Genome-Wide Analysis of mRNA Decay Rates and Their Determinants in Arabidopsis thaliana

Reena Narsai,a Katharine A. Howell,a A. Harvey Millar,a Nicholas O’Toole,a,b Ian Small,a,b and James Whelan*a,1

a Australian Research Council Centre of Excellence for Plant Energy Biology, University of Western Australia, Crawley 6009, Australia
b Centre of Excellence for Computational Systems Biology, University of Western Australia, Crawley 6009, Australia

To gain a global view of mRNA decay in Arabidopsis thaliana, suspension cell cultures were treated with a transcriptional inhibitor, and microarrays were used to measure transcript abundance over time. The deduced mRNA half-lives varied widely, from minutes to >24 h. Three features of the transcript displayed a correlation with decay rates: (1) genes possessing at least one intron produce mRNA transcripts significantly more stable than those of intronless genes, and this was not related to overall length, sequence composition, or number of introns; (2) various sequence elements in the 3’ untranslated region are enriched among short- and long-lived transcripts, and their multiple occurrence suggests combinatorial control of transcript decay; and (3) transcripts that are microRNA targets generally have short half-lives. The decay rate of transcripts correlated with subcellular localization and function of the encoded proteins. Analysis of transcript decay rates for genes encoding orthologous proteins between Arabidopsis, yeast, and humans indicated that yeast and humans had a higher percentage of transcripts with shorter half-lives and that the relative stability of transcripts from genes encoding proteins involved in cell cycle, transcription, translation, and energy metabolism is conserved. Comparison of decay rates with changes in transcript abundance under a variety of abiotic stresses reveal that a set of transcription factors are downregulated with similar kinetics to decay rates, suggesting that inhibition of their transcription is an important early response to abiotic stress.

INTRODUCTION

When the existence of an mRNA-type molecule was proposed, it was immediately recognized that the stability of this intermediate could be an important point in the control of gene expression (Jacob and Monod, 1961). Transcript abundance is the equilibrium between the rate of mRNA synthesis, catalyzed by polymerases, and the rate of degradation, catalyzed by nucleases (Yang et al., 2003). Although it is widely accepted that the rate of synthesis and degradation determine transcript abundance, the majority of studies only measure steady state transcript levels, largely due to the technical ease with which they are determined, compared with measuring the rates of synthesis or degradation (Abier and Green, 1996; Shu and Hong-Hui, 2004). Changes in transcript abundance are often assumed to be due to changes in the rate of transcription; thus, regulation by changes in promoter activity is inferred. Significant research effort is invested in determining the factors that are responsible for the proposed coregulation of genes that display similar patterns of transcript abundance (Pilpel et al., 2001; Zhu et al., 2002). This type of approach has produced significant outcomes in identifying regulatory networks, such as those associated with developmental processes in animals (Ochoa-Espinosa and Small, 2006), cell cycle and rRNA transcription and processing in yeast (Sudarsanam et al., 2002; Bahler, 2005), and light regulation in plants (Casal and Yanovsky, 2005). However, an understanding of transcript decay rates provides further insight into the global regulation of gene expression and may reveal previously unknown coregulatory gene networks.

Studies examining mRNA decay rates in mammalian or yeast cells using chemical or genetic means to inhibit transcription have shown that the decay rate of mRNAs is correlated with the function of the encoded protein (Wang et al., 2002; Yang et al., 2003). It was observed in both organisms that transcripts encoding proteins involved in central metabolism had longer half-lives compared with those involved in regulatory processes. It was also reported that the median half-life correlated with the length of the cell cycle of Escherichia coli, yeast, and human HepG2/Bud8 cells, ranging from 5 to 21 to 600 min, respectively (Yang et al., 2003). In-depth analysis of mRNA decay rates in yeast also revealed that changes in transcript abundance after inhibiting transcription were similar to those following heat shock and that control of ribosome biogenesis factors was posttranscriptional (Grigull et al., 2004). These studies confirm that knowledge of transcript decay rates can yield insights into diverse biological processes that are not evident by measuring transcript abundance alone.

Approximately 100 transcripts have been defined as being rapidly degraded in Arabidopsis thaliana with half-lives of <60 min (Gutierrez et al., 2002). As was found in yeast and human studies, these unstable transcripts were enriched in genes coding for regulatory functions. Additional studies in Arabidopsis revealed that transcripts encoding proteins associated with circadian control were unstable and that this was associated with the.

1 Address correspondence to seamus@cyllene.uwa.edu.au.
"Online version contains Web-only data. www.plantcell.org/cgi/doi/10.1105/tpc.107.055046
presence of instability elements, such as the destabilizing downstream (DST) element in the 3' untranslated region (UTR; Newman et al., 1993; Gutierrez et al., 2002; Lidder et al., 2005). Experiments involving a mutant defective in DST-mediated decay revealed that this element was involved in the changes in transcript half-life which in turn affected circadian regulation at a whole-plant level, indicating that plant circadian rhythms are in part regulated by a sequence-specific mRNA degradation pathway (Lidder et al., 2005). To gain a genome-wide perspective on transcript instability in plant systems, this study makes use of the availability of microarrays and the improved annotation status of the Arabidopsis genome to investigate mRNA decay rates on a whole-genome level in Arabidopsis cell culture.

RESULTS

Actinomycin D Inhibits Transcription in Arabidopsis Cell Culture

To determine the decay rates of transcripts in Arabidopsis, the transcriptional inhibitor Actinomycin D (ActD) was used on Arabidopsis suspension cell cultures. To confirm that transcription was efficiently inhibited, we tested the ability of ActD to prevent the induction of transcripts upon addition of salicylic acid, known to result in an increase in transcript abundance of specific transcripts, such as NDB2, which encodes a mitochondrial external NADH dehydrogenase (Clifton et al., 2005, 2006). A 1-h pretreatment of cell cultures with ActD prevented the rapid four- to fivefold induction of NDB2 transcript abundance observed upon addition of salicylic acid (see Supplemental Figure 1A online). Analysis of total RNA by agarose gel electrophoresis before ActD treatment (0 h) and 1, 3, 6, 12, 24, and 48 h after ActD treatment indicated that the RNA bands were intact up to 24 h; however, degradation of the rRNA bands was evident after 48 h (see Supplemental Figure 1B online). Furthermore, the cRNA quality was assessed up to 24 h using the Agilent Bioanalyzer, and the 3′/5′ Actin ratios were found to be <2.5 up to 24 h. Typically, 3′/5′ ratios of <3.0 indicate that the cRNA is not degraded and is of sufficient quality to proceed with analysis of abundance by microarrays. Also, the yield of total RNA per wet cell weight at each time point, following ActD treatment, was not significantly different up to 24 h (data not shown). Cell viability was also examined, and it was indicated that >90% cell viability was maintained up to 24 h following ActD treatment, whereas at 48 h, cell viability was reduced to 60 to 70% (data not shown). Thus, we assessed the decay rates of transcripts over a 24-h period of ActD treatment, when cells were viable, and total RNA was intact and good quality cRNA could be generated. It is important to consider, however, that in any transcription inhibition experiment, specific cellular functions may be hindered at the later time points.

Genome-Wide View of mRNA Decay

Effective transcriptional inhibition by ActD was further confirmed by the observed decrease in cRNA synthesized from total RNA during the microarray target preparation process (see Supplemental Figure 1C online). However, this presented a challenge given that standard normalization procedures for analysis of microarray data attempt to compensate for differences in total mRNA levels, dampening real changes in transcript levels due to transcriptional inhibition. For this reason, it was necessary to modify data normalization as outlined in the Methods. This preliminary analysis produced a genome normalized data set consisting of mRNA decay profiles of 13,012 genes, referred to from this point on, as the whole-genome set. The microarray data were validated by comparison to transcript profiles generated by quantitative RT-PCR (QRT-PCR), and a Pearson correlation coefficient was calculated for each of the 72 transcripts that were analyzed by both methods (Figure 1A; see Supplemental Table 1 online). These 72 transcripts were chosen as they reflected a range of transcript characteristics, including differing base composition, transcript lengths, starting quantities, and intron numbers, and the encoded proteins also represented a range of functional catalogues (FunCats). An average correlation of 0.95 confirmed that the normalization method employed was valid. Furthermore, mock-treated cells, which were not analyzed using microarrays, were examined by QRT-PCR and showed that transcript abundance either remained stable or increased and then stabilized to be similar to pretreatment levels (Figure 1A). The 13,012 transcript decay profiles generated from microarray analysis of the three independent time courses were regressed using SAS (version 9.1). Nonlinear least-squares regression was used to fit the decay profiles to an exponential decay model and the regression parameters, including the decay rate (k), the SE of the decay rate, the 95% confidence interval, and the mRNA half-life (t1/2), were calculated for each of the transcripts analyzed (see Supplemental Table 2 online). This method effectively regressed decay profiles that varied widely, with the abundance of some transcripts seen to decrease rapidly after ActD treatment (e.g., WRKY6 transcription factor; At1g62300; t1/2 = 1.13 h), while some transcripts decreased more slowly (e.g., hydroxypyruvate reductase; At1g68010; t1/2 = 14.75 h; Figure 1B). This method enabled comparison between transcripts despite differences in decay rates, and in most cases the transcript abundance and regression models tended to come down to an asymptote by 24 h. However, as it cannot be ruled out that inhibiting transcription may differently affect the decay rates of transcripts, the half-lives of long-lived transcripts may be overestimated; thus, the differences in decay rates need to be treated in a relative manner.

The mRNA half-lives varied greatly across the genome, ranging from 0.2 to 24 h, with a mean of 5.9 h and median of 3.8 h (see Supplemental Table 3 online). Hierarchical clustering of transcript abundance of the whole-genome set allowed visualization of the variation in decay rates, with differences in the transition from red to green on the heat map indicating different rates of decay (Figure 2). After just 1 h, it was evident that some transcripts had decreased significantly as seen by an early transition from red to green at this time point (top of heat map, Figure 2). By contrast, some transcripts represented at the bottom of the heat map were still red at 12 h, indicating slow decay rates. Even at 24 h, some transcripts were still quite abundant, as evidenced by the fact that at 24 h, visualization on the heat map was not a uniform green color, as would be the case if all transcript decay had occurred at the same rate.
Relationship of mRNA Decay Rates with Structural and Regulatory Gene Features

Studies on individual genes have concluded that mRNA decay rates may be affected by features within the mRNA sequence itself, such as intron number/presence or microRNA (miRNA) binding sites (Newman et al., 1993; Ross, 1995; Wang et al., 2007). Subsets of transcripts were generated based on these features, and to visualize any differences between the mRNA half-lives in each of the subsets of genes compared with the genome, cumulative distribution plots were generated. This displayed the additive percentage frequency of transcripts over a range of mRNA half-lives, and the Kolmogorov-Smirnov statistic was used to determine if differences in distributions were significant (Davis, 1986). This method calculates the maximum distance difference between the two distribution data sets (D value) and then calculates a P value, $p(same)$, for that distance that indicates the probability of two data sets being distributed in the same manner. A stringent P value of 1.0E-05 or less was defined as significant.

mRNA binding factors, such as interfering RNAs, including miRNAs, are known to be involved in influencing mRNA stability. Studies in eukaryotes have revealed that the interfering RNAs work by binding specific targets and preventing full translation and ultimately lead to the degradation of the target transcript (Reinhart et al., 2002). To determine the influence of miRNA targeting on mRNA decay rates, a list of miRNA targets were generated based on several miRNA databases (see Methods and Supplemental Table 2 online), and their mRNA half-lives were examined. It was evident that there were a significantly higher proportion of miRNA target transcripts with shorter half-lives, compared with the whole genome set, with >50% of these transcripts having half-lives of <3 h (Figure 3A).

Another factor thought to be involved in influencing mRNA stability is the presence of introns in the transcript. There appeared to be no correlation between mRNA half-life and the number of introns present in each transcript (see Supplemental Figure 2A online). However, upon closer examination, it was clear that for intronless genes (no introns in the .1 model) there were a

(A) Transcript abundance of some genes analyzed by both QRT-PCR and microarray analysis with the calculated correlation coefficient given for each transcript. A value of 1.0 was assigned to the level of transcript prior to treatment, and the transcript abundance at all other time points was expressed relative to this value. Each data point represents the mean ± SE. The right panel displays transcript abundance of the same genes analyzed by QRT-PCR following mock treatment with ethanol.

(B) Examples of mRNA decay profiles (using the microarray data) and the nonlinear least-squares regression curve produced for each gene. Data were regressed using the relative values from all three biological replicates over all the time points using SAS, which allowed the rate of decay for each gene to be determined. The blue line represents the actual data points, and the red line represents the fitted exponential model.

WRKY6, WRKY6 transcription factor (At1g62300); OM64, outer mitochondrial membrane protein of 64 kD (At5g09420); MDHAR, monodehydroascorbate reductase (At1g63940); UCP, uncoupling protein (At3g54110); HPR, hydroxypyruvate reductase (At1g68010); TIM14, translocase of the inner membrane protein 14 (At2g35795).
significantly greater proportion of transcripts with shorter half-lives compared with the genome \( (P = 1.81 \times 10^{-66}) \), with >60% having half-lives of <3 h (Figure 3B). To confirm that stability could be attributed to the presence of introns and not the additional length of the transcript, the lengths of the 3' UTR, the coding sequence, and the 5' UTR were all calculated, and total mRNA length was compared with the mRNA half-lives, revealing no correlation (see Supplemental Figure 2B online) and thus confirming that the presence of introns has an independent positive effect on mRNA stability.

Considering that the absence of introns appears to have a pronounced effect on mRNA half-life, it was important to consider the effect of alternative splicing on mRNA decay. To address this, The Arabidopsis Information Resource (TAIR) 7.0 database was used to generate a list of genes that contain introns in the first TAIR gene model and all of its splice variants (always intron containing) and a list of genes for which the first TAIR gene models contained at least one intron, but at least one subsequent splice variant was intronless (intronless splice variants). There are only 91 genes that encoded intronless splice variants in this study, and it is important to note that apart from this analysis, these transcripts have generally been designated as containing introns due to the fact that their first gene model contains at least one intron. Interestingly, there was a greater proportion of transcripts with shorter half-lives for genes containing intronless splice variants among other (2, .3, etc.) gene models compared with the genes without any intronless splice variants (always intron containing); however, this was not statistically significant (Figure 3B).

Previous studies have suggested that the length of the 3' UTR or the presence of an intron in the 3' UTR contributes to mRNA stability of transcripts by acting as cis-elements (Kertesz et al., 2006). This was examined here, and it was observed that transcripts with 3' UTRs longer than 300 bp showed no significant difference in the distribution of mRNA half-lives of these transcripts compared with the genome (Figure 3C). Similarly, there did not appear to be a significant difference between the distribution of mRNA half-lives in the genome compared with (1) genes containing an intron in the 3' UTR, (2) genes that contained at least one 3' UTR intron retention splice variant with respect to the first gene model (56 genes), and (3) genes containing at least one 3' UTR intron, but no 3' UTR intron retention splice variant genes (Figure 3C).

It is widely accepted that transcript abundance in the cell is a function of the rate of decay and the rate of synthesis. Various studies have examined the kinetics of this relationship and have suggested that there is a proportional relationship between the rate of decay and the concentration (Hargrove, 1993; Yugi et al., 2005). It has been suggested that as half-life of a transcript increases, the steady state concentration increases as it becomes a higher multiple of the synthesis rate (Hargrove, 1993). Looking at 0 h signal intensity as a representation of starting quantity of mRNA abundance, it was observed that transcripts that had lower starting signal intensities (<100) had a significantly higher proportion of transcripts with shorter half-lives \( (P = 5.20 \times 10^{-49}) \), while transcripts with high starting signal intensities (>1000) had a significantly higher proportion of transcripts with longer half-lives \( (P = 1.45 \times 10^{-156}) \), and transcripts with signal intensities between 100 and 1000 showed no significant difference compared with the genome (Figure 3D). This suggests that transcripts with longer half-lives are more abundant, possibly due to their longer half-lives, while transcripts with shorter half-lives are likely to be less abundant due to their greater rate of decay. However, it is important to observe this data with caution as the accuracy of the relationship between signal intensity and actual mRNA concentration is not a direct relationship, although a comparison of microarray and TaqMan gene expression assays concluded that microarray platforms could be used for quantitative measurement of gene expression (Canales et al., 2006).
Figure 3. The Distribution of mRNA Half-Lives of the Whole Genome Compared with Subsets of Transcripts with Specific Sequence Features.
It is established that cis-acting sequence elements within the 3’ UTR of transcripts can influence mRNA stability (Garneau et al., 2007). This has been observed particularly for AU-rich elements where the presence of these motifs is associated with highly unstable transcripts (Ross, 1995; Abler and Green, 1996; Yang et al., 2003). To determine whether the AU content of transcripts affect mRNA decay rates, the AU content within the 3’ UTR, coding sequence, 5’ UTR, and whole mRNA was calculated (see Supplemental Table 2 online). However, no significant correlation was found (see Supplemental Figure 2C online). To investigate if any sequence elements were enriched in fast- or slow-decaying transcripts, we searched for the occurrence of all possible 6-mers in a set of transcripts that were the 1000 fastest decaying (half-life ~1 h or less) or the 1000 slowest decaying (half-life ~12 h or greater). Among the 4096 possible 6-mers, 167 were found to be significantly enriched at the P < 0.01 level in the 3’ UTR of the fast-decaying transcripts, and 283 were found to be enriched within the 3’ UTR of slow-decaying transcripts (see Supplemental Table 5 online). As a control, nonoverlapping sets of 1000 unique, randomly selected transcript 3’UTRs were generated, and an identical analysis was performed to test for enrichment or depletion of 6-mers in one set with respect to the other. This test was performed 20 times. In these trials, an average of 37 motifs (SD of 6) was significantly enriched at the P < 0.01 level, compared with the 450 found enriched in either the fast-decaying transcript set with respect to the slow or vice versa. Using a more strict cutoff for significance of P < 0.001 (|Z| > 3.89), we found 117 motifs significantly enriched in the fast-decaying transcript set with respect to the slow or vice versa. By contrast, among the 20 trials of random transcript sets, 13 did not produce any enriched motifs at this significance level. Five trials detected one enriched motif, and two trials detected two enriched motifs at this significance level. All data on the enrichment or depletion of 6-mers are available in Supplemental Table 5 online, including the proportions of occurrences of the motifs in all transcripts in this study for which there is 3’ UTR sequence data. As expected, the proportions of those motifs significantly enriched or depleted with respect to fast- or slow-decaying transcripts among all genes are between those of the fast and slow transcript data sets, whereas the proportions of the entire transcript set are similar to those of the fast and slow-decaying transcripts for those motifs not significantly enriched or depleted in either set.

The top 10 elements ranked by significance as determined by the Z-score for each category are shown in Table 1. For the elements defined to be enriched in the fast-decaying transcripts, elements ranked 3 (Table 1), 38, 45, and 103 (see Supplemental Table 5 online) have been previously defined as instability elements. It is likely that the elements defined display degeneracy, such as elements 2, 4, 7, and 9, which may all constitute a core element that can tolerate some base variation (Table 1). However, even taking this into account, there are several distinct elements that are enriched in both groups, suggesting a number of different binding factors (Table 1; see Supplemental Table 5 online). The occurrence of the 50 most significant putative destabilization and stabilization motifs present in the 3’ UTRs of the fastest and slowest decaying transcripts reveals that transcripts contain several different motifs, with any single motif often present more than once (see Supplemental Table 5 online).

Another important factor worth mentioning is the presence of elements within 5’ UTR of transcripts. To address this, we calculated the enrichment or depletion of 6-mers, including the proportions of occurrences of the motifs in all transcripts in this study for which there is 5’ UTR sequence data (see Supplemental Table 5 online). Interestingly, some of the AU-rich motifs in the 3’ UTR analysis are also found among the top of the list, enriched...

**Figure 3.** (continued).

(A) to (C) The proportion of transcripts in each mRNA half-life category for the whole-genome set and for subsets, including transcripts that are (A) known miRNA targets; (B) contain no introns (light blue), has one or more splice variant/s that does not have an intron (black), or has a one or more splice variants that contain an intron in all the splice variants (i.e., always intron containing [red]) (C); contain a 3’ UTR >300 bp, contain a 3’ UTR intron, has one or more splice variants that contain a retained 3’ UTR, or has one or more splice variants that never contains a 3’ UTR intron in all the splice variants. (D) Transcripts that had 0 h signal intensities of <100, >100 to < 1000, and >1000. The cumulative percentage plots graphically display how the mRNA half-lives are distributed in each category. The D-value represents the distance between the distributions and was used to calculate the Kolmogorov-Smirnov statistic [p(same)] to determine if there was a statistically significant difference (asterisk) between each subset compared with the genome. The number of genes in each data set is indicated in parentheses.

**Table 1.** Putative Destabilization/Stabilization Motifs

<table>
<thead>
<tr>
<th>Number</th>
<th>Motif</th>
<th>Fast (%)</th>
<th>Slow (%)</th>
<th>Z-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AATT TT</td>
<td>41.1</td>
<td>20.7</td>
<td>9.9</td>
</tr>
<tr>
<td>2</td>
<td>TTTT TG</td>
<td>50.8</td>
<td>31.8</td>
<td>8.6</td>
</tr>
<tr>
<td>3</td>
<td>TTTT TT</td>
<td>48.1</td>
<td>31.1</td>
<td>8.2</td>
</tr>
<tr>
<td>4</td>
<td>AGTT TT</td>
<td>37.7</td>
<td>21.1</td>
<td>8.2</td>
</tr>
<tr>
<td>5</td>
<td>TTTTG T</td>
<td>58.7</td>
<td>43.4</td>
<td>6.8</td>
</tr>
<tr>
<td>6</td>
<td>AT TT TG</td>
<td>39.3</td>
<td>25.3</td>
<td>6.7</td>
</tr>
<tr>
<td>7</td>
<td>ATT TT T</td>
<td>43.1</td>
<td>28.8</td>
<td>6.7</td>
</tr>
<tr>
<td>8</td>
<td>GTTTT T G</td>
<td>36.2</td>
<td>22.7</td>
<td>6.6</td>
</tr>
<tr>
<td>9</td>
<td>GTTTT T T</td>
<td>43.4</td>
<td>31.2</td>
<td>5.6</td>
</tr>
<tr>
<td>10</td>
<td>TAGG AT T</td>
<td>7.7</td>
<td>2.5</td>
<td>5.3</td>
</tr>
<tr>
<td>4087</td>
<td>TT GCT T</td>
<td>14.6</td>
<td>24.0</td>
<td>-5.3</td>
</tr>
<tr>
<td>4088</td>
<td>TG TGT G</td>
<td>11.1</td>
<td>19.7</td>
<td>-5.3</td>
</tr>
<tr>
<td>4089</td>
<td>TG CTT T</td>
<td>12.3</td>
<td>21.2</td>
<td>-5.3</td>
</tr>
<tr>
<td>4090</td>
<td>AA TA AG</td>
<td>7.9</td>
<td>15.8</td>
<td>-5.5</td>
</tr>
<tr>
<td>4091</td>
<td>AT CTC T</td>
<td>11.1</td>
<td>19.9</td>
<td>-5.5</td>
</tr>
<tr>
<td>4092</td>
<td>AT TT GG</td>
<td>11.5</td>
<td>20.8</td>
<td>-5.7</td>
</tr>
<tr>
<td>4093</td>
<td>TT AT CT</td>
<td>13.8</td>
<td>23.7</td>
<td>-5.7</td>
</tr>
<tr>
<td>4094</td>
<td>AA TA AA</td>
<td>14.9</td>
<td>25.4</td>
<td>-5.9</td>
</tr>
<tr>
<td>4095</td>
<td>GA TA A</td>
<td>7.5</td>
<td>16.0</td>
<td>-5.9</td>
</tr>
<tr>
<td>4096</td>
<td>TC TT TT</td>
<td>17.9</td>
<td>30.2</td>
<td>-6.4</td>
</tr>
</tbody>
</table>

The 10 most significant putative destabilization (t1/2 <1.148 h) and stabilization (t1/2 >12.487 h) 6-mers in the 3’ UTRs (numbered 1 to 10 and 4087 to 4096, respectively), the percentage of fast/slow-decaying transcripts that contain the motif, and the respective Z-scores for each 6-mer. An asterisk next to the motif indicates that it has been previously identified as involved in mRNA stability.
for fast-decaying transcripts. We also searched for a known 5’ UTR mRNA stability element present in yeast [poly(G) sequence] (Muhlrad et al., 1995), although there was no correlation between mRNA half-life and this motif. However, it is important to note that accurate statistical treatment of this data is made difficult by the short and probably truncated 5’ UTR sequence data used for many transcripts.

**Relationship of mRNA Decay with Function and Location of the Encoded Proteins**

Previous studies in eukaryotes revealed that transcripts that encoded proteins involved in highly regulated processes decayed relatively in contrast with those involved in central metabolic functions that were observed to decay relatively slowly (Ross, 1995; Wang et al., 2002; Yang et al., 2003). To determine whether this general relationship existed for Arabidopsis, a broad annotation, such as the FunCat (Ruepp et al., 2004), was employed for the annotation of genes based on the function of the encoded proteins. This was clearly not feasible for all genes; however, using various databases and manual annotation methods, 63% of the 13,012 genes were annotated into 26 FunCats (see Supplemental Table 4 online), allowing statistical analysis of the half-lives of the transcripts in each group to be determined (see Supplemental Table 3 online).

The distribution of the whole-genome mRNA half-lives was found to be significantly different from transcripts classified in metabolism (P = 1.3E-18), energy (P = 7.1E-31), protein synthesis (P = 5.7E-29), and subcellular localization (P = 4.8E-08) categories (Figure 4A). This is particularly evident for the transcripts classified into energy and protein synthesis functions where a much higher percentage of transcripts have longer half-lives compared with the whole genome (Figure 4A). By contrast, some FunCats showed a greater distribution of transcripts with shorter half-lives, including transcripts classified in transcription (P = 4.9E-21) and communication (P = 9.1E-11) categories (Figure 4A; see Supplemental Table 3 online). These results indicated that, in Arabidopsis, there are differences in the distribution of half-lives of transcripts in specific FunCats compared with the whole genome, and specific groups of transcripts merited further investigation.

Sets of genes encoding transcription factors, pentatricopeptide repeat (PPR) proteins, kinases, and ribosomal proteins were also compared. Transcripts that encode transcription factors, PPR proteins, and kinases had a significantly higher proportion of transcripts with short half-lives compared with the genome (Figure 4B). By contrast, transcripts encoding ribosomal proteins had a significantly higher proportion of transcripts with longer half-lives, with almost 30% having half-lives longer than 12 h in contrast with the genome where only 9% had half-lives of 12 h or more, consistent with the fact that they are involved in the basic cellular process of protein synthesis and have little regulatory function (Figure 4B).

It has been established that there is a relationship between mRNA half-life and function of the encoded protein, but to determine if location of the protein product has any relationship with transcript half-life, we analyzed sets of transcripts on the basis of experimental localization of their translated products. It was found that transcripts encoding mitochondrial, chloroplast, and peroxisomal proteins have a significantly higher proportion of transcripts with longer half-lives (Figure 4C). This may be largely due to the fact that many of the proteins known to be located in these organelles are associated with metabolism and energy. However, exceptions to this clearly apply given that some transcripts encoding organellar proteins have short half-lives, as pointed out above for transcripts encoding PPR proteins (Figure 4B). Similarly, it was found that transcripts encoding the proteins in the major membrane systems of the cell, the vacuolar tonoplast, and plasma membrane, as well as the major metabolic machinery in the cytosol, had a significantly higher proportion of transcripts with longer half-lives with 55, 42, and 64% of transcripts having half-lives longer than 6 h compared with the genome (31%; Figure 4C). By contrast, transcripts encoding nuclear proteins were found to have a higher proportion of transcripts with shorter half-lives; 48% were found to have half-lives <3 h compared with the genome (40%) or to the average of the other organelle and membrane systems (25%; Figure 4C).

**Analysis of mRNA Decay in Arabidopsis Orthologs**

Given that large data sets of yeast and human mRNA decay profiles already exist (Wang et al., 2002; Yang et al., 2003), a comparison of these with the data set derived from this study was performed. Summary statistics, including minimum, maximum, and median half-life, as well as the cell cycle length (CCL) were compared for all three species (Figure 5A). Previous mRNA decay studies have observed a roughly proportional relationship between the median mRNA half-life and the CCL, where the median half-life appears to be 20 to 25% of the CCL. This has been found for a range of organisms, including E. coli (median half-life = 5 min, CCL = 20 min; Bernstein et al., 2002), yeast (median half-life = 20 min, CCL = 90 min; (Wang et al., 2002), and human HepG2/Bud8 cells (median half-life = 10 h, CCL = 50 h; Yang et al., 2003). Interestingly, this was also observed for Arabidopsis, where the median half-life was found to be 3.8 h and CCL, for cell culture derived from the same source, was previously reported to be ~19 h (Menges and Murray, 2002).

The InParanoid database was used to obtain a list of orthologous genes between Arabidopsis and human as well as Arabidopsis and yeast (Remm et al., 2001; O’Brien et al., 2005). Filtration methods were then used to generate a list of 1830 Arabidopsis genes for which there are orthologs in humans (712), yeast (1407), or both (289; see Supplemental Table 6 online). To compare across the three species, the mRNA half-lives were calculated relative to the CCL, and the frequencies were illustrated as cumulative distribution plots (Figure 5B). The Kolmogorov-Smirnov statistic was then used to determine differences in the distribution of mRNA half-lives across the species, and the distribution of the Arabidopsis mRNA half-lives was found to be significantly different to the distribution of mRNA half-lives in the orthologs in yeast (P = 5.39E-30) and humans (P = 5.96E-09; Figure 5B). It was also evident that the mRNA half-life distributions of Arabidopsis and human (D value = 14%) were more closely matched to each other than with yeast (D value = 20%; Figure 5B). Although, there appears to be a higher proportion of transcripts with very short half-lives (<10% CCL) in
Figure 4. The Distribution of mRNA Half-Lives of the Whole Genome Compared with Subsets of Transcripts Encoding Proteins within Specific Functional and Localization Categories.
humans and Arabidopsis compared with yeast, this was not statistically significant (Figure 5B). However, there was a significantly higher percentage of transcripts with relatively short half-lives in humans (D = 14%) and yeast (D = 20%), where 51 and 74% of the transcripts had half-lives ≤20 and ≤30% of the length of the CCL, respectively, compared with 37 and 54% in Arabidopsis, respectively (Figure 5B). However, for the human mRNA half-lives, there was also a higher percentage of transcripts (~6%) with very long mRNA half-lives of >50 h (>100% CCL) compared with Arabidopsis, where <2% of transcripts had mRNA half-lives longer than the CCL (Figure 5B).

After consideration of the number of transcripts analyzed, the CCL and the minimum, maximum, and median transcript half-lives for each of the genome datasets, four transcript decay rate categories were formed based on the CCL, and these were defined as “fast” – t1/2 ≤ 10% of the CCL, “moderate1” – t1/2 10 to 22% of the CCL, “moderate2” – t1/2 22 to 50% of the CCL, and “slow” – t1/2 ≥ 50% of the CCL (Figure 5A). Analysis was then performed to compare how many transcripts had half-lives that fell into these decay rate categories across Arabidopsis versus yeast and Arabidopsis versus human data sets. From all the orthologous genes analyzed, 627 were found to fall in the same decay rate category (Figure 5C; see Supplemental Table 6 online). Of the 79 categorized as “fast”, almost 50% were involved in nucleic acid binding, cell cycle, transcription, or RNA processing compared with 25, 16, and 3% for the “moderate 1”, “moderate 2”, and “slow” categories, respectively.

---

**Figure 5.** Comparison of the Half-Lives of Transcripts of Orthologous Genes among Arabidopsis, Human, and Yeast.

**A** General features of all three mRNA decay studies, including summary statistics for each data set.

**B** The distribution of mRNA half-lives across all three species. As there were several transcripts with short mRNA half-lives in all three species, there is a scale break where the scale from 0 to 0.1 is smaller (0.02).

**C** A comparison of the 627 transcripts classified as “fast”, “moderate 1,” “moderate 2,” or “slow” in yeast and/or humans compared with the respective Arabidopsis orthologous transcripts in the same categories. The number of transcripts in each category is indicated in red, and the number of genes in each category was expressed as a percentage to allow comparison of the proportion of transcripts encoding proteins in different FunCats.

**Figure 4.** (continued).

**A** Distribution of the half-lives of transcripts for genes grouped into various functional categories.

**B** Distribution of the half-lives of transcripts for genes encoding transcription factors, PPR proteins, kinases, and ribosomal proteins.

**C** Distribution of the half-lives of transcripts for genes encoding proteins targeted to various subcellular locations. The number of genes in each data set is indicated in parentheses. Significant differences (asterisks) in distributions when compared with the whole-genome set were determined as outlined in Figure 3.
“moderate 2,” and “slow” categories, respectively (Figure 5C). By contrast, only 5% of the “fast” transcripts were found to be involved in metabolism or energy, while 33% of the transcripts categorized as “slow” fell into this FunCat and included components of glycolysis and the trichloroacetic acid cycle. For the protein synthesis FunCat, the percentage in the “fast” through “slow” groupings increased from 6 to 24% (Figure 5C). Thus, it appears that transcripts encoding orthologous proteins involved in cell cycle, transcription, translation, and energy metabolism have relatively conserved rates of mRNA decay across diverse species.

A Comparison of Transcript Decay Profiles and the Transcriptional Stress Response

A previous study in yeast revealed there was similarity between mRNA stability profiles and transcript profiles under heat stress, suggesting that one of the effects of heat stress in yeast was an inhibition of transcription of a subset of genes (Grigull et al., 2004). To determine if any stress treatment in Arabidopsis may be acting to inhibit transcription, microarray experiments using shoot tissue and involving stress treatments over a 24-h time period (Kilian et al., 2007) with time points identical to those used in this study were examined to see if patterns of transcript change correlated with those observed with the inhibition of transcription by ActD. An overview of the comparisons performed to identify genes that contained similar patterns of change is outlined in Supplemental Figure 3 online. Using self-organizing map (SOM) analysis (Tamayo et al., 1999), 8878 transcript profiles were grouped into five clusters that showed variation over time, from relatively stable levels (Clusters 1 and 2) to rapid decreases (Cluster 5) over the 24-h time period (Figure 6A). From these clusters, genes for which the transcript profiles under the stress treatments resembled the decay profile in response to ActD could be identified (SCA; Figure 6A). A total of 711 genes were identified in this manner, and the greatest overlap with the ActD profiles was observed with osmotic stress followed by wounding (Figure 6B; see Supplemental Figure 3 online). The six most abundant FunCats in this SCA set and are displayed as blocks of color next to the corresponding genes (Figure 6B; see Supplemental Table 7 online). Previous studies of mRNA decay in Arabidopsis (Kilian et al., 2007) with time points identical to those used in our study, five had mRNA half-lives of <1.5 h, which correlates with the expected high rate of decay of these transcripts (Newman et al., 1993). This was complemented by our finding that identical (ATAAGAT) and highly similar (e.g., CATAGA) 6-mers to those within the DST motif were found to be significantly enriched in transcripts with shorter half-lives (see Supplemental Table 5 online). However, it cannot be excluded that inhibiting transcription affects the decay of transcripts differentially; for instance, the decay rates of long transcripts may be slowed if the machinery to decay mRNA is compromised. Although this cannot be overcome in such studies, the relative differences in transcript abundance could be observed using a single exponential decay model for all transcripts with a good fit, in a time frame where cell viability was not significantly different to mock-treated cells and the quality of cRNA in terms of 3′ to 5′ labeling ratio was sufficient for microarray analyses.

An Intron Stabilizes mRNA

In Arabidopsis, ~20% of genes lack introns, and of the 13,012 transcripts for which half-lives were calculated, 1571 of these lacked introns. This allowed a meaningful comparison to be performed to analyze the effect of this physical feature of a gene on transcript stability. Notably, this correlation did not depend on the number of introns, the length of the genes or introns, the AU content, or the length of the 5′ or 3′ untranslatable region, but simply on the presence of one or more introns. An analysis of transcript decay rates in mammalian cell lines with a smaller data set of <100 genes that lacked introns revealed a similar trend (Wang et al., 2007). Analysis of the effect of introns in both mammalian and plant systems has been investigated using changes in the activity of reporter proteins (Bourdon et al., 2001; Nott et al., 2003), and a recent study in mammalian cells revealed that the presence of a single intron stabilized a transcript irrespective of length or position, even though it contained two instability elements (Zhao and Hamilton, 2007). This is consistent with our findings and suggests that intron-associated stabilization may be a feature common to eukaryotes. It is unclear how the presence of an intron in a gene can affect
Figure 6. Comparison of the Patterns of Transcript Changes Observed upon Treatment with ActD and Several Stresses.

(A) SOM analyses of transcript abundance data from microarray analysis following ActD treatment. The transcripts were grouped by a 1 × 5 SOM (labeled C1-C5) using GeneCluster 2.0 software. In each of the graphs, the solid blue line represents the data mean, and the red lines indicate the range of the data fitted to form each cluster. For each of the clusters, \( n \) = the number of genes in that cluster, \( A \) = the number of decay profiles (following ActD treatment), and \( SCA \) = the number of genes whose expression profiles under one or more stress(es) clustered with their respective ActD decay profile. This produced a list of 711 transcripts (total number of SCA) whose expression profile, under at least one stress, clustered with the respective ActD decay profile of that transcript.
stability of a processed transcript, but the process of splicing may result in binding of additional proteins that stabilize the transcript or prevent the formation of secondary structures that may induce degradation. Alternatively, splicing may increase the efficiency of other modifications, such as capping or polyadenylation, thereby increasing stability.

**Sequence Features of Transcripts That Determine Decay Rate**

Further analysis of the physical features of a gene that affect transcript stability indicated that there was a significant enrichment of specific sequence elements in both transcripts with short half lives (< ~ 1 h) and long half-lives (> ~ 12 h) (Table 1; see Supplemental Table 5 online). The presence of instability elements has been previously documented in Arabidopsis and other systems (Ohme-Takagi et al., 1993; Meisner et al., 2004; Barreau et al., 2006), and three of these previously identified instability elements, TTTTTT, ATAGAT, and ATTTAn, were found to be significantly enriched in transcripts with short half-lives in this study. The motifs found to be most enriched among fast decayers are thymine rich, accompanied by one or more purines, and their RNA equivalents include and resemble closely the classic AU-rich element AUUUU found to be destabilizing in mammals and yeast. On the other hand, motifs found to be enriched among slow decayers are typically pyrimidine rich. The significance of this is unclear; however, pyrimidine-rich sequence elements in the 3' UTR have been demonstrated to confer stability to mRNA transcripts in a number of human genes (Lindquist et al., 2004). Interestingly, the polyadenylation motif AATAAA (Keller, 1995) was found to be significantly enriched among slow-decaying transcripts. While this motif signals for polyadenylation for the majority of animal transcripts, it is present in <20% of 3' UTR sequences in plants, for which the process of polyadenylation appears more complex (Loke et al., 2005). However, it has been shown that substitution of the AATAAA motif for the native polyadenylation motif in the pea (Pisum sativum) rbcS-E9 gene enhanced polyadenylation efficiency (Li and Hunt, 1995). Since poly(A) tails are understood to stabilize mRNA, it is plausible, and our results would suggest, that genes possessing the AATAAA motif tend to produce more stable transcripts. Interestingly, it was also observed that the reverse of several motifs was also significantly enriched (e.g., ATAGAT and TAGATA; and the reverse of the polyadenylation motif [i.e., AAATAA]), suggesting that trans-elements that may bind the motif/motifs may not necessarily be direction specific (see Supplemental Table 5 online). Together, our initial results on sequence motifs influencing mRNA stability provide a fruitful ground for further experimental and computational studies.

Nevertheless, for many of the fast-decaying transcripts in this study, no 3' UTR data were available in the TAIR7 annotation of their genes. Indeed, genes without an annotated 3' UTR are significantly more unstable than transcripts of the genome as a whole (data not shown). The explanation for this is simple: transcripts with low half-lives are more likely to be expressed at low levels and are thus less likely to yield EST or full-length cDNA data, which is the most common source of UTR annotation. Therefore, a somewhat incomplete data set was used for the identification of putative destabilizing sequence motifs in fast-decaying transcripts as 219 transcripts that were classified as “fast decayers” lacked annotated 3' UTRs. To further investigate these transcripts, artificial 3' UTRs were constructed by extracting the average length of Arabidopsis 3' UTRs (238 bases) of genomic sequence downstream of their stop codons. As before, the occurrence of all possible 6-mers among this set was compared with the set of slow decayers (see Supplemental Table 5 online). Many of the putative stabilization/distabilization motifs found in the initial comparison (e.g., ATAGAT, AATTTT, pyrimidine-rich motifs, and AATAAA) are also found using these artificial UTRs. It is clear that more accurate knowledge of the 3' UTRs of unstable UTRs would enhance this analysis.

It is well known that the 5' Gppp cap plays a role in stabilizing transcripts in mammals and plants (Ross, 1995; Abler and Green, 1996). Also, studies in humans and yeast have revealed 5' UTR elements thought to be involved in mRNA stability (Decker and Parker, 1993; Meng et al., 2005; Wang et al., 2005). Thus, we examined the 5' UTRs of the transcripts in this study for enriched 6-mers in the fast- and slow-decaying transcripts; however, there were limitations for statistical analysis in this data set (see Supplemental Table 5 online). Nevertheless, we did notice some of the AU-rich motifs in the 3' UTR analysis were also found to be enriched in the 5' UTR analysis for fast-decaying transcripts. This is intriguing as a recent study in humans (ovarian carcinoma cells) revealed that a protein (ELAV RNA stability factor HuR) known to bind AU-rich elements within the 3' UTR was also found to bind to the 5' UTR (of the insulin-like growth factor receptor), thereby illustrating how some trans-factors can bind elements in both the 3' and 5' UTR (Peng et al., 1998; Meng et al., 2005). Therefore, it is important to note that mRNA stability as a whole is likely to be regulated by several factors, including elements within both UTRs, and in some cases, these elements may be similar and bind the same factor(s).

Furthermore, it was determined that transcripts that were a target for miRNAs had significantly shorter half-lives (Figure 3A). miRNAs mediate the cleavage of mRNA via specific base pairing and cleavage by the RNA-induced silencing complex, and fragments are subsequently degraded by an exoribonuclease (Souret et al., 2004). The targets of miRNAs in Arabidopsis

---

**Figure 6.** (continued).

(B) Heat map of the mRNA decay profiles and stress transcript expression profiles that clustered together. Rapidly decaying transcripts can be seen by a change from red to green at the early time points (indicated in hours at the top of the heat map). An increase in the brightness of green indicates a decrease in transcript abundance. By contrast, slow-decaying transcripts only changed from red to green at later times. For each gene, the transcript abundance at 0 h was set to 1, and the transcript abundance at all other times were expressed in a relative manner. Next to the heat map is a colored distribution representing the six most abundant functional categories in decreasing order from left to right.

(C) The 12 transcripts that were found to cluster with three or more stresses. The transcripts that encode transcription factors are indicated in red.
were found to have significantly shorter half-lives, suggesting that binding of miRNA acts as a control step in the degradation of mRNA.

Together, these analyses have presented several sequence features involved in the control of mRNA decay. For some transcripts, features such as the presence/absence of introns cannot be regulated, and their role in mRNA decay is predetermined for a given spliced transcript. Similarly, other sequence features, such as the presence of miRNA target sequences, allow regulation of mRNA decay by controlling the trans-factors, such as the presence/amount of miRNAs and RNA-induced silencing complex in the system. Similarly, the presence of motifs facilitates complex regulation as it enables several levels of control based not only on the total number of motifs but also on the number of unique motifs that may bind unique trans-factors. The presence of multiple different motifs and multiples of single motifs in any transcript suggests complex combinatorial control of transcript degradation (see Supplemental Table 5 online), and we propose that combinatorial regulation of both transcript production and degradation may determine transcript steady state levels.

mRNAs in eukaryotes can be degraded via several partially independent pathways (Newbury, 2006), but the primary pathways appear to be degradation in cytoplasmic P-bodies (via the 5′-3′ exonuclease XRN4 in Arabidopsis) or degradation via the exosome (a complex of 3′-5′ exonucleases). Recruitment of mRNAs into the degradation pathways is often linked to translational defects and can be enhanced by binding of small interfering RNAs or specific protein factors. The best-known of these factors are UPF1, UPF2, and UPF3, which can act to rapidly destabilize mRNAs when bound to them. The UPF factors are implicated in nonsense-mediated decay (NMD), so-named because mRNAs with premature termination codons are often rapidly degraded (reviewed in Conti and Izaurralde, 2005). The manner by which UPF factors recognize premature termination codons is controversial and may differ between yeast, animals, and plants. In plants, both exceptionally long 3′ UTRs and introns within the 3′ UTR have been proposed to be signals for UPF binding and thus NMD (Kertesz et al., 2006; Schwartz et al., 2006; HorI and Watanabe, 2007). However, the evidence for this is based on analyses of only a few transgene transcripts. The results of our genome-wide study of mRNA decay indicate no evidence that decay rates are influenced either by 3′ UTR length or the presence of an intron in the 3′ UTR (Figure 3B). Hence, natural Arabidopsis transcripts can either avoid NMD or the triggers proposed so far are too simplistic.

Effects of Alternative Splicing

It is important to note that in this study, sequence features of transcripts related to mRNA decay rates have been analyzed on the basis of the first gene models provided by the TAIR database (i.e., those ending with .1). Additional gene models (.2, .3, etc.) have been constructed on the basis of alternatively spliced transcripts. In Arabidopsis, ~20% of genes produce alternatively spliced transcripts (Iida et al., 2004; Ner-Gaon and Fluhr, 2006). Therefore, of the 13,012 genes in this study, we expected 2602 genes to be subject to alternative splicing, and in agreement with this, we found 2788 genes known to be alternatively spliced. These genes were found to have a significantly higher proportion of transcripts with longer half-lives (P = 6.64E-10) compared with the genome (see Supplemental Figure 4 online). It is unclear why this subset has higher proportion of longer half-lives, but it may be related to RNA processing required to produce alternatively spliced transcripts. We focused on alternative splicing in the context of intron retention as this appears to be the major form (~14%) of alternative splicing in Arabidopsis and other plants, in contrast with mammalian systems for which exon skipping is more predominant (Wang and Brendel, 2006). In plants, altered mRNA stability has been demonstrated for a small number of genes that have transcript variants as a result of alternative splicing (Lejeune and Maquat, 2005; Thiele et al., 2006). However, very little is known about the relative abundance of splice variants. We chose to use the first gene models in our analysis because anecdotal evidence suggests that they are generally the most abundant, and these gene models formed the basis for the design of the ATH1 genome array used in this study. It is possible, however, that splice variants with altered mRNA stability compared with transcripts of the first gene models are present in sufficient abundance within our pool of mRNA to affect the results of this work.

As this study found that the absence of an intron had a pronounced affect on transcript stability and intron retention alternative splicing occurs in ~14% of genes in Arabidopsis (Wang and Brendel, 2006), we examined the mRNA half-lives of (1) genes for which the first TAIR gene models contained at least one intron, but at least one subsequent splice variant was intronless (there are only 91 such genes in the list of 13,012 genes), and (2) genes for which the first gene model contains at least an intron and all of its splice variants, if they exist, also contain at least one intron. We found a no significant difference between the first subset and the genome. By contrast, we found that the transcripts that contain introns in the first gene model and all of its splice variants had a higher proportion of transcripts with longer half-lives; however, this was not statistically significant (Figure 3B). The small number of genes in the first set means limits the statistical testing capability, but the result does (1) confirm our earlier results on the instability of intronless transcripts, (2) show that alternatively spliced transcripts possess differing decay profiles, and (3) show that these splice variants are in sufficient abundance to have an affect, albeit small, on the analysis.

Recently, a study performed in humans showed that transcripts with retained introns in the 3′ UTR contain more putative miRNA target sites, suggesting that such splice variants may be more unstable (Tan et al., 2007). Therefore, we examined the genes in this study that (1) contained at least one 3′ UTR, (2) contained at least one 3′ UTR intron retention splice variant with respect to the first gene model (56 genes), and (3) genes that have no 3′ UTR intron retention splice variants compared their cumulative mRNA half-life distribution to the genome (Figure 3C). For these subsets, there was no significant difference between their distributions and the genome. Therefore, no conclusions could be drawn on 3′ UTR intron-retention/miRNA-mediated destabilization, probably due to the fact that the abundances of 3′ UTR intron-retention splice variants are not high enough in
comparison with other splice variants to detect the effect. However, even if this was the case, the results of this study suggest that a variety of factors influence mRNA stability. Another study in humans has revealed significant differences in the translation efficiency and mRNA stability of 5' UTR splice variants of the human SP-A1 and SP-A2, which highlighted the role of the 5' UTR and the importance of considering alternative splicing in mRNA stability (Wang et al., 2005). Comprehensive studies on the relationship between alternative splicing and mRNA stability could possibly be performed through the use of exon/tiling arrays that are designed to detect alternative isoform or genomic deletions and not only to measure transcript abundance. The decay profiles of the splice variants could then be compared. The results of this work provide a starting reference point for such studies.

Transcript Decay Rates and the Function of Encoded Proteins

Significant correlations were detected between decay rates and the functional categorization of the encoded proteins (Figure 4A), the intracellular location of the encoded protein (Figure 4C), and the physical characteristics of the transcript/gene (Figure 3). Correlation of mRNA half-life with function has been previously reported in other systems (Wang et al., 2002; Yang et al., 2003), and this study indicates that this is also true for Arabidopsis. Specifically, we found that transcripts encoding transcription factors, PPR proteins, and kinases have short half-lives. Rapid control of the abundance of transcription factors and PPR proteins could act as a switch for global regulation of expression of other genes controlled by these factors, and, similarly, kinases are involved in regulating protein activity, cellular communication, and signal transduction. Thus, rapid degradation of mRNA of these regulatory factors may be an important control mechanism to ensure quick responses to stimuli and rapid reversion to equilibrium. Furthermore, comparison of mRNA decay data sets for yeast and human, with the data obtained for Arabidopsis from this study, indicates that transcripts encoding orthologous proteins involved in cell cycle, transcription, translation, and energy metabolism have conserved rates of mRNA decay.

The comparison of patterns of transcript changes observed upon various stress treatments with ActD treatment revealed that inhibition of transcription and/or transcript degradation occurs in response to stress (Figure 6). Either of these mechanisms would account for the decrease in transcript abundance observed. The similarity in profiles observed for some genes suggests that inhibition of transcription may be an important step in the stress response. Furthermore, given that many of the genes that consistently fell into this category across a variety of stresses were annotated as transcription factors, this suggests that the stress response signals may act to reduce transcriptional activation by these factors and/or inactivate factors that may have a repressor function. Notably, a common response to four of the stress treatments examined was a decrease in transcript abundance of two ethylene response factor transcription factors, reported to be involved in the plant stress response as well as growth and development (Riechmann, 2002; Gutterson and Reuber, 2004), suggesting that these processes may be an early target of stress-induced transcriptional repression. Two other types of regulatory proteins, RAV1 (for Related to ABI3/VP1) and RAV2 and ARGOS-like (for Auxin-Responsive Gene Inducing Organ Size), also showed transcript decreases in three or more stresses (Figure 6C). Both of these types of proteins have been reported to play a regulatory role in organ or cell expansion and development (Hu et al., 2004, 2006). As both RAV and ethylene response factor proteins can be negative or positive regulators (Hu et al., 2004; McGrath et al., 2005), the rapid downregulation of their transcript abundance may have multiple effects in response to a number of stresses. Furthermore, it was found for both wounding and osmotic stresses that a reduction in protein synthesis may be a secondary target and may impair the ability to repair damage caused by stress.

Conclusion

Overall, genome-wide analysis of the transcript degradation profiles in Arabidopsis, determined using transcription inhibition combined with a microarray approach, revealed that several physical features of a transcript and the function of the encoded protein correlated with its half-life. We have identified three features of a gene that appear to play a role in determining the half-life of mRNA: the presence or absence of an intron, the occurrence of specific sequence elements, and if the mRNA is a miRNA target. Thus, our analysis reveals that regulation of mRNA stability in Arabidopsis is likely to occur at multiple levels. In addition, the general relationship between mRNA half-lives and function appears to be conserved across Arabidopsis, yeast, and humans. Furthermore, there is evidence to suggest that decline in transcript abundance, particularly for those transcripts encoding proteins with regulatory roles, may be a commonly overlooked response to stress, where increases in transcript abundance are routinely the focus. It will be of interest with future studies to determine if decay rates are altered under development or environmental signals and to identify the factors that bind to the elements that are enriched in transcripts with different decay rates. In addition to providing an alternative view and greater insights into the overall process of regulation of gene expression, it would provide a platform to design constructs with altered stability, thus allowing alternative approaches to altering expression of individual transcripts than simply overexpression or knockout.

METHODS

Arabidopsis Cell Culture Growth, ActD Treatment, and RNA Isolation

Arabidopsis thaliana (ecotype Landsberg erecta) cell suspension was grown at 22°C under long-day conditions (16 h of 100 μE m⁻² s⁻¹ light and 8 h dark) with orbital shaking at 100 to 150 rpm. For confirmation of ActD effectiveness, cells were treated for 1 h with ActD (Sigma-Aldrich) to a final concentration of 100 μg/mL and then treated with salicylic acid to a final concentration of 100 μM. Cells were collected prior to salicylic acid treatment (–1 h), prior to ActD treatment (0 h), and then 1, 3, 6, 12, 24, and 48 h after treatment. For the global mRNA decay study, triplicate experiments were performed where, 4 d after subculturing, cells were collected prior to ActD treatment (0 h) and then 1, 3, 6, 12, 24, and 48 h
after treatment. Cells were immediately frozen in liquid nitrogen, and total RNA was isolated using the RNeasy plant mini kit (Qiagen), with genomic DNA removed using both the RNase-free DNase set (Qiagen) and the DNA-free kit (Ambion). Cell viability was also examined at each time point using acridine orange and ethidium bromide staining.

Quantitative RT-PCR Analysis

cDNA was prepared from 1 μg of each RNA sample using the iScript cDNA synthesis kit (Bio-Rad), and quantitative RT-PCR analysis was performed using an iCycler instrument and iQ SYBR Green Supermix (Bio-Rad) under conditions optimized to maximize amplification efficiency and minimize primer-dimer formation, using primers listed in Supplemental Table 8 online. For every transcript, each cDNA sample was analyzed in duplicate, and transcript abundance was expressed as a ratio relative to pretreatment levels (0 h), which was set to a value of 1.0.

Microarray Analysis

Microarray analysis was performed on samples collected prior to ActD treatment (0 h) and 1, 3, 6, 12, and 24 h after treatment using the Affymetrix GeneChip Arabidopsis ATH1 genome arrays. Prior to target preparation, RNA quality was assessed using a Bioanalyzer (Agilent Technologies) and spectrophotometric analysis to determine A260 to A280 ratios. cRNA was prepared from 5 μg of total RNA almost exactly as described in the Affymetrix GeneChip Expression Analysis Technical Manual. However, as a result of inhibition of transcription and subsequent mRNA decay, the yield of cRNA decreased over time. Therefore, rather than using the same amount of cRNA to hybridize to each microarray, the same volume of the purified in vitro transcription reaction as was used for the 0 h samples (to give 15 μg of cRNA) was used in the hybridization cocktail. Array washing, staining, and scanning was performed as described in the Affymetrix GeneChip Expression Analysis Technical Manual, using an Affymetrix Hybridization Oven 640, an Affymetrix Fluidics Station 450, and an Affymetrix GeneChip Scanner 3000 7G. The quality of the data was assessed using GCOS 1.4 (Affymetrix) before CEL files were imported into Avadis 4.3 (Strand Genomics) for further analysis. The MAS 5.0 algorithm was applied to determine present or absent calls, and correlation between replicates was checked and found to be >0.95. Transcripts were removed from subsequent analysis if they were called absent in two out of three of the 0 h (pretreatment) samples, and to further eliminate false present calls, the transcripts that had the bottom 2.5% average signal intensity at 0 h were also removed. The data were then normalized against the Bacillus subtilis thrB and thrC genes (present as controls on the ATH1 GeneChip), and the pretreatment (0 h) abundance of each transcript was assigned a value of 1 and transcript abundance at all other time points was expressed relative to this. Ambiguous probe sets and probes encoding organelle encoded genes were removed before retrieval of locus number from the TAIR database (www.arabidopsis.org/), for each probe ID, resulting in a final data set of 13,012 Arabidopsis transcripts. All microarray data are available from ArrayExpress (http://www.ebi.ac.uk/arrayexpress) under accession number E-ATMX-21.

Analysis of mRNA Decay Profiles and Half-Life Calculations

SAS version 9.1 (SAS Institute) was used to fit a nonlinear least-squares model to the QT-PCR data and the normalized, relative microarray data. mRNA decay has generally been found to obey first-order kinetics (Ross, 1995; Gutierrez et al., 2002); therefore, an exponential regression model (A = A0 e−kt) was applied, allowing a kdecay (with 95% confidence intervals) to be calculated for each transcript (see Supplemental Table 2 online). The mRNA half-life was then calculated using the following equation: 

$$t_{1/2} = \ln(2)/k_{\text{decay}}$$

Statistical analysis of the regression was performed in SAS whereby a t value and probt was calculated for each kdecay value to test the null hypothesis that the data fits the exponential model. Approximately 400 (~3%) transcripts were found to increase by 25% or more from 0 to 1 h before decreasing from 3 h onwards, and most of these genes were found to encode for heat shock proteins and stress-related functions as seen in the yeast global mRNA decay study (Wang et al., 2002). The decay profiles for these transcripts were regressed using the equation A = be−kt, where b represents the value at 1 h, allowing the data to be regressed more accurately.

Assignment of FunCats

To determine if there was any correlation between mRNA half-life and function, the Functional Catalogue (Ruepp et al., 2004), based on the fusion of the gene ontology (GO) biological process and the GO molecular function, was used. Using the accession numbers for all 13,012 genes to interrogate the FunCats detailed in the Munich Information Center for Protein Sequences (MIPS) database resulted in 3284 of these being annotated into different FunCats, while 9728 were classified either as “miscellaneous function,” “classification not yet clear-cut,” or “unclassified proteins.” This limited further analysis of transcript decay data and function; therefore, different means were used to annotate as many of the 9728 genes remaining as possible. Analysis of genes annotated in several databases including the Arabidopsis Mitochonndrial Protein Database (AMPDDB; http://www.plantenergy.uwa.edu.au/applications/ampdb/index.html), the Database of Arabidopsis Transcription Factors (DATF; http://datf.cbi.pku.edu.cn), the A. thaliana Transcription Factor Database (ATFDB; http://arabidopsis.med.ohio-state.edu/atfdb/), the RIKEN Arabidopsis Transcription Factor database (RARTF; http://rarge.gsc.riken.jp/rartf/), TAIR (http://www.arabidopsis.org), and one broad transcription factor study (Czechowski et al., 2004) allowed annotation of a further 1101 genes into appropriate FunCats. As described, the FunCat was produced from analysis of the GO biological process and GO molecular function; therefore, manual annotation of genes was performed following analysis of the GO annotations, comparison to the FunCats of members of the same family (where appropriate), and the broad annotations (if available) (see Supplemental Table 4 online). In this way, a further 3857 genes were annotated, giving a final annotated set of 8242 genes.

Generation of Specific Lists

To closely examine coordination of mRNA decay with function, gene lists were generated based on previous definition into gene families (PPR proteins and ribosomal proteins), groups of transcripts encoding similar functions (kinases and transcription factors), or transcripts known to be subject to specific forms of posttranscriptional regulation (miRNA targets). The transcription factor list was generated by analysis of the genes annotated in the AMPDB, DATF, ATFDB, RARTF, and TAIR databases (described above) as well as one broad transcription factor study (Czechowski et al., 2004) to produce a list of 1137 transcription factors from the 13,012 genes. Similarly, genes that are the targets for mRNA were compiled from several databases to produce a list of 270 mRNA targets, both experimental and predicted, and the list of these targets and corresponding references is in Supplemental Table 2 online (Rhoades et al., 2002; Bonnet et al., 2004; Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004; Wang et al., 2004; Adai et al., 2005; Jones-Rhoades et al., 2006; Poethig et al., 2006). Subsets consisting of 258 kinases, 263 PPR proteins, and 325 ribosomal proteins were generated by examination and manual selection of the relevant annotations to produce each list.
www.suba.bcs.uwa.edu.au; Heazlewood et al., 2005, 2007), which contains data for 6743 Arabidopsis proteins on localization to the following cellular compartments: cell plate, chloroplast, cytoskeleton, cytosol, endoplasmic reticulum, extracellular space, Golgi, mitochondria, nucleus, peroxisome, plasma membrane, and vacuole. Occasionally, two or more of the information sources used by the SUBA database annotated it as belonging to that compartment. Using the SUBA localization data assigned a single subcellular localization to 5832 Arabidopsis proteins of the 13,012 set, and a winner-takes all approach was only required to resolve location in 358 cases.

Analysis and Presentation of Gene Expression Data

The whole-genome mRNA decay profiles were hierarchically clustered using complete linkage and Euclidean distance. For all mRNA half-lives and the mRNA half-lives in each of the FunCats, summary statistics, including the number of genes in each data set, the minimum, maximum, mean, SD, and the number of outliers, was calculated (see Supplemental Table 3 online). The summary statistics also included calculation of the mRNA half-life at different percentiles for that particular data set (e.g., for the metabolism category) at percentile 25 the mRNA half-life is 2.70, which indicates that 75% of transcripts in that data set have half-lives of 2.70 h or more.

Sequence Analysis of mRNAs

To determine whether sequence features of mRNAs correlate with half-life, information on the UTR and coding sequence length and AU content, as well as intron numbers and length, were extracted for all Arabidopsis genes from the gene-finding format and chromosome sequence files from the TAIR website (ftp://ftp.Arabidopsis.org/home/tair/Genes/TAIR7_genome_release) using Perl scripts.

Motif Searches: Enrichment/Depletion of 6-Mers

For all genes in this study, annotated 3′ UTR sequences were searched for the presence of one or more of all possible 4096 6-mers of the bases A, T, G, and C. From the mRNA half-life data, two sets of 1000 transcripts each were generated: fast decayers, comprising genes with t1/2 < 1.148 h, and slow decayers, comprising genes with t1/2 > 12.487 h. The proportions \( p_f \) and \( p_s \) of genes that contain each 6-mer in the fast and slow sets, respectively, were compared using the two-sample z-statistic:

\[
Z = \frac{p_f - p_s}{\sqrt{p(1-p)/(1/n_f + 1/n_s)}},
\]

where \( p \) is the pooled sample proportion. Statistical significance was assessed at \( P = 0.01 \) (|z| > 2.58) and \( P = 0.001 \) (|z| > 3.89).

Analysis of mRNA Decay in Arabidopsis Orthologs

To compare genome-wide mRNA half-lives across the yeast, human, and Arabidopsis genomes, a list of orthologous genes was compiled using the InParanoid: Eukaryotic Ortholog Groups database (Remm et al., 2001; O’Brien et al., 2005). The orthologous group files were downloaded for whole-genome comparison of Arabidopsis versus Homo sapiens (6302 orthologs) and Arabidopsis versus Saccharomyces cerevisiae (5400 orthologs). For the human orthologs, the G-profiler database was used to convert GenBank accession numbers to Ensembl protein identifiers (Reimand, 2006; Reimand et al., 2007) to allow comparison to the human mRNA decay study (Wang et al., 2002; Yang et al., 2003). Avadis 4.3 was then used to filter for orthologs where half-life data for both species was available, which produced a list of 712 orthologs for Arabidopsis versus humans and 1407 orthologs for Arabidopsis versus yeast. By assessing overlap of these lists, a final list of 1830 Arabidopsis genes was generated, of which 423 are orthologous to humans only, 1118 to yeast only, and 289 to both humans and yeast (see Supplemental Table 6 online). After consideration of the minimum, maximum, median mRNA half-life, and CCL for each species data set, all transcripts were then categorized into four groups based on mRNA half-lives relative to CCL for each species. These categories were defined as short (<10% of CCL), moderate 1 (10 to 22% of CCL), moderate 2 (22 to 50% of CCL), and long (>50% CCL).

Stress Arrays and GeneCluster 2.0 Analysis

To determine whether changes in transcript abundance following stress treatments correlated with the decrease in transcript abundance following transcriptional inhibition with ActD, microarray data consisting of a time series following stress treatments (Kilian et al., 2007) were compared with the ActD decay profiles. A flow chart for the handling of the data is shown in Supplemental Figure 3 online. Briefly, all array data were filtered and normalized to produce a list of 9061 genes that were represented in all data sets. A Pearson correlation of 0.7 was used to define ActD transcript profiles that were similar to their profiles in each stress, producing a redundant list of 5001 stress profiles that matched with 3877 ActD profiles (total of 8878 profiles). These profiles were analyzed by SOM clustering (Tamayo et al., 1999) to determine which stress transcript profiles clustered with the ActD transcript profiles. Genes whose transcript decay profiles clustered with five, four, three, two, or only one stress profile/s were identified (see Supplemental Table 7 online), and these data were visualized in a heat map, arranged in descending order so genes that overlapped with the most number of stresses were positioned at the top.

Accession Numbers

Sequence data from this article can be found in the GenBank/EMBL data libraries under the following accession numbers: WRKY6 (WRKY family transcription factor), At1g62300; OM64 (mitochondrial outer membrane protein of 64 kD), At5g09420; MDHAR (monodehydroascorbate reductase), At1g62300; UCP1 (uncoupling protein UCP/PUMP), At3g54110; HPR (hydroxyopyruvate reductase), At1g68010; and TIM14-1 (translocase of the inner mitochondrial membrane), At2g35795.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Analysis of RNA after Treatment with ActD in Arabidopsis Cell Culture.

Supplemental Figure 2. Comparison of mRNA Half-Lives with Intron Number, Length, and AU% of All 13,012 Transcripts.

Supplemental Figure 3. Flow Diagram of Analysis Performed to Generate the List of Genes Showing Stress Profiles Similar to Decay Profiles.

Supplemental Figure 4. The Distribution of mRNA Half-Lives of the Whole Genome Compared with Alternatively Spliced Transcripts.

Supplemental Table 1. Comparison of mRNA Decay Profiles Using QRT-PCR and Microarray Analysis.

Supplemental Table 2. The Decay Rates and mRNA Half-Lives of the 13,012 Transcripts.

Supplemental Table 3. Descriptive Statistics of the mRNA Half-Lives for the Whole Genome and All Analyzed Subsets.
Supplemental Table 4. Functional Categorization Assignments Used.

Supplemental Table 5. A List of All 4096 6-Mers and the Significance of Their Occurrence in the 3′ UTR Regions of Genes Defined as Fast or Slow and the Total Number of Occurrences of the Top and Bottom 50 Motifs in These Genes.

Supplemental Table 6. The 1830 Arabidopsis Genes for Which There Are Orthologs in Yeast and Humans with mRNA Half-Life Data.

Supplemental Table 7. The 711 Genes That Display Similar Changes in Transcript Abundance upon Treatment with ActD and in Response to One or More Stresses.

Supplemental Table 8. Sequences of Primers Used for QRT-PCR.

ACKNOWLEDGMENTS

This work was supported by the Australian Research Council Centre of Excellence Program CE0561495. We thank Kevin Murray from the University of Western Australia Department of Mathematics and Statistics for general statistical advice and, in particular, for help with the exponential regression of the transcript abundance data. We thank Joshua Heazlewood for assistance with using the MIPS database for annotation of FunCats.

Received August 10, 2007; revised October 17, 2007; accepted October 21, 2007; published November 16, 2007.

REFERENCES


Global mRNA Decay in Arabidopsis thaliana 3435


Genome-Wide Analysis of mRNA Decay Rates and Their Determinants in *Arabidopsis thaliana*
Reena Narsai, Katharine A. Howell, A. Harvey Millar, Nicholas O'Toole, Ian Small and James Whelan
*Plant Cell* 2007;19;3418-3436; originally published online November 16, 2007; DOI 10.1105/tpc.107.055046

This information is current as of November 6, 2017

<table>
<thead>
<tr>
<th>Supplemental Data</th>
<th>/content/suppl/2007/10/26/tpc.107.055046.DC1.html</th>
</tr>
</thead>
<tbody>
<tr>
<td>References</td>
<td>This article cites 74 articles, 28 of which can be accessed free at: /content/19/11/3418.full.html#ref-list-1</td>
</tr>
<tr>
<td>eTOCs</td>
<td>Sign up for eTOCs at: <a href="http://www.plantcell.org/cgi/alerts/ctmain">http://www.plantcell.org/cgi/alerts/ctmain</a></td>
</tr>
<tr>
<td>CiteTrack Alerts</td>
<td>Sign up for CiteTrack Alerts at: <a href="http://www.plantcell.org/cgi/alerts/ctmain">http://www.plantcell.org/cgi/alerts/ctmain</a></td>
</tr>
<tr>
<td>Subscription Information</td>
<td>Subscription Information for <em>The Plant Cell</em> and <em>Plant Physiology</em> is available at: <a href="http://www.aspb.org/publications/subscriptions.cfm">http://www.aspb.org/publications/subscriptions.cfm</a></td>
</tr>
</tbody>
</table>

© American Society of Plant Biologists
ADVANCING THE SCIENCE OF PLANT BIOLOGY