Unexpected Diversity of Chloroplast Noncoding RNAs as Revealed by Deep Sequencing of the Arabidopsis Transcriptome

Amber M. Hotto,* Robert J. Schmitz,* Zhongjun Fei,*† Joseph R. Ecker,*§ and David B. Stern*,1

*Boyce Thompson Institute for Plant Research and †Robert W. Holley Center for Agriculture and Health, Agricultural Research Service, United States Department of Agriculture, Ithaca, New York 14853, †Plant Biology and Genomic Analysis Laboratory, and §Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, California 92037

ABSTRACT Noncoding RNAs (ncRNA) are widely expressed in both prokaryotes and eukaryotes. Eukaryotic ncRNAs are commonly micro- and small-interfering RNAs (18–25 nt) involved in posttranscriptional gene silencing, whereas prokaryotic ncRNAs vary in size and are involved in various aspects of gene regulation. Given the prokaryotic origin of organelles, the presence of ncRNAs might be expected; however, the full spectrum of chloroplast ncRNAs has not been determined systematically. Here, strand-specific RNA-Seq analysis was used to identify 107 candidate ncRNAs from Arabidopsis thaliana chloroplasts, primarily encoded opposite protein-coding and tRNA genes. Forty-eight ncRNAs were shown to accumulate by RNA gel blot as discrete transcripts in wild-type (WT) plants and/or the pnp1-1 mutant, which lacks the chloroplast ribonuclease polynucleotide phosphorylase (cpPNPase). Ninety-eight percent of the ncRNAs detected by RNA gel blot had different transcript patterns between WT and pnp1-1, suggesting cpPNPase has a significant role in chloroplast ncRNA biogenesis and accumulation. Analysis of materials deficient for other major chloroplast ribonucleases, RNase R, RNase E, and RNase J, showed differential effects on ncRNA accumulation and/or form, suggesting specificity in RNase-ncRNA interactions. 5' end mapping demonstrates that some ncRNAs are transcribed from dedicated promoters, whereas others result from transcriptional read-through. Finally, correlations between accumulation of some ncRNAs and the symmetrically transcribed sense RNA are consistent with a role in RNA stability. Overall, our data suggest that this extensive population of ncRNAs has the potential to underpin a previously underappreciated regulatory mode in the chloroplast.

Both prokaryotes and eukaryotes express a large number of noncoding RNAs (ncRNA) antisense to coding regions, ranging from ~9% (Arabidopsis) to 29% (mouse) of identified genes (Jen et al. 2005; Wang et al. 2005; Zhang et al. 2006). Although the term ncRNA refers to abundant transcripts, such as rRNAs and tRNAs, it has more recently incorporated regulatory RNAs, such as bacterial small RNAs (sRNA; <400 nt), eukaryotic micro- and small-interfering RNAs (miRNA and siRNA), antisense RNAs (asRNA) and long noncoding RNAs (lncRNA). The present study focuses on asRNAs, which can be divided into two groups based on the target interaction: cis-encoded asRNAs bind to and regulate the complementary sense RNA, and trans-encoded asRNAs act on one or more unlinked loci through short regions of complementarity. Base pairing of these asRNAs to their targets can elicit translational inactivation/activation, mRNA stabilization/destabilization, or differential transcription termination (Storz et al. 2005; Repoil and Darfeuille 2009).

The occurrence of plant organelar ncRNAs has been established by limited analysis of specific cDNA populations. Studies have revealed mitochondrial ncRNAs from wild-type (WT) and mitochondrial...
polynucleotide phosphorylase (PNPase)-deficient Arabidopsis (Holec et al. 2006; Lung et al. 2006) and short (~<500 nt) ncRNAs from tobacco and Arabidopsis chloroplasts (Marker et al. 2002; Lung et al. 2006). Subsequent studies have attempted to elucidate the regulatory roles of chloroplast ncRNAs. In one case, an ndhB asRNA was hypothesized to stabilize or regulate maturation of the cognate sense transcript, whereas a psbT asRNA was proposed to regulate accumulation of PsbT protein through occlusion of the psbT ribosomal binding site (Georg et al. 2010; Zghidi-Abouzid et al. 2011). Our own work suggested a role for an asRNA complementary to the SS rRNA, ASS, in regulating the processing and accumulation of SS rRNA (Hotto et al. 2010; Sharwood et al. 2011).

There are at least two indications that the organelar ncRNA population might be significantly more complex than elucidated to date. First, in cyanobacteria, which represent the chloroplast progenitor, numerous ncRNAs have been identified, some of which accumulate differentially in response to stress or developmental stage (Steglich et al. 2008; Georg et al. 2009). Second, transcription termination in chloroplasts has long been known to be inefficient (Stern and Grussem 1987), suggesting that intergenic and antisense regions may be readily transcribed.

Taken together, it is clear that ncRNAs accumulate in chloroplasts, and available evidence favors functional roles in gene expression, at least for some. To gain a more complete picture of this population, we have used strand-specific, high-throughput RNA sequencing (RNA-Seq) of total RNA. We present data extracted from this sequencing, demonstrating that accumulating ncRNAs are derived from much of the chloroplast genome. The biogenesis and regulation of these RNAs are further examined through analysis of chloroplast ribonuclease mutants and 5′ end mapping.

MATERIALS AND METHODS

Plant growth conditions and material

Arabidopsis thaliana Columbia ecotype (Col-0) was used as the WT for this study. The three T-DNA mutants used contain insertions in the genes At3g03710 (pnp1-1; SALK_013306), At5g02250 (rne1-1; SALK_090294), and At2g04270 (rne1-1; SALK_093546) and have been previously characterized (Alonso et al. 2003; Bollenbach et al. 2005; Mudd et al. 2008; Marchive et al. 2009). WT and pnp1-1 plants were germinated and grown on soil with a 16-h light/dark photoperiod (Georg et al. 2010). Affected tissue was harvested after 40 days from two WT and two pnp1-1 plants for subsequent analysis. RNase III was depleted of rRNAs using the plant RiboMinus-Kit (Invitrogen). The ncRNA cDNA was extracted using TRI reagent (Molecular Research Center) with minor modifications in this study. Mature leaf tissue was ground in liquid nitrogen, and total RNA was depleted of rRNAs using the plant RiboMinus-Kit (Invitrogen). The resulting reads were aligned to the Arabidopsis genome (TAIR9) using Tophat (version 1.0.13/Bowtie (version 0.12.3) with the following commands: -F 0 -g 2 -i 5000 (Langmead et al. 2009; Trapnell et al. 2010). Up to two locations were accepted for placement of sequenced reads to allow mapping to the large inverted repeat of the chloroplast genome. Sequence data can be downloaded from National Center for Biotechnology Information Sequence Read Archive SRA046998.

RNA-Seq

Total RNA was extracted using the RNasy Plant Mini Kit (Qiagen) from two WT and two pnp1-1 samples. Ten micrograms total RNA were depleted of rRNAs using the plant RiboMinus-Kit (Invitrogen). Strand-specific RNA-Seq libraries were prepared following the “Directional mRNA-Seq Library Prep Pre-Release” protocol by Illumina. Each sample was sequenced for 85 cycles on an Illumina GAIIx. Image analysis and base calling were performed with the standard Illumina pipeline (Firecrest v1.3.4 and Bustard v1.3.4). The resulting reads were aligned to the Arabidopsis genome (TAIR9) using Tophat (version 1.0.13/Bowtie (version 0.12.3) with the following commands: -F 0 -g 2 -i 5000 (Langmead et al. 2009; Trapnell et al. 2010). Up to two locations were accepted for placement of sequenced reads to allow mapping to the large inverted repeat of the chloroplast genome. Sequence data can be downloaded from National Center for Biotechnology Information Sequence Read Archive SRA046998.

RNA isolation, RT-PCR, and RNA gel blots

Mature leaf tissue was ground in liquid nitrogen, and total RNA extracted using TRI reagent (Molecular Research Center) with minor modifications to the manufacturer’s instructions. RNA was precipitated overnight with isopropanol at −20°C, and the pellet was washed with 75% ethanol and dissolved in water. Primers were designed for cDNA synthesis, PCR, and RNA blot probe synthesis to amplify a ≤100 nt ncRNA section from the RNA-Seq data using Primer3 (Rozen and Skaletsky 2000). For strand-specific cDNA synthesis, 1 μg of DNase-treated RNA (Promega) was reverse-transcribed with SuperScript III (Invitrogen) using 2 μM of the 3′ ncRNA gene-specific primers (supporting information, Table S1). The PCR reaction contained 1X Master Mix, 0.2 mM each dNTP, 400 nM each 5′ and 3′ primer, 1.25 U GoTaq DNA polymerase (Promega), and 100 ng of cDNA in a 25 μL reaction volume. Amplification was completed with the following protocol: initial denaturation at 94°C for 3 min, then 30 cycles at 94°C for 30 s, 60°C for 30 s, and 72°C for 30 s, and a final extension of 72°C for 7 min. Amplicons were visualized after migration in 2% agarose gels.

For RNA gel blot analysis, 5 μg of total RNA was separated in 1.2% agarose/formaldehyde gels, which were blotted overnight onto Hybond-N+ (GE Healthcare) in 25 mM sodium phosphate buffer. Membranes were probed with single-stranded DNA or double-stranded DNA probes as indicated in the figure legends. The ncRNA templates for probe synthesis were amplified with the 5′ primer and the 3′ primer containing a T7 promoter, and the sense strand templates for probe synthesis were amplified with the corresponding ncRNA 3′ primer and 5′ primer containing a T7 promoter. The only exception was the sense strand tRNA template, which was amplified with tRNA-specific primers given in Table S1. RNA probes were made from 100 ng of template using T7 RNA polymerase (Promega) and 40 μCi α-32P-UTP, and purification through a Sephadex G-25 column. Membrane hybridization and washing were performed as previously described (Hotto et al. 2010). Where indicated in Figure 4, dsDNA probes synthesized from 100 ng of template were used, with hybridization according to Church and Gilbert (1984).

5′ RACE

5′ RACE used the GeneRacer Kit (Invitrogen) with minor modifications. DNase-treated (Ambion) total RNA (4 μg) was incubated with and without tobacco acid phosphatase (TAP), followed by ligation to the GeneRacer RNA Oligo with T4 RNA ligase (treatment with calf intestinal phosphatase was omitted). The ncRNA cDNA was
PnPase has an active role in bacterial ncRNA regulation, and dopsis mitochondrial ncRNAs overaccumulate in plants due to identify potential nc- and asRNAs. We examined chloroplast PNPase (cpPNPase) null mutant Total RNA isolated from mature leaves of Arabidopsis the mitochondrial isozyme (Holec et al. 2008). To achieve strand-specific sequencing of the chloroplast transcriptome, 5′ and 3′ oligonucleotides were sequentially added to total RNA after depletion of rRNA and metal hydrolysis, a method that retains transcripts longer than approximately 80 nt. The Illumina Genome Analyzer IIx platform was used to sequence this pool, resulting in a strand-specific RNA-Seq dataset. Sequences that aligned to the chloroplast genome were extracted, resulting in an average of 10,545,033 reads per sample and coverage of greater than 99% of both strands of the WT chloroplast genome, including unannotated regions. For this study, we focused on identification of putative asRNAs with at least partial complementarity to a known sense transcript and identification of ncRNAs that are complementary to intergenic regions within known gene clusters, thus excluding possible ncRNAs derived from cleavage of known functional transcripts. To limit identification of false positives, a minimum of 50× coverage at any nucleotide in the WT and/or pnp1-1 was defined as a transcription peak corresponding to an ncRNA candidate.

One challenge was identifying the 5′ and 3′ ends of novel transcripts, particularly those of low abundance. Since RNA-Seq relies on ligation of oligonucleotides to the RNA ends, those ends engaged in secondary structures will be underrepresented, resulting in an uneven distribution of reads across contiguous transcripts and low coverage of ends (Wang et al. 2009; Roberts et al. 2011). For the purpose of this study, an end was assigned when the sequencing coverage fell below 10×. As an example of the application of these criteria, Figure 1A shows two candidate asRNAs, as-psbK and as-psbl, and Figure 1B shows four potential asRNAs, as-ndhD1-4. In these cases, the occurrence of an asRNA candidate was prompted when the depth of coverage exceeded 50× at a single nucleotide, and it was assumed to continue in both directions until the coverage dropped below 10×. In Figure 1A, this resulted in the identification of two asRNAs in this region with a gap between them, as annotated in Table 1. Under such circumstances, however, two peaks might actually correspond to a single ncRNA with poor sequence coverage in the intervening region, as shown by the continuous dashed red line. The ndhD situation was more complex, given that the four individual asRNA peaks might represent as few as one, or as many as four, actual ncRNA species. This type of ambiguity arose rather frequently, with the extreme example being the 10 transcripts defined as antisense to ycf21. The same phenomenon was also seen when sense strand transcripts were examined (insits in Figure 1). Therefore, the fragmentation of transcripts appears to be a limitation to the RNA-Seq method, rather than a peculiarity of ncRNAs themselves.

When the RNA-Seq datasets from both genotypes were analyzed using these criteria, 107 ncRNA candidates were identified (Table 1), ranging in size from 48 to 1,300 nt, with an average size of 217 nt. Of this pool, 12 candidates were defined strictly as ncRNAs because they were complementary to intergenic regions within operons, whereas the remaining 95 candidates were asRNAs as they were antisense to known coding regions. Of the 107 candidates, 29 exceeded the 50× coverage only in pnp1-1, although most were also visible as peaks in the WT reads (e.g. as-psbI and as-ndhD1; Figure 1). Additionally, both sense and antisense transcript abundances were estimated by bimming reads according to their start position against the TAIR9 genome annotation. Overall, asRNAs contained ~4-fold more sequencing reads in pnp1-1 compared with the WT (Figure 1 and Table S2). This observation explains, at least in part, why some candidates were identified from the mutant. It also suggests that, as in Arabidopsis mitochondria (Holec et al. 2006), cpPNPase may globally modulate ncRNA abundance.

The 95 asRNA candidates mentioned above can be classified by the function of the complementary strand. As shown in Figure 2, the distribution of sense strand gene function was broad, with the largest group being miscellaneous protein-coding genes. These include several ycf (hypothetical coding frame) genes, which are genes of unknown function that are conserved between species. However, all gene classes were well represented except rRNAs. In total, more than half of the annotated chloroplast genes had one or more predicted asRNA counterparts. Putative locations of sense-antisense RNA pairing can also be used to predict function, including 5′ end pairing (~31 asRNAs) and 3′ end pairing (~39 asRNAs), and pairing within the coding region (~40 asRNAs; see Discussion). Overall, this suggests a variety of possible regulatory functions and targets.

**RESULTS**

**Analysis of the Arabidopsis chloroplast transcriptome reveals 107 ncRNA candidates**

Total RNA isolated from mature leaves of Arabidopsis WT and the chloroplast PNPase (cpPNPase) null mutant pnp1-1 was sequenced to identify potential nc- and asRNAs. We examined pnp1-1 because PNPase has an active role in bacterial ncRNA regulation, and Arabidopsis mitochondrial ncRNAs overaccumulate in plants deficient for the mitochondrial isoform (Holec et al. 2006; Viegas et al. 2007; Bollenbach et al. 2008). To achieve strand-specific sequencing of the chloroplast transcriptome, 5′ and 3′ oligonucleotides were sequentially added to total RNA after depletion of rRNA and metal hydrolysis, a method that retains transcripts longer than approximately 80 nt. The Illumina Genome Analyzer IIx platform was used to sequence this pool, resulting in a strand-specific RNA-Seq dataset. Sequences that aligned to the chloroplast genome were extracted, resulting in an average of 10,545,033 reads per sample and coverage of greater than 99% of both strands of the WT chloroplast genome, including unannotated regions. For this study, we focused on identification of putative asRNAs with at least partial complementarity to a known sense transcript and identification of ncRNAs that are complementary to intergenic regions within known gene clusters, thus excluding possible ncRNAs derived from cleavage of known functional transcripts. To limit identification of false positives, a minimum of 50× coverage at any nucleotide in the WT and/or pnp1-1 was defined as a transcription peak corresponding to an ncRNA candidate.

One challenge was identifying the 5′ and 3′ ends of novel transcripts, particularly those of low abundance. Since RNA-Seq relies on ligation of oligonucleotides to the RNA ends, those ends engaged in secondary structures will be underrepresented, resulting in an uneven distribution of reads across contiguous transcripts and low coverage of ends (Wang et al. 2009; Roberts et al. 2011). For the purpose of this study, an end was assigned when the sequencing coverage fell below 10×. As an example of the application of these criteria, Figure 1A shows two candidate asRNAs, as-psbK and as-psbl, and Figure 1B shows four potential asRNAs, as-ndhD1-4. In these cases, the occurrence of an asRNA candidate was prompted when the depth of coverage exceeded 50× at a single nucleotide, and it was assumed to continue in both directions until the coverage dropped below 10×. In Figure 1A, this resulted in the identification of two asRNAs in this region with a gap between them, as annotated in Table 1. Under such circumstances, however, two peaks might actually correspond to a single ncRNA with poor sequence coverage in the intervening region, as shown by the continuous dashed red line. The ndhD situation was more complex, given that the four individual asRNA peaks might represent as few as one, or as many as four, actual ncRNA species. This type of ambiguity arose rather frequently, with the extreme example being the 10 transcripts defined as antisense to ycf21. The same phenomenon was also seen when sense strand transcripts were examined (insits in Figure 1). Therefore, the fragmentation of transcripts appears to be a limitation to the RNA-Seq method, rather than a peculiarity of ncRNAs themselves.

When the RNA-Seq datasets from both genotypes were analyzed using these criteria, 107 ncRNA candidates were identified (Table 1), ranging in size from 48 to 1,300 nt, with an average size of 217 nt. Of this pool, 12 candidates were defined strictly as ncRNAs because they were complementary to intergenic regions within operons, whereas the remaining 95 candidates were asRNAs as they were antisense to known coding regions. Of the 107 candidates, 29 exceeded the 50× coverage only in pnp1-1, although most were also visible as peaks in the WT reads (e.g. as-psbI and as-ndhD1; Figure 1). Additionally, both sense and antisense transcript abundances were estimated by bimming reads according to their start position against the TAIR9 genome annotation. Overall, asRNAs contained ~4-fold more sequencing reads in pnp1-1 compared with the WT (Figure 1 and Table S2). This observation explains, at least in part, why some candidates were identified from the mutant. It also suggests that, as in Arabidopsis mitochondria (Holec et al. 2006), cpPNPase may globally modulate ncRNA abundance.

The 95 asRNA candidates mentioned above can be classified by the function of the complementary strand. As shown in Figure 2, the distribution of sense strand gene function was broad, with the largest group being miscellaneous protein-coding genes. These include several ycf (hypothetical coding frame) genes, which are genes of unknown function that are conserved between species. However, all gene classes were well represented except rRNAs. In total, more than half of the annotated chloroplast genes had one or more predicted asRNA counterparts. Putative locations of sense-antisense RNA pairing can also be used to predict function, including 5′ end pairing (~31 asRNAs) and 3′ end pairing (~39 asRNAs), and pairing within the coding region (~40 asRNAs; see Discussion). Overall, this suggests a variety of possible regulatory functions and targets.
size, or abundance of transcripts often differed between WT and pnp1-1 (Figure 3A and Figure S1). Transcript sizes ranged from ~100 nt in WT and ~200 nt in pnp1-1 to >3 kb in both. In all but five cases, the transcript size was underestimated by the RNA-Seq data for all transcripts observable by RNA blot in both WT and pnp1-1. In one case, RNA-Seq overestimated the transcript size (asycf2.1-5, detected in WT only; Figure S1); in two cases, there was a transcript of approximately the predicted size in WT (asycf9, astrnR.1); and in two cases, the transcript was smaller than the predicted transcript in WT (as-trnfM, nc-trnL-ndhBint).

The representative examples in Figure 3A show that transcript abundance, as estimated by hybridization signal intensity, ranged from low (e.g. as-psbB) to high (e.g. as-psbK and as-ndhD-5’-2). Additionally, the transcript patterns varied from simple (e.g. as-psbK and as-ndhD-2) to more complex (e.g. as-ycf1.2-4), suggestive of varying biogenesis pathways. Additional examples are shown in Figure S1. For the 21 ncRNAs not detected by gel blot (Table 1), we presume that they are of even lower abundance than species such as as-psbB.

The differences in ncRNA patterns between the WT and pnp1-1 were in some instances quite dramatic. For example, some ncRNAs were only detectable in the pnp1-1 sample (e.g. as-rps16i-1 and as-psbB), although they were originally verified by RT-PCR from WT. This suggests that their abundance is normally strongly limited by cpPNPase activity and that, in effect, pnp1-1 is an overexpressor of these species, a finding consistent with RNA-Seq data (Table S2). In other cases, such as nc-rbcL-accD and as-ycf1.2-4, there were additional, shorter transcripts in pnp1-1. These could be degradation intermediates that persist in the absence of cpPNPase activity. Many ncRNAs were slightly longer in pnp1-1 compared with the WT (e.g. as-trnfM and as-psbK). Because of the known functions of cpPNPase, the longer transcripts presumably include 3’ extensions. Taken together, these data point to cpPNPase as a key player for ncRNA processing and accumulation.

**Differential accumulation of sense-strand RNAs corresponding with altered ncRNA patterns in WT vs. pnp1-1**

Because ncRNA transcript abundance and form differed between WT and pnp1-1, we decided to examine the corresponding sense strand transcripts for possible correlations between sense-antisense pairs for the 10 examples shown in Figure 3. In 7 of these 10 cases, the genotypes varied in ncRNA abundance and/or transcript number. In other cases, pnp1-1 ncRNAs appeared to have 3’ extensions but similar accumulation. To facilitate the comparison, single-stranded probes for the sense and antisense strands were created from identical regions of the genome, except for tRNAs where the complete tRNA sequence was used regardless of the antisense probe location.
The three tRNA probes revealed an estimated 2- to 3-fold increased tRNA abundance in pnp1-1 compared with the WT. This result is consistent with an earlier report implicating cpPNPase in the regulation of tRNAs (Walter et al. 2002). Interestingly, the abundance of the corresponding antisense tRNAs increased in pnp1-1 at least slightly (trnQ) or markedly (trnT.1 and trnS.3). In addition to tRNA regulation, cpPNPase has been implicated in the degradation of group II introns following splicing (Germain et al. 2011). We observed here that the rps16 intron overaccumulates in pnp1-1 as multiple larger transcripts and a single smaller transcript, in addition to the WT transcripts. The increase in rps16 intron abundance correlated with an increase in abundance of as-rps16i-1, which is barely detectable in the WT. An analogous positive correlation was observed for psbB.

An opposite phenomenon was observed for nc-rbcl-accDint and the corresponding sense strand intergenic region. In this case, a smaller and less abundant ncRNA correlated with accumulation of the complementary region, which does not accumulate in the WT. Another inverse correlation was observed for the ndhF 5′ end. Although these correlations are intriguing, and a causal relationship would be consistent with known functions of ncRNAs in prokaryotes, the behaviors of complementary transcripts could also be independent effects of cpPNPase depletion. Additional experimentation will be required to make this determination.

**Chloroplast ncRNAs exist as primary transcripts and processed species, and are transcribed by two RNA polymerase types**

Chloroplast genes can be transcribed from their own promoter and/or cotranscribed with an upstream gene by either a nuclear-encoded phage-like RNA polymerase (NEP) or the bacterial-like PEP (Liere and Borner 2008). PEP-regulated genes often encode proteins involved in photosynthesis, whereas many NEP-regulated genes maintain gene expression functions. Additionally, PEP- and NEP-regulated genes are expressed differentially throughout plastid development (Liere and Maliga 2001). We elected to examine several ncRNAs from this viewpoint to begin to define ncRNA biogenesis pathways in the chloroplast.

**Transcription by PEP vs. NEP was determined by germinating plants on media with and without spectinomycin.**

Spectinomycin inhibits plastid translation and, therefore, PEP synthesis, creating chlorophyll-deficient plants with only NEP activity (Swiatecka-Hagenbruch et al. 2007). Confirmation of PEP inhibition can be seen from the etiobium bromide−stained gel in which the plastid rRNAs are greatly reduced or absent (Figure 4A, left panel). Analysis of these samples by RNA gel blot revealed as-pskB and as-ndhD-2 to be PEP dependent, whereas as-accD is transcribed by NEP. For the two PEP-dependent ncRNAs, upstream genes on the same strand could give rise to the ncRNAs by read-through transcription. We found, however, that trnS.1 is PEP independent, unlike the downstream ncRNA. On the other hand, ycS5 and as-ndhD-2 were both PEP dependent, consistent with a cotranscription model.

To ascertain whether these ncRNAs have dedicated promoters, their 5′ ends were mapped using 5′ RACE with or without prior treatment by tobacco acid phosphatase (TAP; Figure S2). Chloroplast primary transcripts are triphosphorylated and amenable to RNA ligase−mediated oligonucleotide addition only after TAP treatment removes the 5′ diphosphate (Bensing et al. 1996). Conversely, processed transcripts do not require TAP treatment for oligonucleotide adaptation. RACE analysis revealed two 5′ ends for as-pskB, as diagrammed in Figure 4B, that are not TAP dependent. The longer transcript has its 5′ end precisely at the 3′ end of the upstream trnS.1 coding region (Figure S2, band a), which suggests that the asRNA 5′ end is derived from 3′ processing of a tRNA precursor. The second, shorter as-pskB transcript (Figure S2, band b), with its end between psbK and psbl, is likely a processing product of the larger species.

Analysis of as-accD revealed three 5′ ends, the longest of which was TAP dependent (Figure S2, band c), indicating that it is a primary transcript. The region upstream of as-accD is AT-rich and includes a putative type-1a YRTA NEP consensus motif (TATA) at -8, consistent with expression on spectinomycin media (Hajdukiewicz et al. 1997; Swiatecka-Hagenbruch et al. 2007). Given that there is no immediate upstream gene transcribed in the same direction, the presence of an ncRNA-specific promoter was not unexpected. Lastly, three as-ndhD-2 5′ ends were mapped, none of which were TAP dependent (Figure S2, bands d–h), consistent with the cotranscription model proposed above. The longest transcript extended just past the 3′ end of the complementary ndhD gene, whereas the other two 5′ ends were internal to the ndhD coding region.

**Depletion of chloroplast exo- and endoribonucleases differentially affects ncRNA maturation and accumulation**

In bacteria, the ribonucleases PNPase, RNase E, and RNase III have been shown to regulate ncRNA stability (Viejas et al. 2007; Andrade and Araiaño 2008; Stead et al. 2010). Therefore, we examined the roles of chloroplast ribonucleases in ncRNA accumulation and biogenesis (Figure 5). These were chloroplast PNPase, RNase R, a 3′ → 5′ hydrolytic exoribonuclease; RNase E, an endoribonuclease; and RNase J, whose known activities in bacteria include 5′ → 3′ exoribonuclease and endoribonuclease (Deikus et al. 2008). Published studies for chloroplasts have implicated RNase R in rRNA maturation (Bollenbach et al. 2005) and regulation of the asRNA AS5 (Sharwood et al. 2011), and RNase E in polycistronic RNA cleavage (Walter et al. 2010). Although null mutants for RNase R and RNase E are viable, RNase J deficiency is embryo-lethal, so we used tissue partly depleted for the enzyme by VIGS (Sharwood et al. 2011).

The left column of Figure 5 shows analysis of as-pskB and psbK. In the absence of RNase R, neither as-pskB nor psbK was detected, and in the absence of RNase E, both transcripts underaccumulated. Thus, there is a positive correlation between psbK and as-pskB accumulation. One possibility is that RNase R and RNase E are part of a maturation pathway for psbK and/or as-pskB, and in their absence, misprocessed transcripts may be readily degraded.

The center column of Figure 5 shows nc-rbcl-accDint and the corresponding sense intergenic region. We found that nc-rbcl-accDint underaccumulates in pnp1-1, as shown in Figure 3, and in rnel-1. In both cases, underaccumulation of the ncRNA correlated with accumulation of the sense strand. As rbcl and accD are transcribed from individual PEP and NEP promoters, respectively, the intergenic region detected is likely an extension of an rbcl precursor rather than an unprocessed dicistron (Gruissem and Zurawski 1985; Hirata et al. 2004). PNPase and RNase E appear to have a role in rbcl maturation that leads to degradation of the intergenic region, as this transcript is not detectable in the WT. We also noted a larger form of nc-rbcl-accDint in rnr1-3 and RNJ(--), implying that these ribonucleases are partially involved in the biogenesis of this ncRNA.

Finally, we analyzed the 5′ end of the ndhF gene (Figure 5, right column). The asRNA was present as a doublet in WT, which was slightly longer in pnp1-1, likely due to a 3′ extension. In rnel-1, a much
Table 1 Noncoding RNAs predicted by RNA-Seq data

| ncrRNA designation | Antisense Gene | Start$^a$ | Stop$^b$ | Strand$^c$ | RT-PCR | RNA Blot Estimated Band Sizes (kb) |
|--------------------|----------------|---------|---------|---------|--------|-----------------------------------|
| 1                  | tmH            | 18      | 125     | +       | +      | 1.6, 1.8, 3.7                     |
| 2                  | matK           | 2946    | 3086    | +       | +      | 0.8, 1.1, 1.3                     |
| 3                  | rps16          | 4814    | 5371    | +       | +      | 0.8, 1.2, 1.3                     |
| 4                  | rps16i-1       | 5426    | 5581    | +       | +      | 0.8, 1.1, 1.3                     |
| 5                  | rps16i-2       | 5632    | 5875    | NT      | NT     | 0.8, 0.9, 1.1, 1.3                |
| 6                  | trnQ           | 6521    | 6646    | +       | +      | 1.3                               |
| 7                  | psbK           | 6882    | 7462    | +       | +      | 1.2, 3.3, 3.4                     |
| 8                  | psbI           | 7611    | 7691    | -       | -      | 1.3, 3.4                          |
| 9                  | trnS.1         | 7817    | 7984    | +       | +      | 1.4                               |
| 10                 | trnG-5’        | 8644    | 9004    | +       | +      | 1.1, 1.8                          |
| 11                 | trnG-3’        | 9244    | 9484    | +       | +      | 0.5, 0.8, 1.0, 1.3                |
| 12                 | trnR.1         | 9555    | 9858    | +       | +      | 0.2, 0.3                          |
| 13                 | atpA           | 10342   | 10536   | -       | NT     | 0.6                               |
| 14                 | atpI           | 14088   | 14182   | +       | +      | 0.2, 0.5, 1.0, 1.2                |
| 15                 | rpoC2-1        | 16984   | 17163   | +       | NT     | 0.2                               |
| 16                 | rpoC2-2        | 17363   | 17410   | +       | NT     | 0.2                               |
| 17                 | trnC           | 27446   | 27666   | -       | +      | 0.6, 1.1, 1.7                     |
| 18                 | trnC-ycfA2     | 27807   | 27956   | +       | +      | 0.5, 0.9                          |
| 19                 | trnD           | 29627   | 29825   | +       | +      | 0.3, 0.9                          |
| 20                 | trnD-trnYnt    | 29939   | 30070   | +       | +      | 0.3, 0.9                          |
| 21                 | trnT.1         | 31408   | 31518   | -       | +      | 0.3, 0.9                          |
| 22                 | psbD           | 33106   | 33456   | -       | +      | 0.3, 0.9                          |
| 23                 | trnS.2         | 35312   | 35408   | +       | 3.3, 3.4 | 3.1, 3.4, 3.7              |
| 24                 | ycfP           | 35820   | 36116   | +       | 3.0, 3.4 | 3.0, 3.4, 3.7              |
| 25                 | trnM           | 36642   | 36862   | +       | 1.3, 3.4 | 1.3, 3.4, 3.7              |
| 26                 | rps4-ycfB1     | 37250   | 37327   | +       | +      | 0.4, 0.7                          |
| 27                 | ycf3-3’        | 42542   | 42629   | +       | +      | 0.4, 0.7                          |
| 28                 | trn5.3         | 44827   | 45050   | +       | +      | 0.4, 0.7                          |
| 29                 | rps4-3’        | 45294   | 45358   | +       | +      | 0.4, 0.7                          |
| 30                 | rps4-5’        | 45747   | 45929   | +       | +      | 0.4, 0.7                          |
| 31                 | trnT.2         | 46026   | 46338   | +       | +      | 0.4, 0.7                          |
| 32                 | trnL.3’        | 47417   | 47670   | -       | NT     | 0.4, 0.7                          |
| 33                 | ndhK           | 49260   | 49386   | +       | +      | 0.4, 0.7                          |
| 34                 | ndhC-5’        | 50269   | 50469   | +       | +      | 0.4, 0.7                          |
| 35                 | ndhC-trnV.1nt  | 50762   | 50916   | +       | +      | 0.4, 0.7                          |
| 36                 | trnVi          | 51688   | 51884   | +       | NT     | 0.4, 0.7                          |
| 37                 | trnM           | 51950   | 52128   | -       | NT     | 0.4, 0.7                          |
| 38                 | rbcL-accD1     | 56584   | 56822   | -       | +      | 0.4, 0.7                          |
| 39                 | accD           | 57949   | 58056   | -       | +      | 0.4, 0.7                          |
| 40                 | accD-psaB1     | 58776   | 58869   | -       | +      | 0.4, 0.7                          |
| 41                 | accD-psaB1-2   | 59030   | 59096   | -       | +      | 0.4, 0.7                          |
| 42                 | ycf10-5’       | 60756   | 60963   | -       | +      | 0.4, 0.7                          |
| 43                 | ycf10          | 60983   | 61159   | +       | +      | 0.4, 0.7                          |
| 44                 | ycf10-petA1    | 61409   | 61529   | +       | +      | 0.4, 0.7                          |
| 45                 | petA           | 62045   | 62271   | +       | +      | 0.4, 0.7                          |
| 46                 | psbA           | 63479   | 63648   | +       | +      | 0.4, 0.7                          |
| 47                 | psbB-psbF      | 63805   | 64018   | +       | +      | 0.4, 0.7                          |
| 48                 | orf131         | 65708   | 65805   | +       | +      | 0.4, 0.7                          |
| 49                 | trnW           | 66248   | 66330   | +       | +      | 0.4, 0.7                          |
| 50                 | trnP           | 66330   | 66701   | +       | +      | 0.4, 0.7                          |
| 51                 | psaJ           | 66920   | 67113   | +       | +      | 0.4, 0.7                          |
| 52                 | psaJ-rpl33nt   | 67154   | 67304   | +       | +      | 0.4, 0.7                          |
| 53                 | rpl33          | 67334   | 67594   | +       | +      | 0.4, 0.7                          |
| 54                 | rps18          | 67784   | 68222   | +       | +      | 0.4, 0.7                          |
| 55                 | rpl20          | 68512   | 68680   | +       | +      | 0.4, 0.7                          |
| 56                 | psbB           | 73677   | 73837   | +       | +      | 0.4, 0.7                          |
| 57                 | psbB-psbT1nt   | 73927   | 74077   | +       | +      | 0.4, 0.7                          |
| 58                 | psbT           | 74136   | 74184   | +       | +      | 0.4, 0.7                          |
| 59                 | psbN           | 74249   | 74380   | +       | +      | 0.4, 0.7                          |
| 60                 | psbH           | 74601   | 74726   | +       | +      | 0.4, 0.7                          |

(continued)
| nRNA designation | Antisense Genea | Startb | Stopb | Strandc | RT-PCR | WT | pnp1-1 |
|------------------|-----------------|--------|-------|---------|--------|----|--------|
| 61               | petB1           | 74967  | 75151 | –       | +      | –  | –      |
| 62               | petD1-3         | 77431  | 77604 | –       | +      | –  | –      |
| 63               | rpoA-3          | 77396  | 78098 | +       | +      | –  | 1.2, 1.3, 3.3 |
| 64               | rpoA            | 78320  | 78768 | +       | NT     | NT | NT     |
| 65               | rpoA-5          | 78827  | 78957 | +       | +      | –  | –      |
| 66               | rps11           | 79166  | 79359 | +       | NT     | NT | NT     |
| 67               | rpl36           | 79502  | 79858 | +       | NT     | NT | NT     |
| 68               | rpl14           | 80723  | 80895 | +       | NT     | NT | NT     |
| 69               | rpl2i           | 85271  | 85419 | +       | NT     | NT | NT     |
| 70               | trnl.1          | 86203  | 86389 | +       | +      | –  | 0.3, 0.35, 0.4, 0.5 |
| 71               | ycf2-1-1        | 86644  | 86764 | –       | –      | +  | 0.7, 1.2, 1.3, 3.5 |
| 72               | ycf2-1-2        | 87948  | 88212 | –       | +      | –  | —      |
| 73               | ycf2-1-3        | 88451  | 88564 | –       | –      | –  | —      |
| 74               | ycf2-1-4        | 88960  | 89328 | –       | –      | –  | —      |
| 75               | ycf2-1-5        | 89445  | 90769 | –       | +      | –  | 1.0    |
| 76               | ycf2-1-6        | 91556  | 91564 | –       | +      | –  | 1.2, 1.7 |
| 77               | ycf2-1-7        | 91703  | 92005 | –       | +      | –  | —      |
| 78               | ycf2-1-8        | 92309  | 92432 | –       | +      | –  | 1.8, 3.7 |
| 79               | ycf2-1-9        | 92979  | 93113 | –       | +      | –  | 1.2, 1.3, 3.5 |
| 80               | ycf2-1-10       | 93178  | 93281 | –       | –      | –  | —      |
| 81               | tml1-ndhBint     | 94377  | 94967 | +       | +      | –  | 0.4, 1.7, 3.7 |
| 82               | ndhB-3          | 95111  | 95707 | +       | +      | –  | 0.6, 1.6, 1.7, 3.7 |
| 83               | ndhB1           | 95910  | 96164 | +       | NT     | NT | NT     |
| 84               | rps12-5         | 98718  | 98822 | +       | NT     | NT | NT     |
| 85               | rm4-5-5nt       | 107905 | 107963| +       | +      | –  | —      |
| 86               | ycf1-1-1        | 109354 | 109553| –       | NT     | NT | NT     |
| 87               | ycf1-1-2        | 109923 | 110387| +       | +      | –  | 0.8, 0.9, 1.3, 3.3 |
| 88               | ndhF-3          | 110389 | 110488| –       | –      | NT | NT     |
| 89               | ndhF-5-1        | 112345 | 112435| +       | +      | –  | 1.7, 1.8, 3.8 |
| 90               | ndhF-5-2        | 112475 | 112665| +       | +      | –  | 1.65, 1.7, 1.75, 1.8 |
| 91               | ycf2-1          | 114738 | 114982| –       | NT     | NT | NT     |
| 92               | ycf2-2          | 115138 | 115220| –       | –      | +  | —      |
| 93               | ndhD-1          | 115897 | 116043| +       | +      | –  | 1.6, 3.3, 1.8 |
| 94               | ndhD-2          | 116137 | 116437| +       | +      | –  | 1.3, 1.35 |
| 95               | ndhD-3          | 116467 | 116567| +       | +      | –  | —      |
| 96               | ndhD-4          | 116627 | 116785| +       | +      | –  | —      |
| 97               | ndhA-3          | 119774 | 119911| +       | NT     | NT | NT     |
| 98               | ndhA1           | 120852 | 120948| +       | +      | –  | 0.3, 0.5 |
| 99               | ndhA1-2         | 121194 | 121344| +       | NT     | NT | NT     |
| 100              | ndhA5-1         | 121398 | 121822| +       | NT     | NT | NT     |
| 101              | ycf1-2-1        | 124129 | 124265| +       | NT     | NT | NT     |
| 102              | ycf1-2-2        | 124605 | 124768| +       | NT     | NT | NT     |
| 103              | ycf1-2-3        | 125263 | 125448| +       | NT     | NT | NT     |
| 104              | ycf1-2-4        | 127146 | 127277| +       | +      | +  | 2.0, 0.8, 2.0, 3.7 |
| 105              | ycf1-2-5        | 127626 | 127788| +       | +      | –  | —      |
| 106              | ycf1-2-6        | 127788 | 128098| –       | NT     | NT | NT     |
| 107              | ycf1-2-7        | 128220 | 129004| +       | NT     | NT | NT     |

**Table 1 Continued**

NT, not tested.

a Gene or genes encoded on complementary strand. 5’ or 3’ indicates which part of the coding region is complementary to the nRNA. Some nRNAs are between two coding regions (int) or opposite an intron (i).

b nRNA termini as predicted by RNA-Seq from WT, except for the NT samples, which were defined from pnp1-1 RNA-Seq data based on the criteria given in the text.

c Strand is + or – for the nRNA according to the GenBank accession for Arabidopsis cpDNA.

longer form was detected (asterisk) and the lower doublet band was absent. This lower band was also absent from rnr1-3. In RNJ(−), however, as-ndhF-5’-2 accumulated in the same forms as in the WT, although the abundance was increased. Together these results indicate that this asRNA undergoes posttranscriptional processing to the mature forms found in the WT. aberrant as-ndhF-5’-2 processing coincided with altered ndhF sense transcript processing and abundance. In the WT, ndhF accumulated as a single transcript, whereas in pnp1-1, this transcript was apparent, along with three smaller transcripts. The sense transcript was reduced in both rnr1-3 and rnr1-1, whereas it overaccumulated in RNJ(−). It remains to be determined whether the correlations between the sense and antisense transcripts are due to direct interactions. However, the data presented reveal roles for many enzymes in determining the accumulation and form of
chloroplast ncRNAs, which likely reflects a variety of biogenesis pathways.

DISCUSSION

Knowledge of the extent and function of ncRNAs in prokaryotes and eukaryotes has grown considerably over the last 20 years, fueled by advances in sequencing technology and bioinformatics. However, information on the occurrence of ncRNAs in organelles, specifically the chloroplast, has been limited in scope and depth (Marker et al. 2002; Lung et al. 2006; Georg et al. 2010; Hotto et al. 2010; Mohorianu et al. 2011; Wang et al. 2011; Zghidi-Abouzid et al. 2011). Here, we employed high-throughput, strand-specific RNA sequencing (Lister et al. 2008) to identify the bulk of stably expressed Arabidopsis chloroplast ncRNAs 100 nt in length. These ncRNAs are transcribed by at least two RNA polymerase types, and their accumulation is affected by a variety of chloroplast ribonucleases. It is likely that some proportion of them plays a role in chloroplast gene expression through pathways that are largely undefined.

Identification of chloroplast ncRNAs using RNA-Seq

Two recent studies using RNA-Seq identified small chloroplast RNAs in cabbage (Wang et al. 2011) and tomato fruit (Mohorianu et al. 2011), excluding transcripts >100 nt. Other studies have used cDNA library sequencing and computational approaches to identify chloroplast ncRNAs (Lung et al. 2006; Georg et al. 2010), and small numbers of polyadenylated ncRNAs are present in EST libraries. However, there are several limitations to these methods, including low sensitivity, deliberate length restrictions, and reliance on specific sequence characteristics.

The RNA-Seq method used here has a large dynamic range, facilitating discovery of low abundance transcripts (Wang et al. 2009), and it identifies expressed transcripts rather than relying on predictions of conserved sequence or structural features (Pichon and Felden 2008). This is important for chloroplast ncRNA identification, as promoters often share little sequence homology and inefficient transcription termination can result in multiple 3’ ends (Bollenbach et al. 2008; Liere and Borner 2008). At the same time, this RNA-Seq method is restricted by strong RNA secondary structures, which can reduce the efficiency at which some oligonucleotides ligate to the oligonucleotides 5’ and 3’ ends. This drawback is mitigated by...
were present. The ncRNA as-fortuitous or functional. In two cases, conserved open reading frames
inevitable that some will have small open reading frames (ORF), either
developmental or stress conditions that were not used in this study.

some chloroplast ncRNAs may only be expressed under particular
expression (Hotto AS5, did not meet the RNA-Seq criteria due to its low endogenous
blot). For instance, a formerly characterized chloroplast asRNA,
72% identity to ORF44, which is encoded in an analogous position


cRNA, however, has been covered by other studies cited above.

Given that RNA-Seq yielded over 10 million sequences that
aligned to the chloroplast genome from each genotype, it was im-
portant to filter this dataset for our purposes. Location was primarily
constrained to regions antisense to known coding sequences to enrich
for possible cis-encoded ncRNAs, and a threshold of ≥50× coverage per transcript was used to trigger follow-up analysis. Indeed, when the
coverage of a potential ncRNA reached 50–100×, 50% were confirmed
by gel blot, and nearly 90% by RT-PCR, whereas >80% could be confirmed by gel blot when >150× coverage was obtained (Figure
6). Given these outcomes, a lower threshold (e.g. 25–50× coverage)
may identify a certain number of additional low abundance ncRNAs
(i.e. those detectable by RT-PCR but less readily detectable by gel
blot). For instance, a formerly characterized chloroplast asRNA, AS5, did not meet the RNA-Seq criteria due to its low endogenous
expression (Hotto et al. 2010; Sharwood et al. 2011). Additionally,
some chloroplast ncRNAs may only be expressed under particular
developmental or stress conditions that were not used in this study.

Although we have called this RNA population ncRNAs, it is
inevitable that some will have small open reading frames (ORF), either
fortuitous or functional. In two cases, conserved open reading frames
were present. The ncRNA as-psbK includes a 49 amino acid ORF with
72% identity to ORF44, which is encoded in an analogous position

fractionation of the total RNA pool prior to oligonucleotide adap-
tation. Also, our method largely excludes RNAs of <80 nt, a tradeoff
that reduces the amount of tRNAs and small rRNAs that would
otherwise be highly sequenced. The pool of <80 nt chloroplast
ncRNAs, however, has been covered by other studies cited above.

ncRNAs are encoded throughout the plastome

At a coverage of at least 1× per nt, RNA-Seq data from both the WT
and pnp1-1 spanned >99% of both chloroplast genome strands. From
this, 107 putative ncRNAs met filtering criteria, and their distribution and
verification status are shown in Figure 7. Of these, only as-rps16-I
was identified in a previous study (Lung et al. 2006). Full symmetric
transcription of the chloroplast genome is consistent with its well-
described inefficient transcription termination, and it implies a heavy
reliance on posttranscriptional regulatory mechanisms. Unwanted
transcripts are likely distinguished from functional ones by the

Figure 5 Analysis of three ncRNAs (top panel) and the complementary
transcripts (lower panel) in chloroplast ribonuclease-deficient samples. The
28S rRNA was ethidium bromide stained to reflect loading (bottom). Null mutants for PNPase (pnp1-1), RNase R (mr1-3), and
RNase E (rne1-1), and VIGS knockdown tissue for RNase J (RNJ(−))
were used. Blots were analyzed as described in Figure 3. The as-psbK
blot was reprobed to detect the sense transcript (the as-psbK species
is marked "as" in both panels). An as-ndhF5′-2 precursor that accumu-
lates in me1-1 is marked by an asterisk.

and stably transcribed in barley chloroplasts (Sexton et al. 1990) al-
though it has no known function. A second example is a 58 amino
acid ORF found within as-ndhD, which has 75–98% identity to pu-
tative proteins encoded by the ndhD antisense strand in numerous chloroplast genomes. Although this ORF also has no attributed func-
tion, the apparent evolutionary selection for ORFs on both strands is
striking.

Figure 6 Percentage of ncRNAs detected by either RT-PCR (black
bars) or RNA gel blot (gray bars) compared with the depth of RNA-Seq
coverage. Total number of ncRNAs in each category prior to
verification is in parentheses.
collective effects of RNA structures, RNA-binding proteins, and their sensitivities to RNases. We speculate that nonfunctional or detrimental ncRNAs are rapidly degraded, because cis-encoded asRNAs are inherently inhibitory to gene expression, a view supported by three published examples for chloroplasts (Nishimura et al. 2004; Hegeman et al. 2005; Hotto et al. 2010). This implies the ncRNAs accumulating to detectable levels may well retain functional roles.

Chloroplast ncRNAs include both primary and processed transcripts emanating from ncRNA-specific promoters (e.g., as-accD; Figure 4) and those generated by cotranscription with an upstream gene. For example, as-rndhD-2 appears to be cotranscribed with ycf5, and as-psbK with trnS.1 (Figure 4B). There is a discrepancy with as-psbK because it fails to accumulate when plants are grown on spectinomycin (Figure 4A), showing it requires PEP, whereas the upstream trnS.1 gene is transcribed by NEP (Figure 4; Gruissem et al. 1986; Wu et al. 1997). We suggest that as-psbK is transcribed from an unidentified PEP promoter upstream of trnS.1, whereas initiation at its NEP promoter leads to transcription termination without read-through into as-psbK. In support of this hypothesis, we note that the as-psbK probe identified a 3.7 kb transcript that may extend upstream of trnS.1 and is PEP dependent (Figures 3A and 4A).

Chloroplast ncRNA maturation and accumulation are affected by ribonucleases

We assessed the effects on ncRNA biogenesis of deficiencies for the two known chloroplast 3′ → 5′ exoribonucleases, cpPNPase and RNase R, the endoribonuclease RNase E, and the putative endoribonuclease and 5′ → 3′ exoribonuclease RNase J. The results were diverse, suggesting that there are several pathways responsible for ncRNA processing and accumulation.

PNPase was previously shown to regulate bacterial ncRNAs, and the eukaryotic exosome (similar in function to PNPase) assists in the maturation and degradation of nuclear ncRNAs (Houseley et al. 2006; Viegas et al. 2007). Because cpPNPase plays a broad role in RNA metabolism (Germain et al. 2011), it was not surprising that it also acts on ncRNAs. Our data suggest two functions, namely, modulation of ncRNA abundance and 3′ end maturation.

Loss of RNase R had a major effect on both sense and antisense psbK RNAs; both were nearly depleted in mutant tissue. Chloroplast RNase R has previously been implicated in RNA maturation and accumulation of the asRNA AS5, transcribed from the rDNA region (Sharwood et al. 2011). In fact, mtr1-3 accumulates only 6% of the WT level of 5S rRNA, giving precedence for reduced RNA accumulation in this mutant. In that case, however, the asRNA increases in abundance, which led to the hypothesis that the asRNA destabilizes the sense transcript. In the case of psbK, the transcripts are coordinately affected, raising the possibility that they protect one another. The proximal cause of instability of one or both in the absence of RNase R remains to be ascertained.

RNase E deficiency appeared to have a destabilizing effect on as-psbK, psbK, and nc-rbcL-accDInt, whereas the rbcL-accDInt sense strand transcript increased in abundance. Some E. coli ncRNAs were destabilized in the absence of RNase E, which may be due to direct RNase E–ncRNA interactions or may result from changes in other mRNAs or unidentified interactors (Stead et al. 2010). RNase E could have a similar role with respect to chloroplast ncRNAs. The appearance of a sense strand rbcL-accD intergenic region is consistent with the postulated role of chloroplast RNase E in polycistrionic transcript cleavage (Walter et al. 2010). Whether this transcript somehow destabilizes its antisense counterpart remains to be determined.

RNase J has been little studied in chloroplasts. RNase E and RNase J have been shown to be partially redundant in Arabidopsis SS-3trnR processing, and RNase J has been shown to replace some RNase E functions in B. subtilis (Britton et al. 2007; Sharwood et al. 2011). We found here that RNase J–deficient material accumulated an nc-rbcL-accDInt precursor and had minor quantitative alterations for other ncRNAs. Taken together, these results indicate that multiple ribonucleases are involved in chloroplast ncRNA maturation and degradation.

Possible functions of chloroplast ncRNAs

The ncRNAs identified in this study had origins throughout the chloroplast genome (Figure 7). Because accumulating asRNAs are complementary to all types of coding regions, including tRNAs, photosynthetic genes, and gene expression machinery, their regulatory roles could be quite diverse, similar to large-scale mechanisms of gene regulation attributed to bacterial ncRNAs (Repoila and Darfeuille 2009). The focus of this study was on cis-encoded asRNAs, although we cannot exclude the possibility that one or more also acts in trans. Potential functions of cis-encoded asRNAs are suggested by the region of complementarity to the corresponding sense RNA, as well as by the target mRNA location within an operon.

At least 31 chloroplast ncRNAs overlap the translation initiation region of a complementary mRNA. In bacteria, base pairing of
ncRNAs to the 5’ end of target transcripts is common. The outcome can be activation or repression of translation by altering ribosome accessibility to the Shine-Dalgarno (SD) sequence and/or start codon (Regnier and Hajnsdorf 2009). However, chloroplast mRNAs frequently lack SD elements and instead require upstream cis-elements and gene-specific trans-factors (Peled-Zehavi and Danon 2007), which may be targets of ncRNAs. The in vitro translation system available for chloroplasts offers one avenue to test these hypotheses (Yukawa et al. 2007).

At least 39 chloroplast ncRNAs are complementary to the sense strand mRNA 3’ end. Bacterial ncRNAs that bind to mRNA 3’ ends often stabilize the mRNA by blocking 3’→5’ exoribonucleases. This is the case for E. coli GadY, which stabilizes the cis-encoded GadX mRNA (Opdyke et al. 2004). Chloroplast transcript 3’ ends are often stabilized by a stem-loop–forming inverted repeat, but not all mRNAs possess predicted 3’ stem loops, and these would be candidates for ncRNA-mediated stabilization. A possible example is as-psbT, which was hypothesized to stabilize the complementary psbT mRNA under oxidative stress conditions (Zghidi-Abouzid et al. 2011).

Approximately 40 chloroplast asRNAs are complementary to the coding region of mRNAs, and some of these putative mRNA targets are within operons. Bacterial ncRNAs that bind to their target within the coding region can alter transcript stability by creating or blocking a ribonuclease binding site (Pfeiffer et al. 2009). Additionally, bacterial ncRNAs can alter transcript accumulation within an operon, one example being RhyB, which downregulates the iscSUA genes within the iscSUA operon to allow independent accumulation of iscR (Desnoyers et al. 2009). Similar scenarios could be envisioned for chloroplast ncRNAs, leading to differential accumulation of transcripts within a gene cluster.

Overall, RNA-Seq analysis proved extremely useful for identifying chloroplast ncRNAs, as approximately 45% were further validated by experimental means. We suspect that this method will also be useful for characterization of other transcriptomes, particularly organellar ones that benefit from the deep coverage of the low-complexity genomes inherent in total RNA analysis.

**ACKNOWLEDGMENTS**

This work was supported by U.S. Department of Energy awards FG02-90ER20015 (D.B.S.) and DE-FG02-08ER15517 (J.R.E.); by National Science Foundation award MCB-0929402 (J.R.E.); by the Howard Hughes Medical Institute (HHMI), the Gordon and Betty Moore Foundation (GBMF), and the Mary K. Chapman Foundation (M.K.C.); and by NIH National Research Service Award postdoctoral fellowship F32-HG004830 (R.J.S.). J.R.E. is an HHMI-GBMF investigator.

**LITERATURE CITED**

Alonso, J. M., A. N. Stepanova, T. J. Leisse, C. J. Kim, H. Chen et al., 2003 Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science 301: 653–657.

Andrade, J. M., and C. M. Arraiano, 2008 PNPase is a key player in the regulation of small RNAs that control the expression of outer membrane proteins. RNA 14: 543–551.

Bensing, B. A., B. J. Meyer, and G. M. Dunny, 1996 Sensitive detection of bacterial transcription initiation sites and differentiation from RNA processing sites in the pheromone-induced plasmid transfer system of Enterococcus faecalis. Proc. Natl. Acad. Sci. USA 93: 7794–7799.

Bollenbach, T., G. Schuster, V. Portnoy, and D. B. Stern, 2007 Processing, degradation and polyadenylation of chloroplast transcripts, pp. 175–211 in Cell and Molecular Biology of Plastids, edited by R. Bock, Topics in Current Genetics (series), Vol. 19. Springer-Verlag, Berlin/Heidelberg, Germany.

Bollenbach, T. J., H. Lange, R. Gutierrez, M. Erhardt, D. B. Stern et al., 2005 RNR1, a 3′-5′ exoribonuclease belonging to the RNR superfamily, catalyzes 3′ maturation of chloroplast ribosomal RNAs in Arabidopsis thaliana. Nucleic Acids Res. 33: 2751–2763.

Britton, R. A., T. Wen, L. Schafer, O. Pellegrini, W. C. Uicker et al., 2007 Maturation of the 5’ end of Bacillus subtilis 16S rRNA by the essential ribonuclease YkC/UrNase J1. Mol. Microbiol. 63: 127–138.

Church, G., and W. Gilbert, 1984 Genomic sequencing. Proc. Natl. Acad. Sci. USA 81: 1991–1995.

Deikus, G., C. Condon, and D. H. Bechhofer, 2008 Role of Bacillus subtilis Rnase J1 endonuclease and 5′-exonuclease activities in trp leader RNA turnover. J. Biol. Chem. 283: 17158–17167.

Desnoyers, G., A. Morissette, K. Prevost, and E. Masse, 2009 Small RNA-induced differential degradation of the polystrictronic mRNA iscsSUA. EMBO J. 28: 1551–1561.

Georg, J., A. Honsel, B. Voss, H. Rennenberg, and W. R. Hess, 2010 A long antisense RNA in plant chloroplasts. New Phytol. 186: 615–622.

Georg, J., B. Voß, I. Scholz, J. Mitschke, A. Wilde et al., 2009 Evidence for a major role of antisense RNAs in cyanobacterial gene regulation. Mol. Syst. Biol. 5: 305.

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. Plant J. 67: 381–394.

Gruissem, W., C. Elsner-Menzel, S. Latshaw, J. O. Narita, M. A. Schaffer et al., 1986 A subpopulation of spinach chloroplast RNA genes does not require upstream promoter elements for transcription. Nucleic Acids Res. 14: 7541–7556.

Gruissem, W., and G. Zurawski, 1985 Analysis of promoter regions for the spinach chloroplast rbcL, atp8 and psbA genes. EMBO J. 4: 3375–3383.

Hadjukiewicz, P. T. J., L. A. Allison, and P. Maliga, 1997 The two RNA polymerases encoded by the nuclear and plastid compartments transcribe distinct groups of genes in tobacco plastids. EMBO J. 16: 4041–4048.

Hegeman, C. E., C. P. Halter, T. G. Owens, and M. R. Hanson, 2005 Expression of complementary RNA from chloroplast transgenes affects editing efficiency of sense and endogenous chloroplast transcripts. Nucleic Acids Res. 33: 1454–1464.

Hirata, N., D. Yonekura, S. Yanagisawa, and K. Iba, 2004 Possible involvement of the 5′-flanking region and the 5′ UTR of plastid accD gene in NER-dependent transcription. Plant Cell Physiol. 45: 176–186.

Holec, S., H. Lange, K. Kuhn, M. Alioua, T. Borner et al., 2006 Relaxed transcription in Arabidopsis mitochondria is counterbalanced by RNA stability control mediated by polyadenylation and polynucleotide phosphorylase. Mol. Cell. Biol. 26: 2869–2876.

Hottò, A. M., Z. E. Huston, and D. B. Stern, 2010 Overexpression of a natural chloroplast-encoded antisense RNA in tobacco destabilizes SS rRNA and retards plant growth. BMC Plant Biol. 10: 213.

Houseley, J., J. LaCava, and D. Tollervey, 2006 RNA-quality control by the exosome. Nat. Rev. Mol. Cell Biol. 7: 529–539.

Jen, C. H., I. Michalopoulos, D. R. Westhead, and P. Meyer, 2005 Natural antisense transcripts with coding capacity in Arabidopsis may have a regulatory role that is not linked to double-stranded RNA degradation. Genome Biol. 6: R51.

Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg, 2009 Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10: R25.

Liere, K., and P. Maliga, 2001 Plastid RNA polymerases in higher plants, pp. 211–174 in Cell and Molecular Biology of Plastids, edited by R. Bock, Topics in Current Genetics (series), Vol. 19. Springer-Verlag, Berlin/Heidelberg, Germany.

Liere, K., and P. Maliga, 2001 Plastid RNA polymerases in higher plants, pp. 29–39 in Regulation of Photosynthesis, edited by E. M. Árbo and B. Andersson, Klüwer, Rotterdam.

Lister, R., R. C. O’Malley, J. Tonti-Filippini, B. D. Gregory, C. C. Berry et al., 2008 Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133: 523–536.
Lung, B., A. Zemann, M. J. Madej, M. Schuelke, S. Techritz et al., 2006 Identification of small non-coding RNAs from mitochondria and chloroplasts. Nucleic Acids Res. 34: 3842–3852.

Marchive, C., S. Yehudai-Resheff, A. Germain, Z. Fei, X. Jiang et al., 2009 Abnormal physiological and molecular mutant phenotypes link chloroplast polynucleotide phosphorylase to the phosphorous deprivation response in Arabidopsis. Plant Physiol. 151: 905–924.

Marker, C., A. Zemann, T. Terhorst, M. Kiefmann, J. P. Kastenmayer et al., 2009 Abnormal physiological and molecular mutant phenotypes link chloroplast polynucleotide phosphorylase to the phosphorous deprivation response in Arabidopsis. Plant Physiol. 151: 905–924.

Mohorianu, I., F. Schwach, R. Jing, S. Lopez-Gomollon, S. Moxon et al., 2011 Profiling of short RNAs during fleshy fruit development reveals stage-specific sRNAsome expression patterns. Plant J. 67: 232–246.

Mudd, E. A., S. Sullivan, M. F. Gisby, A. Mironov, C. S. Kwon et al., 2008 A 125 kDa RNase E/G-like protein is present in plastids and is essential for chloroplast development and autotrophic growth in Arabidopsis. J. Exp. Bot. 59: 2597–2610.

Nishimura, Y., E. A. Kikis, S. L. Zimmer, Y. Komine, and D. B. Stern, 2004 Antisense transcript and RNA processing alterations suppress instability of polyadenylated mRNA in Chlamydomonas reinhardtii. Plant Cell 16: 2849–2869.

Opdyke, J. A., J. G. Kang, and G. Storz, 2004 GadY, a small-RNA regulator of acid response genes in Escherichia coli. J. Bacteriol. 186: 6698–6705.

Peled-Zehavi, H., and A. Danon, 2007 Translation and translational regulation in plastids, pp. 249–282 in Cell and Molecular Biology of Plastids, edited by R. Bock. Topics in Current Genetics (series), Vol. 19. Springer-Verlag, Berlin/Heidelberg, Germany.

Pfeiffer, V., K. Papenfort, S. Lucchini, J. C. D. Hinton, and J. Vogel, 2008 Small RNA gene identification and target predictions in bacteria. Bioinformatics 24: 2807–2813.

Pichon, C., and B. Felden, 2008 Small RNA gene identification and mRNA target predictions in bacteria. Bioinformatics 24: 840–847.

Regnier, P., and E. Hajnsdorf, 2009 Chapter 4 poly(a)- assisted RNA decay and modulators of RNA stability. Prog. Nucleic Acid Res. Mol. Biol. 85: 137–185.

Repola, F., and F. Darfeuille, 2009 Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects. Biol. Cell 101: 117–131.

Roberts, A., C. Trapnell, J. Donaghey, J. Rinn, and L. Pachter, 2011 Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol. 12: R22.

Rozen, S., and H. J. Skaletsky, 2000 Primer3 on the WWW for general users and for biologist programmers, pp. 365–386 in Bioinformatics Methods and Protocols: Methods in Molecular Biology, edited by S. Krawetz and S. Misener. Humana Press, Totowa, New Jersey.

Sexton, T. B., J. T. Jones, and J. E. Mullet, 1990 Sequence and transcriptional analysis of the barley cDNA region upstream of psbD-psbc encoding trnK(UUU), rps16, trnQ(UUG), psbK, psbI, and trnS(GCU). Curr. Genet. 17: 445–454.

Sharwood, R. E., A. M. Hotto, T. J. Bollenbach, and D. B. Stern, 2011 Overaccumulation of the chloroplast antisense RNA AS5 is correlated with decreased abundance of SS rRNA in vivo and inefficient SS rRNA maturation in vitro. RNA 17: 230–243.

Stead, M. B., S. Marshburn, B. K. Mohanty, J. Mitra, L. P. Castillo et al., 2011 Analysis of Escherichia coli RNase E and RNase III activity in vivo using tiling microarrays. Nucleic Acids Res. 39: 3188–3203.

Steglich, C., M. E. Futschik, D. Lindell, B. Voss, S. W. Chisholm et al., 2008 The challenge of regulation in a minimal photoautotroph: non-coding RNAs in Prochlorococcus. PLoS Genet. 4: e1000173.

Storn, D. B., and W. Gruissem, 1987 Control of plastid gene expression: 3’ inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. Cell 51: 1145–1157.

Storz, G., S. Altuvia, and K. M. Wassarman, 2005 An abundance of RNA regulators. Annu. Rev. Biochem. 74: 199–217.

Swiatecka-Hagenbruch, M., K. Liere, and T. Borner, 2007 High diversity of plastidial promoters in Arabidopsis thaliana. Mol. Genet. Genomics 277: 725–734.

Trapnell, C., B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan et al., 2010 Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28: 511–515.

Viegas, S. C., V. Pfeiffer, A. Sittka, I. J. Silva, J. Vogel et al., 2007 Characterization of the role of ribonucleases in Salmonella small RNA decay. Nucleic Acids Res. 35: 7651–7664.

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: 6905–6914.

Walter, M., K. Piepenburg, M. A. Schöttler, K. Petersen, S. Kahla et al., 2010 Knockout of the plastid RNase E leads to defective RNA processing and chloroplast ribosome deficiency. Plant J. 64: 851–863.

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.

Wang, X. J., T. Gaasterland, and N. H. Chua, 2005 Genome-wide prediction and identification of cis-natural antisense transcripts in Arabidopsis thaliana. Genome Biol. 6: R30.

Wang, Z., M. Gerstein, and M. Snyder, 2009 RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10: 57–63.

Wu, C. Y., C. H. Lin, and L. J. Chen, 1997 Identification of the transcription start site for the spinach chloroplast serine tRNA gene. FEBS Lett. 418: 157–161.

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

Zghidi-Abouzid O., L. Merendino, F. Buhr, M. Malik Ghulam, and S. Lerbs-Mache, 2011 Characterization of plastid psbD sense and antisense RNAs. Nucleic Acids Res. 39: 5379–5387.

Zhang, Y., X. S. Liu, Q. R. Liu, and L. Wei, 2006 Genome-wide in silico identification and analysis of cis natural antisense transcripts (cis-NATs) in ten species. Nucleic Acids Res. 34: 3465–3475.

Communicating editor: J. K. Kim