 doi:10.20944/preprints202108.0571.v1

Article

Full length transcriptome highlights the coordination of plastid transcript processing

Marine Guilcher[1,2], Arnaud Liehrmann[1,2,3], Chloé Seyman[3], Thomas Blein[1,2], Guillem Rigaill[1,2,3], Benoit Castandet[1,2], Etienne Delannoy[1,2]*

[1] Université Paris-Saclay, CNRS, INRAE, Univ Evry, Institute of Plant Sciences Paris-Saclay (IPS2), 91405, Orsay, France
[2] Université de Paris, CNRS, INRAE, Institute of Plant Sciences Paris-Saclay (IPS2), 91405, Orsay, France
[3] Laboratoire de Mathématiques et de Modélisation d'Evry (LaMME), Université d'Evry-Val-d'Essonne, UMR CNRS 8071, ENSIIE, USC INRAE
* Correspondence: etienne.delannoy@inrae.fr

Abstract: Plastid gene expression involves many post-transcriptional maturation steps resulting in a complex transcriptome composed of multiple isoforms. Although short read RNA-seq has considerably improved our understanding of the molecular mechanisms controlling these processes, it is unable to sequence full-length transcripts. This information is however crucial when it comes to understand the interplay between the various steps of plastid gene expression. Here, the study of the Arabidopsis leaf plastid transcriptome using Nanopore sequencing showed that many splicing and editing events were not independent but co-occurring. For a given transcript, maturation events also appeared to be chronologically ordered with splicing happening after most sites are edited.

Keywords: Arabidopsis thaliana; plastid; co-maturation; post-transcriptional; Nanopore

1. Introduction

Plastids are derived from the endosymbiosis between photosynthetic organisms and an ancestral Eukaryote. Although most of the initial symbiont genes have been transferred to the nucleus during the course of evolution, plastids of land plants and other photosynthetic Eukaryotes still maintain a small but essential genome. It mainly encodes subunits of each of the photosynthetic complexes (Photosystem I and II, cytochrome b6/f, ATP synthase and Rubisco) and some of the plastid gene expression (PGE) machinery [1]. Most of the proteins involved in PGE are however encoded in the nucleus and need to be targeted back to plastids. As a consequence, PGE retains characteristics from both eukaryotes and bacterial systems, resulting in a sophisticated interplay between nucleus and plastid encoded factors [2–4].

A striking feature of PGE is the importance and complexity of the post-transcriptional maturation steps. In addition to the intron removal by RNA splicing [5] and the specific conversion of cytosines into uridines by RNA editing [6], complete maturation also requires intergenic cleavage of the multigenic transcripts and the generation of 5′ and 3′ ends through RNA processing [7,8]. Most of the RNA binding proteins (RBP) or ribonucleases known to be involved in PGE are localized in a membrane less structure surrounding the plastome, the nucleoid [9]. This close association between RNA maturation factors might be an explanation to the multiple pleiotropic effects observed in chloroplast mutants [7].

Various investigations, both *in vitro* and in organellar gene expression mutant plants, have indeed revealed situations where the different maturation events can influence each other. For example, intron removal is a prerequisite for editing in the *ndhA* second exon [10] and *atpF* splicing is severely reduced in the *aef1* mutant in which the editing of *atpF_92* (C12707) is abolished [11]. *Arabidopsis thaliana* chloroplast RNA editing

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2021 doi:10.20944/preprints202108.0571.v1

is affected in a mutant deficient for the ribonucleases PNPase [12] while correct processing of the potato mitochondrial tRNA Phe requires RNA editing [13]. Editing sites can even influence each other. For example, in *A. thaliana*, editing of mitochondrial *ccmB_566* (M17869) by MEF19 depends on the editing of *ccmB_551* (M17884) by MEF37 [14]. Similarly, in *Physcomitrium patens*, editing of the mitochondrial *ccmFc-C103* by PpPPR_65 controls editing of *ccmFc-C122* by PpPPR_71 [15].

Most of these studies, however, focus on a limited set of transcripts or RNA maturation events. This precludes any general conclusions and illustrates the need for the development of global approaches capable to simultaneously study all the RNA maturation processes, at the genomic level. This issue has recently been tackled by the increasing use of Illumina based RNA-seq strategies to study PGE from transcription to translation [16–22].

Although this has considerably increased the power and sensitivity of PGE analyses, it is ill-suited to study the potential coordination between maturation steps. The short reads used by the Illumina technology (the maximum insert size of Illumina TRUseq RNA libraries reaches around 350 base-pairs) make it impossible to monitor the co-occurrence of these events on single RNA transcripts that can be several kilo bases long. An alternative would be to take advantage of other sequencing technologies such as PacBio or Oxford Nanopore. They theoretically allow the sequencing of full length cDNAs or RNA and should therefore overcome the current technical limitations [23]. A major issue, however, is that most of the available library preparation protocols only capture polyadenylated RNA transcripts therefore excluding plastid transcripts. A recent protocol analyzing chromatin-bound transcripts also captures non polyadenylated transcripts but was not applied to the analysis of plastid transcripts [24,25].

In this work, we describe the analysis of the *A. thaliana* plastid transcriptome by sequencing of full length non polyadenylated cDNAs using the Oxford Nanopore technology. This analysis identified all known post-transcriptional maturation events and provided an overview of their coordination in normal growth conditions.

2. Results

2.1. A protocol to sequence the full length plastid transcriptome

Starting with a 3′ end RNA ligation step, we were able to sequence the non polyadenylated plastid transcripts. We mapped between 1,55M and 2,69M stranded reads (mapping rate between 98.5% and 99.8%) to the *A. thaliana* genome including between 10% and 40% to the plastid genome and between 0.3% and 0.8% to the mitochondrial genome. The median error rate was between 4 and 4.4%. The rRNA depletion was very efficient with less than 0.1% of reads mapping to rRNA loci. Based on the nuclear genome annotation, for the 3 biological replicates, the 3′-5′ bias was negligible for transcripts below 1500nt (22853 genes) and only moderate above (17985 genes; Figure S1). More than 99.5% of the reads mapped to the annotated nuclear genes corresponded to the sense orientation, a proportion similar to Illumina stranded RNA-seq. Most of the reads (99%) were between 195 and 2141 bases long with a median size of 852 bases and a maximum size of 4805 bases. In *A. thaliana*, more than 390 genes (including the plastid *ycf2* gene) are producing transcripts longer than 4800 nucleotides (nt). These results confirm that our Nanopore reads were mostly full-length and stranded but suggest that the longest transcripts are missing from the sequencing libraries.

2.2. A representative picture of the plastid transcriptome

With at least 275000 reads mapped on the plastid genome for each biological replicate, the coverage is deep enough to have a good representation of the plastid transcriptome. To verify that the sequencing data is correctly capturing the plastid transcriptome, we looked at the complex transcriptional profile of the *psbB* to *petD* genomic region (Figure 1).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2021 doi:10.20944/preprints202108.0571.v1

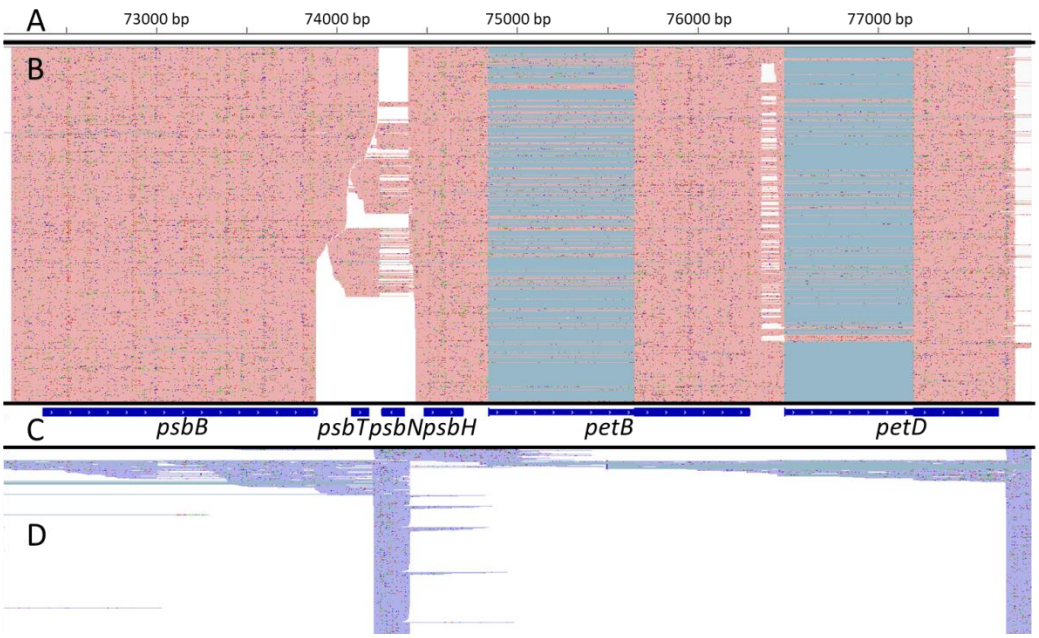

Figure 1. Complexity of the *psbB-petD* locus. Screenshots of IGV displaying Nanopore reads mapping to the *psbB-petD* locus. A. plastid genomic position. B. screenshot of reads mapping on the Watson strand. Matching bases are shown in red. Split reads are joined by blue lines. C. Annotation of the locus. Introns are shown as thinner segments. D. screenshot of reads mapping on the Crick strand. Matching bases are shown in purple. Split reads are joined by blue lines.

Following transcription, transcripts from this multigenic locus are processed into multiple poly- or monocistronic isoforms on both genomic strands [26,27]. A rapid overview of the reads showed the transcription of *psbN* on the Crick strand while *psbB*, *psbT*, *psbH*, *petB* and *petD* were transcribed from the Watson strand as expected. The spliced *petD* and *petB* transcripts were also found. Taking advantages of long read sequencing, it is possible to emulate Northern blots by selecting reads which map on specific positions and plotting the distribution of the read lengths. We generated virtual Northern blots for *psbN*, *psbH*, *petB* and *petD* (Figure 2) using virtual probes equivalent to the probes used for figure 4 I, C, E and H of Felder et al. [26].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2021 doi:10.20944/preprints202108.0571.v1

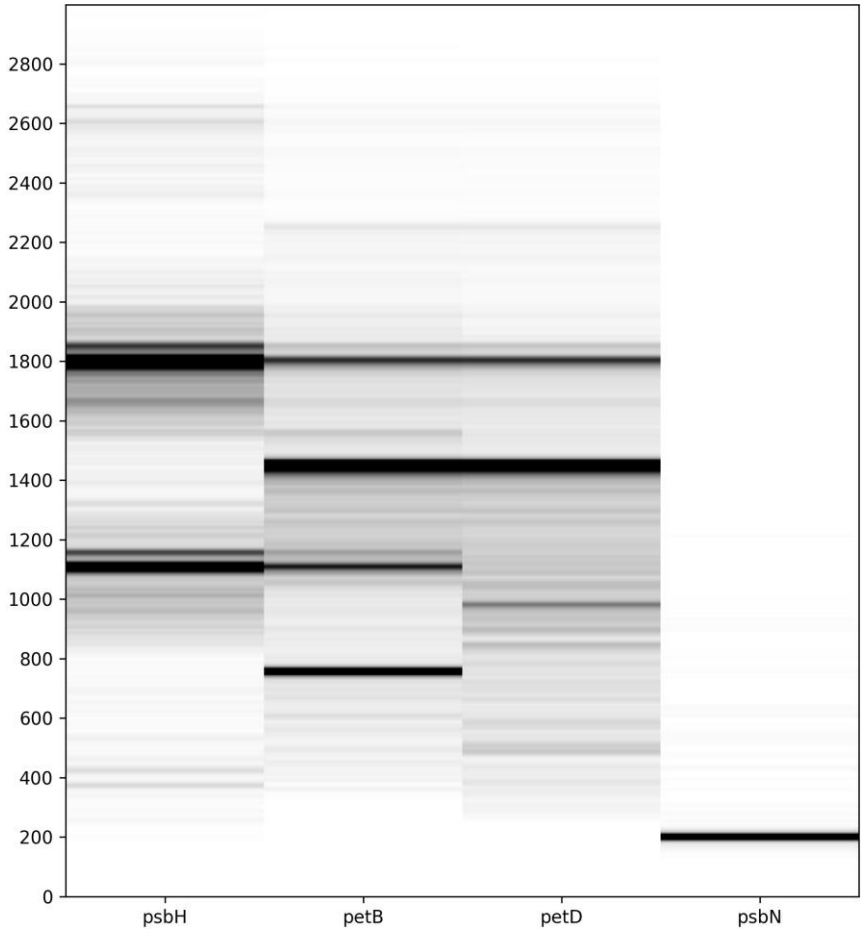

Figure 2. Virtual Northern blots derived from the Nanopore sequencing. Northern blots were emulated from Nanopore reads mapping to *psbN*, *psbH*, the second exon of *petB* or the second exon of *petD*. The size (in nt) is shown on the left.

Reads mapping to *psbN* were almost exclusively 200nt long which is compatible with the signal detected by the regular Northern blot. Reads mapping to *psbH* showed two major isoforms around 1100 nucleotides (nt) and 1800 nt but also two minor isoforms around 370 nt and 2600 nt. This profile is also compatible with the regular Northern blot. However Felder et al. [26] also detected larger isoforms at 3300, 4100, 4900 and 5600 nt that were not captured in our sequencing libraries. The virtual Northern blot for *petB* showed four major isoforms at 750, 1100, 1450 and 1800 nt. A faint isoform may be present at 2250 nt. These isoforms were also detected by Felder et al. [26] who found additional isoforms at 2600, 3300, 4100, 4900 and 5600 nt. Finally, for *petD*, we found two major isoforms around 1450 and 1800 nt and minor isoforms around 990 and 2225 nt. We missed the larger isoforms detected by Felder et al. [26] but also a 1200 nt described as an unspliced *petD* transcript which seemed to be replaced by our 990 nt isoform. This result confirms that our sequencing protocol missed the longest transcripts. However, in these complex loci it is sometimes difficult to identify all the bands on a regular Northern blot. For example, Felder et al. did not associate the 2.2kb transcript of their *petB* and *petD* Northern blots to a particular isoform.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2021 doi:10.20944/preprints202108.0571.v1

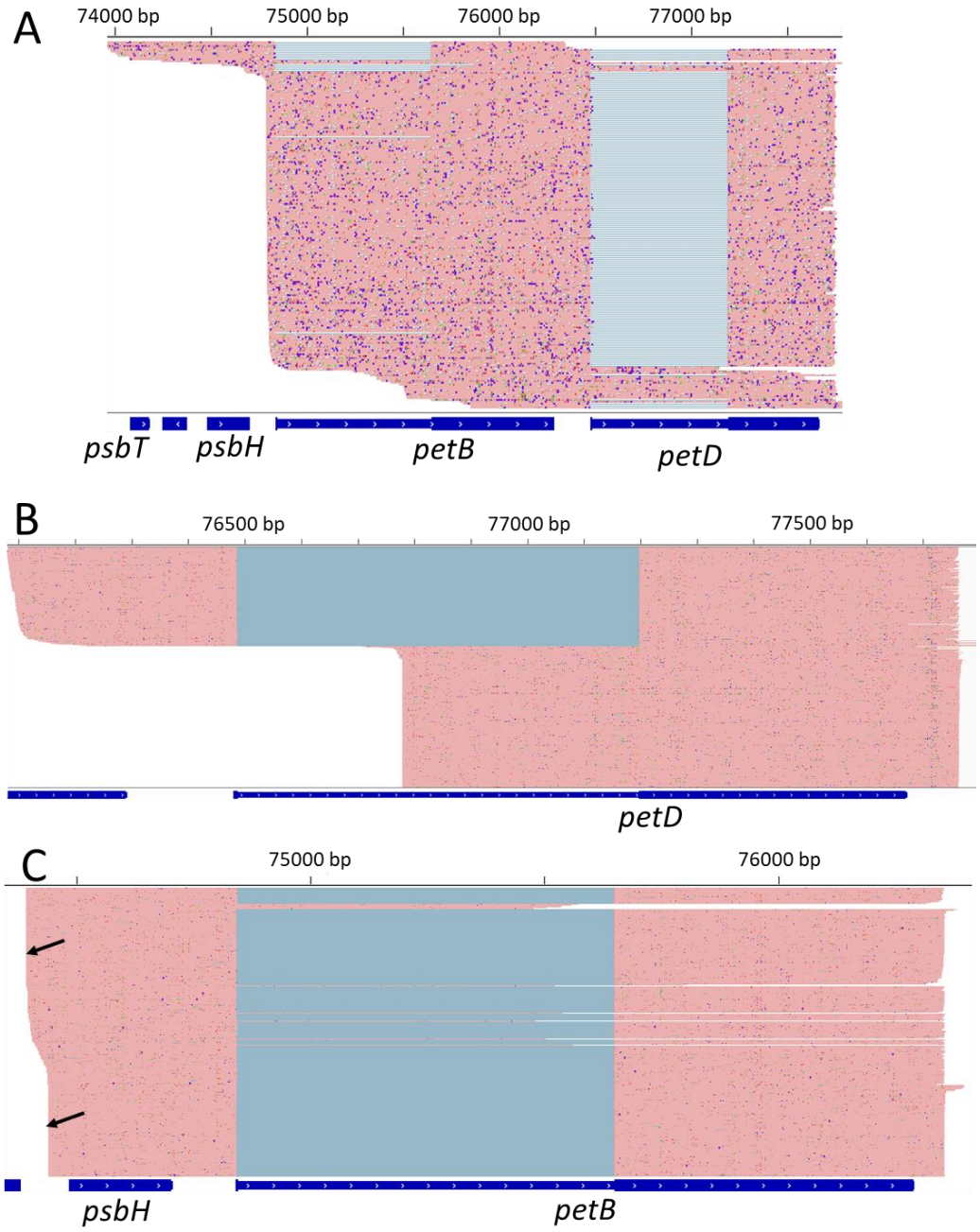

Figure 3. Identification of transcripts isoforms. Screenshots of IGV displaying the reads corresponding to various virtual Northern isoforms. A. Reads corresponding to the 2.2kb isoform of the *petB* and *petD* virtual Northern blots. B. Reads corresponding to the 990 nt isoform of the *petD* virtual Northern blot. C. Reads corresponding to the 1100-1150 isoform of the *psbH* virtual Northern blot. The two 5′ ends are shown by black arrows.

Our sequencing showed that this transcript is most likely a polycistronic intermediate containing an unspliced *petB* with a spliced *petD* (Figure 3A) which Felder et al. associated with a 2600nt isoform [26]. For *petD*, we detected a minor isoform around 990nt. The associated transcripts actually corresponded to 2 distinct isoforms (Figure 3B) transcripts. The first one corresponded to spliced *petD* transcripts but with 5′ ends within the second *petB* exon. The second one had a 5′ end in the *petD* intron at position 76780 and included the second *petD* exon. Position 76780 was identified as a transcription start site and multiple 5′ ends were mapped in this area [19]. Similarly, because of their poor resolution, regular Northern blots can miss isoforms of similar sizes. Our virtual Northern blot for *psbH* showed that the 4 peaks are actually double peaks: the main isoforms are each associated with isoforms which are 50 nt longer. When mapping these isoforms, we could show that the short and long isoforms are associated with different 5′ ends, the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2021 doi:10.20944/preprints202108.0571.v1

long one around the genomic position 74393 and the short one around 74441 (Figure 3C). According to Castandet et al. [19], position 74441 corresponds to the major processed extremity of *psbH* while position 74393 is a transcription start site.

Finally, post-transcriptional maturations events can be quantitatively analyzed. Known editing sites could be detected with rates comparable to leaf datasets (Table 1, Table S1) previously published by Guillaumot et al. [28] ($R^2 = 0.97$; p-value<$2.2\ 10^{-16}$) and Ruwe et al. [12] ($R^2 = 0.94$; p-value<$2.2\ 10^{-16}$).

Table 1. Quantification of known editing and splicing events

name	type	maturation rate	maturation rate (Guillaumot et al. 2017)	maturation rate (Ruwe et al. 2013)
int_RPS16	splicing	4%	4%	NA
int_ATPF	splicing	89%	82%	NA
int_RPOC1	splicing	64%	19%	NA
int_YCF3_i2	splicing	79%	42%	NA
int_YCF3_i1	splicing	63%	45%	NA
int_CLP_i2	splicing	60%	71%	NA
int_CLP_i1	splicing	69%	62%	NA
int_PETB	splicing	91%	58%	NA
int_PETD	splicing	97%	62%	NA
int_RPL16	splicing	69%	12%	NA
int_RPL2.1	splicing	66%	52%	NA
int_NDHB.1	splicing	68%	55%	NA
int_RPS12C	splicing	92%	81%	NA
int_NDHA	splicing	68%	27%	NA
matK(2931)	editing	53%	79%	93%
atpF(12707)	editing	89%	91%	95%
atpH_UTR(13210)	editing	5%	3%	4%
rpoC1(21806))	editing	33%	21%	15%
rpoB(23898)	editing	87%	82%	85%
rpoB(25779)	editing	64%	83%	86%
rpoB(25992)	editing	69%	76%	94%
psbZ(35800)	editing	93%	90%	95%
rps14(37092)	editing	89%	93%	94%
rps14(37161)	editing	92%	97%	96%
ycf3_i2(43350)	editing	16%	10%	12%
rps4_UTR(45095)	editing	6%	3%	10%
ndhK_ndhJ(49209)	editing	4%	4%	6%
accD(57868)	editing	90%	95%	99%
accD(58642)	editing	76%	75%	83%
psbF(63985)	editing	90%	98%	98%
psbE(64109)	editing	95%	100%	100%
petL(65716)	editing	79%	91%	86%
rps18_UTR(68453)	editing	3%	4%	NA
rps12(69553)	editing	21%	26%	27%

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2021 doi:10.20944/preprints202108.0571.v1

clpP(69942)	editing	82%	72%	81%
rpoA(78691)	editing	78%	76%	91%
rpl23(86055)	editing	34%	74%	75%
ycf2_as(91535)	editing	3%	4%	NA
ndhB_UTR(94622)	editing	8%	0%	NA
ndhB(94999)	editing	88%	93%	94%
ndhB(95225)	editing	95%	98%	99%
ndhB(95608)	editing	87%	84%	80%
ndhB(95644)	editing	78%	87%	81%
ndhB(95650)	editing	88%	91%	84%
ndhB(96419)	editing	75%	94%	92%
ndhB(96439)	editing	6%	4%	6%
ndhB(96457)	editing	6%	3%	5%
ndhB(96579)	editing	90%	89%	90%
ndhB(96698)	editing	81%	88%	82%
ndhB(97016)	editing	94%	94%	95%
ndhF(112349)	editing	85%	93%	96%
ndhD(116281)	editing	76%	83%	92%
ndhD(116290)	editing	77%	84%	90%
ndhD(116494)	editing	88%	90%	93%
ndhD(116785)	editing	94%	97%	98%
ndhD(117166)	editing	35%	33%	45%
ndhG(118858)	editing	69%	78%	85%

NA: Not Analyzed

It should be noted that the analysis of poorly edited sites by Nanopore sequencing must be done carefully because of the relatively high error rate of this technology. We detected 123 sites with an editing rate above 10% but only 44 were also detected by Guillaumot et al. [28] using Illumina Sequencing. Similarly intron splicing efficiency could be measured (Table 1, Table S1) and it varied from 4 to 97% depending of the intron. Most values are higher (by 22 points on average) than the efficiencies measured by Guillaumot et al. [28] probably because the abundance of unspliced transcripts is difficult to estimate with Illumina sequencing and the required approximations may over-estimate it.

2.3. Some post-transcriptional events are coordinated and ordered

As editing and splicing events are well defined (a single genomic position, either processed or not), we focused on the possible coordination between these events.

With 14 splicing and 43 editing events analyzed, up to 1596 co-occurrences of events could theoretically be expected. Only 138 co-occurrences were detected at least once which is expected because all events are not found on a single transcript (Table S2). Out of these 138 pairs of maturation events, 42 were not found to occur independently (Figure 4). We observed dependencies between splicing events (*clpP* introns, *petD* and *petB* introns), editing and splicing events (in the *atpF*, *clpP* and *ndhB* transcripts) and between editing events (in the *rps14*, *ndhD* and *ndhB* transcripts). They also occurred between different genes (*petD* and *petB*; *psbE* and *psbF*) belonging to polycistronic transcripts. The sites of coordinated events like *ndhD(116290)* and *ndhD(116281)*, *ndhB(95650)* and *ndhB(95644) or ndhB(95419)* and the *ndhB* intron could be very close but the others were separated by more than 100 nt.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2021 doi:10.20944/preprints202108.0571.v1

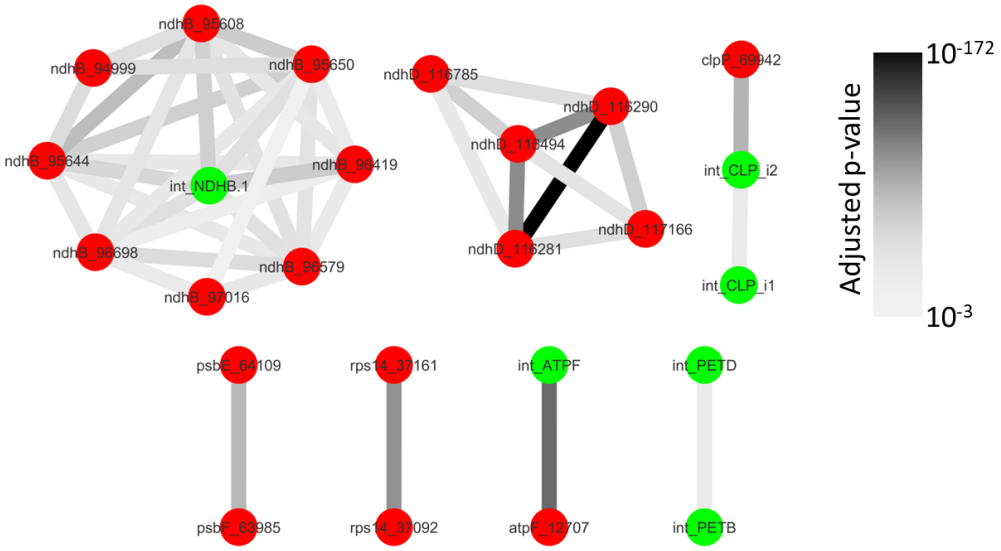

Figure 4. Network of splicing and editing coordination. Splicing events are shown in green and editing events in red. Dependent events are joined by an edge. The darkness of the edge is proportional to the adjusted p-value of the Exact Fisher test for the pool of the three replicates.

A more detailed analysis shows that intermediates of maturation (TF and FT columns of supp. Table 2) were always less frequent than expected for independent events for the 42 pairs of dependent events showing that when one site was processed the second one was more processed than expected randomly. In other words, there was co-maturation but no incompatibility. Furthermore, comparing the abundance of the intermediates of maturation offers the opportunity to order the maturation events (Figure 5).

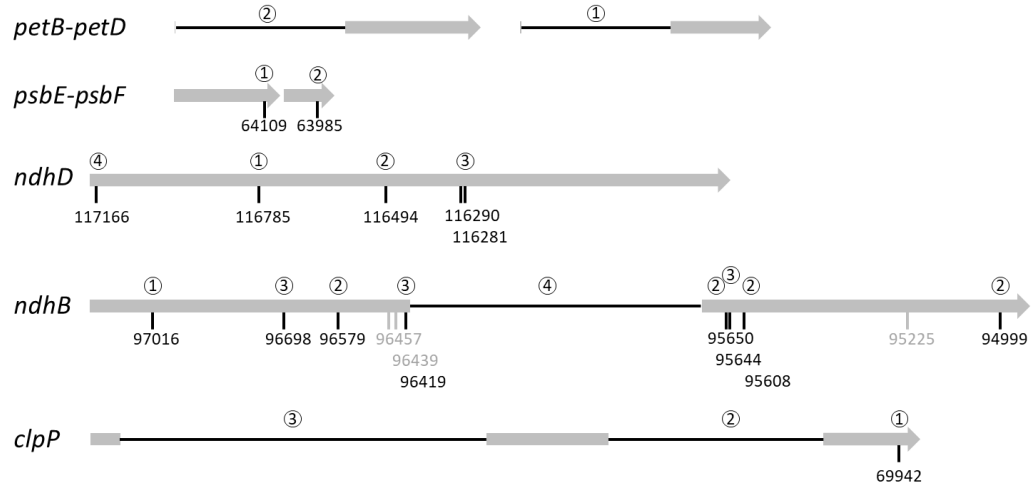

Figure 5. Proposed chronology of maturation events. Exons are shown as grey bars and introns as black lines. The editing sites are indicated by their genomic position. Grey editing sites are processed independently and thus are not included in the chronology. The order of the maturation events is indicated by the numbers above the editing sites or introns.

This analysis suggests that RNA editing at *psbE(64109)* generally occurred before RNA editing at *psbF(63985)* and that the splicing of *petD* occurred before the splicing of *petB*. The maturation of *clpP* started with RNA editing at *clpP(69942)* followed by the splicing of the second intron and finished by the splicing of the first intron. For *ndhD*, the maturation started with RNA editing at *ndhD(116785)* followed by *ndhD(116494)* then both *ndhD(116290)* and *ndhD(116281)* to finish with *ndhD(117166)*, the editing site creating the start codon of *ndhD*. For the *ndhB* transcript, the chronology of the maturation seemed more convoluted as three sites (96457, 96439 and 95225) were edited inde-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2021 doi:10.20944/preprints202108.0571.v1

pendently of the other maturation events. RNA editing at *ndhB(97016)* seemed to occur first followed by editing at the four sites 96579, 95650, 95608 and 94999. The maturation of *ndhB* ended with RNA editing at sites 96698, 96419, 95644 and, probably slightly later, its splicing. To confirm the order deduced from the co-occurrence analysis for transcripts requiring more than 3 maturation events (i.e. *ndhD* and *ndhB*), we identified the reads covering all the maturation events and counted the frequency of the various intermediates (Table S3). Out of 413 intermediate reads, 311 were compatible with the proposed chronology of *ndhD* maturation. For *ndhB*, only 63 intermediate reads were identified. This number is too small to estimate the frequency of the 4096 possible intermediates (12 maturation events) but 35 were compatible with the proposed chronology.

3. Discussion

Following transcription, plastid transcripts undergo a complex array of modifications and maturation and the recent massive use of RNA-Seq based strategies has led to an unprecedented knowledge about its different steps. What is sorely lacking, however, is a global understanding of the interplay between RNA editing, splicing and processing.

Initially thought to be mainly independent [29,30] there is now growing evidences for a crosstalk between the different maturation steps [10,31–34]. Most of these results have however been obtained from experiments based on Sanger sequencing of a cDNA of interest, therefore limiting any potential generalization. Taking advantage of the development of nanopore sequencing, we here systematically studied the link between individual RNA splicing and RNA editing events, at the plastome level.

Our results show that even if we could detect many intermediates of maturation, they were less abundant than expected for independent events suggesting that the various events of maturation are occurring at the same time or in a relatively short succession. This implies that all the actors of these different events are grouped and probably translates the co-localization of numerous maturation factors in the nucleoid [9].

Looking at specific links, splicing of the *atpF* intron and RNA editing at the *atpF_92* site are clearly dependent (Figure 4). This was expected as AEF1, the PPR protein responsible for *atpF_92* editing in *A. thaliana*, also facilitates *atpF* splicing [11]. Similarly, *clpP* intron 2 and *ndhB* splicing is enhanced by RNA editing in the cognate transcripts (Figure 5). Earlier studies have shown that some unspliced or unprocessed transcripts can already be fully edited [29,30] and this was interpreted as the evidence that RNA editing is an early process, mainly occurring before splicing. Although RNA editing can be a prerequisite for splicing when it restores sequences or structures within the intron [35,36], this is an unlikely explanation here as the sites are located far from the identified splicing key elements [37]. A possibility put forward by Yap et al. [11] is that the binding of the RNA editing factor itself could have an indirect effect on splicing through the modification of RNA secondary structure or accessibility.

In agreement with the idea that RNA editing is an early maturation step, we only found marginal evidence that specific RNA editing sites could be influenced by splicing (Figure 4). This result is however probably dependent on our experimental model, *A. thaliana*. In various plants, *ndhA* intron removal was shown to be necessary for an *ndhA* editing site located close to the 3' splice site. In this case, splicing is thought to create the RNA sequence necessary for the recognition of the RNA editing site [10], a site that is absent in *A. thaliana*. As shown for *clpP*, splicing of one intron can also influence splicing of another intron located on the same transcript (Figure 5). Experiments with intron deletions in tobacco have previously shown that the second intron in the *ycf3* transcript needs to be spliced before the first intron. In this case, splicing of the first intron was hypothesized to create a sequence masking essential structural elements of the second intron [38]. Although *A. thaliana ycf3* structure is similar to tobacco, our analysis did not confirm such dependence in this transcript.

The dependence between RNA editing sites themselves has long been debated. For example, *in vitro* results on short fragments of the mitochondrial *atp4* RNA suggested that editing of individual sites had no influence on others while *in organello* experiments with

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2021 doi:10.20944/preprints202108.0571.v1

longer *cox2* transcripts showed pattern of dependencies [39,40]. The identification of distal elements able to enhance RNA editing was also a strong argument against complete stochastic independence of the editing site recognition [34,41]. Our results show that both cases exist in the chloroplast. Editing site *ndhD(117166)* requires earlier editing of the 4 other *ndhD* sites and *ndhB(97016)* editing strongly influences editing at *ndhB(96698)* and *ndhB(96579)* sites. On the other hand editing at *ndhB(95225)* seems autonomous and barely influences any other editing site (Figure 5).

Editing and splicing of organellar transcripts are required to get mRNA translated into functional proteins as editing often restores conserved amino-acids [42] and splicing preserves the translation frame. However, the study of the translational landscape of *A. thaliana* mitochondria[43] or maize chloroplasts [20] showed that ribosomes were associated to partially edited transcripts and a small fraction of ribosomes were even associated to intronic sequences. Earlier chloroplast polysome purification experiments also showed that transcripts of the *psbB* gene cluster containing the *petB* or *petD* intron could still be translated for other genes [44]. This suggests that partially mature (especially partially edited) transcripts can access the organelle translation machinery. In addition to the dependence of some maturations events, our results showed that they could be ordered (Figure 5). In this chronology, splicing events seemed to occur later than editing events: the splicing of *ndhB* occurred after editing at most sites and splicing of *clpP* occurred after its editing. Even if the chronology was not clear from our results, Yap et al. also showed that *atpF* editing probably occurs before its splicing [11]. In addition, events located at the 5′ end of the transcripts tended to be later than the others. That is clearly the case for *clpP* and *ndhD*. In *ndhD*, RNA editing at *ndhD(117166)* is the last maturation event and is required to create the start codon and thus to allow the translation of the transcript. This succession of the maturation events where splicing and 5′ end events tend be last could be a way to ensure the complete (or at least a better) maturation of the transcripts before initiating their translation.

Despite the modest size of the dataset and its rather simple analysis, the results presented in this study highlight the potential of long read RNA-seq for the analysis of plastid transcriptome. It provides access to the full complexity of this transcriptome and already showed numerous links between splicing and editing. For analytical reasons, we did not include the analysis of processing in this study but Nanopore RNA-seq is suited for this type of analysis (Figure 3) and we are developing the required bioinformatical and statistical tools. With this complete toolbox, we anticipate it will be possible to explore the impact of growth conditions and/ or mutants or compare the nucleoid- or polysome associated transcriptome to further decipher the molecular mechanisms controlling plastid but also mitochondrial gene expression.

4. Materials and Methods

4.1. Plant growth and RNA extraction

Col-0 plants were grown in soil in growth chambers with 16h of light per day at 20°C for 5 weeks. Fifteen minutes before the onset of lights, 2 adult leaves were flash-frozen in liquid nitrogen. Total RNA was extracted using Nucleozol (Macherey-Nagel) followed by a purification with AMPure RNA XP beads (Beckman Coulter). Three independent experiments were performed to get three biological replicates.

4.2. Nanopore sequencing

The detailed protocol for the construction of the sequencing library is available online at https://forgemia.inra.fr/guillem.rigaill/nanopore_chloro. Briefly, 10 fmoles of an RNA oligo was ligated to the 3′ end of 100ng of total RNA using 10 U of T4 RNA ligase (NEB). Ligated RNA was depleted of rRNA using the QIAseq FastSelect -rRNA Plant Kit (QIAGEN) before a full-length cDNA synthesis using the SMARTScribe™ Reverse Transcriptase (Takara). Full-length cDNAs were amplified with the SeqAmp DNA Polymerase (Takara) and purified with AMPure XP beads (Beckman-Coulter). 35 fmoles of amplified cDNAs were converted to a Nanopore sequencing library with the PCR

barcoding kit (Oxford Nanopore Technologies, UK) and then sequenced on a R10.3 MinIon flow-cell (Oxford Nanopore Technologies, UK).

4.3. Bioinformatics and statistical analyses

The raw data were base-called and demultiplexed with Guppy v5.0.7 (Oxford Nanopore Technologies) using the dna_r10.3_450bps_hac model. Reads were then oriented using the in-house script "fastq_processing.sh" which uses LAST v1179 [45] and CUTADAPT v2.10 [46] and is available online at https://forgemia.inra.fr/guillem.rigaill/nanopore_chloro. They were mapped on the col-0 genomic sequence with Minimap2 v2.1 [47]. Gene body coverage and strandness were measured with the RSeQC v3.0 package [48].

The maturation events analyzed in this study are listed in Table S1. They include the editing sites detected by Ruwe et al. [12] and the introns of protein coding genes. The tRNA introns were excluded because the matured tRNAs are excluded from the sequencing library during sizing. This information is used to annotate each read for every maturation event according to three modalities: mature site, not mature site, not read site. The latter allows taking insertions/deletions into account which are frequent in Nanopore datasets. For each pair of events jointly observed the following configurations are listed and counted in a contingency table: mature/mature, mature/immature, immature/mature, and immature/immature. The dependency of two events, based on the contingency table, is tested using a Fisher exact test and the p-values were adjusted with a FDR [49]. Only pairs of events characterized by an adjusted p-value < 0.1 in at least 2 of the 3 replicates and an adjusted p-value < 0.005 on the pool of the 3 replicates were considered significant. Commented R scripts to annotate reads, create contingency table, perform Fisher's exact tests and generate the result table are available online at https://forgemia.inra.fr/guillem.rigaill/nanopore_chloro.

The splicing and editing rates were measured from the pool of the reads of the 3 replicates.

Virtual Northern blots were generated by extracting the length of the reads mapping from position 75700 to position 76000 on the Watson strand (*petB*), from position 77200 to position 77500 on the Watson strand (*petD*), from position 74487 to position 74706 on the Watson strand (*psbH*) or from position 74254 to position 74378 on the Crick strand (*psbN*) using samtools [50] and bedtools [51]. The size distributions were normalized by setting the setting the value of the most abundant read length to 100. These distributions were converted into virtual Northern blots with the "vNB.py" python script available on line at https://forgemia.inra.fr/guillem.rigaill/nanopore_chloro.

Supplementary Materials: Figure S1: 5′ to 3′ coverage of nuclear genes by Nanopore reads, is available online at www.mdpi.com/xxx/s1. The supplementary tables Table S1: Maturation rate of known editing sites and introns, Table S2: Co-occurrence of maturation events, Table S3: identification and quantification of the maturation intermediates of *ndhD* and *ndhB* are available online at https://doi.org/10.15454/73GAAV

Author Contributions: Conceptualization, Marine Guilcher, Benoit Castandet and Etienne Delannoy; Data curation, Arnaud Liehrmann, Guillem Rigaill and Etienne Delannoy; Formal analysis, Arnaud Liehrmann, Chloé Seyman and Guillem Rigaill; Funding acquisition, Guillem Rigaill, Benoit Castandet and Etienne Delannoy; Investigation, Marine Guilcher and Etienne Delannoy; Methodology, Guillem Rigaill, Benoit Castandet and Etienne Delannoy; Project administration, Etienne Delannoy; Resources, Etienne Delannoy; Software, Marine Guilcher, Arnaud Liehrmann, Chloé Seyman, Thomas Blein and Guillem Rigaill; Supervision, Etienne Delannoy; Writing – original draft, Marine Guilcher, Benoit Castandet and Etienne Delannoy; Writing – review & editing, Arnaud Liehrmann, Guillem Rigaill, Benoit Castandet and Etienne Delannoy.

Funding: The IPS2 benefits from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007). This work was supported by a grant from the Université Evry-Val d'Essonne to E.D., by the ANR-20-CE20-0004 JOAQUIN to B.C. and by the Evry Genopole to G.R.

Data Availability Statement: The fastq files are available from the NCBI SRA database under the accession number PRJNA748959. The reviewer's link is https://dataview.ncbi.nlm.nih.gov/object/PRJNA748959?reviewer=qjphugo1dod5ftako6ltrdmjai

Acknowledgments: We thank Etienne Sandré-Chardonnal for the python script generating the picture of the virtual Northern blot.

Conflicts of Interest: "The authors declare no conflict of interest."

References

1. Green, B.R. Chloroplast genomes of photosynthetic eukaryotes. *Plant J.* **2011**, *66*, 34–44.
2. Maier, U.G.; Bozarth, A.; Funk, H.T.; Zauner, S.; Rensing, S.A.; Schmitz-Linneweber, C.; Börner, T.; Tillich, M. Complex chloroplast RNA metabolism: just debugging the genetic programme? *BMC Biol.* **2008**, *6*, 36.
3. Stern, D.B.; Goldschmidt-Clermont, M.; Hanson, M.R. Chloroplast RNA metabolism. *Annu. Rev. Plant Biol.* **2010**, *61*, 125–55.
4. Barkan, A. Expression of plastid genes: organelle-specific elaborations on a prokaryotic scaffold. *Plant Physiol.* **2011**, *155*, 1520–1532.
5. de Longevialle, A.F.; Small, I.D.; Lurin, C. Nuclearly-encoded splicing factors implicated in RNA splicing in higher plant organelles. *Mol. Plant* **2010**, *3*, 691–705.
6. Sun, T.; Bentolila, S.; Hanson, M.R. *The Unexpected Diversity of Plant Organelle RNA Editosomes*; 2016; Vol. 21, pp. 962–973;.
7. Germain, A.; Hotto, A.M.; Barkan, A.; Stern, D.B. RNA processing and decay in plastids. *Wiley Interdiscip. Rev. RNA* **2013**, *4*, 295–316.
8. MacIntosh, G.C.; Castandet, B. Organellar and Secretory Ribonucleases: Major Players in Plant RNA Homeostasis. *Plant Physiol.* **2020**, *183*, 1438.
9. Majeran, W.; Friso, G.; Asakura, Y.; Qu, X.; Huang, M.; Ponnala, L.; Watkins, K.P.; Barkan, A.; van Wijk, K.J. Nucleoid-enriched proteomes in developing plastids and chloroplasts from maize leaves: A new conceptual framework for nucleoid functions. *Plant Physiol.* **2012**, *158*, 156–189.
10. Schmitz-Linneweber, C.; Tillich, M.; Herrmann, R.G.; Maier, R.M. Heterologous, splicing-dependent RNA editing in chloroplasts: Allotetraploidy provides trans-factors. *EMBO J.* **2001**, *20*.
11. Yap, A.; Kindgren, P.; Colas des Francs-Small, C.; Kazama, T.; Tanz, S.K.; Toriyama, K.; Small, I. AEF1/MPR25 is implicated in RNA editing of plastid atpF and mitochondrial nad5, and also promotes atpF splicing in Arabidopsis and rice. *Plant J.* **2015**, *81*, 661–669.
12. Ruwe, H.; Castandet, B.; Schmitz-Linneweber, C.; Stern, D.B. Arabidopsis chloroplast quantitative editotype. *FEBS Lett.* **2013**, *587*, 1429–1433.
13. Marechal-Drouard, L.; Cosset, A.; Remacle, C.; Ramamonjisoa, D.; Dietrich, A. A single editing event is a prerequisite for efficient processing of potato mitochondrial phenylalanine tRNA. *Mol. Cell. Biol.* **1996**, *16*.
14. Malbert, B.; Burger, M.; Lopez-Obando, M.; Baudry, K.; Launay-Avon, A.; Härtel, B.; Verbitskiy, D.; Jörg, A.; Berthomé, R.; Lurin, C.; et al. The analysis of the editing defects in the dyw2 mutant provides new clues for the prediction of rna targets of arabidopsis e+-class ppr proteins. *Plants* **2020**, *9*.
15. Ichinose, M.; Sugita, C.; Yagi, Y.; Nakamura, T.; Sugita, M. Two DYW subclass PPR proteins are involved in RNA editing of ccmFc and atp9 transcripts in the moss Physcomitrella patens: first complete set of PPR editing factors in plant mitochondria. *Plant Cell Physiol.* **2013**, *54*, 1907–16.

16. Castandet, B.; Hotto, A.M.; Strickler, S.R.; Stern, D.B. ChloroSeq, an Optimized Chloroplast RNA-Seq Bioinformatic Pipeline, Reveals Remodeling of the Organellar Transcriptome Under Heat Stress. *G3 (Bethesda).* **2016**, *6*, 2817–27.

17. Michel, E.J.S.S.; Hotto, A.M.; Strickler, S.R.; Stern, D.B.; Castandet, B. A Guide to the Chloroplast Transcriptome Analysis Using RNA-Seq. *Methods Mol. Biol.* **2018**, *1829*, 295–313.

18. Malbert, B.; Rigaill, G.; Brunaud, V.; Lurin, C.; Delannoy, E. *Bioinformatic analysis of chloroplast gene expression and rna posttranscriptional maturations using RNA sequencing*; 2018; Vol. 1829;.

19. Castandet, B.; Germain, A.; Hotto, A.M.; Stern, D.B. 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.* **2019**, *47*, 11889–11905.

20. Chotewutmontri, P.; Barkan, A. Dynamics of Chloroplast Translation during Chloroplast Differentiation in Maize. *PLOS Genet.* **2016**, *12*, e1006106.

21. Ruwe, H.; Wang, G.; Gusewski, S.; Schmitz-Linneweber, C. Systematic analysis of plant mitochondrial and chloroplast small RNAs suggests organelle-specific mRNA stabilization mechanisms. *Nucleic Acids Res.* **2016**, *44*, 7406–7417.

22. Zhelyazkova, P.; Sharma, C.M.C.M.; Forstner, K.U.; Liere, K.; Vogel, J.; Borner, T.; Förstner, K.U.; Liere, K.; Vogel, J.; Börnera, T. The Primary Transcriptome of Barley Chloroplasts: Numerous Noncoding RNAs and the Dominating Role of the Plastid-Encoded RNA Polymerase. *Plant Cell* **2012**, *24*, 123–136.

23. Cui, J.; Shen, N.; Lu, Z.; Xu, G.; Wang, Y.; Jin, B. Analysis and comprehensive comparison of PacBio and nanopore-based RNA sequencing of the Arabidopsis transcriptome. *Plant Methods* **2020**, *16*.

24. Long, Y.; Jia, J.; Mo, W.; Jin, X.; Zhai, J. FLEP-seq: simultaneous detection of RNA polymerase II position, splicing status, polyadenylation site and poly(A) tail length at genome-wide scale by single-molecule nascent RNA sequencing. *Nat. Protoc.* **2021**.

25. Jia, J.; Long, Y.; Zhang, H.; Li, Z.; Liu, Z.; Zhao, Y.; Lu, D.; Jin, X.; Deng, X.; Xia, R.; et al. Post-transcriptional splicing of nascent RNA contributes to widespread intron retention in plants. *Nat. Plants* **2020**, *6*, 780–788.

26. Felder, S.; Meierhoff, K.; Sane, A.P.; Meurer, J.; Driemel, C.; Plücken, H.; Klaff, P.; Stein, B.; Bechtold, N.; Westhoff, P. The Nucleus-Encoded HCF107 Gene of Arabidopsis Provides a Link between Intercistronic RNA Processing and the Accumulation of Translation-Competent psbH Transcripts in Chloroplasts. *Plant Cell* **2001**, *13*, 2127.

27. Stoppel, R.; Meurer, J. Complex RNA metabolism in the chloroplast: an update on the psbB operon. *Planta* **2013**, *237*, 441–449.

28. Guillaumot, D.; Lopez-Obando, M.; Baudry, K.; Avon, A.; Rigaill, G.; Falcon De Longevialle, A.; Broche, B.; Takenaka, M.; Berthomé, R.; De Jaeger, G.; et al. Two interacting PPR proteins are major Arabidopsis editing factors in plastid and mitochondria. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 8877–8882.

29. Freyer, R.; Hoch, B.; Neckermann, K.; Maier, R.M.; Kössel, H. RNA editing in maize chloroplasts is a processing step independent of splicing and cleavage to monocistronic mRNAs. *Plant J.* **1993**, *4*.

30. Ruf, S.; Zeltz, P.; Kössel, H. Complete RNA editing of unspliced and dicistronic transcripts of

the intron-containing reading frame IRF170 from maize chloroplasts. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*.

31. Maréchal-Drouard, L.; Kumar, R.; Remacle, C.; Small, I. RNA editing of larch mitochondrial tRNA His precursors is a prerequisite for processing. *Nucleic Acids Res.* **1996**, *24*, 3229–3234.

32. Tillich, M.; Hardel, S.L.S.L.S.L.; Kupsch, C.; Armbruster, U.; Delannoy, E.; Gualberto, J.M.J.M.; Lehwark, P.; Leister, D.; Small, I.D.I.D.; Schmitz-Linneweber, C. Chloroplast ribonucleoprotein CP31A is required for editing and stability of specific chloroplast mRNAs. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 6002–6007.

33. Karcher, D.; Bock, R. Site-selective inhibition of plastid RNA editing by heat shock and antibiotics: A role for plastid translation in RNA editing. *Nucleic Acids Res.* **1998**, *26*, 1185–1190.

34. Takenaka, M.; Neuwirt, J.; Brennicke, A. Complex cis-elements determine an RNA editing site in pea mitochondria. *Nucleic Acids Res.* **2004**, *32*, 4137–4144.

35. Castandet, B.; Choury, D.; Bégu, D.; Jordana, X.; Araya, A. Intron RNA editing is essential for splicing in plant mitochondria. *Nucleic Acids Res.* **2010**, *38*, 7112–7121.

36. Farré, J.-C.J.-C.C.; Aknin, C.; Araya, A.; Castandet, B. RNA Editing in Mitochondrial Trans-Introns Is Required for Splicing. *PLoS One* **2012**, *7*, e52644.

37. Vogel, J.; Börner, T. Lariat formation and a hydrolytic pathway in plant chloroplast group II intron splicing. *EMBO J.* **2002**, *21*.

38. Petersen, K.; Schöttler, M.; Karcher, D.; Thiele, W.; Bock, R. Elimination of a group II intron from a plastid gene causes a mutant phenotype. *Nucleic Acids Res.* **2011**, *39*, 5181–5192.

39. Verbitskiy, D.; Takenaka, M.; Neuwirt, J.; Van Der Merwe, J.A.; Brennicke, A. Partially edited RNAs are intermediates of RNA editing in plant mitochondria. *Plant J.* **2006**, *47*.

40. Castandet, B.; Araya, A. The RNA editing pattern of cox2 mRNA is affected by point mutations in plant mitochondria. *PLoS One* **2011**, *6*.

41. Staudinger, M.; Bolle, N.; Kempken, F. Mitochondrial electroporation and in organello RNA editing of chimeric atp6 transcripts. *Mol. Genet. Genomics* **2005**, *273*, 130–136.

42. Small, I.D.; Schallenberg-Rüdinger, M.; Takenaka, M.; Mireau, H.; Ostersetzer-Biran, O. Plant organellar RNA editing: what 30 years of research has revealed. *Plant J.* **2019**.

43. Planchard, N.; Bertin, P.; Quadrado, M.; Dargel-Graffin, C.; Hatin, I.; Namy, O.; Mireau, H. The translational landscape of Arabidopsis mitochondria. *Nucleic Acids Res.* **2018**, *46*, 6218.

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

45. Kiełbasa, S.M.; Wan, R.; Sato, K.; Horton, P.; Frith, M.C. Adaptive seeds tame genomic sequence comparison. *Genome Res.* **2011**, *21*, 487–493.

46. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **2011**, *17*, 10–12.

47. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **2018**, *34*, 3094–3100.

48. Wang, L.; Nie, J.; Sicotte, H.; Li, Y.; Eckel-Passow, J.E.; Dasari, S.; Vedell, P.T.; Barman, P.; Wang, L.; Weinshiboum, R.; et al. Measure transcript integrity using RNA-seq data. *BMC Bioinformatics* **2016**, *17*, 1–16.

49. Benjamini, Y.; Hochberg, Y. Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing. *J. R. Stat. Soc. B* **1995**, *57*, 289–300.

50. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.;

 doi:10.20944/preprints202108.0571.v1

Durbin, R.; Subgroup, 1000 Genome Project Data Processing The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078.

51. Quinlan, A.R.; Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841.