))
  res2<-res1[, 5:13]
  res2[res2<0.05]<-0
  res2[res2>=0.05]<-1
}

ids<-list.files(pattern="-Step2.csv")
res<-read.csv(ids[1], header=T) [, 8:11]
for(id in ids){
  print(id)
  temp<-read.csv(id, header=T) [, 3]
  res<-cbind(res, temp)
}; res[is.na(res)]<- -5; colnames(res) [5:13]<-ids
res2<-res[, 5:13]

colData<-
data.frame(cond=factor(cond), rep=as.factor(rep), geno=geno, dev=dev); rownames(colData)<-
colnames(counts)
dds <- DESeqDataSetFromMatrix(countData = counts, colData = colData, design = ~
cond+rep+geno+dev)
dds1 <- DESeq(dds, fitType="local")
vsd <- varianceStabilizingTransformation(dds1, fitType="local")

```

```
write.csv(cbind(coor, as.data.frame(assay(vsd))), file=file.path("ALLGenesVSDNormPV.csv"))
fin<-read.csv("ALLGenesVSDNormPV.csv");rownames(fin)<-fin$X;fin$X<-NULL

ids<-colnames(fin)
samples<-ids[grepl("sample",ids)]
controls<-sub("sample","control",samples)

res<-as.data.frame(lapply(1:length(samples),function(x){print(x);fin[,samples[x]]-
fin[,controls[x]]}))
colnames(res)<-samples;rownames(res)<-rownames(fin)
temp<-strsplit(colnames(res),"_\\.bam")
genotype<-unlist(lapply(temp,function(x)x[1]))
dev<-unlist(lapply(temp,function(x)x[2]))
cond0<-unlist(lapply(temp,function(x)x[3]));cond<-cond0
cond[grepl("sample",cond)]<- "sample";cond[grepl("control",cond)]<- "control"
colData<-data.frame(cond,dev,genotype);rownames(colData)<-colnames(res)
```

### Klassifizierung mit dem statistischen Modell

```
tssLocations <- read.csv("Targets-FDR-1FC1.5.csv", header = T, sep = ",")
counts <- read.csv("counts.csv", header = T, sep = ",")
sample_counts_early <- select(counts,contains("early_sample"))
poscounts <- counts$X
row.names(sample_counts_early) <- poscounts
sample_counts_early <- sample_counts_early[which(row.names(sample_counts_early) %in%
tssLocations$pos1),]
sample_counts_early <- log(sample_counts_early+1)
geneNames <- sample_counts_early$tssPos <- with(tssLocations,paste0(pos," ",tss," ",strand1))
sample_counts_early <- sample_counts_early[-13]
row.names(sample_counts_early) <- geneNames
pos <- rownames(sample_counts_early)

get_mu <- function(sample_counts_early,rowlength,range0,range1){
  tableMeans <- data.frame()
  if(is.null(range1)) {
    for(i in 1:rowlength) {
      tableMeans[i,1] = mean(as.numeric(sample_counts_early[i,range0]))
    }
  } else {
    for(i in 1:rowlength) {
      tableMeans[i,1] = mean(as.numeric(sample_counts_early[i,range0]))
      tableMeans[i,2] = mean(as.numeric(sample_counts_early[i,range1]))
      if(tableMeans[i,1] < tableMeans[i,2]) {
        tableMeans[i,1] = mean(as.numeric(sample_counts_early[i,]))
        tableMeans[i,2] = mean(as.numeric(sample_counts_early[i,]))
      }
    }
  }
  colnames(tableMeans) <- c("mu0","mu1")
  return(tableMeans)
}

get_sig <- function(sample_counts_early,rowlength,mu,range0,range1){
  tableSig <- data.frame()
  if(is.null(range1)) {
    for(i in 1:rowlength) {
      tableSig[i,1] = sqrt(mean(as.numeric(sample_counts_early[i,range0]-mu[i,1])^2))
    }
  } else {
    for(i in 1:rowlength) {
      tableSig[i,1] = sqrt(mean(c(as.numeric(sample_counts_early[i,range0]-
mu[i,1])^2,as.numeric(sample_counts_early[i,range1]-mu[i,2])^2)))
    }
  }
  return(tableSig)
}

get_likelihood <- function(sample_counts_early,rowlength,mu,sigma,range0,range1) {
  likelihood <- data.frame()
  if(is.null(range1)) {
    for(i in 1:rowlength) {
      likelihood[i,1] <- prod(dnorm(as.numeric(sample_counts_early[i,]),mu[i,1], sigma[i,1]))
    }
  } else {

```

## Anhang

```
for(i in 1:rowlength) {
  likelihood[i,1] <-
  prod(dnorm(as.numeric(sample_counts_early[i,range0]),mu[i,1],sigma[i,1]),dnorm(as.numeric(sample_counts_early[i,range1]),mu[i,2],sigma[i,1]))
}
}
return(likelihood)
}

get_logLikelihood <- function(sample_counts_early,rowlength,mu,sigma,range0,range1) {
  logLikelihood <- data.frame()
  if(is.null(range1)) {
    for(i in 1:rowlength) {
      logLikelihood[i,1] <- sum(dnorm(as.numeric(sample_counts_early[i,]),mu[i,1],sigma[i,1],log=T))
    }
  } else {
    for(i in 1:rowlength) {
      logLikelihood[i,1] <-
      sum(dnorm(as.numeric(sample_counts_early[i,range0]),mu[i,1],sigma[i,1],log=T),dnorm(as.numeric(sample_counts_early[i,range1]),mu[i,2],sigma[i,1],log=T))
    }
  }
  return(logLikelihood)
}

modelfinal <- function(sample_counts_early,range0) {
  rowlength <- nrow(sample_counts_early)
  n = ncol(sample_counts_early)
  k = 3

  range1 <- c(1:12)[-range0]
  if(length(range1)==0) {
    range1 <- NULL
    k = 2
  }
  mu <- get_mu(sample_counts_early,rowlength,range0,range1)
  sigma <- get_sig(sample_counts_early,rowlength,mu,range0,range1)
  likelihoodResults <- get_likelihood(sample_counts_early,rowlength,mu,sigma,range0,range1)
  logLikelihoodResults <-
  get_logLikelihood(sample_counts_early,rowlength,mu,sigma,range0,range1)
  bic <- logLikelihoodResults-(k/2 * log(n))
  result <-
  list("mu"=mu,"sigma"=sigma,"likelihood"=likelihoodResults,"logLikelihood"=logLikelihoodResults
, "BIC" = bic)
  return(result)
}

fall1 <- modelfinal(sample_counts_early, (1:12))
fall2 <- modelfinal(sample_counts_early, (4:12))
fall3 <- modelfinal(sample_counts_early, c(1:6,10:12))
fall4 <- modelfinal(sample_counts_early, c(4:6,10:12))
fall5 <- modelfinal(sample_counts_early, c(1:3,7:12))
fall6 <- modelfinal(sample_counts_early, (7:12))
fall7 <- modelfinal(sample_counts_early, c(1:3,10:12))
fall8 <- modelfinal(sample_counts_early, (10:12))

BIC <- cbind(fall1$BIC,fall2$BIC,fall3$BIC,fall4$BIC,fall5$BIC,fall6$BIC,fall7$BIC,fall8$BIC)
colnames(BIC) <- c("Fall1","Fall2","Fall3","Fall4","Fall5","Fall6","Fall7","Fall8")

likelihood <-
cbind(fall1$likelihood,fall2$likelihood,fall3$likelihood,fall4$likelihood,fall5$likelihood,fall6$likelihood,fall7$likelihood,fall8$likelihood)
colnames(likelihood) <- c("Fall1","Fall2","Fall3","Fall4","Fall5","Fall6","Fall7","Fall8")

logLikelihood <-
cbind(fall1$logLikelihood,fall2$logLikelihood,fall3$logLikelihood,fall4$logLikelihood,fall5$logLikelihood,fall6$logLikelihood,fall7$logLikelihood,fall8$logLikelihood)
colnames(logLikelihood) <- c("Fall1","Fall2","Fall3","Fall4","Fall5","Fall6","Fall7","Fall8")

n = ncol(sample_counts_early)
a.priori = c(.125,.125,.125,.125,.125,.125,.125,.125)
bayes.scewed.probabilities <- likelihood
bayes.scewed.probabilities$Fall1 <- bayes.scewed.probabilities$Fall1/(sqrt(n)^2)
bayes.scewed.probabilities[,-1] <- bayes.scewed.probabilities[,-1]/(sqrt(n)^3)
```

```

likehoodsData.a.posteriori.bayes <- matrix(0,
dim(bayes.scewed.probababilities) [1],dim(bayes.scewed.probababilities) [2])
for(i in 1:nrow(sample_counts_early)){
  likehoodsData.a.posteriori.bayes[i,] <-
as.numeric(bayes.scewed.probababilities[i,]*a.priori/sum(a.priori*bayes.scewed.probababilities[i,]
))
}

for(i in 1:nrow(sample_counts_early)){
  BIC[i,9] <- which.max(BIC[i,])
}
colnames(BIC) [9] <- "Winner"
row.names(BIC) <- geneNames

for(i in 1:nrow(sample_counts_early)){
  bayes.scewed.probababilities[i,9] <- which.max(bayes.scewed.probababilities[i,])
}
colnames(bayes.scewed.probababilities) [9] <- "Winner"
row.names(bayes.scewed.probababilities) <- geneNames

likehoodsData.a.posteriori.bayes <- as.data.frame(likehoodsData.a.posteriori.bayes)
for(i in 1:nrow(sample_counts_early)){
  likehoodsData.a.posteriori.bayes[i,9] <- which.max(likehoodsData.a.posteriori.bayes[i,])
}
colnames(likehoodsData.a.posteriori.bayes) <-
c("Fall1", "Fall2", "Fall3", "Fall4", "Fall5", "Fall6", "Fall7", "Fall8", "Winner")
row.names(likehoodsData.a.posteriori.bayes) <- geneNames

```

### Klassifizierung mit pamK

```

SAMPLESCORETHRESHOLD <- 10
for(FDR_THRESHOLD in c(0.05)){
  for(CONTROLSCORETHRESHOLD in c(5)){
    for(FC in c(1.5)){

setwd(here("runfdr_0.05_pseudocount_5"))
outdir <- paste0('run_', 'fdr_', FDR_THRESHOLD, '_pseudocount_', CONTROLSCORETHRESHOLD, '_late')
dir.create(showWarnings = F, outdir)

targetpeaks<-read.csv(paste(here("runfdr_0.05_pseudocount_5/Targets-FDR-
lFC"), FC, ".csv", sep=""))
targetpeaks$Geneid = targetpeaks$X%>%str_extract('\\d+')%>%as.character

countfiles<-list.files(pattern=".counts$", countsfolder, full=TRUE)
countfileinfo<-read.table(countfiles[1], header=T) [, -7]

res0<-read.csv(paste(here("runfdr_0.05_pseudocount_5/Targets-FDR-lFC"), FC, ".csv", sep=""))
fdr<-read.csv(paste(here("runfdr_0.05_pseudocount_5/Targets-FDR-
lFC"), FC, ".csv", sep="")); rownames(fdr)<-fdr$X; fdr$X<-NULL

counts = countfiles%>%
lapply(.%>%read_tsv(skip=1, col_types=cols())%>%select(7)%>%.[[1]])%>%
bind_cols%>%
set_colnames(basename(countfiles))%>%
as.matrix

rownames(counts)<-countfileinfo$Geneid
ww<-which(rowSums(counts[,grep("sample", colnames(counts))] >SAMPLESCORETHRESHOLD)>0)
for(i in grep("control", colnames(counts))){
  counts[counts[[i]]<CONTROLSCORETHRESHOLD, i]<-CONTROLSCORETHRESHOLD
}
colData = data.frame(
geno=colnames(counts)%>%str_extract('(clb19|ptac2| rpoTmp| rpoTp|wt)'),
dev=colnames(counts)%>%str_extract('(early|late)'),
rep=colnames(counts)%>%str_extract('\\d+(?=\\.bam)'),
time=colnames(counts)%>%str_extract('0\\+|05|dry|etio')%>%paste0('time_', .)
)%>%set_rownames(colnames(counts))

contrastnames = c('rpoTp', 'rpoTmp', 'ptac2')
resultslist = list(
results(dds1, contrast=c("geno", "rpoTp", "wt")),
results(dds1, contrast=c("geno", "rpoTmp", "wt")),
results(dds1, contrast=c("geno", "ptac2", "wt"))
)%>%setNames(contrastnames)
}

```

## Anhang

---

```
foldchangematrix =
resultslst%>%lapply(.%>%.$log2FoldChange)%>%bind_cols%>%as.matrix%>%set_rownames(rownames(dds
1))
foldchange_pvalues =
resultslst%>%lapply(.%>%.$padj)%>%bind_cols%>%as.matrix%>%set_rownames(rownames(dds1))
foldchangematrix = foldchangematrix[targetpeaks$Geneid,]
foldchange_stderr = foldchange_stderr[targetpeaks$Geneid,]
foldchange_pvalues = foldchange_pvalues[targetpeaks$Geneid,]

{
anno<-read.csv(here("ext_data/PlastidAnnotation.csv"),stringsAsFactors=F)
symbol<-
sapply(as.character(targetpeaks$target),function(x)c(anno$symbol[anno$TAIR_id==x],NA)[1])
type<-sapply(as.character(targetpeaks$target),function(x)c(anno$type[anno$TAIR_id==x],NA)[1])

foldchangematrix%>%dim
targetpeaks%>%dim
rownames(foldchangematrix)<-paste(targetpeaks$pos,targetpeaks$strand,symbol)
rownames(foldchangematrix)<-paste(targetpeaks$pos,targetpeaks$strand,symbol)

type[is.na(type)]<-"No_assigned"
rowData<-data.frame(type);rownames(rowData)<-rownames(foldchangematrix)
ann_colors=list(type=c(other="light
pink",Photosynthesis="green",Ribosome="burlywood4",RNA_polymerase="dark
red",rRNA="red",tRNA="dark blue",No_assigned="yellow"))
}
{
pdf(paste("LateFCvsWT-pamk-1FC",FC,".pdf",sep=""))
res2=foldchangematrix
cluster1=pamk(res2[,1:3],3:7,scaling=F)[[1]][[3]]
table(cluster1)
par(mfrow=c(3,3))
plotcluster(res2[,1:3],cluster1,main=paste("psbA is cluster",as.integer(cluster1["1515-
PSBA"])))
boxplot(res2[,1:3],main="All");abline(h=0,lwd=3,col="green")
for(i in 1:max(cluster1)){boxplot(res2[cluster1==i,1:3],main=paste("cluster",i,"\n
n=",sum(cluster1==i)));abline(h=0,lwd=3,col="green")}
write.csv(data.frame(res2,kmers=cluster1),file=paste("LateFCvsWT-pamk-FC",FC,".csv",sep=""))
dev.off()
normalizePath(paste("LateFCvsWT-pamk-1FC",FC,".pdf",sep=""))

write.csv(data.frame(foldchangematrix,kmers=cluster,pamk=cluster1),file="LateFCvsWT-
2Classifiers.csv")
cluster1%>%table
cluster1[names(cluster1)%>%str_detect('100867')]
write.csv(foldchangematrix,file="LateFCvsWT-all.csv")
}
```

### DANKSAGUNG

An dieser Stelle möchte ich mich bei all denen bedanken, ohne die die Fertigstellung dieser Arbeit nicht möglich gewesen wäre.

Zuerst möchte ich mich bei Prof. Dr. Kristina Kühn bedanken, die mir die Chance gegeben hat, dieses bereits fast auslaufende Projekt zu übernehmen und zu Ende zu führen. Ich danke Kristina für die Unterstützung in wissenschaftlichen und privaten Fragen, für die Diskussionsbereitschaft auch zwischen Tür und Angel und dass ihre Hilfe immer nur ein Türklopfen entfernt war.

Ich danke Prof. Dr. Christian Schmitz-Linneweber und Prof. Dr. Thomas Pfannschmidt für die Bereitschaft, diese Arbeit zu begutachten. Für die finanzielle Unterstützung danke ich dem FP7 People Programm von Marie Curie Actions.

Mein großer Dank gilt Antje Krippner, die das letzte Jahr meine rechte Hand im Labor war und mich kompetent und sorgfältig in vielen Experimenten unterstützt hat. Ihr Wissen zu Laborsicherheit und präzise Ausführung der Experimente hat unsere ganze AG bereichert. Für die bioinformatische Hilfe bedanke ich mich bei Dr. Jose Muino, Dr. Dermot Harnett, Frederic Ehlers und Prof. Dr. Ivo Große. Ich danke Jose und Dermot für die Analyse meiner Daten und für die vielen Lehrstunden in „Bioinformatics for Dummies“. Ich danke Frederic, der die finale Datenanalyse abgeschlossen hat und unermüdlich an den Parametern gedreht hat, bis ich zufrieden war. Großer Dank an Ivo für die tollen Diskussionen. Dr. Gerd Hause danke ich für die elektromikroskopischen Aufnahmen und dafür, dass Chloroplasten seine Lieblingsobjekte sind.

Ich möchte mich auch bei meinen Kollegen in Berlin und Halle für diese tollen Jahre zusammen bedanken. Ich danke Maria für die gemeinsamen Mittagessen und die Shoppingtouren, Lea für immer lustige Atmosphäre in unserem kleinen Lab 14 mit Musik und Klopsi. Rebekka danke ich für große und kleine Hilfen im Laboralltag. Ebenfalls gilt mein Dank Dr. Andreas Richter und Dr. Hannes Ruwe für Ihre Hilfestellungen in computertechnischen und bioinformatischen Fragen. Ich danke Theresa für die mühsame Aufgabe des Korrekturlesens dieser Arbeit und die konstruktiven Verbesserungsvorschläge. Für den regen wissenschaftlichen Austausch zu NGS und spannenden Sachen in der Organellentranskription danke ich Dr. Omar Saleh und Sarlita. Herzlichst bedanke ich mich bei Micha „MacGyver“ Röser, der viele Stunden mit der Rettung meines Computers verbracht hat und auch bei Internetproblemen immer zur Hilfe eilte.

Großer Dank geht an meine Familie. Ich danke Philipp, der all diese Jahre an meiner Seite war und meine Entscheidungen unterstützt hat, der mir nach Halle gefolgt ist und mit geholfen hat, dieses aufregende Lebenskapitel abzuschließen, sei es durch Einkaufen gehen oder immer für eine Umarmung da sein. Meinen Schwiegereltern danke ich für die beste Kinderbetreuung, die man sich vorstellen kann. Ich danke Papa für seine Gene für Sturheit und Durchsetzungsvermögen, die es mir ermöglicht haben, diese Arbeit zu Ende zu führen. Auch wenn er mich nicht mehr auf diesem Weg begleiten konnte, haben doch seine Ratschläge und sein Glaube an meine Fähigkeiten mir immer zur Seite gestanden.



LEBENS LAUF UND PUBLIKATIONS LISTE

|                               |  |
|-------------------------------|--|
| <b>Persönliche Daten</b>      |  |
| Name                          | Julia Rott   |
| Geburtsdaten                  | 29.09.1988   |
| <b>Beruflicher Werdegang</b>  |  |
| 01/2019-10/2021               | Martin-Luther-Universität Halle (Saale),<br>Wissenschaftliche Mitarbeiterin  |
| 06/2018-01/2019               | Elternzeit   |
| 09/2015-05/2018               | Humboldt Universität Berlin,<br>Wissenschaftliche Mitarbeiterin  |
| <b>Akademische Ausbildung</b> |  |
| 10/2012-10/2014               | Freie Universität Berlin, Studiengang: Biologie<br>Schwerpunkt: Molekular- und Zellbiologie<br>Abschluss: Master of Science mit Note 1,4<br>Thema der Masterarbeit:<br>„Versuche zur Verbesserung der Löslichkeit der<br>Transposase des transponierbaren Elements Activator<br>(Ac) aus Mais“ |
| 10/2009-07/2012               | Universität Potsdam, Studiengang: Biowissenschaften<br>Schwerpunkt: Molekular- und Zellbiologie<br>Abschluss: Bachelor of Science mit Note 1,8<br>Thema der Bachelorarbeit:<br>„Verbesserung der Genomannotation in <i>Arabidopsis thaliana</i> durch RNA-Sequenzierung“                       |
| <b>Schulabschluss</b>         |  |
| 08/2002-06/2009               | Hermann-Ehlers-Gymnasium in Berlin, Abschluss:<br>Abitur   |
| <b>Publikationen</b>          |  |
|                               | Omar Saleh, Sarlita Dwiani, <u>Julia Rott</u> , and Kristina Kühn<br><br>Nascent transcript sequencing for the mapping of promoters in <i>Arabidopsis thaliana</i> mitochondria (Methods in Molecular Biology, 2021, im Druck)   |

## EIDESSTATTLICHE ERKLÄRUNG

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe und nur unter Verwendung der angegebenen Quellen und Hilfsmittel verfasst habe. Die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen wurden als solche kenntlich gemacht.

Mit dieser Arbeit bewerbe ich mich erstmals um die Erlangung des Doktorgrades. Diese Arbeit wurde an keiner anderen Universität zur Begutachtung eingereicht.

Halle/Saale, 26.03.2021