The presence of multiple introns is essential for *ERECTA* expression in *Arabidopsis*

RUCHA KARVE,¹ WUSHENG LIU,¹ SPENCER G. WILLET,¹ KEIKO U. TORII,² and ELENA D. SHPAK^{1,3}

¹Department of Biochemistry, Cellular and Molecular Biology, University of Tennessee, Knoxville, Tennessee 37996, USA

²Department of Biology, University of Washington, Seattle, Washington 98195, USA

ABSTRACT

Gene expression in eukaryotes is often enhanced by the presence of introns. Depending on the specific gene, this enhancement can be minor or very large and occurs at both the transcriptional and post-transcriptional levels. The *Arabidopsis ERECTA* gene contains 27 exons encoding a receptor-like kinase that promotes cell proliferation and inhibits cell differentiation in above-ground plant organs. The expression of *ERECTA* very strongly depends on the presence of introns. The intronless *ERECTA* gene does not rescue the phenotype of *erecta* mutant plants and produces about 500–900 times less protein compared with the identical construct containing introns. This result is somewhat surprising as the region upstream of the *ERECTA* coding sequence effectively promotes the expression of extraneous genes. Here, we demonstrate that introns are essential for *ERECTA* mRNA accumulation and, to a lesser extent, for mRNA utilization in translation. Since mRNA produced by intronless *ERECTA* is degraded at the 3' end, we speculate that introns increase mRNA accumulation through increasing its stability at least in part. No individual intron is absolutely necessary for *ERECTA* expression, but rather multiple introns in specific locations increase *ERECTA* expression in an additive manner. The ability of introns to promote *ERECTA* expression might be linked to the process of splicing and not to a particular intron sequence.

Keywords: gene expression; intron-mediated enhancement; splicing; intron; *ERECTA*

INTRODUCTION

Most genes of higher eukaryotes contain introns, noncoding DNA sequences that are transcribed and then removed from pre-mRNA molecules through a process called splicing. Frequently introns are critical for gene function as they might be necessary for alternative splicing or could contain vital noncoding RNAs and ORFs. In addition, introns have the ability to promote gene expression. For example, maize alcohol dehydrogenase-1 expression in a transient assay is increased 50–100 times when introns are included (Callis et al. 1987). Introns are critical for rabbit β -globin expression in animal cell cultures (Buchman and Berg 1988), for expression of the mouse thymidylate synthase gene (Deng et al. 1989), and for human purine nucleoside phosphorylase (Jonsson et al. 1992). While endogenous introns can enhance the expression of many genes, in most cases the effect is rather modest, and only in a limited number of cases, such as the

rabbit β -globin gene or maize alcohol dehydrogenase-1, it is dramatic (Lu and Cullen 2003).

The most well understood mechanism of gene expression regulation by introns is through the presence of *cis*-regulatory elements in intronic sequences. For example, the second intron of the *Arabidopsis* gene *AGAMOUS* contains binding sites for several transcription factors, including LEAFY and WUSCHEL (Sieburth and Meyerowitz 1997; Hong et al. 2003). Alternatively, introns can also promote gene expression through the recruitment of splicing factors. During the past decade, it has become clear that the different steps of mRNA biogenesis such as transcription, 5' end capping, splicing, and the cleavage and polyadenylation of the 3' end are mutually dependant (Moore and Proudfoot 2009; Hocine et al. 2010). Thus, introns can promote transcription by recruiting spliceosomal proteins that directly interact with the RNA polymerase II complex and stimulate transcriptional elongation (Fong and Zhou 2001; Lin et al. 2008). There are multiple reports suggesting that splicing and polyadenylation can be linked: Several splicing factors interact directly with the 3' end processing machinery (Millevoi et al. 2002; Kyburz et al. 2006), specific intronic sequences can positively regulate polyadenylation (Lou

³Corresponding author.

E-mail eshpak@utk.edu.

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.2825811>.

et al. 1996), the 3' splice site of the last intron is able to enhance poly(A) site cleavage (Rigo and Martinson 2008), and spliced mRNAs are often more efficiently polyadenylated than are identical intronless mRNAs (Huang and Gorman 1990; Niwa et al. 1990; Lu and Cullen 2003). As the 3' end influences mRNA stability, transport, and translation, the splicing-induced changes to its structure should alter the level of gene expression.

Finally, during splicing a complex of proteins is deposited on mRNA 20–24 nucleotides (nt) upstream of an exon–exon junction (Le Hir et al. 2000). An exon junction complex (EJC) consists of four core proteins (eIF4A-III, MLN51, Y14, and Magoh) and a variable number of proteins associated with it in a transient manner (Tange et al. 2005). The EJC is well known for its role in nonsense-mediated mRNA decay (Tange et al. 2004). For some intron-containing genes, the EJC is required for efficient mRNA export to the cytoplasm as it is able to recruit mRNA export factors (Zhou et al. 2000; Le Hir et al. 2001). In other cases, the EJC is believed to increase the translation of spliced transcripts, possibly through interaction with the small ribosomal subunit. The human PYM protein interacts with both the EJC and the small ribosomal subunit, and its knockout results in reduced translational efficiency of spliced mRNAs but does not change the translation of intronless mRNAs (Diem et al. 2007).

Why is the presence of introns extremely significant for expression of some genes but not others? In mammals only a small proportion of genes are naturally intronless (~5% in *Homo sapiens* and ~8% in *Mus musculus*). It is speculated that for most mammalian genes, splicing is essential for efficient 3' end formation of mRNA and its export from the nucleus. Naturally intronless genes and some intron-containing genes have *cis*-elements that promote these processes, and thus, their expression becomes independent of splicing. Such elements have been identified in human c-Jun gene (Guang and Mertz 2005), mouse H2a gene (Huang and Carmichael 1997), and several viral genes (Malim et al. 1989; Ernst et al. 1997). In plants the proportion of intronless genes is higher than in mammals: 21.7% in *Arabidopsis* and 19.9% in rice (Jain et al. 2008). The removal of introns from many plant genes is not detrimental to their function, based on effective complementation of numerous *Arabidopsis* mutants by cDNA sequences. Thus, for most plant genes, the 3' end formation of mRNA and its export from the nucleus is probably not dependent upon splicing.

At the same time, there are still quite a few examples when introns are essential for plant gene expression (Callis et al. 1987; Dean et al. 1989; Rose and Last 1997; Maas et al. 1991; Rose 2008). The mechanism by which they function is not clear except in cases when *cis*-regulatory elements are found in intronic sequences. Historically, stimulation of plant gene expression by introns has been called intron-mediated enhancement (IME) (Mascarenhas et al. 1990). In described

cases of IME, introns often increase mRNA accumulation (Callis et al. 1987; Dean et al. 1989; Rose and Last 1997; Wang et al. 2002), but sometimes they also have an effect on translation (Bourdon et al. 2001; Curi et al. 2005; Lu et al. 2008). In several cases, nuclear run-on transcription assays suggested that the increase in mRNA accumulation was not due to changes in transcription initiation or elongation but instead occurred at a later step (Dean et al. 1989; Rose and Last 1997; Rose and Beliakoff 2000). For efficient IME, introns have to be within the transcribed sequences and in their natural orientation (Callis et al. 1987; Mascarenhas et al. 1990; Clancy et al. 1994; Curi et al. 2005). The position of introns in a gene is significant, and often IME is stronger when introns are located close to the 5' UTR (Rose 2004). While IME requires at least a partial assembly of the spliceosome, the completion of splicing is not always essential (Rose and Beliakoff 2000; Rose 2002). Recently it was suggested that promoter proximal introns are enriched in sequences that elevate gene expression, and an algorithm for finding such sequences called IMEter was proposed and tested in *Arabidopsis* (Rose et al. 2008). However, in a later study in rice, the IMEter predictions were not supported (Morello et al. 2011). While one can expect that recruitment of splicing factors and the EJC proteins have the same effect on gene expression in plants as it does in mammals, and therefore should be involved in IME, this effect has not been demonstrated experimentally.

To gain more knowledge of gene expression regulation in plants, we have investigated the *Arabidopsis* gene *ERECTA*. The *ERECTA* gene encodes a receptor-like kinase that is involved in regulation of plant size and epidermis development (Torii et al. 1996; Shpak et al. 2004, 2005). Orthologs of *ERECTA* can be found in many different plant species, including Bryophytes, and in most plants, the family is represented by several paralogs. In *Arabidopsis*, *ERECTA* belongs to a family of three genes (the others are *ERL1* and *ERL2*) that have overlapping functions and are critical for plant growth. Mutation of all three genes leads to severe dwarfism and dramatically increases the proportion of stomata in the epidermis (Shpak et al. 2004, 2005). The phenotype of single *erecta* mutants is not so striking but is still very distinct. *Erecta* plants have compact inflorescence with flowers clustered on the top; short and thick stems; small, round flowers; and short siliques (Torii et al. 1996).

Here we demonstrate that the intronless *ERECTA* gene cannot rescue this phenotype, while intron-containing *ERECTA* can. The presence of introns is required for mRNA accumulation, and in this way, IME of *ERECTA* is similar to IME of most other plant genes. Based on our data, we speculate that IME of *ERECTA* expression does not require specific intron sequences but is related to some aspect of splicing and operates at least in part through promotion of 3' end mRNA formation.

RESULTS

The intronless *ERECTA* gene produces a dramatically reduced amount of mRNA

The 9.4-kb genomic fragment of *ERECTA*, including the intron-containing coding sequence (here will be referred to as *gERECTA*), and the 1.7-kb 5' and 1.9-kb 3' regions rescued the *erecta-105* knockout mutant in all 14 independent transgenic T1 *Arabidopsis* plants we examined (Fig. 1A). However, when the intronless sequence of *ERECTA* (here will be referred to as *cERECTA*) was placed between the same promoter and terminator regions, the construct did not complement *erecta-105* even partially in all 43 transgenic lines that we obtained (Fig. 1A). We confirmed the presence of the *ER::cERECTA* construct by PCR analysis in 11 of the transgenic lines.

FIGURE 1. The intronless *ERECTA* gene does not rescue the *erecta* knockout mutant and produces dramatically reduced amount of mRNA. (A) Complementation of *erecta-105* with *ER::gERECTA* and *ER::cERECTA*. Bar, 3 mm. (B–E) The amount of *ERECTA* mRNA in *ER::cERECTA* plants is lower and in *ER::gERECTA* plants is equal or higher compared with the wild type as determined by qRT-PCR with *ACTIN-2* as an internal control. The total RNA was isolated from T1 inflorescences (B,C) and 15-d-old gentamicin selected T2 seedlings (D,E) of multiple independent transgenic lines. Graphs C and E demonstrate that *ERECTA* mRNA is expressed at a higher level in *ER::cERECTA* plants compared with the *erecta-105* mutant. The mean of three technical replicates is plotted; error bars, 1 SD.

In a previous study, *ERECTA* RNA gel blot analysis detected only one RNA species (Torii et al. 1996). All 97 expression sequenced tags (ESTs) available for the *ERECTA* gene correspond to the cDNA structure used in our experiments. An extensive reverse transcription PCR (RT-PCR) analysis of the *ERECTA* gene segments detected that introns 6, 20, and 21 are less efficiently spliced than are others (data not shown). However, they contain multiple stop codons and will most likely lead to nonsense-mediated decay of mRNA. Thus, we were not able to detect any alternatively spliced *ERECTA* mRNA.

To determine whether introns are necessary for transcription and/or mRNA stability or impact the later stages of expression, we compared *ERECTA* levels in transgenic lines to those in the wild type and in *erecta-105* using quantitative RT-PCR (qRT-PCR). T1 *cERECTA* inflorescences of the five analyzed transgenic lines had at least 10 times less *ERECTA* than did the wild type (Fig. 1B). Subsequent T2 generation transgenic seedlings had at least 5.5 times less (Fig. 1D). We detected higher levels of *ERECTA* in *cERECTA* lines (two to three times) compared with the *erecta-105* mutant, suggesting that some *ERECTA* is expressed from this construct (Fig. 1C,E). T1 *gERECTA* inflorescences had up to four times more *ERECTA* compared with the wild type, and T2 gentamicin selected seedlings had from 2.8–9.5 times more (Fig. 1B,D). All *gERECTA* lines in which we measured *ERECTA* mRNA levels contained only one T-DNA insert based on their 3:1 segregation ratio on selection plates. During cloning we made sure to place both *gERECTA* and *cERECTA* in the vector oriented such that the terminator, but not the promoter of, *ERECTA* was adjacent to the *d35S* promoter regulating the *aacC1* gentamicin resistance gene. Therefore it is not likely that the *d35S* promoter is able to influence *ERECTA* expression in our constructs. Higher *ERECTA* expression in the *gERECTA* lines could be due to negative regulation of expression in its endogenous location by elements outside of the 9.4-kb region or by the chromatin structure. While *ERECTA* mRNA is expressed at five to nine times higher levels in *gERECTA* lines 3 and 5, the phenotype of plants was indistinguishable from the wild type (Supplemental Fig. S1). We were not able to determine whether higher expression at the mRNA level is translated into a larger amount of protein produced. *ERECTA* mRNA consistently accumulated at considerably lower levels in the *cERECTA* lines and at considerably higher levels in the *gERECTA* lines, strongly suggesting that the *ERECTA* gene experiences IME of gene expression occurring at the stage of mRNA accumulation.

Importance of introns in the *ERECTA* gene for protein production

As the *cERECTA* lines did produce some mRNA, we wanted to investigate whether introns are necessary for any later stages of mRNA metabolism. It was not possible to directly

compare the accumulation of ERECTA protein in our transgenic plants as the endogenous ERECTA protein is produced at a very low level (Shpak et al. 2003), and reliable measurements of its quantity by Western blotting analysis are extremely challenging. To overcome this problem, we transformed *erecta-105* with the *ER::gERECTA-RLUC* and *ER::cERECTA-RLUC* constructs where the ERECTA sequences were fused to the Renilla luciferase (RLUC) coding region and expressed under control of the endogenous ERECTA promoter and the 35S terminator. *ER::gERECTA-RLUC* complemented the mutation, while *ER::cERECTA-RLUC* did not. Complementation by *ER::gERECTA-RLUC* was partial in more than half of the T1 plants, presumably due to inhibition of ERECTA function by the addition of luciferase to its C terminus.

We characterized two independent transgenic lines for each construct with qRT-PCR analysis and Renilla luciferase assay using homozygous transgenic seedlings carrying a single T-DNA insert. In *cERECTA* seedlings we detected around 8.9 and 14.1 times less and in *gERECTA* 3.1 and 1.2 times more ERECTA compared to the wild type (Fig. 2A). Similar to the *cERECTA* lines, the *cERECTA-RLUC* lines produced ERECTA mRNA above the *erecta-105* level (Fig. 2B). A RT-PCR analysis with primers annealing upstream of the first ATG and downstream from the stop codon demonstrated that the length of the ERECTA-RLUC coding region was of the same size in *cERECTA-RLUC* and *gERECTA-RLUC* transgenic plants and only one species of mRNA was present (Supplemental Fig. S2). The sequencing results confirmed that the sequence of *cERECTA-RLUC* and *gERECTA-RLUC* mRNA is identical and the presence of introns does not result in alternative splicing.

By measuring luciferase activity, we detected a high level of ERECTA-RLUC protein in the *gERECTA-RLUC* lines (Fig. 2C). Two *cERECTA-RLUC* lines exhibited statistically significantly higher luciferase activity compared with that of the wild type, suggesting that they also produced some ERECTA-RLUC protein (Fig. 2C). However, there was 486–941 times less protein in the *cERECTA-RLUC* lines compared with the *gERECTA-RLUC* plants (Fig. 2B), which is considerably larger than the difference in mRNA levels (11–43 times), suggesting that the small amount of *cERECTA-RLUC* mRNA produced in transgenic plants is not translated efficiently.

The mRNA produced by the intronless ERECTA gene is less stable at the 3' end

The difference in mRNA accumulation could be due to a different rate of transcription initiation/elongation or unequal mRNA stability. As mRNA degradation often begins with deadenylation at the 3' end followed by degradation in a 3' → 5' direction by the cytoplasmic exosome (Houseley and Tollervey 2009), we checked whether there is any difference in the stability of the 3' end of *cERECTA* and *gERECTA* mRNA.

FIGURE 2. The *cERECTA-RLUC* gene produces 11–43 times less mRNA and 486–941 times less protein compared with the *gERECTA-RLUC* gene. (A,B) The amount of ERECTA mRNA in *ER::cERECTA-RLUC* plants is dramatically lower compared with the wild type and *ER::gERECTA* as determined by qRT-PCR with *ACTIN-2* as an internal control. The total RNA was isolated from T5 seedlings of two independent transgenic lines. Graph B demonstrates that ERECTA is expressed at a higher level in *ER::cERECTA-RLUC* plants compared with the *erecta-105* mutant. The mean of two biological replicates is plotted; error bars, 1 SD. (C) Measurements of luciferase activity per milligram of total protein isolated from seedlings. RLU indicates relative light units. The mean of three biological replicates is plotted; error bars, 1 SD.

Semiquantitative RT-PCR analysis with primers annealing at different regions along the length of ERECTA or ERECTA-RLUC cDNA suggested that the 3' end is less stable in both *cERECTA* and *cERECTA-RLUC* compared with *gERECTA* and *gERECTA-RLUC* (Fig. 3). When we used primers annealing closer to the 3' end of ERECTA, there was a higher discrepancy in mRNA accumulation in the *cERECTA* versus *gERECTA* lines compared with when we used primers that annealed closer to the 5' end (Fig. 3B). We observed the same phenomena in the *cERECTA-RLUC* versus *gERECTA-RLUC* transgenic lines (Fig. 3C). The decreased stability of the 3' end was not dependant on a particular 3' UTR sequence as two different UTRs were used in the ERECTA and ERECTA-RLUC constructs (Fig. 3A).

The ligation-mediated poly(A) test (LM-PAT) (Salles and Strickland 1999) and RT-PCR analysis of 5' to 3' end-ligated mRNAs (Kuhn and Binder 2002) were used to examine the presence and length of the poly(A) tail. LM-PAT detected a poly(A) tail in the *gERECTA-RLUC* lines but not in

FIGURE 3. Semiquantitative RT-PCR detects decreased stability of *cERECTA* and *cERECTA-RLUC* 3' UTR. (A) A schematic diagram of primer annealing to *ERECTA* and *ERECTA-RLUC* cDNA. EPF and ERR primers were used for the qRT-PCR presented in Figures 1 and 2. All other primers were used for semiquantitative RT-PCR. The diagram also indicates 3' UTR regions that were used in the constructs: Endogenous *ERECTA* terminator in the constructs with *ERECTA* sequences and the 35S terminator in the constructs with *ERECTA-RLUC* sequences. (B,C) RT-PCR analysis to investigate 3' end stability. While a similar amount of total RNA was used for cDNA synthesis, a different amount of cDNA was used in PCR reactions (as indicated below in the figure) to provide for differences in the *cERECTA* and *gERECTA* mRNA expression. When cDNA was used at the indicated ratios, PCR with primers in the center of the cDNA (P1 and P2 for *gERECTA* and *cERECTA*; P6 and P7 for *gERECTA-RLUC* and *cERECTA-RLUC*) detected similar levels of the template. At the same time in PCR reactions with primers near the 3' UTP (P3 and P4 and P3 and P5 for *gERECTA* and *cERECTA*; P8 and P9 and P8 and P10 for *gERECTA-RLUC* and *cERECTA-RLUC*), the amplification was decreased in *cERECTA* and *cERECTA-RLUC* lines compared with *gERECTA* and *gERECTA-RLUC* lines.

cERECTA-RLUC (Fig. 4A). The measured length of the poly(A) tail in *gERECTA-RLUC* was up to 120 nt based on the size of the amplified fragment. Sequencing confirmed that the poly(A) addition site is 13 nt downstream from the AATAAA sequence of the 35S terminator as reported earlier (Sanfaçon et al. 1991), and we concurrently detected 120–130 adenines. LM-PAT assays were not successful in determining the length of the poly(A) tail in *gERECTA*, possibly due to the shortness of the tail. But sequencing of a fragment obtained in a RT-PCR reaction with the ERg7044 and PAT.rc primers indicated a tail of 17–25 adenines located 184 nt downstream from the stop codon after the sequence GTTTTCT. RT-PCR analysis of 5' to 3' end-ligated mRNAs identified the same site of polyadenylation in endogenous *ERECTA* and the same length of a poly(A) tail. Multiple *ERECTA* ESTs possess the same poly(A) addition site; however, there are also two additional poly(A) addition sites 155 and 169 nt downstream from the stop codon that we did not observe. We were unable to detect any poly(A) tail in *cERECTA* or *cERECTA-RLUC* by LM-PAT. In addition, RT-PCR analysis of 5' to 3' end-ligated mRNAs was not successful in determining the 3' end sequence of those mRNAs, suggesting a heterogeneous nature of the 3' end.

The binding of specific mRNAs to the oligo-dT primer depends on the polyadenylation status of an individual RNA, where mRNAs with a longer poly(A) tail bind more efficiently than do those having a shorter tail. To determine whether mRNA produced by the intronless *ERECTA-RLUC* gene is efficiently polyadenylated or not, we measured the mRNA level with qRT-PCR before and after selection for polyadenylated mRNAs (Fig. 4B). The relative fold difference of *cERECTA-RLUC* to *ACTIN-2* did not change significantly after selection for polyadenylated mRNAs, suggesting similar levels of polyadenylation for these two genes. In contrast, the ratio of *gERECTA-RLUC* to *ACTIN-2* increased five to seven times after selection, indicating a higher level of *gERECTA-RLUC* polyadenylation compared with both *cERECTA-RLUC* and *ACTIN-2*. These data imply that *cERECTA-RLUC* has a poly(A) tail similar to *ACTIN-2* but that *gERECTA-RLUC* is polyadenylated more efficiently, which can mean both a longer poly(A) tail or a higher percentage of mRNA molecules that have a poly(A) tail. To our knowledge, the length of the poly(A) tail in *Arabidopsis ACTIN-2* mRNA is not known. Our attempts to determine it by LM-PAT were not successful. Therefore, introns

in the *ERECTA* gene might increase mRNA accumulation through increasing its stability at the 3' end, and in this model, their function would resemble the role introns play in mammalian β -globin gene expression, where mRNA expression is enhanced by splicing and splicing is necessary for efficient 3' end processing (Lu and Cullen 2003).

The length of the poly(A) tail is important for *ERECTA* translation in vitro

Since our analysis demonstrated that endogenous *ERECTA* has a fairly short poly(A) tail, we investigated whether the extent of *ERECTA* polyadenylation is critical for its translation and whether a short poly(A) tail might be one of the reasons for low protein expression. To verify whether the poly(A) tail is necessary for *ERECTA* translation, an RNaseH treatment in the presence of the oligo-dT primer was used to remove it from *gERECTA-RLUC* mRNA and other mRNAs in the total RNA sample. RNaseH specifically degrades the RNA in RNA:DNA hybrids and does not degrade unhybridized RNA. Removal of the poly(A) tail completely abolished *gERECTA-RLUC* translation in vitro in competitive conditions (Fig. 5A). *ERECTA-RLUC* RNA produced by in vitro

FIGURE 4. *cERECTA-RLUC* mRNA has a significantly lower level of polyadenylation compared to *gERECTA-RLUC*. (A) Poly(A) tail-length assay of total RNA detected about 120 adenines (the amplified fragment is ~400 bp long, of which 280 bp encodes a segment of *RLUC*, the 3' UTR sequence, and the oligo-dT anchor primer) at the end of *gERECTA-RLUC* but none in *cERECTA-RLUC*. The sequencing determined that the two weak bands in *cERECTA-RLUC* are unspecific products of amplification. The identity of the major band and the poly(A) length in the *gERECTA-RLUC* sample was confirmed by sequencing. (B) *ERECTA-RLUC* to *ACTIN-2* ratio before and after elution from oligo(dT) beads as determined by qRT-PCR. The selection for polyadenylated mRNAs does not change *cERECTA-RLUC*/ *ACTIN-2* ratio but significantly increases the *gERECTA-RLUC*/ *ACTIN-2* ratio. RNA was isolated from T3 homozygous seedlings. The mean of three technical replicas is plotted.

transcription does not have a poly(A) tail, but it is efficiently translated in our reaction, most likely because it does not have to compete for ribosomes with other mRNAs. The fact that its translation is not changed by oligo-dT/RNaseH treatment indicated that this treatment did not lead to internal degradation of *ERECTA-RLUC* mRNA (Fig. 5B).

Next, we examined whether lengthening the poly(A) tail can increase *cERECTA-RLUC* and *gERECTA-RLUC* in vitro translation. For this analysis we used mRNA that was eluted from an oligo-dT column and already had some poly(A) tail. After treatment with poly(A) polymerase, *cERECTA-RLUC* and *gERECTA-RLUC* were translated 7.5 and 2.3 times better, respectively, compared with untreated samples (Fig. 5C). Thus, the length of a poly(A) tail is critical for *ERECTA* translation in competitive conditions in vitro, and the decrease in *cERECTA-RLUC* translation in transgenic plants is likely the result of changes in the 3' end mRNA structure.

The sequence in front of the *ERECTA* gene functions as a strong promoter

The first intron of *ERECTA* is larger in size, at 346 nt, than the other introns at 69–164 nt. Introns located near the translational start site are known to alter transcription (Callis et al. 1987; Gidekel et al. 1996). However, inclusion of the first intron in the *ERECTA* cDNA (the construct pESH212) did not result in complementation of *erecta-105* based on phenotypic analysis of multiple T1 plants (Table 1; Fig. 6A). Moreover, analysis of *ERECTA* expression by a promoter GUS construct (Shpak et al. 2004) was consistent with in situ hybridization (Yokoyama et al. 1998) and microarray data (analysis of available data by Genevestigator and eFP browser) suggesting that elements upstream of the translational start site are sufficient for proper *ERECTA* expression. *ERECTA* promoter was successfully used for expres-

sion of *CLAVATA1* and *LhG4* in *Arabidopsis* (Dievart et al. 2003; Song et al. 2006) as it is considered to be a relatively strong promoter for shoot meristematic tissues. In addition, replacement of the *ERECTA* promoter with the 35S promoter in front of the *cERECTA* sequence does not result in a functional construct (data not shown). All these data suggest that introns might be important for transcription elongation, 3' end formation, or mRNA stability but not for transcription initiation.

While the presence of introns is important, no individual intron is absolutely necessary for *ERECTA* expression

Of the 26 introns in the *ERECTA* gene, 19 separate exons encoding individual leucine-rich repeats (LRRs) (Torii et al. 1996). To determine which specific introns are essential for *ERECTA* function, we expressed several constructs with varying sets of introns in *erecta-105* and analyzed the phenotype of T1 transgenic plants (Table 1; Fig. 6A). The last three introns separate exons that encode the kinase domain of *ERECTA*. Their removal does not decrease the functionality of the construct (pKUT 199), nor does their inclusion into *cERECTA* (pDLH 101) restore the construct's complementation capability. However, the presence of introns between exons encoding LRRs was significant for the functionality of the *ERECTA* gene. Two constructs, one with introns 1–14 (pESH210) and the other with introns 15–23 (pESH217), were able to rescue the *erecta-105* mutant (Fig. 6A,B), although the first construct was slightly less effective (Fig. 6D). These data show that none of the individual *ERECTA* introns are absolutely obligatory for the functionality of the gene.

Further removal of introns from both groups reduced or abolished the ability of constructs to complement the

FIGURE 5. The length of the poly(A) tail is important for *ERECTA* translation in vitro. (A,B) Removal of the poly(A) tail decreased *ERECTA* translation in vitro in competitive conditions. (A) When the poly(A) tail is removed by oligo(dT)/RNaseH treatment, g*ERECTA*-RLUC mRNA is not translated in vitro. (B) RNaseH does not change translation of *ERECTA*-RLUC RNA produced in vitro without a poly(A) tail in noncompetitive conditions. (C) Elongation of the poly(A) tail increases in vitro translation of both c*ERECTA*-RLUC and g*ERECTA*-RLUC. Plant produced mRNA eluted from oligo(dT) beads was used as a starting material.

mutation. For example, while a construct with the first 14 introns (pESH 210) was able to complement the mutant, constructs with introns 1–10 (pESH 222) or with 11–14 (pESH 221) were not able to do so (Table 1). Our data also suggest that a functional *ERECTA* gene is not characterized by the presence of a minimal number of introns but by the presence of introns in the correct locations. For example, *ERECTA* with five introns (introns 19–23; pESH 225) rescued the mutant, while *ERECTA* with 10 introns (introns 1–10; pESH222) did not (Fig. 6A).

We chose two constructs, pESH 210 and pDHL 101, for further analysis in T3 generation. In T1 and T2 generations, the first one partially complemented *erecta* and the second did not. Analysis of mRNA expression in T3 transgenic seedlings established that the ability of constructs to complement the *erecta* mutant (Fig. 6B,D) is correlated with mRNA abundance (Fig. 6C), just as in the case with the c*ERECTA* and g*ERECTA* constructs (Fig. 1B,D), thus again confirming the importance of introns for mRNA accumulation. These data suggest that *ERECTA* expression depends on

the presence of introns in multiple specific locations and that introns function in an additive manner to ensure a high level of mRNA accumulation.

Expression of individual *ERECTA* exons is dependent on the presence of an intron

A search for introns that are required for *ERECTA* expression through generation of transgenic plants proved to be very time consuming. So a different approach was used to dissect the individual introns that can affect *ERECTA* expression. We created constructs in which different segments of the *ERECTA* gene with or without introns were fused with Renilla luciferase and placed under control of 35S promoter and terminator. These constructs, together with a construct encoding firefly luciferase (FLUC; as an internal control), were transiently expressed in *Arabidopsis* seedlings. By measuring luciferase activity in protein extracts, the expression of constructs with introns was compared to expression of corresponding constructs without introns.

This assay demonstrated that small segments of the *ERECTA* gene can change their expression in an intron-dependent manner (Fig. 7A). While the expression level of one construct (exons 23–27) was not affected by intron presence, the expression of a majority of the constructs was. For example, the expression of exons 17–21 and exons 2–7 increased 11 times when they were separated by introns. In other cases, the presence of introns moderately increased the expression of constructs: Introns increased expression of exons 8–13 1.6 times and of exons 14–19 2.6 times. The results of this experiment were similar to the data we obtained from the analysis of transgenic plants. Thus, introns 23–26 did not confer functionality to *ERECTA* (pKUT199) (Table 1) and were not able to increase expression of exons 23–27 (Fig. 7A). A construct with introns 19–23 (pESH225) (Table 1) was able to complement the *erecta*-105 mutant, and introns increased expression of constructs with exons 17–21 and exons 20–23 (Fig. 7A).

More constructs were created with the intent of finding the smallest segment with expression dependent on the presence of an intron; some of them are presented in Figure 7B. In general, decreasing the size of constructs reduced the difference in expression with and without introns. For example, the presence of introns in the construct with exons 20–23 increased its expression 5.5 times (Fig. 7A). Breaking down this segment of *ERECTA* into smaller pieces containing only two exons (constructs with exons 20 and 21, 21 and 22, or 22 and 23) reduced the effect to only a 2–2.8 times boost of expression (Fig. 7B). In no cases have we observed that the inclusion of an intron in a construct leads to decreased expression. Further analysis of the construct with exons 22–23 demonstrated that the increased expression imparted by an intron occurs at both the protein (Fig. 7D) and mRNA levels (Fig. 7E) and, in both cases, leads to roughly doubling the expression.

TABLE 1. Ability of an *ERECTA* gene with varying number of introns to complement the *erecta-105* mutant

<i>ERECTA</i> gene structure in the construct	Introns in the construct	Intron count	Ability to rescue the <i>erecta</i> mutation	# of fully or partially complemented plants/total T1 plants analyzed	Name of the construct
	1-26	26	+++	14/14	pKUT 196
	0	0	-	0/43	pKUT 524
	1	1	-	0/30	pESH 212
	1-23	23	+++	20/20	pKUT 199
	1-14	14	++	27/29	pESH 210
	1-10	10	-	0/38	pESH 222
	11-14	5	-	0/36	pESH 221
	11-23	13	+++	32/36	pESH 218
	15-23	9	+++	29/36	pESH 217
	15-18	4	-	0/38	pESH 224
	19-23	5	++	54/54	pESH 225
	24-26	3	-	0/27	pDLH 101

Exons are represented as boxes and introns as lines in the structure of created *ERECTA* constructs. Red, blue, orange, dark blue, and green boxes represent exons encoding a signal sequence, segments with paired cysteines, LRR repeats, a transmembrane domain, and a kinase domain, respectively. The size of exons and introns boxes is scaled. The ability to rescue the *erecta* mutation is as follows: +++, complete complementation; ++, partial complementation (final plant height and pedicel length are intermediate between those of wt and *erecta-105*); and -, no complementation.

Based on our data, we were able to analyze whether the ability of *ERECTA* introns to promote gene expression correlates with their IMETER score (Rose et al. 2008). Several *ERECTA* introns gave a positive IMETER score: intron 1 (score 24), intron 16 (6.4), intron 2 (6), intron 19 (4), intron 26 (3.5), intron 15 (1.3), and intron 4 (0.3). All other introns had a negative score. These scores did not correspond to the introns' ability to enhance gene expression. While introns 20, 21, and 22 had a high ability to enhance expression in our experimental system, they had negative IMETER scores.

Our data show that multiple introns increase *ERECTA* expression in an additive manner and that the presence of introns is significant for mRNA accumulation. However, not all introns are able to influence *ERECTA* expression. The placement of introns between exons 23–26, 4–5, or 13–14 does not alter expression of the corresponding constructs.

The ability of *ERECTA* introns to increase expression is not sequence specific

As most *Arabidopsis* genes do not require introns to be expressed, we reasoned that there may be sequences in the *ERECTA* exons that inhibit expression unless introns are

present. We examined exons 22 and 23 to gain some insight into the mechanism by which introns increase gene expression. First, we hypothesized that some aspect of splicing is important for *ERECTA* expression, but not the nature of the intron. Consistent with this hypothesis, intron 13, which does not change expression of exons 13 and 14, increased the expression of exons 22 and 23 to the same degree as did intron 22 (Fig. 7C). We then searched for an expression inhibiting sequence that requires the presence of an intron. Replacement of exon 22 with exon 13 abolished dependence of construct expression on the presence of an intron (13 + 23) (Fig. 7C), while replacement of exon 23 with exon 14 did not (22 + 14) (Fig. 7C), suggesting that an inhibiting sequence might be located in exon 22. Furthermore, placing the intron in front of exon 22 did not increase expression of the construct (23 + 22) (Fig. 7C), suggesting that a splice site needs to be located behind a putative inhibiting sequence. We hypothesize that exon 22 has a sequence that is able to inhibit some process that is important for mRNA accumulation, such as transcription elongation, 3' end formation, or post-transcriptional mRNA stability.

The assembly of spliceosome or deposition of EJC proteins might somehow obstruct the function of such a sequence.

DISCUSSION

The reduced ability of cDNA constructs to complement mutants, especially when expressed under their native promoters, is a widespread, but unfortunately not well documented, phenomenon due to the negative result nature of these experiments. In pursuit of increased transgene expression, an intron is routinely inserted into commercially available mammalian expression vectors, but it is not known whether its presence in these vectors increases expression of the majority of cDNA inserts or improves expression only under selected circumstances. Deeper understanding of intron involvement in regulation of gene expression is critical both for the basic knowledge of gene function and, at a practical level, for increased protein production in research and industry.

Here, we examined the function of introns in the regulation of *ERECTA* gene expression. The vital role introns play in *ERECTA* expression makes this gene an excellent model system. The intronless *ERECTA* gene is not able to rescue the

FIGURE 6. The presence of multiple introns is critical for *ERECTA* gene function and mRNA accumulation, but no particular intron is essential. (A,B) Complementation of the *erecta-105* mutant with *ERECTA* constructs containing selected introns; bar, 3 mm. The structure of constructs is described in Table 1. (C) The ability of a construct to complement *erecta-105* correlates with the amount of mRNA produced. qRT-PCR analysis demonstrates that pDLH101 plants produce dramatically less *ERECTA* mRNA and pESH201 plants produce an equal amount compared with the wild type. The total RNA was isolated from T3 gentamicin selected seedlings. (D) Morphometric analysis of fully grown 8-wk-old plants of the wild type, *erecta-105*, and of T3 transgenic plants expressing pDLH101 and pESH210 in an *erecta-105* background. Twenty-one plants were analyzed for plant (inflorescence) height. Lengths of 40 mature pedicels and siliques on the main inflorescence stem (eight measurements per stem) were analyzed. Bars, average values $\pm$ 1 SD.

erecta mutant, and it produces at least 10 times less mRNA compared with the endogenous gene. The endogenous *ERECTA* protein is produced in very small amounts that are difficult to detect, but by fusing *ERECTA* with Renilla luciferase, we were able to study protein accumulation. We observed that the intronless *ERECTA* gene generates 500–900 times less protein compared with an identical construct containing introns. Therefore, the presence of introns in the *ERECTA* gene is required both for mRNA accumulation and for its efficient translation. Inefficient translation of mRNA produced by the intronless gene is likely due to its degradation at the 3' end, as we were unable to detect the poly(A) tail and found decreased stability of the 3' UTR. This instability is not related to the structure of the 3' UTR as we detected it

using both the endogenous *ERECTA* 3' UTR and the 35S terminator.

Having observed defects of the poly(A) tail, we investigated the importance of poly(A) for *ERECTA* translation. Normally, the Poly(A)-Binding Protein complex (PABP) binds the poly(A) tail and has a stimulatory effect on translation due to its interaction with eIF4G (Le et al. 1997; Imataka et al. 1998), a part of the cap-binding complex that mediates the attachment of the 43S pre-initiation complex. PABP–eIF4G interactions confer an advantage under competitive conditions, ensuring a stronger association of the cap binding complex with mRNA (Proweller and Butler 1997; Borman et al. 2000). Removal of the poly(A) tail from *ERECTA* mRNA demonstrated that its presence was critical for *ERECTA* translation in competitive conditions in vitro. In accordance, elongation of the poly(A) tail enhanced translation of *ERECTA*, and this effect was stronger for mRNA produced by the intronless gene, suggesting an altered nature of its tail.

The decreased accumulation of *ERECTA* mRNA produced by the intronless gene theoretically can be due to impairment in transcription initiation or elongation, inefficient formation of the 3' end, or formation of defective mature ribonucleoproteins (mRNPs) that are less stable. The high activity of the *ERECTA* promoter suggests that it is able to successfully initiate transcription on its own. While hypothetically it is possible that the *ERECTA* exons carry multiple elements that negatively regulate transcription initiation and that the presence of introns can counterbalance them, this

possibility seems to be far-fetched. Thus, we presume that inefficiency in the later steps of biogenesis of mRNPs is responsible for reduced expression of the intronless *ERECTA* gene. Not only does the intronless *ERECTA* gene produce less mRNA, but the mRNA produced has a deficient 3' end. As polyadenylation of mRNA is known to be important for mRNA stability by protecting mRNA from degradation by 3' exonuclease (Houseley and Tollervey 2009), its decline could be one of the reasons for decreased *ERECTA* accumulation. However, other scenarios involving abortive transcription or changed composition of mRNPs leading to mRNA degradation are possible as well.

While multiple introns promote *ERECTA* expression in an additive manner, none of the introns is essential. Introns

FIGURE 7. The level of expression for some *ERECTA* exons depends on the presence of introns. (A–C) RLUC/FLUC luminescence ratio measured in protein extracts from transiently transformed *Arabidopsis* seedlings expressing selected *ERECTA* exons fused to RLUC (35S::*ERECTA* exons-RLUC) and FLUC as an internal control. The level of protein expression is compared for constructs encoding identical proteins but distinguished by the presence or absence of introns. * $P < 0.05$ based on Student *t*-test. (D,E) The higher expression of intron-containing exons 22 + 23-RLUC construct versus the identical construct but without intron is evident on both protein (D) and mRNA (E) levels. The amount of mRNA was determined by qRT-PCR with FLUC as an internal control and amount of protein by measurements of RLUC/FLUC luminescence ratio. The mean of three biological replicates is plotted; error bars, 1 SD for A though E.

retain the ability to promote gene expression even when small segments of the *ERECTA* gene are expressed under control of the 35S promoter. Our data suggest that it is not the sequence of introns but their specific location that is essential for efficient gene expression. It has been proposed that in plants, promoter proximal introns are enriched in signals that enhance gene expression (Rose et al. 2008). However, many *ERECTA* introns that contribute to a high level of expression are not promoter proximal and do not have a high IMEter score.

Here, we propose an additional mechanism for IME in plants. Our data suggest that the ability of introns to promote *ERECTA* gene expression might be independent of their sequence and instead be related to recruitment of spliceosomal or EJC proteins. These proteins are capable of influencing multiple stages of mRNA biogenesis. For example, the assembly of a spliceosome might promote transcription elongation or efficient cleavage and polyadenylation of the *ERECTA* 3' end. Alternatively, due to their association with EJC proteins, spliced mRNAs might have a different rate of cytoplasmic polyadenylation/deadenylation and, as a result, experience increased stability.

Recently it has been reported that *ERECTA* expression is decreased in an *emu* mutant (Furumizu et al. 2010). EMU encodes a protein homologous to yeast HPR1, a component of the THO complex. This complex plays a role in mRNA biogenesis and is needed for efficient transcriptional elongation, 3' end processing, and mRNA export (Rondon et al. 2010). It is recruited to genes during transcription and is involved in formation of optimal mRNPs. In yeast, THO mutant phenotypes are suppressed by slowing down transcription (Jimeno et al. 2008), and THO might be critical for highly expressed genes and genes that require a fast assembly of mRNPs. Whether THO functions in an analogous manner in plants and what its role is in *ERECTA* expression are interesting questions for future research. One possibility is that both splicing and THO contribute to efficient assembly of *ERECTA* mRNPs that are critical for this gene's expression.

Why does *ERECTA* expression depend so strongly on the presence of introns? While it is hypothesized that for most mammalian genes efficient 3' end formation and export from the nucleus relies on recruitment of spliceosome (Guang and Mertz 2005), this is obviously not the case for plants. Plants have a much higher proportion of intronless genes (~20% in

Arabidopsis versus ~5% in *H. sapiens*), and the removal of introns alters gene expression only in selected cases. Thus, the formation of the 3' end and mRNA export from nucleus should be less dependent on splicing for the majority of plant genes. We speculate that exon sequences of some plant genes, including *ERECTA*, have *cis*-elements preventing efficient mRNP biogenesis. For example, exons might have sequences that inhibit transcription elongation, titrate out polyadenylation factors, or lead to mRNA instability. During splicing these sequences could be concealed by spliceosomal proteins or by an EJC. The degree to which gene expression depends on the presence of introns could hinge on the number and strength of those sequences. Based on the facts that the *ERECTA* gene functions best when multiple introns are present and that the degree of complementation of the *erecta* mutant increases gradually with intron number, we expect to find multiple sequences in exons inhibiting mRNPs biogenesis. So far we have not identified such elements by exon sequences analysis. On the one hand, at this point we do not have complete information about which exons might carry such sequences, and on the other hand, the predicted

RNA *cis*-acting elements might be determined partially by secondary structure.

The *ERECTA* gene has a rather distinctive structure that is shared by many other *Arabidopsis* LRR receptor-like kinases. Of the 27 exons in the *ERECTA* gene, 20 encode tandem LRRs of 24 amino acids each. Every LRR is encoded by a separate exon, and introns of similar sizes are present at the exact same “phase 2” position, between the second and third nucleotides of the codon for leucine at position 13 (underlined) in the consensus P**LG*L**L**L*L**N*L*G*I (asterisks represent nonconserved amino acids) (Torii et al. 1996). The identical gene structure is found in the *Arabidopsis* *ERL1* and *ERL2* paralogs of *ERECTA*, as well as *ERECTA* genes in other species as evolutionarily distant as *Physcomitrella patens* (data not shown). Analysis of the published gene structure of 220 *Arabidopsis* LRR receptor-like kinases (Shiu and Bleecker 2003) revealed that about half of them (111) did not possess a single intron in the gene segment encoding LRRs. The other half (93) had most LRRs encoded by separate exons with introns placed in exactly the same position as in *ERECTA*, between the second and the third nucleotides of the codon for leucine that is analogous to Leu13 in *ERECTA*. The consistent and repeated location of introns in LRR-containing receptors suggests that they may have evolved by exon duplication. In some genes, introns might have later become functionally important and involved in regulation of gene expression. An analysis of IME occurrence in *ERECTA* family genes in *Arabidopsis* and other species and in other LRR receptors could help determine how common IME is in these gene families.

It is also possible that the importance of introns for *ERECTA* expression is gene specific and related to *ERECTA* function in the regulation of cell proliferation (Shpak et al. 2003, 2004; Pillitteri et al. 2007). Endogenous *ERECTA* mRNA has a poly(A) tail of 17–25 adenines. The shorter length of the poly(A) tail in mRNAs that regulate the cell cycle and a specific control of its deadenylation has been proposed to be responsible for delimiting their expression during the cell cycle (Beilharz and Preiss 2007). In addition to the 3' UTR, the presence of introns in *ERECTA* might influence the length of poly(A) and thus contribute to tight regulation of *ERECTA* expression.

The ability of *ERECTA* to regulate plant morphology and transpirational efficiency (Masle et al. 2005) makes it an attractive target of genetic manipulation in crop plants. Expression of *ERECTA* in *erecta-105* or the wild type using the 35S promoter and the endogenous *ERECTA* terminator rescues the mutant phenotype but does not result in bigger or faster growing plants compared to the wild type (data not shown). In this article, we described several transgenic lines where expression of *ERECTA* under control of its own promoter and terminator resulted in a three to nine times increase of RNA levels, but plants were still phenotypically identical to the wild type. One possibility is that *ERECTA* expression is tightly regulated at the post-transcriptional

level and that to manipulate plant size one has to address it. For example, expression of *ERECTA* under control of terminators that are able to promote formation of a longer poly(A) tail might be advantageous for increased protein accumulation. Knowledge of *ERECTA* mRNP biogenesis will further contribute to our understanding of this gene's expression and might indicate novel means to manipulate its expression and, therefore, plant size.

MATERIALS AND METHODS

Generation of transgenic plants

The plasmid carrying *gERECTA* (pKUT 196) has been described previously (Godiard et al. 2003). To construct a plasmid carrying *cERECTA* (pKUT 524), the cDNA sequence of *ERECTA* was cloned as a *SpeI* fragment into a pZP222 vector carrying the *ERECTA* promoter and terminator sequences (pKUT 522). For cloning purposes, an *NcoI* restriction site was introduced in front of the first ATG in pKUT 524. To construct plasmids carrying *gERECTA-RLUC* (pESH 427) and *cERECTA-RLUC* (pWSL123), the Renilla luciferase (*RLUC*) sequence and the 35S terminator from pAVA 696 (Subramanian et al. 2006) were cloned behind *ERECTA* into pKUT 196 or pKUT 524, replacing the endogenous *ERECTA* terminator. The stop codon at the end of the *ERECTA* sequence was changed to the coding sequence of Ala in pESH427 and to Gly-Ser-Ala in pWSL123.

The structures of the *ERECTA* gene in the following constructs (pESH 212, pKUT 199, pESH 210, pESH 222, pESH 221, pESH 218, pESH 217, pESH 224, pESH 225, and pDLH 101) are described in Table 1. The backbone of all these plasmids was the vector pZP222, and all constructs carried the endogenous *ERECTA* promoter and terminator in the same orientation as in pKUT 522. These plasmids were constructed by combining different *ERECTA* segments from pKUT196 (*gERECTA*) and pKUT524 (*cERECTA*). We did not introduce any novel restriction sites inside of the *ERECTA* sequence but, instead, used the endogenous *ERECTA* restriction sites during cloning.

The described plasmids were introduced into an *Agrobacterium tumefaciens* strain GV3101/pMP90 by electroporation and into *erecta-105* plants by the floral dip method (Clough and Bent 1998). The transgenic plants were selected based on gentamicin resistance. The *Arabidopsis erecta-105* knockout mutant (Colombia ecotype) has been described previously (Torii et al. 1996). The conditions for plant growth were as described earlier (Shpak et al. 2003).

RNA isolation and cDNA synthesis

Total RNA was isolated from 7-d-old seedlings grown on MS plates or from floral buds of plants grown in soil using the Spectrum Plant RNA Isolation Kit (Sigma-Aldrich). The RNA quantification was done by both spectrophotometry and non-denaturing electrophoresis.

Poly(A) mRNA was isolated using the PolyATtract mRNA Isolation System III (Promega) as follows: Total RNA (~35 µg) was heat denatured, incubated with a biotinylated oligo-dT probe in the presence of 0.5× SSC, and then incubated for 15 min with prewashed MagneSphere streptavidin beads. The beads were

washed three times with $0.1\times$ SSC solution before elution of poly(A) mRNA with 35 μ L of water.

The total RNA was treated with RNase-free RQ1 DNase (Promega). First-strand cDNA was synthesized from 1 or 2 μ g of total RNA or 10 μ L of eluted poly(A) mRNA with random hexamer primers using a ProtoScript M-MuLV Taq RT-PCR kit (NEB). Use of total RNA and random hexamer primers for cDNA synthesis was significant as it allowed measuring the level of *ERECTA* mRNA independently of its polyadenylation status. As a control for genomic DNA contamination, we used primers flanking intron-containing regions and analyzed products by gel electrophoresis or used RNA directly for PCR.

qRT-PCR and semiquantative RT-PCR

qPCR was performed with a MyiQ Real-Time PCR Detection System (BioRad) using Sso Fast EvaGreen Supermix (Bio-Rad). Each experiment contained three technical replicates and was performed in a total volume of 20 μ L with 2 μ L of the five times diluted first-strand reaction. Cycling conditions were 1 min at 95°C and then 50 repeats of 30 sec at 95°C, 30 sec at 61.6°C for the *ERECTA* or 59.3°C for the *ACTIN-2*, and 30 sec at 72°C followed by the melt curve analysis. The qPCR ERF and qPCR ERR primers were used to amplify *ERECTA* and *ERECTA-RLUC*; the *sbrActin2F* and *sbrActin2R* primers were used to amplify *ACTIN-2* (At3g18780); the *Fluc891* and *35Sterm.rc* primers were used to amplify *ERECTA* exons 22-23-*RLUC*; and the *Fluc* qPCR F and *Fluc* qPCR R primers were used to amplify *FLUC*. The Ct values were calculated using the iQ5 software (Bio-Rad). The fold difference in the gene expression was calculated using relative quantification or the $2^{-\Delta\Delta C_t}$ method (Livak and Schmittgen 2001).

Semiquantative RT-PCR was used to assay the stability along the length of *ERECTA* mRNA. PCR was performed for 2 min at 93°C and then with varying numbers of cycles for 30 sec at 93°C, for 30 sec at 58°C, for 60 sec at 72°C, and then for 3 min at 72°C. The amount of cDNA used for reaction was as in the figure legend. Sets of primers used are described in the figure legend, and sequences are given in Supplemental Table S1.

Renilla luciferase activity assay

ERECTA-RLUC protein expression was measured by monitoring Renilla luciferase activity with a 20/20n single tube luminometer in transgenic seedlings using the Renilla Luciferase Reporter Assay (Promega). The protein concentration in each sample was determined using the Bradford reagent (Sigma).

Poly(A) tail length assays

The LM-PAT assay was performed according the method described by Salles and Strickland (1999). One microgram of total RNA was heat denatured for 7 min at 65°C in the presence of 20 ng of the phosphorylated oligo-dT primer (NEB), and then the ligation reaction was performed for 30 min at 42°C using the following conditions: $1\times$ Reverse Transcriptase Reaction Buffer (NEB), 1 mM DDT, 2 mM dNTPs, 0.1 mM ATP, and 10 U T4 DNA ligase (NEB). The ligation was continued for 2 h at 12°C in the presence of 200 ng of the oligo-dT anchor primer (Invitrogen). The first-strand cDNA was synthesized for 1 h at 42°C with 10 U of MuLV-RT (NEB). The amplification reaction was done with the *Fluc891* and *PAT.rc* primers.

The RT-PCR analysis of 5' to 3' end-ligated mRNAs was performed according the method described by Kuhn and Binder (2002) with slight modifications. Ten microliters of total RNA was de-capped with 5 U of tobacco acid phosphatase (TAP; Epicentre) in supplied buffer for 2 h at 37°C and then phenol-chloroform extracted, ethanol precipitated, and dissolved in 10 μ L of water. Five microliters of decapped RNA was 5' to 3' end-ligated with 40 U of T4 RNA Ligase 1 (NEB) in supplied buffer supplemented with 1 mM ATP for 24 h at 4°C and then again phenol-chloroform extracted, ethanol precipitated, and dissolved in 10 μ L of water. The cDNA synthesis reaction was performed with 1–3 μ L of the circularized RNA and *ERECTA*-specific primer ERg2550.rc. The 5' and 3' end sequences of mRNA were established by sequencing the amplified fragment obtained in a PCR reaction with the produced cDNA and the ERg7201 and ERg1790.rc primers. To verify that cDNA synthesis was successful, a control PCR reaction with the ERg1761 and ERg2339.rc primers was performed.

In vitro transcription

The *ERECTA* gene was amplified by Phusion high-fidelity polymerase (NEB) using the ER-T7 and 35Sterm.rc primers and pWSL123 as a template. The ER-T7 primer contained a T7 promoter in front of the *ERECTA* 5' UTR sequence. The 35-Sterm.rc primer annealed to *ERECTA* downstream from the stop codon but upstream of the putative polyadenylation site. PCR product was used for in vitro transcription reactions with T7 RNA Polymerase (NEB) and rNTPs (NEB).

Poly(A) tail removal by RNase H and poly(A) tail elongation

Total RNA (8 μ g) from T3 *gERECTA-RLUC* transgenic plants was heat denatured for 5 min at 70°C and then incubated for 10 min at 42°C. Poly(A) tail digestion was carried out for 1 h at 37°C in the presence of oligo-dT (NEB) and 5 U of RNaseH (NEB). Two control reactions were performed: the first without addition of oligo-dT and the second with 1 μ g of in vitro transcribed *ERECTA* RNA (a poly(A) tail is absent) instead of total RNA. After the digestion, RNA was extracted with phenol-chloroform-isoamyl alcohol (25:24:1), precipitated with ethanol, and then resuspended in 10 μ L of water. The RNA (3.25 μ L) was subsequently used for in vitro translation and luciferase activity assay.

Seventeen microliters of poly(A) eluted mRNA (as described above) from T3 *gERECTA-RLUC* and *cERECTA-RLUC* transgenic plants was incubated with 10 U of poly(A) polymerase (NEB) in the presence of $1\times$ poly(A) polymerase buffer, 1mM ATP, and 20–40 U of RNase inhibitor in the 25 μ L volume for 20 min at 37°C. Control reactions were performed without the poly(A) polymerase. Then RNA was extracted with phenol-chloroform-isoamyl alcohol (25:24:1), precipitated with ethanol in the presence of polyacryl carrier (1:400 dilution; MRC, Inc.), and then resuspended in 10 μ L of water. The RNA (3.25 μ L) was subsequently used for in vitro translation and luciferase activity assay.

Prior to the in vitro translation reaction, the RNA was denatured for 3 min at 65°C. In vitro translation was done for 90 min at 30°C using a rabbit reticulocyte lysate system (Promega) with total volume of 12.5 μ L. To measure in vitro *ERECTA-RLUC*

production, 2 μ L of translation mixture were used in a Renilla luciferase assay system (Promega).

Transient transformation of *Arabidopsis* seedlings with plasmids carrying short *ERECTA* segments fused to Renilla luciferase (*RLUC*)

To create plasmids for the transient expression experiment, we amplified short segments of the *ERECTA* gene using as templates plasmids carrying *gERECTA* (such as pKUT 196) or *cERECTA* (such as pKUT524) and cloned these segments in a frame with C-terminal *RLUC* into *Nco*I digested pAVA 696 (Subramanian et al. 2006). This cloning placed the fusion proteins under control of the 35S promoter and the 35S terminator. To create plasmids with novel arrangements of *ERECTA* introns and exons (such as exon22-intron13-exon23; exon13-intron13-exon23; exon13-exon23; exon22-intron13-exon14; exon22-exon14; exon23-intron22-exon22), we used overlap extension PCR. We generated PCR products encoding individual exons and introns and then placed them together in the PCR reaction, where they annealed and primed off each other to give a full-length product. The sequence of all amplified fragments and their orientation in pAVA 696 were confirmed.

Arabidopsis 10-d-old seedlings were transiently transformed with created vectors and pAVA 315 (35S::*FLUC*) as an internal control. As we compared expression of fusion proteins from genes that are identical except for the presence or absence of introns, we took special care to use the same amount of plasmids encoding such genes (1 μ g, the concentration determined by both spectrophotometric measurements and confirmed by gel electrophoresis). For transient transformation, we used 1.1 μ m tungsten microcarriers (BioRad) and a PSD-1000/He particle bombardment system (Bio Rad). Three independent transformation experiments were performed for each plasmid. *Arabidopsis* seedlings were harvested 24 h after bombardment and assayed by the Dual Luciferase Reporter Assay System (Promega) or 16 h after bombardment and used for RNA extraction. The *RLUC* activity was divided by the *FLUC* activity to normalize for transfection efficiencies.

Oligo DNA sequences

Primer sequences are available as supplementary data (Table S1).

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank Albrecht von Arnim for comments on the manuscript and DanYu Kong for confirming the sequence of *ERECTA*-*RLUC* mRNAs. We also thank Albrecht von Arnim for plasmid constructs (pAVA 696 and pAVA 315). This work was supported in part by a start-up fund from University of Tennessee and in part by the National Science Foundation (IOS-0843340 to E.S.). The initial part of the work was supported by Department of Energy (grant no. DE-FG02-03ER15448 to K.U.T.).

Received May 23, 2011; accepted July 30, 2011.

REFERENCES

- Beilharz TH, Preiss T. 2007. Widespread use of poly(A) tail length control to accentuate expression of the yeast transcriptome. *RNA* 13: 982–997.
- Borman AM, Michel YM, Kean KM. 2000. Biochemical characterization of cap-poly(A) synergy in rabbit reticulocyte lysates: the eIF4G-PABP interaction increases the functional affinity of eIF4E for the capped mRNA 5'-end. *Nucleic Acids Res* 28: 4068–4075.
- Bourdon V, Harvey A, Lonsdale DM. 2001. Introns and their positions affect the translational activity of mRNA in plant cells. *EMBO Rep* 2: 394–398.
- Buchman AR, Berg P. 1988. Comparison of intron-dependent and intron-independent gene expression. *Mol Cell Biol* 8: 4395–4405.
- Callis J, Fromm M, Walbot V. 1987. Introns increase gene expression in cultured maize cells. *Genes Dev* 1: 1183–1200.
- Clancy M, Vasil V, Curtis HL, Vasil IK. 1994. Maize Shrunken-1 intron and exon regions increase gene expression in maize protoplasts. *Plant Sci* 98: 151–161.
- Clough SJ, Bent AF. 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16: 735–743.
- Curi GC, Chan RL, Gonzalez DH. 2005. The leader intron of *Arabidopsis thaliana* genes encoding cytochrome c oxidase subunit 5c promotes high-level expression by increasing transcript abundance and translation efficiency. *J Exp Bot* 56: 2563–2571.
- Dean C, Favreau M, Bond-Nutter D, Bedbrook J, Dunsmuir P. 1989. Sequences downstream of translation start regulate quantitative expression of two petunia *rbcS* genes. *Plant Cell* 1: 201–208.
- Deng T, Li Y, Johnson LF. 1989. Thymidylate synthase gene expression is stimulated by some (but not all) introns. *Nucleic Acids Res* 17: 645–658.
- Diem MD, Chan CC, Younis I, Dreyfuss G. 2007. PYM binds the cytoplasmic exon-junction complex and ribosomes to enhance translation of spliced mRNAs. *Nat Struct Mol Biol* 14: 1173–1179.
- Diévarat A, Dalal M, Tax FE, Lacey AD, Huttly A, Li J, Clark SE. 2003. *CLAVATA1* dominant-negative alleles reveal functional overlap between multiple receptor kinases that regulate meristem and organ development. *Plant Cell* 15: 1198–1211.
- Ernst RK, Bray M, Rekosh D, Hammariskjold ML. 1997. A structured retroviral RNA element that mediates nucleocytoplasmic export of intron-containing RNA. *Mol Cell Biol* 17: 135–144.
- Fong YW, Zhou Q. 2001. Stimulatory effect of splicing factors on transcriptional elongation. *Nature* 414: 929–933.
- Furumizu C, Tsukaya H, Komeda Y. 2010. Characterization of *EMU*, the *Arabidopsis* homolog of the yeast THO complex member *HPRI*. *RNA* 16: 1809–1817.
- Gidekel M, Jimenez B, Herrera-Estrella L. 1996. The first intron of the *Arabidopsis thaliana* gene coding for elongation factor 1 β contains an enhancer-like element. *Gene* 170: 201–206.
- Godiard L, Sauviac L, Torii KU, Grenon O, Mangin B, Grimsley NH, Marco Y. 2003. *ERECTA*, an LRR receptor-like kinase protein controlling development pleiotropically affects resistance to bacterial wilt. *Plant J* 36: 353–365.
- Guang S, Mertz JE. 2005. Pre-mRNA processing enhancer (PPE) elements from intronless genes play additional roles in mRNA biogenesis than do ones from intron-containing genes. *Nucleic Acids Res* 33: 2215–2226.
- Hocine S, Singer RH, Grünwald D. 2010. RNA processing and export. *Cold Spring Harb Perspect Biol* 2: a000752. doi: 10.1101/cshperspect.a000752.
- Hong RL, Hamaguchi L, Busch MA, Weigel D. 2003. Regulatory elements of the floral homeotic gene *AGAMOUS* identified by phylogenetic footprinting and shadowing. *Plant Cell* 15: 1296–1309.
- Houseley J, Tollervey D. 2009. The many pathways of RNA degradation. *Cell* 136: 763–776.
- Huang Y, Carmichael GG. 1997. The mouse histone H2a gene contains a small element that facilitates cytoplasmic accumulation

- of intronless gene transcripts and of unspliced HIV-1-related mRNAs. *Proc Natl Acad Sci* **94**: 10104–10109.
- Huang MTF, Gorman CM. 1990. Intervening sequences increase efficiency of RNA 3' processing and accumulation of cytoplasmic RNA. *Nucleic Acids Res* **18**: 937–947.
- Imataka H, Gradi A, Sonenberg N. 1998. A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. *EMBO J* **17**: 7480–7489.
- Jain M, Khurana P, Tyagi A, Khurana J. 2008. Genome-wide analysis of intronless genes in rice and *Arabidopsis*. *Funct Integr Genomics* **8**: 69–78.
- Jimeno S, Garcia-Rubio M, Luna R, Aguilera A. 2008. A reduction in RNA polymerase II initiation rate suppresses hyper-recombination and transcription-elongation impairment of *THO* mutants. *Mol Genet Genomics* **280**: 327–336.
- Jonsson JJ, Foresman MD, Wilson N, McIvor RS. 1992. Intron requirement for expression of the human purine nucleoside phosphorylase gene. *Nucleic Acids Res* **20**: 3191–3198.
- Kuhn J, Binder S. 2002. RT-PCR analysis of 5' to 3'-end-ligated mRNAs identifies the extremities of *cox2* transcripts in pea mitochondria. *Nucleic Acids Res* **30**: 439–446.
- Kyburz A, Friedlein A, Langen H, Keller W. 2006. Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing. *Mol Cell* **23**: 195–205.
- Le H, Tanguay RL, Balasta ML, Wei CC, Browning KS, Metz AM, Goss DJ, Gallie DR. 1997. Translation initiation factors eIF-iso4G and eIF-4B interact with the poly(A)-binding protein and increase its RNA binding activity. *J Biol Chem* **272**: 16247–16255.
- Le Hir H, Izaurralde E, Maquat L, Moore M. 2000. The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J* **19**: 6860–6869.
- Le Hir H, Gatfield D, Izaurralde E, Moore M. 2001. The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J* **20**: 4987–4997.
- Lin S, Coutinho-Mansfield G, Wang D, Pandit S, Fu XD. 2008. The splicing factor SC35 has an active role in transcriptional elongation. *Nat Struct Mol Biol* **15**: 819–826.
- Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C(T)}$ method. *Methods* **25**: 402–408.
- Lou H, Gagel RF, Berget SM. 1996. An intron enhancer recognized by splicing factors activates polyadenylation. *Genes Dev* **10**: 208–219.
- Lu S, Cullen BR. 2003. Analysis of the stimulatory effect of splicing on mRNA production and utilization in mammalian cells. *RNA* **9**: 618–630.
- Lu J, Sivamani E, Azhakanandam K, Samadder P, Li X, Qu R. 2008. Gene expression enhancement mediated by the 5' UTR intron of the rice *rubi3* gene varied remarkably among tissues in transgenic rice plants. *Mol Genet Genomics* **279**: 563–572.
- Maas C, Laufs J, Grant S, Korfhage C, Werr W. 1991. The combination of a novel stimulatory element in the first exon of the maize *Shrunken-1* gene with the following intron 1 enhances reporter gene expression up to 1000-fold. *Plant Mol Biol* **16**: 199–207.
- Malim MH, Hauber J, Le SY, Maizel JV, Cullen BR. 1989. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* **338**: 254–257.
- Mascarenhas D, Mettler IJ, Pierce DA, Lowe HW. 1990. Intron-mediated enhancement of heterologous gene expression in maize. *Plant Mol Biol* **15**: 913–920.
- Masle J, Gilmore SR, Farquhar GD. 2005. The *ERECTA* gene regulates plant transpiration efficiency in *Arabidopsis*. *Nature* **436**: 866–870.
- Millevoi S, Geraghty F, Idowu B, Tam J, Antoniou M, Vagner S. 2002. A novel function for the U2AF 65 splicing factor in promoting pre-mRNA 3'-end processing. *EMBO J* **3**: 869–874.
- Moore MJ, Proudfoot NJ. 2009. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136**: 688–700.
- Morello L, Giani S, Troina F, Breviaro D. 2011. Testing the IMET on rice introns and other aspects of intron-mediated enhancement of gene expression. *J Exp Bot* **62**: 533–544.
- Niwa M, Rose SD, Berget SM. 1990. In vitro polyadenylation is stimulated by the presence of an upstream intron. *Genes Dev* **4**: 1552–1559.
- Pillitteri LJ, Bemis SM, Shpak ED, Torii KU. 2007. Haploinsufficiency after successive loss of signaling reveals a role for *ERECTA*-family genes in *Arabidopsis* ovule development. *Development* **134**: 3099–3109.
- Proweller A, Butler JS. 1997. Ribosome concentration contributes to discrimination against poly(A)⁺ mRNA during translation initiation in *Saccharomyces cerevisiae*. *J Biol Chem* **272**: 6004–6010.
- Rigo F, Martinson HG. 2008. Functional coupling of last-intron splicing and 3'-end processing to transcription in vitro: the poly(A) signal couples to splicing before committing to cleavage. *Mol Cell Biol* **28**: 849–862.
- Rondon AG, Jimeno S, Aguilera A. 2010. The interface between transcription and mRNP export: From THO to THSC/TREX-2. *Biochim Biophys Acta* **1799**: 533–538.
- Rose AB. 2002. Requirements for intron-mediated enhancement of gene expression in *Arabidopsis*. *RNA* **8**: 1444–1453.
- Rose AB. 2004. The effect of intron location on intron-mediated enhancement of gene expression in *Arabidopsis*. *Plant J* **40**: 744–751.
- Rose AB. 2008. Intron-mediated regulation of gene expression. In *Nuclear pre-mRNA processing in plants* (ed. ASN Reddy, M Golovkin), pp. 277–290. Springer, Berlin-Heidelberg.
- Rose AB, Beliakoff JA. 2000. Intron-mediated enhancement of gene expression independent of unique intron sequences and splicing. *Plant Physiol* **122**: 535–542.
- Rose AB, Last RL. 1997. Introns act post-transcriptionally to increase expression of the *Arabidopsis thaliana* tryptophan pathway gene *PAT1*. *Plant J* **11**: 455–464.
- Rose AB, Elfersi T, Parra G, Korf I. 2008. Promoter-proximal introns in *Arabidopsis thaliana* are enriched in dispersed signals that elevate gene expression. *Plant Cell* **20**: 543–551.
- Salles FJ, Strickland S. 1999. Analysis of poly(A) tail lengths by PCR: the PAT assay. *Methods Mol Biol* **118**: 441–448.
- Sanfaçon H, Brodmann P, Hohn T. 1991. A dissection of the cauliflower mosaic virus polyadenylation signal. *Genes Dev* **5**: 141–149.
- Shiu SH, Bleecker AB. 2003. Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in *Arabidopsis*. *Plant Physiol* **132**: 530–543.
- Shpak ED, Lakeman MB, Torii KU. 2003. Dominant-negative receptor uncovers redundancy in the *Arabidopsis* *ERECTA* Leucine-rich repeat receptor-like kinase signaling pathway that regulates organ shape. *Plant Cell* **15**: 1095–1110.
- Shpak ED, Berthiaume CT, Hill EJ, Torii KU. 2004. Synergistic interaction of three *ERECTA*-family receptor-like kinases controls *Arabidopsis* organ growth and flower development by promoting cell proliferation. *Development* **131**: 1491–1501.
- Shpak ED, McAbee JM, Pillitteri LJ, Torii KU. 2005. Stomatal patterning and differentiation by synergistic interactions of receptor kinases. *Science* **309**: 290–293.
- Sieburth LE, Meyerowitz EM. 1997. Molecular dissection of the *AGAMOUS* control region shows that *cis* elements for spatial regulation are located intragenically. *Plant Cell* **9**: 355–365.
- Song SK, Lee MM, Clark SE. 2006. POL and PLL1 phosphatases are CLAVATA1 signaling intermediates required for *Arabidopsis* shoot and floral stem cells. *Development* **133**: 4691–4698.
- Subramanian C, Woo J, Cai X, Xu X, Servick S, Johnson CH, Nebenfuhr A, von Arnim AG. 2006. A suite of tools and application notes for *in vivo* protein interaction assays using bioluminescence resonance energy transfer (BRET). *Plant J* **48**: 138–152.

- Tange TO, Nott A, Moore MJ. 2004. The ever-increasing complexities of the exon junction complex. *Curr Opin Cell Biol* **16**: 279–284.
- Tange TO, Shibuya T, Jurica MS, Moore MJ. 2005. Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. *RNA* **11**: 1869–1883.
- Torii KU, Mitsukawa N, Oosumi T, Matsuura Y, Yokoyama R, Whittier RF, Komeda Y. 1996. The Arabidopsis *ERECTA* gene encodes a putative receptor protein kinase with extracellular Leucine-rich repeats. *Plant Cell* **8**: 735–746.
- Wang H, Lee MM, Schiefelbein JW. 2002. Regulation of the Cell Expansion Gene *RHD3* during Arabidopsis Development. *Plant Physiol* **129**: 638–649.
- Yokoyama R, Takahashi T, Kato A, Torii KU, Komeda Y. 1998. The Arabidopsis *ERECTA* gene is expressed in the shoot apical meristem and organ primordia. *Plant J* **15**: 301–310.
- Zhou Z, Luo MJ, Straesser K, Katahira J, Hurt E, Reed R. 2000. The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature* **407**: 401–405.

RNA

A PUBLICATION OF THE RNA SOCIETY

The presence of multiple introns is essential for *ERECTA* expression in *Arabidopsis*

Rucha Karve, Wusheng Liu, Spencer G. Willet, et al.

RNA 2011 17: 1907-1921 originally published online August 31, 2011

Access the most recent version at doi:[10.1261/rna.2825811](https://doi.org/10.1261/rna.2825811)

Supplemental Material

<http://rnajournal.cshlp.org/content/suppl/2011/08/10/rna.2825811.DC1>

References

This article cites 69 articles, 34 of which can be accessed free at:
<http://rnajournal.cshlp.org/content/17/10/1907.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *RNA* go to:
<http://rnajournal.cshlp.org/subscriptions>