

# ChloroSeq, an Optimized Chloroplast RNA-Seq Bioinformatic Pipeline, Reveals Remodeling of the Organellar Transcriptome Under Heat Stress

Benoît Castandet, Amber M. Hotto, Susan R. Strickler, and David B. Stern<sup>1</sup>

Boyce Thompson Institute, Ithaca, New York 14853-1801

ORCID IDs: 0000-0002-0121-0048 (S.R.S.); 0000-0002-0653-6602 (D.B.S.)

**ABSTRACT** Although RNA-Seq has revolutionized transcript analysis, organellar transcriptomes are rarely assessed even when present in published datasets. Here, we describe the development and application of a rapid and convenient method, ChloroSeq, to delineate qualitative and quantitative features of chloroplast RNA metabolism from strand-specific RNA-Seq datasets, including processing, editing, splicing, and relative transcript abundance. The use of a single experiment to analyze systematically chloroplast transcript maturation and abundance is of particular interest due to frequent pleiotropic effects observed in mutants that affect chloroplast gene expression and/or photosynthesis. To illustrate its utility, ChloroSeq was applied to published RNA-Seq datasets derived from *Arabidopsis thaliana* grown under control and abiotic stress conditions, where the organellar transcriptome had not been examined. The most appreciable effects were found for heat stress, which induces a global reduction in splicing and editing efficiency, and leads to increased abundance of chloroplast transcripts, including genic, intergenic, and antisense transcripts. Moreover, by concomitantly analyzing nuclear transcripts that encode chloroplast gene expression regulators from the same libraries, we demonstrate the possibility of achieving a holistic understanding of the nucleus-organellar system. ChloroSeq thus represents a unique method for streamlining RNA-Seq data interpretation of the chloroplast transcriptome and its regulators.

## KEYWORDS

*Arabidopsis thaliana*  
chloroplast RNA-Seq  
heat stress transcriptome  
noncoding RNAs  
introns  
RNA splicing  
RNA editing

Expression of the chloroplast genome is the result of a complex interplay of nuclear and organelle-encoded proteins. As many as three RNA polymerases and six sigma factors create the primary transcriptome, which is further processed by an array of ribonucleases, RNA-binding proteins, and RNA splicing and editing factors, ultimately producing mature transfer RNAs, ribosomal RNAs, and messenger RNAs (tRNAs, rRNAs, and mRNAs, respectively) (Barkan 2011; Germain *et al.* 2013; Stern *et al.* 2010). Disruption of any of these posttranscriptional steps, whether initially affecting one or multiple genes, often leads to pleiotropic effects. These may be at the whole-plant level, such as chlorosis due

to photosynthetic deficiency, and/or at the level of gene expression, for example because of global transcriptional or translational impairment. Therefore, most stresses and mutations affecting chloroplast gene expression warrant analysis at the genome rather than gene-by-gene scale.

The first global strategies for chloroplast transcript analysis utilized microarrays, most recently tiling arrays, which have yielded breakthroughs in our understanding of RNA targets of specific regulatory factors, developmental variations in the transcriptome, and more recently translation (Allorent *et al.* 2013; Nagashima *et al.* 2004; Schmitz-Linneweber *et al.* 2005; Zoschke *et al.* 2013). On the other hand, they have limited value as quantitative tools, are rarely developed to cover both strands of an entire genome, and do not reveal features such as splicing and editing status (Cahoon *et al.* 2008; Kahlau and Bock 2008; Nagashima *et al.* 2004; Nakamura *et al.* 2003; Schmitz-Linneweber *et al.* 2005; Valkov *et al.* 2009; Zmienko *et al.* 2011; Zoschke *et al.* 2013). These limitations have been largely overcome by high-throughput cDNA sequencing (RNA-Seq), which has rapidly become preeminent for plants since its initial application to *Arabidopsis thaliana* (Weber *et al.* 2007), and now for many model and nonmodel organisms (Martin *et al.* 2013). Currently, some 30,000 RNA-Seq libraries are archived for the Streptophyta taxon in the Sequence Read Archive (SRA) database ([www.ncbi.nlm.nih.gov/sra](http://www.ncbi.nlm.nih.gov/sra)).

Copyright © 2016 Castandet *et al.*

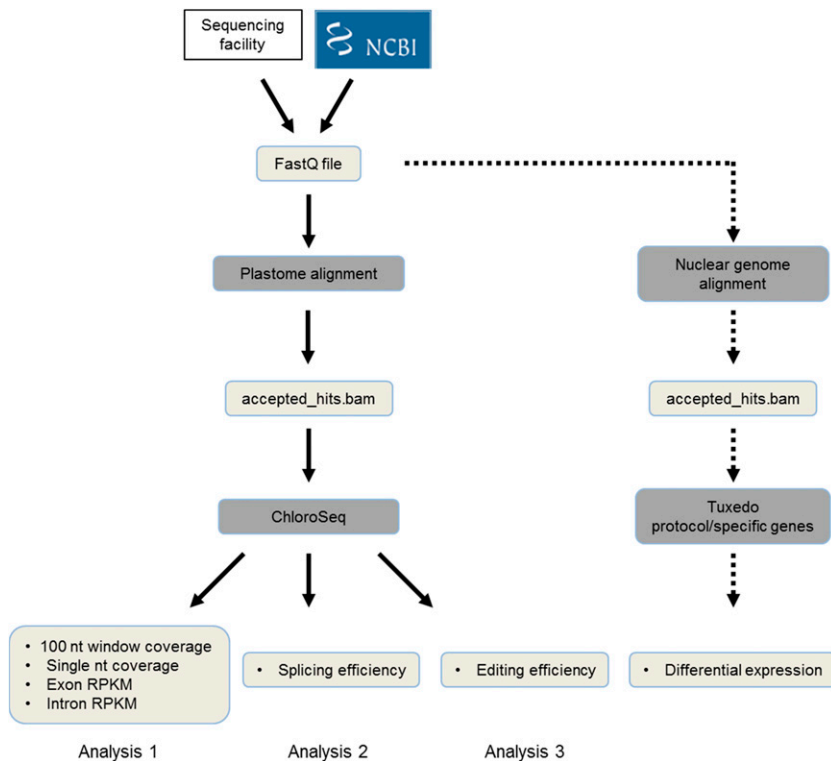
doi: 10.1534/g3.116.030783

Manuscript received May 2, 2016; accepted for publication July 1, 2016; published Early Online July 6, 2016.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplemental material is available online at [www.g3journal.org/lookup/suppl/doi:10.1534/g3.116.030783/-/DC1](http://www.g3journal.org/lookup/suppl/doi:10.1534/g3.116.030783/-/DC1)

<sup>1</sup>Corresponding author: Boyce Thompson Institute, 533 Tower Rd., Ithaca, NY 14853-1801. E-mail: [ds28@cornell.edu](mailto:ds28@cornell.edu)



**Figure 1** Flow chart of the ChloroSeq strategy. RNA-Seq reads (obtained from public databases or a sequencing facility) in the FastQ format are aligned to the plastome genome using TopHat2. ChloroSeq is then applied to the bam file containing aligned reads, giving a complete view of the chloroplast transcriptome. Using the same data, it is possible to concomitantly analyze the nuclear transcriptome using, for example, the Tuxedo protocol. NCBI, National Center for Biotechnology Information; nt, nucleotide; RNA-Seq, RNA sequencing; RPKM, reads per kilobase per million mapped reads.

RNA-Seq libraries are generally obtained from total RNA, and as a significant proportion of plant cellular RNA comes from organelles (Loening and Ingle 1967), they can be an invaluable resource to study chloroplast and mitochondrial RNA metabolism (Smith 2013). The major caveat is that most RNA-Seq libraries rely on enrichment of polyadenylated transcripts, representing mRNAs. Since plant organellar transcripts become unstable upon polyadenylation (Rorbach *et al.* 2014), library preparation protocols developed for organelles cannot use such a step. Fortunately, plant organelle-friendly libraries are becoming more prevalent, for example through work specifically designed to include organellar transcripts (Hotto *et al.* 2015, 2011; Zhelyazkova *et al.* 2012b), or to analyze nuclear noncoding RNAs (Di *et al.* 2014). Additionally, previously published RNA-Seq data can often be used to gain information aside from the original aim (Castandet *et al.* 2013; Ruwe and Schmitz-Linneweber 2012; Zhelyazkova *et al.* 2012a). Thus, published and novel RNA-Seq data can be harnessed to increase our understanding of how the chloroplast transcriptome responds to stress, genotype, and developmental factors.

In spite of the growing opportunities to discover properties of plant organellar transcriptomes, most compatible datasets have not been mined for their organellar component, in part because most of the analytical pipelines developed for nuclear transcriptomes are not adapted to the peculiarities of chloroplast gene expression. To facilitate cell-wide transcriptomic analyses, we developed ChloroSeq, a pipeline yielding a comprehensive view of chloroplast RNA processing, editing, splicing, and relative abundance. As applied to published data on *Arabidopsis* stress responses (Di *et al.* 2014), ChloroSeq reveals substantial effects at various steps of chloroplast gene expression under heat stress, some of which could be validated using gene-by-gene approaches. In addition, integration of these results with the nuclear transcriptome revealed RNA abundance changes for chloroplast-targeted enzymes that may contribute to alterations of the chloroplast transcriptome in response to heat stress, opening the way to a better understanding of nucleus-organelle crosstalk.

## MATERIALS AND METHODS

### Hardware and software

All analyses described here were performed on a 64-bit laptop running the Ubuntu 14.04 LTS operating system with 8 GB (gigabytes) of RAM, with the following software installed: Bowtie2 version 2.2.3 (<http://bowtie-bio.sourceforge.net/index.shtml>), SAMtools (<http://www.htslib.org/>), TopHat2 version 2.0.11 (<http://ccb.jhu.edu/software/tophat/index.shtml>), Cufflinks version 2.2.1 (<http://cole-trapnell-lab.github.io/cufflinks/>), and BED tools version 2.25.0 (<https://github.com/arq5x/bedtools2>).

RNA-Seq data (Accession SRP028304; Di *et al.* 2014) was downloaded from the SRA database (Kodama *et al.* 2012) (<http://www.ncbi.nlm.nih.gov/sra?term=SRP028304>), and then converted into the corresponding FastQ file using fastq-dump from the SRA toolkit ([http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=toolkit\\_doc&f=fastq-dump](http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=toolkit_doc&f=fastq-dump)). Reads were initially quality trimmed using the fastq-mcf software (<https://code.google.com/p/ea-utils/wiki/FastqMcf>), and only bases with a quality score higher than 30 and reads with a minimum length of 60 were retained.

Before starting, ChloroSeq and the accompanying annotation files must be downloaded (<https://github.com/BenoitCastandet/chloroseq>) and ungrouped in the working directory. The annotation files specify genomic intervals to be used by ChloroSeq (see below). The TAIR10\_Chrc\_files directory contents further expand in TAIR10\_Chrc\_bowtie2\_index, and contain the indexed files for the chloroplast genome used during the alignment.

### Genome alignment and differential expression analysis

Alignment to the chloroplast genome was performed using TopHat2, allowing up to two alignments to account for the chloroplast inverted repeats ( $-g$  2). The GFF file containing the gene coordinates and the indexed genome files used can be found in the TAIR10\_Chrc\_files and TAIR10\_Chrc\_files/TAIR10\_Chrc\_bowtie2\_index directories, respectively, and the novel splice site discovery was disabled ( $-no-novel-juncs$ ).

■ **Table 1 Summary of chloroplast alignment rates for several published libraries**

Name	Tissue	RNA Preparation	Mappable Reads	Mapped Reads	Mapped/Mappable (%)	Reference
<b>Group A reads</b>						
Control (SR63)	12-day-old seedlings	PolyA(-), rRNA subtraction	24057306	6088444	25.3	Di et al. 2014
Heat_3h (SR66)	12-day-old seedlings	PolyA(-), rRNA subtraction	36308522	10984310	30.3	Di et al. 2014
Heat_3h (SR67)	12-day-old seedlings	PolyA(-), rRNA subtraction	33484374	10205141	30.5	Di et al. 2014
<b>Group B reads</b>						
Col-0	Chloroplasts	Total, rRNA subtraction	12306665	9541442	77.5	Hotto et al. 2015
mnc3/4	Chloroplasts	Total, rRNA subtraction	13433730	9941425	74	Hotto et al. 2015
Col-0	2-day-old seedlings	Total, rRNA subtraction	26945257	3711755	13.8	Woodson et al. 2013
gun1	2-day-old seedlings	Total, rRNA subtraction	37022286	3823419	10.3	Woodson et al. 2013
<b>Group C reads</b>						
Col_15mn-1	5-day-old seedlings	Total, oligo(dT) RT	5067731	36746	0.7	Chen et al. 2015
Col_15mn-2	5-day-old seedlings	Total, oligo(dT) RT	4686444	36680	0.8	Chen et al. 2015
Col_15mn-3	5-day-old seedlings	Total, oligo(dT) RT	5133265	37018	0.7	Chen et al. 2015

In all cases, mappable reads represent reads with a quality higher than 30, and mapped reads represent the mappable reads that aligned to the *Arabidopsis* chloroplast genome (TAIR10) using TopHat2. Up to two locations were accepted for sequence reads to account for the large inverted repeat in the chloroplast genome. Group A reads from Di et al. (2014) were produced from poly(A) depleted total RNA followed by rRNA subtraction; Group B reads were obtained from total RNA from isolated chloroplasts or seedlings, followed by rRNA subtraction (Hotto et al. 2015; Woodson et al. 2013); and Group C reads were obtained following oligo(dT)-primed reverse transcription (Chen et al. 2015). Poly(A), poly(adenosine); rRNA, ribosomal RNA; oligo(dT) RT, oligo(dT)-primed reverse transcription.

In addition, differential expression of nuclear genes was assayed following the step-by-step Tuxedo protocol using TopHat and Cuffdiff (Trapnell et al. 2012). We provide an additional file (Supplemental Material, File S1) containing the genes known to be involved in chloroplast RNA metabolism, to simplify the extraction of their expression values.

### Running ChloroSeq

ChloroSeq is made of four bash shell script files (.sh files) that prepare the read file for processing, then address coverage, splicing, and editing, and one Perl script (.pl file) that runs the pipeline and calls the bash shell scripts as needed. All of these files need to be made executable before use. Here is an overview of ChloroSeq:

```
chloroseq.pl [-h] -a <analysis> -b <accepted_hits.bam> -e <exon.gff3> -i <intron.gff3> -g <genome_size> -n <name> -v <editing_sites.gff3> -f <fasta> -s <splice_sites.gff3> -k <keep_files>
```

To obtain details on running the pipeline and available options, the following command may be used:

```
perldoc chloroseq.pl
```

### Preparing the bam file

When the pipeline is first run, the shell script prep\_bam.sh is called to index the bam file. The desired reads corresponding to the genome of interest (here ChrC) are then extracted from the aligned reads bam file. The resulting bam file is then used for all other steps in the pipeline. This gives ChloroSeq the capacity to extract and analyze any subgenome of interest (for example chloroplast or mitochondrion) when the alignment has been performed against the complete genome. One simply needs to know the size of the genome (-g option) and the name of the genome of interest (-n option). The name of the genome of interest must match the one used for alignment.

### Analysis 1: coverage

By choosing option analysis 1 (-a 1) in the chloroseq.pl pipeline, the get\_coverage.sh script is called and produces four different outputs: (1) single nucleotide depth of coverage of the complete chloroplast genome,

(2) genome coverage using a 100 nt window, where each window overlaps the next by 50 nt, and RPKM values for (3) exons and (4) introns. RPKM values are calculated according to the formula  $RPKM = C \frac{R_i}{T_i L}$  where  $C$  is a constant ( $C = 10^9$ ),  $R_i$  is the number of reads that intersect the region of interest,  $T_i$  is the total number of reads of the sample, and  $L$  is the length of the region.

Once completed, six text files containing the four different outputs are found in the working directory: counts, bam\_name, nt\_coverage.txt, window\_coverage.txt, exon\_rpkms.txt, and intron\_rpkms.txt. Intermediate files created may be kept with the -k option and are named coverage\_plus.csv, coverage\_minus.csv, window\_plus.csv, window\_minus.csv, exp\_exon.txt, and exp\_intron.txt.

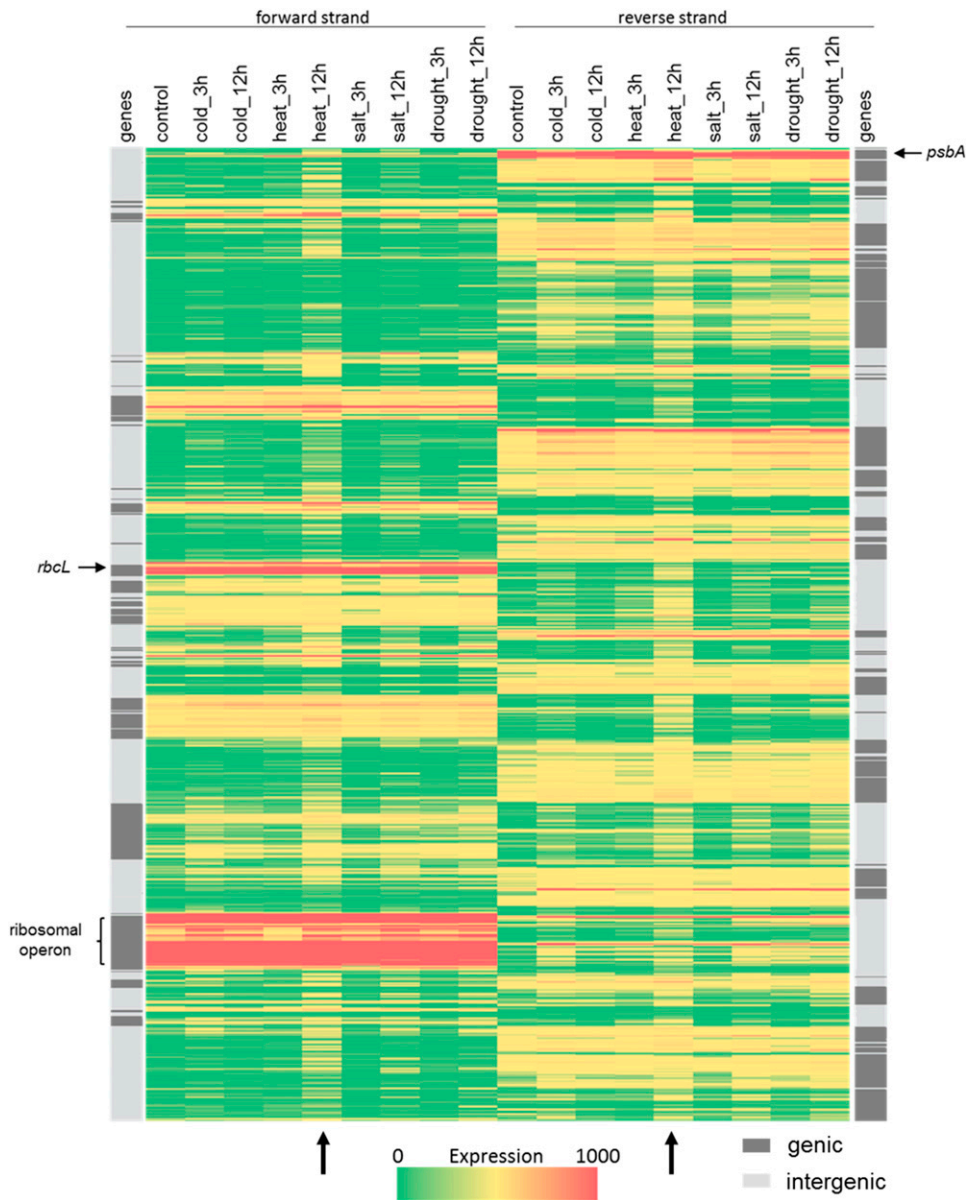
### Analysis 2: splicing efficiency

Analysis 2 calls the get\_splicing\_efficiency script and produces one output file that contains the coordinates, the number of spliced and unspliced reads, and the calculated splicing efficiency for each intron. Splicing efficiency ( $SE$ ) was calculated according to the formula  $SE = \frac{jR}{jR + (liR + riR)/2}$  where  $jR$  is the number of reads spanning the exon/exon junction,  $liR$  is the number of reads spanning the 5' exon/intron boundary, and  $riR$  is the number of reads spanning the 3' intron/exon boundary.

The output file is named splicing\_efficiency.txt and is found in the current directory. Intermediate files are named splice\_junctions.bam, spliced\_coverage.txt, and unspliced\_coverage.txt, and may be retained with the -k option.

### Analysis 3: editing efficiency

Analysis 3 calls the shell script get\_editing\_efficiency.sh. This step produces two output files, editing\_pileup and editing\_efficiency.txt. The latter contains the coordinates of known editing sites for the *Arabidopsis* chloroplast (Ruwe et al. 2013), the plus strand identity of the edited nucleotide (C if plus strand, G if minus strand), the total coverage of the editing site, and the coverage for each of the four possible nucleotides. If a gff3 file of known editing sites is provided (-v option), the ratio of edited to (unedited + edited) reads is calculated to obtain the editing efficiency. The script uses SAMtools to create a pileup file containing the variants (Li et al. 2009).



**Figure 2** Heat map representation of chloroplast transcript levels following application of abiotic stresses. The window coverage data given by ChloroSeq analysis 1 was used to create the heat map. Low to high expression is represented by a green to red transition. Results are split between the forward (left) and reverse (right) strands. Dark gray stripes (outside lanes) represent known genic areas. The highly expressed *psbA* and *rbcL* genes, and the ribosomal operon are marked for reference. Vertical black arrows indicate the 12 hours heat stress treatment discussed in the text.

### Plant growth conditions

*A. thaliana* ecotype Columbia was grown according to Di *et al.* (2014). Briefly, seeds were plated on 1/2 MS agar (1% sucrose, 0.1% MES, pH 5.7, 1.2% agar) under sterile conditions, and, after vernalization, were grown at room temperature under a 16 hr/8 hr light/dark cycle with 100  $\mu\text{mol}/\text{m}^2/\text{sec}$  light. After 12 d, seedlings were exposed to 37° heat stress for 0, 3, or 12 hr. Seedlings were harvested in microfuge tubes, frozen in liquid nitrogen, and stored at -80°.

### RNA extraction and analysis

Total RNA was extracted using TRI reagent (Molecular Research Center <http://mrcgene.com/>) according to the manufacturer's instructions. For RNA blot analysis, 1 or 5  $\mu\text{g}$  of total RNA was separated in 1.2% agarose/formaldehyde gels, blotted overnight onto Hybond-N+ (GE Healthcare), and hybridized with either  $\alpha$ -<sup>32</sup>P-dCTP-labeled double-stranded DNA or  $\alpha$ -<sup>32</sup>P-UTP-labeled single-stranded RNA probes as previously described (Hotto *et al.* 2011). Amounts of RNA analyzed and type of probe used are indicated in the Figure Legends.

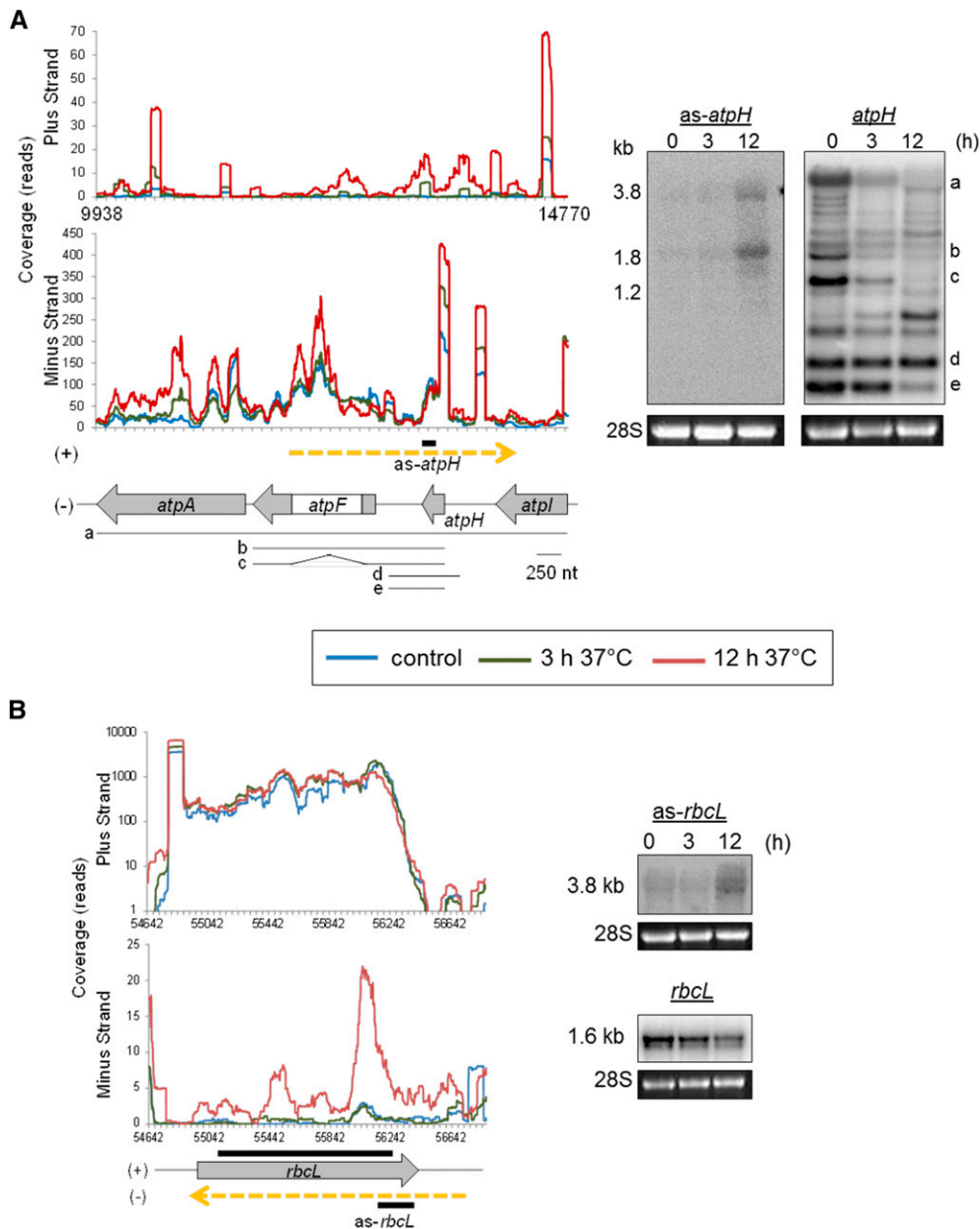
### Data availability

The RNA-Seq data used in this article was obtained from the SRA database (Accession SRP028304; <http://www.ncbi.nlm.nih.gov/sra?term=SRP028304>). All bioinformatic scripts developed for this project can be found on GitHub (<https://github.com/BenoitCastandet/chloroSeq>).

## RESULTS AND DISCUSSION

### ChloroSeq overview

Most published RNA-Seq libraries have only been used to study differential gene expression (DGE), and step-by-step protocols have been published to facilitate such analyses (Ghosh and Chan 2016; Trapnell *et al.* 2012). DGE does not, however, capture the range of transcriptome modifications that occur in organelles. For example, the widespread transcript abnormalities that occur in mutants lacking chloroplast polynucleotide phosphorylase (PNPase; (Castandet *et al.* 2013; Germain *et al.* 2011; Walter *et al.* 2002), spanning 3' end maturation, editing, and splicing, cannot be identified by DGE. To overcome this limitation, we created ChloroSeq (<https://github.com/BenoitCastandet/>



**Figure 3** Accumulation of *as-atpH* and *as-rbcL* under heat stress. (A) Normalized RNA-Seq read coverage in the *atpH* coding region (left panel) for 0 (blue line), 3 (green line), or 12 hr (red line) at 37° for either the plus or minus strands, with the gene model shown below depicting the extent of sense (gray arrows) and antisense *atpH* transcripts (dotted yellow arrow). The thick black line represents the strand-specific RNA probe used for gel blot analysis (right panel) to detect either *as-atpH* or sense *atpH* with 5 μg of total RNA. The *atp* operon model shows the major *atpH*-containing transcripts (lettered a–e correlating to bands identified in Germain *et al.* 2011). Stained 28S rRNA is included to reflect loading and size markers are shown at left. (B) Analysis of the *rbcL* coding region by RNA-Seq and gel blot as described for (A). For the *rbcL* sense gel blot, 1 μg of RNA was analyzed using a double-stranded DNA probe. nt, nucleotide; RNA-Seq, RNA sequencing; rRNA, ribosomal RNA.

chloroSeq), a pipeline designed to reveal these subtler, yet critical facets of chloroplast gene expression. The strategy is depicted in Figure 1, beginning with reads in the FastQ format from either a sequencing facility or downloaded from a database. The reads are then aligned to the chloroplast genome using TopHat2 (Kim *et al.* 2013). Following this, ChloroSeq extracts and counts the number of aligned reads (from the accepted\_hit.bam file) that intersect the genomic intervals of interest, using the BEDTools suite (Quinlan 2014). Three different analyses can be completed with ChloroSeq, which can be run as a group or individually. Coverage analysis produces genome coverage data, including exon and intron Reads Per Kilobase per Million mapped reads (RPKM) information; splicing analysis generates splicing efficiency; and editing analysis yields RNA editing levels. The analyses all generate text files, which can be read by any statistical software to produce graphical representations of the data. An ancillary pipeline yields nuclear DGE. All the annotation files used here are based on the TAIR10 version of the *A. thaliana* genome, modified to add the first exon of the chloroplast gene *ycf3*. ChloroSeq has been optimized for single-

end strand-specific RNA-Seq data, but can be adapted to work with paired-end reads. Additional details are given in *Materials and Methods*.

### rRNA subtraction increases chloroplast transcriptome coverage

Because rRNA constitutes the vast majority of cellular RNA, RNA-Seq library preparation protocols either selectively target poly(A+) RNA using oligo(dT) to prime cDNA synthesis, or deplete rRNA through hybridization to complementary oligonucleotides. Only the latter method preserves chloroplast transcripts, which are primarily free of poly(A) tails, and was applied to nearly 200 plant RNA-Seq libraries deposited in the SRA database. rRNA subtraction is becoming more widely used, however, because of the recognition that important components of the nuclear transcriptome are not polyadenylated, such as noncoding RNAs (ncRNAs) (Di *et al.* 2014).

The effect of rRNA depletion on the chloroplast genome alignment rate is illustrated in Table 1, where we compare RNA-Seq libraries

prepared from poly(A)-depleted total RNA followed by rRNA subtraction (Group A); total RNA from isolated chloroplasts or seedlings, followed by rRNA subtraction (Group B); and an oligo(dT)-primed reverse transcription method (Group C). Approximately 30% of the reads aligned to the plastome from the rRNA-depleted libraries, whereas less than one percent of the oligo(dT)-selected RNAs could be aligned. When chloroplasts are purified before library preparation, the alignment rate reaches 75%, while young seedlings (2 d post emergence) contained less chloroplast RNA than older ones (12 d), on a total RNA basis. The low alignment rate for oligo(dT)-selected RNAs may mask quantitative and qualitative information, and place many transcripts below the limit of detection, skewing downstream gene expression analysis. On the other hand, such transcript pools could be useful for studying polyadenylation-mediated RNA degradation. In general, however, applying rRNA subtraction, without selection for poly(A)-tailed RNA, is essential for the comprehensive study of chloroplast RNA metabolism via RNA-Seq.

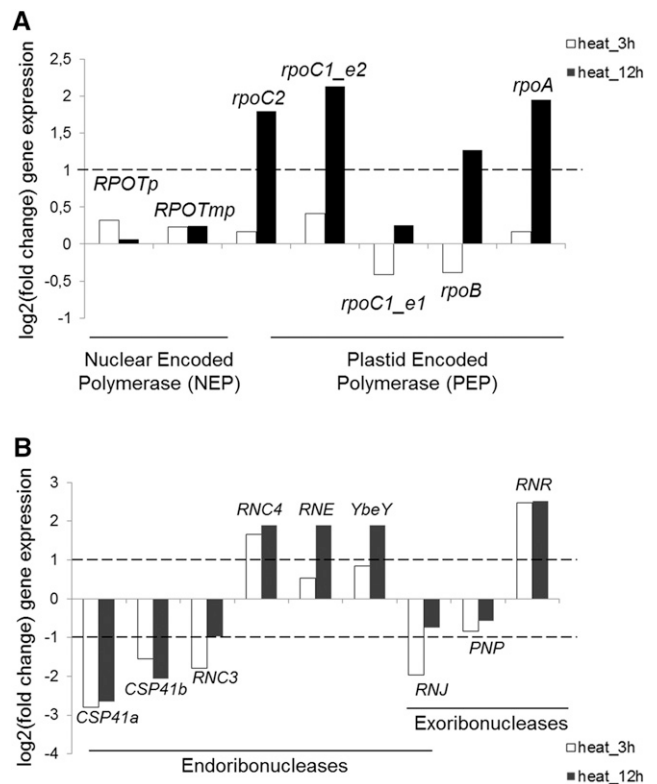
### Coverage analysis

To give a rapid overview of the chloroplast transcriptional landscape from both DNA strands, ChloroSeq analysis 1 yields a 100 nucleotide (nt) window coverage overlapping by 50 nt, normalized to the total number of reads. Graphical representation of the window coverage, for example, can be used for quick identification of regions that merit further investigation or validation using molecular techniques. Additionally, a more precise view, for instance where coverage varies dramatically within a region, can be achieved using the single nucleotide coverage output. Importantly, either coverage analysis does not require any *a priori* knowledge of a region of interest, and in our experience is particularly valuable for the study of mutants or growth conditions whose effects are largely unknown. Such a strategy has already proven useful to identify a plastid noncoding (pnc) RNA specifically regulated by chloroplast Mini-ribonuclease III, and previously unidentified 3' extensions and cleaved tRNA leaders found in the PNPase mutant (Castandet *et al.* 2013; Hotto *et al.* 2015).

In the present study, ChloroSeq 100 nt coverage was completed using RNA-Seq data from *A. thaliana* plants grown under control, heat, salt, cold, and drought stress conditions (Figure 2 and Figure S1). Figure S1 illustrates how the complete genome can be rapidly scanned by eye to reveal relative expression levels, while Figure 2 is a heat map representation of the same coverage. The heat map shows a strong correlation between RNA accumulation under control conditions, and the positions of annotated genes. The high expression of rRNAs, even after rRNA subtraction, is evidence that defects in rRNA processing can still be readily identified (Hotto *et al.* 2015). Comparison of heat maps between the control and four different stress conditions readily revealed that 12 hr of heat stress (vertical arrows at bottom) produced the strongest effect on the transcriptional landscape. We therefore focused our downstream analysis around the heat stress phenomenon.

### pncRNA levels increase under heat stress

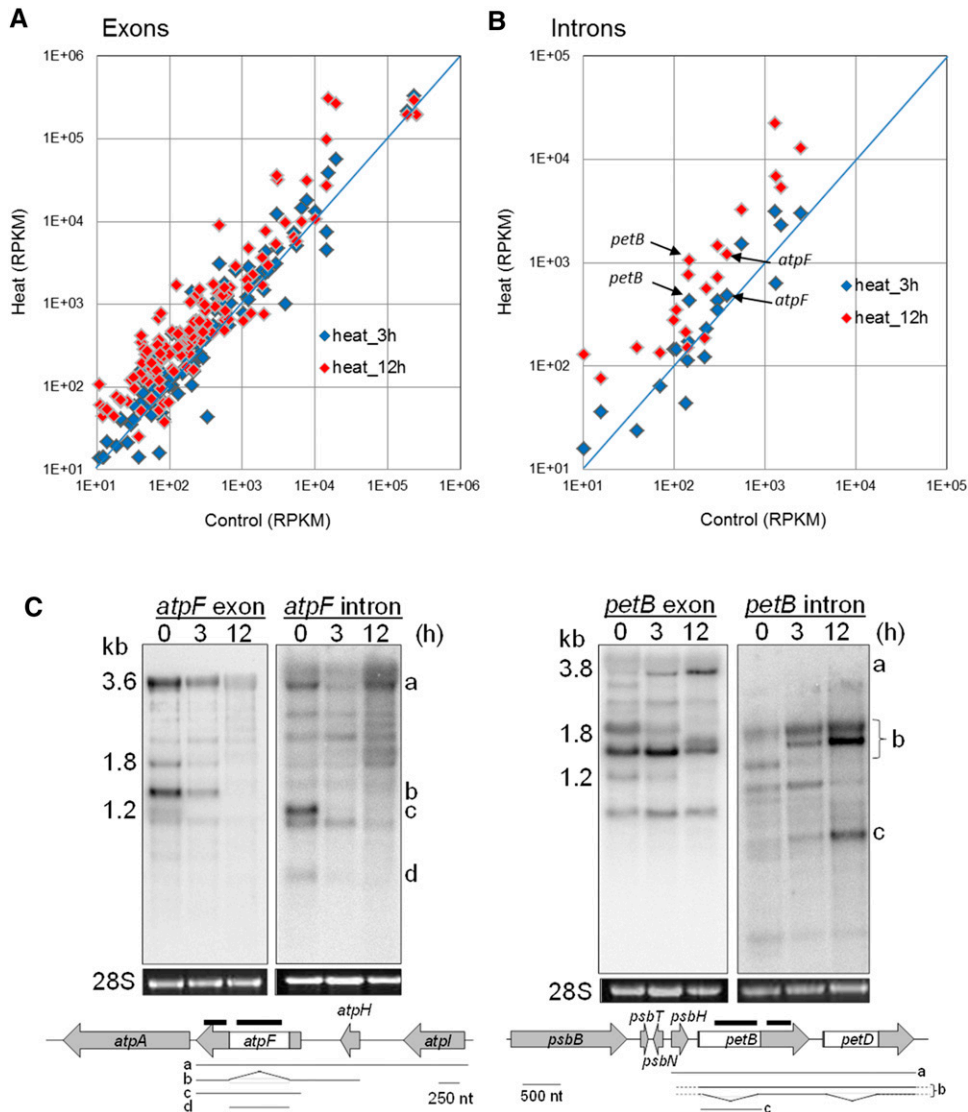
Earlier RNA-Seq analysis by ourselves and others had revealed the presence of more than 100 accumulating noncoding RNAs in chloroplasts, which have been termed pncRNAs (Hotto *et al.* 2012, 2011; Wang *et al.* 2011; Zhelyazkova *et al.* 2012b). An increase in the pncRNA population after 12 hr of heat stress could be recognized either by inspection of the 100 nt window coverage or heat maps. In the former case, numerous coverage peaks outside of genic regions became substantially higher upon heat stress (asterisks in Figure S1A). In the heat map, many areas outside the genic regions (light gray bars at edges of Figure 2) exhibited low expression (green color) under control con-



**Figure 4** Accumulation of transcripts encoding the plastid transcriptional machinery and ribonucleases under heat stress. (A) Fold gene expression compared to the control is on the log<sub>2</sub> scale for 3 hr (white bars) and 12 hr (black bars) of heat stress, with the dotted line representing a twofold difference. *RPOThp* and *RPOThmp* encode the phage-type RNA polymerases; *rpoA-C* encode the bacterial-like RNA polymerase core subunits. (B) The graph description is as in (A). *CSP41a* and *b* encode the paralogous chloroplast stem-loop binding proteins/ribonucleases *CSP41a* and *CSP41b*; *RNC3* and *RNC4* encode redundant mini-ribonuclease III proteins; *YbeY* encodes endoribonuclease Y; *RNJ* encodes ribonuclease J; *PNP* encodes polynucleotide phosphorylase; and *RNR* encodes ribonuclease II/R. NEP, nuclear encoded polymerase; PEP, plastid encoded polymerase.

ditions that increased under heat stress, as indicated by a shift toward yellow or orange hues. Variation in the pncRNA population was expected, as heat stress decreased pncRNA accumulation in *Brassica rapa* (Wang *et al.* 2011). The two studies are not directly comparable, however, due to differences in heat treatment ( $\geq 44^\circ$  vs.  $37^\circ$ ), type of pncRNA detected ( $< 36$  nt, in contrast to our cutoff of  $> 100$  nt for *Arabidopsis*), and species examined, but rather corroborates the influence of heat on pncRNA accumulation. To see whether apparent pncRNA increases would be reflected in RNA gel blots, we examined transcripts antisense to *atpH* (*as-atpH*; Figure 3A) and *rbcl* (*as-rbcl*; Figure 3B). In both cases, the pncRNA increased after 12 hr of heat stress, while little to no transcript was detectable at 0 and 3 hr of heat, confirming the RNA-Seq result.

RNA-Seq data plots for the *atpH* and *rbcl* coding regions exhibited minor differences following 12 hr of heat stress. The *rbcl* mRNA appeared to decline slightly by RNA gel blot, whereas the *atpH* transcript pattern changed qualitatively, due to altered transcript processing or stability within the polycistronic transcription unit. Such shifts are generally not visible in RNA-Seq profiles, and the findings in this instance illustrate certain strengths and limitations of each



**Figure 5** Increased intron accumulation under heat stress. Fold gene expression change of exons (A) and introns (B) between control and heat stress conditions by comparison of RPKM values. (C) RNA gel blot analysis of exons and introns for *atpF* (left) and *petB* (right) analyzed with double-stranded DNA probes (thick black bars in operon models) and 1  $\mu$ g of total RNA. Letters on the right identify transcripts that are depicted in gene models under the blots. Stained 28S rRNA is included to reflect loading, and size markers are at the left. nt, nucleotide; RPKM, reads per kilobase per million mapped reads.

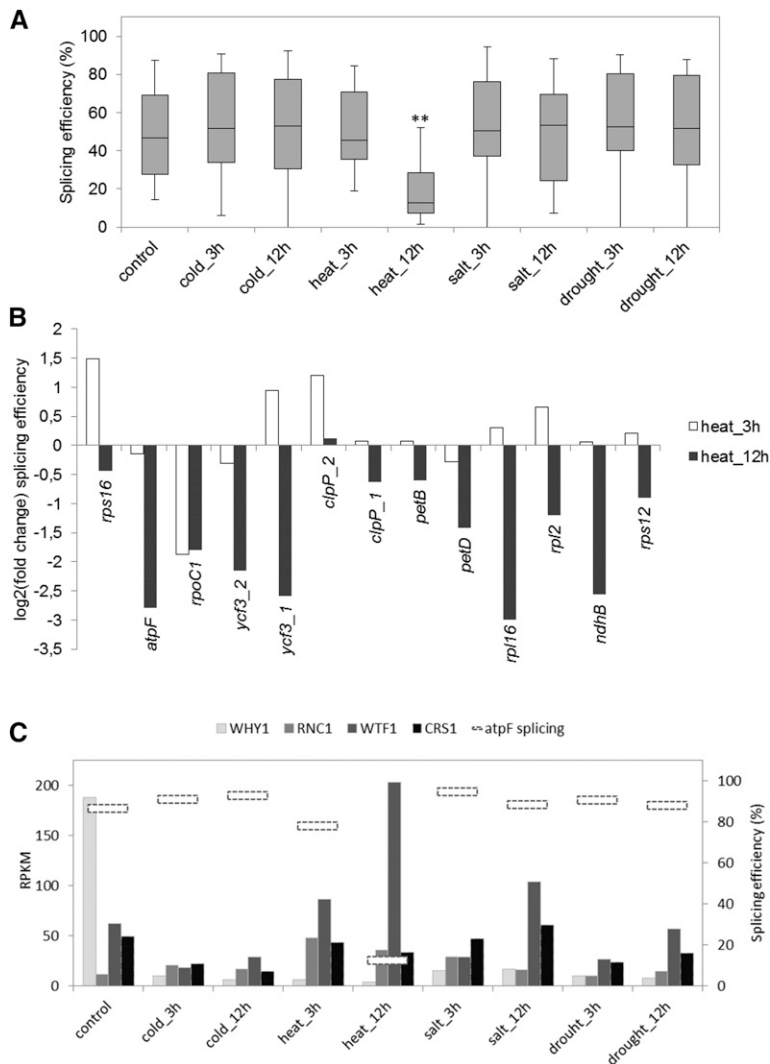
approach. Whether the observed increases in the pncRNAs are related to the alterations in sense transcripts cannot be determined, especially given the  $\sim$ 10–40-fold differences in coverage levels when the two strands are compared.

The mechanism underlying pncRNA proliferation here may be the consequence of increased transcriptional activity and/or pncRNA stability. Little is known about the effect of heat stress on chloroplast transcription, apart from experiments in maize showing that chloroplasts were 5–10 times more transcriptionally active in plants grown at 37° for 24 hr compared to plants kept at 20° (Nakajima and Mulligan 2001). Here, the genes encoding components of the plastid-encoded RNA polymerase exhibited an increase after 12 hr of heat stress, suggesting that transcriptional activity may partially contribute to the increase in pncRNAs (Figure 4A). Differences in ribonuclease activity at elevated temperatures could also underlie altered pncRNA stability. Candidates would include RNase J, which prevents the overaccumulation of antisense RNAs (asRNAs), including an *as-rbcL* similar to the one observed here (Sharwood *et al.* 2011a), as well as PNPase, RNase II/RNR1, and Mini-III, which have all been shown to affect the pncRNA population (Sharwood *et al.* 2011b; Hotto *et al.* 2011, 2015). As shown in Figure 4B, transcripts from nuclear genes encoding chloroplast RNases exhibited both declines

(CSP41, RNC3, and RNase J) and increases (RNC4, RNase E, YbeY, and RNase II/R) upon heat stress. In particular, the RNase J transcript decreased, raising the possibility that this could contribute to the accumulation of pncRNAs. However, possible heat effects on enzyme abundance and catalytic activity also need to be taken into account.

### Intronic and some exonic reads accumulate under heat stress

ChloroSeq analysis 1 also produces exon and intron RPKM values used to identify differential accumulation of reads for these features in the chloroplast. We first used a box plot to inspect the sum total of exon and intron values upon heat stress (Figure S2). Only after 12 hr of heat stress did total exonic RPKMs appear higher than the control condition, but the difference is not statistically significant (Figure S2A). However, this trend is amplified for introns, for which median RPKMs are approximately fivefold higher after 12 hr of heat stress compared to control and other stress conditions (Figure S2B). Examination of individual exons and introns revealed that at 12 hr of heat stress, a small number of exons appeared to be responsible for the overall RPKM increase (Figure 5A), whereas the majority of introns evinced increased accumulation under the same condition (Figure 5B).



**Figure 6** Heat stress inhibits splicing. (A) Box plot representation of splicing efficiencies under different stress conditions. The horizontal bars represent median splicing efficiencies and the top and bottom of the boxes represent 25 and 75% of the distribution, respectively. The top and bottom whiskers represent the highest and lowest efficiencies, respectively. The tRNA introns were omitted from the analysis because of poor coverage in this type of RNA-Seq library. \*\*  $P < 0.01$  in a Student's t-test. (B) Splicing efficiency change after 3 hr and 12 hr of heat stress as compared to control conditions. (C) Relationship between the splicing efficiency of the *atpF* intron (horizontal dashed boxes; Y2 axis) and expression of RNAs encoding protein factors involved in its splicing under different stress conditions (RPKM; Y1 axis). RNA-Seq, RNA sequencing; RPKM, reads per kilobase per million mapped reads.

The *atpF* and *petB* transcripts were further analyzed by RNA gel blot. For both introns, the transcript pattern changed dramatically after 12 hr of heat with a trend toward longer, unprocessed products. Quantitatively, more intron-containing transcripts were detected at 12 hr for *atpF* and *petB*, but less so for *atpF*, consistent with the RNA-Seq result (Figure 5C). Similarly, transcript processing for the corresponding exonic regions drifted toward unprocessed, longer transcripts, although quantitatively *atpF* seemed to decrease while *petB* appeared largely unchanged. Reduced processing would tend to increase intron-containing reads compared to those from exons, and could result from increased transcription overloading the splicing machinery. Reduced RNase activity might also play into the observed results.

### Heat stress impairs splicing

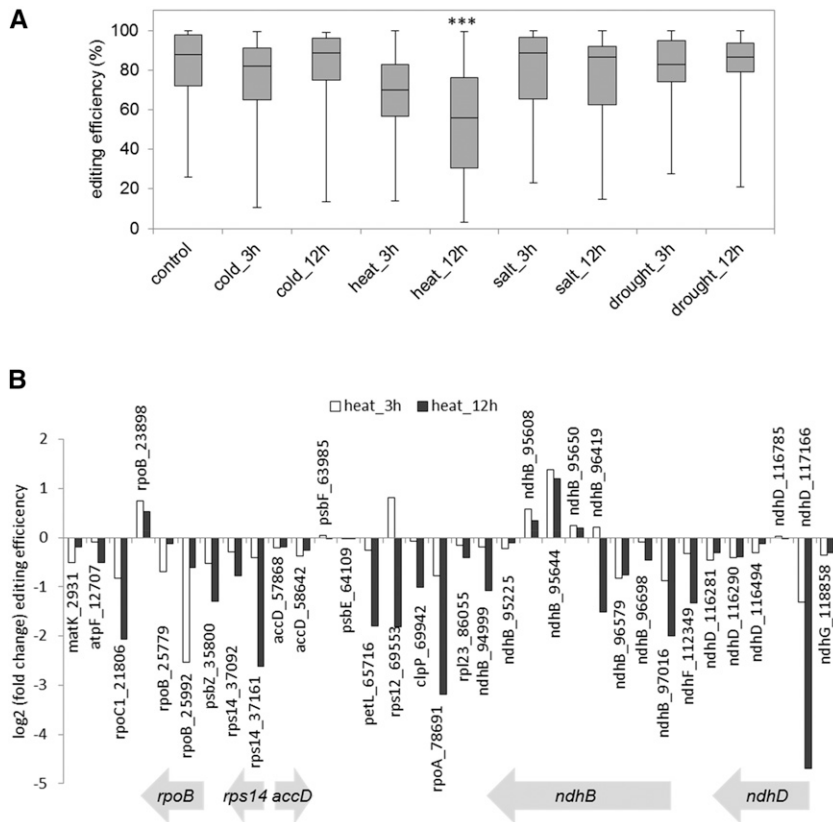
Splicing in the chloroplast was quantified using ChloroSeq analysis 2, which measures the ratio of spliced to total RNA for a given gene. In this case, splicing efficiency for the six tRNA introns in *Arabidopsis* cannot be calculated accurately, as the spliced tRNAs fall below the 100 nt threshold that was excluded from the library preparation protocol used (Di *et al.* 2014). However, other strand-specific libraries may have higher cutoffs that make them amenable to tRNA splicing analysis.

When the effect of abiotic stress on collective chloroplast splicing apart from tRNAs was examined, a strong inhibition specific to heat

stress could be identified (Figure 6A). Individually, all introns are less efficiently spliced after 12 hr of heat stress, with the exception of *clpP* intron 2 (Figure 6B). Heat-sensitive splicing was also documented for *ndhB* in tobacco (Karcher and Bock 2002), and for *ycf3* intron 1 in the barley CL3 mutant, where two point mutations in the intron were proposed to alter its secondary structure in a thermosensitive manner (Landau *et al.* 2009). Whether wild-type intron structures are sensitive to temperature in a functionally significant manner, or whether more rapid transcription impairs proper intron folding (reviewed by Schmitz-Linneweber and Barkan 2007), remains to be determined.

Akin to mechanisms underlying altered RNA accumulation, splicing *trans* factor expression could be related to splicing efficiency in response to heat stress. These nucleus-encoded factors (reviewed in Germain *et al.* 2013) include both gene-specific members and others that service multiple introns. To achieve the general splicing inhibition observed would connote a suppression of at least one factor required for the splicing of each intron, which was not observed at the RNA level (Figure S3A). This is well illustrated by the *atpF* intron, for which four proteins (WHY1, RNC1, WTF1, and CRS1) (Kroeger *et al.* 2009; Prikryl *et al.* 2008; Till *et al.* 2001; Watkins *et al.* 2007) are known to be involved in its splicing. Of these, only transcripts for WHY1 experience a drastic reduction under heat stress (Figure 6C). This reduction is, however, not specific to heat stress and therefore cannot explain by





**Figure 7** Heat stress inhibits RNA editing. (A) Box plot representation of editing efficiencies under different stress conditions. The horizontal bar represents the median editing efficiencies and the top and bottom of the boxes represent 25 and 75% of the distribution, respectively. The top and bottom whiskers represent the highest and lowest editing efficiencies, respectively. Recently-discovered editing sites (Ruwe *et al.* 2013) were not included in the analysis, because they are poorly edited even under control conditions. \*\*\*  $P < 0.001$  in a Student's *t*-test. (B) Change in editing efficiency after 3 hr and 12 hr of heat stress, compared to control conditions. Genes containing more than one editing site are indicated by gene models across the bottom.

itself the *atpF* splicing inhibition. Although WTF1 expression shows an inverse correlation with *atpF* splicing under heat stress, reciprocity is not observed after 12 hr of salt stress. Similarly, the *CRS1* gene is expressed at a higher level in green tissues than in roots, correlating with *atpF* splicing, but is also highly expressed in the leaf base where splicing is less efficient (Till *et al.* 2001). Finally, the chloroplast-encoded splicing factor MatK, which is known to have a role in the removal of several introns (Zoschke *et al.* 2010), appears to be upregulated after 12hr of heat stress, in contrast to the general deficiency in splicing (Figure S3A). Therefore, the results to date do not support a direct link between expression at the RNA level of a given splicing factor and splicing efficiency.

### Heat stress and RNA editing

Editing in the chloroplast was quantified using ChloroSeq analysis 3. Like splicing, editing showed a significant collective reduction after 12 hr of heat stress (Figure 7A). Editing was also reduced for most individual sites, although some displayed little variation in editing efficiency (*psbF\_63985*, *psbE\_64109*, *ndhD\_116785*, *ndhB\_95225*, or *rpl23\_86055*), or modest increases (*rpoB\_23898*, *ndhB\_95608*, and *ndhB\_95644*) (Figure 7B). Different editing sites within a single transcript sometimes displayed opposite behavior, as illustrated by *rpoB*, *ndhB*, and *ndhD*. The reduction of editing that we observed for *rps14* (Figure 7B) also occurs in heat-stressed maize (Nakajima and Mulligan 2001). Site-specific influence of heat on *ndhB* editing has been reported for tobacco (Karcher and Bock 1998, 2002), and our results are in partial concurrence. For example, editing sites 96579 (EII in tobacco) and 95650 (EVI) display the same behavior in both species, being heat-sensitive and insensitive, respectively, whereas site 95644 (EVII), which was described as heat-insensitive in tobacco (Karcher and Bock 2002), is increasingly edited upon heat stress in *Arabidopsis* (Figure 7B).

Some sites show a different behavior. Editing at *rpoB\_25992* displays a stronger reduction after 3 hr of heat stress compared to 12 hr, and the *rps12\_69553* site experiences an increase in editing after 3 hr and a decrease after 12 hr. It is worth noting that two recently identified proteins involved in chloroplast RNA editing, RIP2 and RIP9 (Bentolila *et al.* 2013), show either an opposite behavior between 3 and 12 hr (RIP2 is first downregulated then upregulated) or a stronger downregulation after 3 hr than 12 hr (RIP9) (Figure S3B). In some cases, editing factor expression correlates with editing efficiency (Figure S3B). For example, CRR21 and OTP80 expression remained stable under heat stress, as did their corresponding editing sites *ndhD\_116785* and *rpl23\_86055*. On the other hand, editing sites *rps14\_37161*, *clpP\_69942*, and *rpoA\_78691* displayed a large reduction in editing, while expression of their editing factors OTP86 and CLB19 increased or remained stable, respectively. Finally, editing at *ndhD\_117166*, which creates the AUG initiation codon of the *ndhD* transcript, is strongly inhibited under heat stress. Its editing requires the cooperative association between CRR4 and DYW1 (Boussardon *et al.* 2012; Kotera *et al.* 2005), and the latter shows a lower expression under heat stress. The NDHD protein is a subunit of the NADH dehydrogenase complex, which is postulated to have a role in plant stress acclimation (Rumeau *et al.* 2007). Increasing evidence now directly links RNA editing regulation to chloroplast development and stress responses, independently of editing factor regulation, and the wide range of behavior observed here under heat stress may be understood in this framework (Garcia-Andrade *et al.* 2013; Kakizaki *et al.* 2012; Zhang *et al.* 2014).

### Conclusions

We have presented ChloroSeq, a bioinformatics pipeline developed to investigate the chloroplast transcriptome systematically utilizing RNA-Seq data. Using ChloroSeq, information on chloroplast transcript

accumulation, splicing, and editing can be determined simultaneously, and can be combined with nuclear transcript analysis. Although not shown here, ChloroSeq can also be used to study the mitochondrial transcriptome, which features some of same RNA-Seq analytical challenges as the chloroplast (Stone and Storchova 2015). Moreover, we see no reason why ChloroSeq could not be used outside of the plant kingdom. Most of the challenges posed by the derived organellar genomes of some unicellular eukaryotes, for example those of dinoflagellate chloroplasts (Howe *et al.* 2008), concern the alignment step, upstream of ChloroSeq. Once the alignment is performed, one can then use customized annotation files, making this pipeline easily adaptable to multiple research questions.

Here, we demonstrated the power of ChloroSeq through analysis of previously published datasets on abiotic stress, for which organellar data had not been studied. Among the stresses, heat stress was specifically associated with multifaceted effects on the chloroplast transcriptome. These included increased pncRNA accumulation, and decreases in splicing and editing efficiencies, all of which are phenomena that could potentially impact chloroplast protein synthesis. A global decrease in chloroplast protein synthesis following heat stress was earlier observed in *Phaseolus vulgaris* (Süss and Yordanov 1986). The main exception in that study was the PSII reaction center protein D1, which interestingly requires neither splicing nor editing for proper expression. A thorough review of the interaction of abiotic stresses with photosynthesis, and some of the response mechanisms, was recently published (Gururani *et al.* 2015), but the role of the chloroplast genome was not addressed. This highlights the tremendous opportunity ChloroSeq offers to gain knowledge of an essential participant in plant physiology, the chloroplast transcriptome.

## ACKNOWLEDGMENTS

We thank the Boyce Thompson Institute Bioinformatics Help Desk for assistance with experimental methods. This is a consulting service for early-stage ideas or issues in bioinformatics. This work was supported by grant DE-FG02-10ER20015 from the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences, of the U.S. Department of Energy. The funding supported B.C. and A.M.H. and associated laboratory costs.

Author contributions: B.C. and S.R.S. created and optimized ChloroSeq; B.C. conducted the analyses presented here; A.M.H. and B.C. conducted the wet lab experiments; D.B.S. helped to conceive and supervise the research, and participated along with all other authors in the writing of the manuscript. All authors read and approved the final manuscript. The authors declare no conflict of interest

## LITERATURE CITED

Allorent, G., F. Courtois, F. Chevalier, and S. Lerbs-Mache, 2013 Plastid gene expression during chloroplast differentiation and dedifferentiation into non-photosynthetic plastids during seed formation. *Plant Mol. Biol.* 82(1–2): 59–70.

Barkan, A., 2011 Expression of plastid genes: organelle-specific elaborations on a prokaryotic scaffold. *Plant Physiol.* 155(4): 1520–1532.

Bentolila, S., J. Oh, M. R. Hanson, and R. Bukowski, 2013 Comprehensive high-resolution analysis of the role of an *Arabidopsis* gene family in RNA editing. *PLoS Genet.* 9(6): e1003584.

Boussardon, C., V. Salone, A. Avon, R. Berthome, K. Hammani *et al.*, 2012 Two interacting proteins are necessary for the editing of the *ndhD-1* site in *Arabidopsis* plastids. *Plant Cell* 24(9): 3684–3694.

Cahoon, A. B., E. M. Takacs, R. M. Sharpe, and D. B. Stern, 2008 Nuclear, chloroplast, and mitochondrial transcript abundance along a maize leaf developmental gradient. *Plant Mol. Biol.* 66(1–2): 33–46.

Castandet, B., A. M. Hotto, Z. Fei, and D. B. Stern, 2013 Strand-specific RNA sequencing uncovers chloroplast ribonuclease functions. *FEBS Lett.* 587(18): 3096–3101.

Chen, S., C. Kim, J. M. Lee, H. A. Lee, Z. Fei *et al.*, 2015 Blocking the  $Q_B$ -binding site of photosystem II by tenuazonic acid, a non-host-specific toxin of *Alternaria alternata*, activates singlet oxygen-mediated and EXECUTER-dependent signalling in *Arabidopsis*. *Plant Cell Environ.* 38(6): 1069–1080.

Di, C., J. Yuan, Y. Wu, J. Li, H. Lin *et al.*, 2014 Characterization of stress-responsive lncRNAs in *Arabidopsis thaliana* by integrating expression, epigenetic and structural features. *Plant J.* 80(5): 848–861.

Garcia-Andrade, J., V. Ramirez, A. Lopez, and P. Vera, 2013 Mediated plastid RNA editing in plant immunity. *PLoS Pathog.* 9(10): e1003713.

Germain, A., S. Herlich, S. Larom, S. H. Kim, G. Schuster *et al.*, 2011 Mutational analysis of *Arabidopsis* chloroplast polynucleotide phosphorylase reveals roles for both RNase PH core domains in polyadenylation, RNA 3'-end maturation and intron degradation. *Plant J.* 67(3): 381–394.

Germain, A., A. M. Hotto, A. Barkan, and D. B. Stern, 2013 RNA processing and decay in plastids. *Wiley Interdiscip. Rev. RNA* 4(3): 295–316.

Ghosh, S., and C. K. Chan, 2016 Analysis of RNA-Seq data using TopHat and Cufflinks. *Methods Mol. Biol.* 1374: 339–361.

Gururani, M. A., J. Venkatesh, and L. S. Tran, 2015 Regulation of photosynthesis during abiotic stress-induced photoinhibition. *Mol. Plant* 8(9): 1304–1320.

Hotto, A. M., R. J. Schmitz, Z. Fei, J. R. Ecker, and D. B. Stern, 2011 Unexpected diversity of chloroplast noncoding RNAs as revealed by deep sequencing of the *Arabidopsis* transcriptome. *G3 (Bethesda)* 1(7): 559–570.

Hotto, A. M., A. Germain, and D. B. Stern, 2012 Plastid non-coding RNAs: emerging candidates for gene regulation. *Trends Plant Sci.* 17(12): 737–744.

Hotto, A. M., B. Castandet, L. Gilet, A. Higdon, C. Condon *et al.*, 2015 *Arabidopsis* chloroplast Mini-Ribonuclease III participates in rRNA maturation and intron recycling. *Plant Cell* 27(3): 724–740.

Howe, C. J., R. E. Nisbet, and A. C. Barbrook, 2008 The remarkable chloroplast genome of dinoflagellates. *J. Exp. Bot.* 59(5): 1035–1045.

Kahlau, S., and R. Bock, 2008 Plastid transcriptomics and translomics of tomato fruit development and chloroplast-to-chromoplast differentiation: chromoplast gene expression largely serves the production of a single protein. *Plant Cell* 20(4): 856–874.

Kakizaki, T., F. Yazu, K. Nakayama, Y. Ito-Inaba, and T. Inaba, 2012 Plastid signalling under multiple conditions is accompanied by a common defect in RNA editing in plastids. *J. Exp. Bot.* 63(1): 251–260.

Karcher, D., and R. Bock, 1998 Site-selective inhibition of plastid RNA editing by heat shock and antibiotics: a role for plastid translation in RNA editing. *Nucleic Acids Res.* 26(5): 1185–1190.

Karcher, D., and R. Bock, 2002 Temperature sensitivity of RNA editing and intron splicing reactions in the plastid *ndhB* transcript. *Curr. Genet.* 41(1): 48–52.

Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley *et al.*, 2013 TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14(4): R36.

Kodama, Y., M. Shumway, and R. Leinonen, 2012 The sequence read archive: explosive growth of sequencing data. *Nucleic Acids Res.* 40(D1): D54–D56.

Kotera, E., M. Tasaka, and T. Shikanai, 2005 A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature* 433(7023): 326–330.

Kroeger, T. S., K. P. Watkins, G. Friso, K. J. van Wijk, and A. Barkan, 2009 A plant-specific RNA-binding domain revealed through analysis of chloroplast group II intron splicing. *Proc. Natl. Acad. Sci. USA* 106(11): 4537–4542.

Landau, A. M., H. Lokstein, H. V. Scheller, V. Lainez, S. Maldonado *et al.*, 2009 A cytoplasmically inherited barley mutant is defective in photosystem I assembly due to a temperature-sensitive defect in *ycf3* splicing. *Plant Physiol.* 151(4): 1802–1811.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16): 2078–2079.

Loening, U. E., and J. Ingle, 1967 Diversity of RNA components in green plant tissues. *Nature* 215(5099): 363–367.

- Martin, L. B., Z. Fei, J. J. Giovannoni, and J. K. Rose, 2013 Catalyzing plant science research with RNA-seq. *Front. Plant Sci.* 4: 66.
- Nagashima, A., M. Hanaoka, R. Motohashi, M. Seki, K. Shinozaki *et al.*, 2004 DNA microarray analysis of plastid gene expression in an *Arabidopsis* mutant deficient in a plastid transcription factor sigma, SIG2. *Biosci. Biotechnol. Biochem.* 68(3): 694–704.
- Nakajima, Y., and R. M. Mulligan, 2001 Heat stress results in incomplete C-to-U editing of maize chloroplast mRNAs and correlates with changes in chloroplast transcription rate. *Curr. Genet.* 40(3): 209–213.
- Nakamura, T., Y. Furuhashi, K. Hasegawa, H. Hashimoto, K. Watanabe *et al.*, 2003 Array-based analysis on tobacco plastid transcripts: preparation of a genomic microarray containing all genes and all intergenic regions. *Plant Cell Physiol.* 44(8): 861–867.
- Prikryl, J., K. P. Watkins, G. Friso, K. J. van Wijk, and A. Barkan, 2008 A member of the Whirly family is a multifunctional RNA- and DNA-binding protein that is essential for chloroplast biogenesis. *Nucleic Acids Res.* 36(16): 5152–5165.
- Quinlan, A.R., 2014 BEDTools: the Swiss-Army tool for genome Feature analysis. *Current Protoc. Bioinformatics* 47:11.12.11–11.12.34.
- Rorbach, J., A. Bobrowicz, S. Pearce, and M. Minczuk, 2014 Polyadenylation in bacteria and organelles. *Methods Mol. Biol.* 1125: 211–227.
- Rumeau, D., G. Peltier, and L. Cournac, 2007 Chlororespiration and cyclic electron flow around PSI during photosynthesis and plant stress response. *Plant Cell Environ.* 30(9): 1041–1051.
- Ruwe, H., and C. Schmitz-Linneweber, 2012 Short non-coding RNA fragments accumulating in chloroplasts: footprints of RNA binding proteins? *Nucleic Acids Res.* 40(7): 3106–3116.
- Ruwe, H., B. Castandet, C. Schmitz-Linneweber, and D. B. Stern, 2013 *Arabidopsis* chloroplast quantitative editotype. *FEBS Lett.* 587(9): 1429–1433.
- Schmitz-Linneweber, C., and A. Barkan, 2007 RNA splicing and RNA editing in chloroplasts, pp. 213–248 in *Cell and Molecular Biology of Plastids*, edited by Bock, R. Springer-Verlag, Berlin/Heidelberg.
- Schmitz-Linneweber, C., R. Williams-Carrier, and A. Barkan, 2005 RNA immunoprecipitation and microarray analysis show a chloroplast pentatricopeptide repeat protein to be associated with the 5' region of mRNAs whose translation it activates. *Plant Cell* 17(10): 2791–2804.
- Sharwood, R. E., M. Halpert, S. Luro, G. Schuster, and D. B. Stern, 2011a Chloroplast RNase J compensates for inefficient transcription termination by removal of antisense RNA. *RNA* 17(12): 2165–2176.
- Sharwood, R. E., A. M. Hotto, T. J. Bollenbach, and D. B. Stern, 2011b Overaccumulation of the chloroplast antisense RNA AS5 is correlated with decreased abundance of 5S rRNA in vivo and inefficient 5S rRNA maturation in vitro. *RNA* 17(2): 230–243.
- Smith, D. R., 2013 RNA-Seq data: a goldmine for organelle research. *Brief. Funct. Genomics.* 12: 454–456.
- Stern, D. B., M. Goldschmidt-Clermont, and M. R. Hanson, 2010 Chloroplast RNA metabolism. *Annu. Rev. Plant Biol.* 61: 125–155.
- Stone, J. D., and H. Storchova, 2015 The application of RNA-seq to the comprehensive analysis of plant mitochondrial transcriptomes. *MGG* 290(1): 1–9.
- Süss, K.-H., and I. T. Jordanov, 1986 Biosynthetic cause of *in vivo* acquired thermotolerance of photosynthetic light reactions and metabolic responses of chloroplasts to heat stress. *Plant Physiol.* 81(1): 192–199.
- Till, B., C. Schmitz-Linneweber, R. Williams-Carrier, and A. Barkan, 2001 CRS1 is a novel group II intron splicing factor that was derived from a domain of ancient origin. *RNA* 7(9): 1227–1238.
- Trapnell, C., A. Roberts, L. Goff, G. Pertea, D. Kim *et al.*, 2012 Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7(3): 562–578.
- Valkov, V. T., N. Scotti, S. Kahlau, D. Maclean, S. Grillo *et al.*, 2009 Genome-wide analysis of plastid gene expression in potato leaf chloroplasts and tuber amyloplasts: transcriptional and posttranscriptional control. *Plant Physiol.* 150(4): 2030–2044.
- Walter, M., J. Kilian, and J. Kudla, 2002 PNPase activity determines the efficiency of mRNA 3'-end processing, the degradation of tRNA and the extent of polyadenylation in chloroplasts. *EMBO J.* 21(24): 6905–6914.
- Wang, L., X. Yu, H. Wang, Y. Z. Lu, M. de Ruiter *et al.*, 2011 A novel class of heat-responsive small RNAs derived from the chloroplast genome of Chinese cabbage (*Brassica rapa*). *BMC Genomics* 12: 289.
- Watkins, K. P., T. S. Kroeger, A. M. Cooke, R. E. Williams-Carrier, G. Friso *et al.*, 2007 A ribonuclease III domain protein functions in group II intron splicing in maize chloroplasts. *Plant Cell* 19(8): 2606–2623.
- Weber, A. P., K. L. Weber, K. Carr, C. Wilkerson, and J. B. Ohlrogge, 2007 Sampling the Arabidopsis transcriptome with massively parallel pyrosequencing. *Plant Physiol.* 144(1): 32–42.
- Woodson, J. D., J. M. Perez-Ruiz, R. J. Schmitz, J. R. Ecker, and J. Chory, 2013 Sigma factor-mediated plastid retrograde signals control nuclear gene expression. *Plant J.* 73(1): 1–13.
- Zhang, F., W. Tang, B. Hedtke, L. Zhong, L. Liu *et al.*, 2014 Tetrapyrrole biosynthetic enzyme protoporphyrinogen IX oxidase 1 is required for plastid RNA editing. *Proc. Natl. Acad. Sci. USA* 111(5): 2023–2028.
- Zhelyazkova, P., K. Hammani, M. Rojas, R. Voelker, M. Vargas-Suarez *et al.*, 2012a Protein-mediated protection as the predominant mechanism for defining processed mRNA termini in land plant chloroplasts. *Nucleic Acids Res.* 40(7): 3092–3105.
- Zhelyazkova, P., C. M. Sharma, K. U. Forstner, K. Liere, J. Vogel *et al.*, 2012b The primary transcriptome of barley chloroplasts: numerous noncoding RNAs and the dominating role of the plastid-encoded RNA polymerase. *Plant Cell* 24(1): 123–136.
- Zmienko, A., M. Guzowska-Nowowiejska, R. Urbaniak, W. Plader, P. Formanowicz *et al.*, 2011 A tiling microarray for global analysis of chloroplast genome expression in cucumber and other plants. *Plant Methods* 7: 29.
- Zoschke, R., M. Nakamura, K. Liere, M. Sugiura, T. Borner *et al.*, 2010 An organellar maturase associates with multiple group II introns. *Proc. Natl. Acad. Sci. USA* 107(7): 3245–3250.
- Zoschke, R., K. P. Watkins, and A. Barkan, 2013 A rapid ribosome profiling method elucidates chloroplast ribosome behavior *in vivo*. *Plant Cell* 25(6): 2265–2275.

Communicating editor: J. K. Kim