Endogenous siRNA and miRNA Targets Identified by Sequencing of the Arabidopsis Degradome

Charles Addo-Quaye,1 Tifani W. Eshoo,2 David P. Bartel,4,5 and Michael J. Axtell2,3,*
1Department of Computer Science and Engineering
2Cell and Developmental Biology Graduate Program
Huck Institutes of the Life Sciences
3Department of Biology
Pennsylvania State University
University Park, Pennsylvania 16802
4Whitehead Institute
Cambridge, Massachusetts 02142
5Howard Hughes Medical Institute and Department of Biology
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Summary

MicroRNAs (miRNAs) regulate the expression of target mRNAs in plants and animals [1]. Plant miRNA targets have been predicted on the basis of their extensive and often conserved complementarity to the miRNAs [2–4], as well as on miRNA overexpression experiments [5]; many of these target predictions have been confirmed by isolation of the products of miRNA-directed cleavage. Here, we present a transcriptome-wide experimental method, called “degradome sequencing,” to directly detect cleaved miRNA targets without relying on predictions or overexpression. The 5′ ends of polyadenylated, uncapped mRNAs from Arabidopsis were directly sampled, resulting in an empirical snapshot of the degradome. miRNA-mediated-cleavage products were easily discerned from an extensive background of degraded mRNAs, which collectively covered the majority of the annotated transcriptome. Many previously known Arabidopsis miRNA targets were confirmed, and several novel targets were also discovered. Quantification of cleavage fragments revealed that those derived from TAS transcripts, which are unusual in their production of abundant secondary small interfering RNAs (siRNAs), accumulated to very high levels. A subset of secondary siRNAs are also known to direct cleavage of targets in trans [6]; degradome sequencing revealed many cleaved targets of these trans-acting siRNAs (ta-siRNAs). This empirical method is broadly applicable to the discovery and quantification of cleaved targets of small RNAs without a priori predictions.

Results and Discussion

The Arabidopsis Degradome Generally Reflects Transcript Abundance

Although plant miRNAs can interact with target transcripts without directing their hydrolysis [7, 8], the most common outcome of pairing to known targets is miRNA-directed cleavage at the tenth nucleotide of complementarity relative to the guiding miRNA [9, 10]. The resulting 3′ target fragments have a 5′ monophosphate and a 3′ polyA tail and can be recovered by RNA ligase-mediated 5′ rapid amplification of cDNA ends (RLM 5′-RACE [9, 10]). These experiments require an a priori miRNA target prediction and are limited to testing one candidate target at a time. In an alternative approach, a primer designed to match presumed miRNA complementary sites is used to identify targets without predictions, but this method is also limited by low throughput [11].

We modified the RLM 5′-RACE method to globally sample RNAs with a 5′ monophosphate and a 3′ polyA tail. An RNA adaptor was engineered to contain a 5′ Mmel restriction site and ligated directly to polyA+ RNA; this selected against capped, full-length mRNAs, which lacked the free 5′ monophosphate required for ligation. After reverse transcription and second-strand synthesis, Mmel digestion left a 20–21 nt tag attached to the 5′ adaptor; a dsDNA adaptor was then attached to the 3′ end of the tag. Amplified libraries of these “degradome” 5′ tags were then sequenced. In most cases, the 20–21 nt signatures were sufficient to unambiguously identify the transcripts of origin, with the junction between the signature and the 5′ adaptor indicating the 5′ terminus of the original RNA.

Four distinct libraries were sequenced from either inflorescence or seedling specimens of wild-type Arabidopsis (Table S1 available online). In total, 835,902 unique 20–21 nt signatures (represented by 5,092,568 reads) were found to exactly match the sense strand of one or more annotated transcripts (Table S1). The majority (~73%) of annotated Arabidopsis transcripts had at least one unambiguously matched tag. The libraries of degradome tags were biased toward the 3′ ends of annotated transcripts (Figure 1A). This bias persisted in libraries where random hexamers replaced oligo(dT) as the reverse-transcription primer. Thus, to a large extent, the global 3′ bias probably reflected the distribution of in vivo-polyadenylated degradation intermediates. The repeat-normalized total abundance of all degradome tags for each transcript was summed. Comparison of these degradome abundances to the AtGenExpress-derived expression values [12] for the relevant tissue samples revealed a positive correlation between degradome tag density and steady-state levels of mRNAs (Figures 1B–1D). Thus, degradome tag abundance generally reflected the abundance of the respective mRNAs.

Discovery of miRNA Targets

The 5′ ends of tags derived from miRNA cleavage fragments would precisely correspond to the tenth nucleotide of miRNA complementary sites. To find such tags, we first extracted the potential complementary sites by retrieving query sequences extended 15 nt upstream from the tag 5′ end (Figure 2A). These sequences were aligned with confidently annotated Arabidopsis miRNAs and were scored according to previously described methods for Arabidopsis miRNA target predictions [2, 13]. All alignments with mismatch scores of seven or lower were initially retained. Alignments where the degradome tag 5′ end was precisely opposite the tenth nucleotide of the miRNA constituted potential evidence of miRNA-mediated cleavage (Figure 2A). When such alignments were found, neighboring tags whose 5′ ends aligned with the ninth or eleventh positions were also scored as evidence for
miRNA-mediated cleavage; this accounts for the occasional positional heterogeneity seen for mature miRNAs [13] and their cleavage products [3, 14].

Examination of potential targets revealed a spectrum of evidence for target cleavage. For many targets, degradome tags indicative of miRNA-mediated cleavage were the most abundant tags matching the transcript (Figure 2B); these were classified as category I. For others, tags diagnostic of cleavage were not the most abundant tags matching the transcript, but still formed a clear peak at the complementary site—these were classified as category II (Figure 2C). miRNA-aligned tags, which formed minor peaks, were deemed category III targets (Figure 2D). As a control, alignments were also performed against 30 cohorts of randomized miRNA sequences; the resulting signal-to-noise ratios were used to define appropriate cutoffs for confident target identification (Figure 2E).

Many miRNA hairpins produce positional variants of the annotated mature miRNA, abundant miRNA* sequences, and in

Figure 1. Degradome Tags Were 3' Biased and Correlated with Transcript Abundance
(A) Histogram displaying the 5' positions of degradome tags from the indicated libraries relative to normalized transcript position. Tags were counted in 1% bins. For clarity, one tag that was extremely abundant in libraries three and four was omitted.
(B-D) Relationships between microarray-derived estimates of transcript abundance (y axis) and degradome tag abundance (x axis) for the indicated libraries. Robust multichip average (RMA)-normalized array values were from [12] and were matched according to the tissue source (inflorescence, ATGE_29 [B and C]; seedlings, ATGE_96 [D]). Transcripts with degradome tag abundances equal to or less than one were omitted. Lines represent best-fit linear regressions; r² values represent correlation coefficients.

Figure 2. Experimental Identification of Cleaved miRNA Targets without Predictions
(A) Schematic of methodology to find evidence of miRNA-mediated cleavage from degradome tags. For many targets, degradome tags indicative of miRNA-mediated cleavage were the most abundant tags matching the transcript (Figure 2B); these were classified as category I. For others, tags diagnostic of cleavage were not the most abundant tags matching the transcript, but still formed a clear peak at the complementary site—these were classified as category II (Figure 2C). miRNA-aligned tags, which formed minor peaks, were deemed category III targets (Figure 2D). As a control, alignments were also performed against 30 cohorts of randomized miRNA sequences; the resulting signal-to-noise ratios were used to define appropriate cutoffs for confident target identification (Figure 2E).
Many miRNA hairpins produce positional variants of the annotated mature miRNA, abundant miRNA* sequences, and in
some cases multiple miRNA/miRNA* duplexes from the same hairpin [13, 15–17]. Some of these variants might also direct target cleavage. To find cleavage targets that may have been missed when considering only annotated mature miRNA sequences, we expanded the search to include all experimentally identified small RNAs between 20 and 22 nt in length [13, 18] matching one or more Arabidopsis miRNA hairpins. Randomizations using this expanded set of queries indicated a need for more stringent cutoffs than those used for annotated mature miRNAs only (Figure S1A). Seven additional targets of three miRNA families were found; because one target had two cleavage sites, these represented eight additional distinct cleavage sites (Table S3). This analysis was critical to identify the PPR targets of miR161 and the previously predicted DC1 target of miR822. Altogether, these two analyses confirmed 57 out of 103 previously validated miRNA targets, 14 that were previously predicted but had not been validated, and six for which there was no previous experimental evidence or predictions (Figure 2F, Table S3).

The identified miRNA cleavage targets were biased toward conserved miRNA families: 18 of the 29 miRNA families for which one or more targets were found have been annotated in at least one other plant species besides Arabidopsis [19]. This is consistent with the general correlation between miRNA conservation, expression level, and number of verifiable targets [13, 14]. Newly confirmed or discovered targets were mostly related to already known targets of the same miRNA. For instance, SPL5, -6, -9, -13, and -15 were confirmed as miR156 targets, and MYB13, -20, and -111 were confirmed as miR858 targets (Table S3). Seven MYB targets of miR159 have been described [20], but regulation of only two, MYB33 and MYB65, accounts for the phenotypes observed in mir159ab loss-of-function mutants [21]. That only MYB33 and MYB65 were confirmed in our experiments suggests that the degradome technique might have a propensity to identify targets of higher in vivo relevance.

Discovery of ta-siRNA Targets
Known Arabidopsis secondary small interfering RNA (siRNA)-producing loci include four families of TAS genes, a set of PPR genes targeted by multiple miRNAs, trans-acting siRNAs (ta-siRNAs), and the cleaved targets of miR393 and miR168 [7, 13, 22, 23]. A subset of these secondary siRNAs direct the cleavage of targets distinct from their precursors and thus are trans-acting siRNAs (ta-siRNAs; [6]). Cleaved ta-siRNA targets were identified through the use of all secondary siRNAs between 20 and 22 nt in length whose expression has been documented by sequencing [13, 18] as queries to find degradome tags indicative of cleavage. Alignments indicative of cis interactions (where an siRNA directs cleavage of its precursor transcript) were excluded. After applying empirically determined thresholds (Figure S1B), 38 ta-siRNA-directed cleavage events were confidently identified within 24 distinct transcripts (Figure 2F; Table S3). These 24 transcripts included most of those previously confirmed by gene-specific 5′-RACE (11 out of 14). The dense network of PPR-derived ta-siRNAs directing cleavage of other closely related PPR genes was particularly striking in our analysis and supports the notion of a self-reinforcing cascade of PPR repression initiated by miR173, miR161, and TAS1/2 family ta-siRNAs ([22, 24]; Table S3).

TAS Cleavage Products Accumulate to High Levels
Degradome tags corresponding to cleaved TAS transcripts were among the most abundant (Figure 3). TAS genes are unusual miRNA targets in that miRNA-mediated cleavage serves to stimulate the production of abundant secondary siRNAs from one of the two resulting fragments [2, 25]. High accumulation levels of 3′ TAS cleavage products were observed regardless of whether the 3′ product directly contributed to downstream ta-siRNA formation (TAS1a-c and TAS2) or not (TAS3a and TAS3c; [2, 25]). High accumulation of TAS cleavage products could reflect initially high TAS transcript levels and/or a specific stabilization after miRNA-mediated processing. The observation that SGS3, a factor required for postranscriptional gene silencing and ta-siRNA accumulation [26–28], is required to stabilize TAS1a and TAS2 cleavage products [29] argues in favor of selective stabilization. One recent model for secondary siRNA biogenesis from the TAS3a locus invokes specific stabilization of the 5′ cleavage product, which serves as a template for siRNA production [29]. How this might connect to the high abundance of the 3′ TAS3a cleavage product (Figure 3), which does not serve as a template for siRNA production, is unclear. Perhaps high accumulation of a miRNA cleavage product is a hallmark of the secondary siRNA pathway. In cases like TAS1a-c and TAS2, for which other factors known to promote secondary siRNA formation, such as multiple miRNA complementary sites [7, 22] or AGO7 dependency [29–32], are not evident, high cleavage-product accumulation may promote entrance to the secondary siRNA pathway.

Confirmation of Novel Targets
The accumulation of many Arabidopsis miRNA targets increases in mutants with defects in miRNA biogenesis but remains unchanged in siRNA mutants [2]. Similarly, the accumulation of targets identified through degradome sequencing generally increased in mutants affecting miRNAs (dcl1-7, hent1-1, hst-15, and hyl1-2) but not in siRNA mutants (dcl2-1, dcl3-1, rdr1-1, rdr2-1, and rdr6-15; Figure 4A). This trend was evident for all three categories of degradome-tag-identified miRNA targets, as well as during comparison of previously confirmed to newly confirmed miRNA targets. As expected, ta-siRNA target levels were generally increased in both miRNA mutants and the rdr6-15 mutant (Figure 4A).

We employed signal-to-noise analyses to determine empirical cutoffs for confident target identification using degradome data (Figure 2E, Figure S1). Potential targets that exceeded these cutoffs were not all necessarily false positives—in these
cases, however, additional data are required to confidently demonstrate targeting. Gene-specific 5'-RACE was used to individually test three cases involving potential targets that exceeded the degradome confidence thresholds: one that had previously been predicted but not confirmed (miR164/NAC2 [3]) and two novel, previously unpredicted miRNA/target interactions (miR390/At3G24660, miR396/At1G10120). In two of these cases, gene-specific 5'-RACE supported the hypothesis developed from degradome data (Figures 4B and 4C). Thus, potential targets that cannot be confidently confirmed with degradome data alone can still be validated by secondary, focused experiments. Further supporting this conclusion, we found another three transcripts with at least one tag matching a previously validated or predicted miRNA cleavage site in contexts that exceeded the confidence thresholds for degradome-based target identification (Table S3). Taken together with the ta-siRNA targets, our sampling of the Arabidopsis degradome discerned 121 miRNA and ta-siRNA cleavage sites within 99 distinct transcripts (Figure 2F; Table S3).

Conclusions

Direct sequencing of degradome tags derived from the 5' ends of uncapped mRNAs delivered an empirical overview of cleaved miRNA and ta-siRNA targets without computational predictions or overexpression. This methodology is likely to be broadly applicable to small-RNA target discovery and quantification for other organisms in which target cleavage is a frequent mode of repression. Degradome sequencing also provides information on the relative abundance of cleaved targets. The observation that Arabidopsis TAS cleavage products accumulated to high levels implies that stabilization and/or high-level expression of ta-siRNA precursors might play a role in promoting ta-siRNA biogenesis.

Our initial analysis of the degradome data focused on miRNA and ta-siRNA targets. However, most of the tags represented degradation intermediates of messages that were not targets of these small RNAs. We anticipate that analyses of these tags with a focus on other degradation pathways might provide general insights into mRNA turnover in plants and non-plant species, including those that do not undergo small-RNA-dependent mRNA cleavage. For example, degradome analysis of mutants in RNA metabolism, such as the XRN family of 5'-to-3' exonucleases [33, 34], are likely to yield rich insights into RNA turnover in a wide variety of species. Of particular interest is the observation that the 3' bias of degradome tags persisted even when random hexamers were used to prime reverse transcription. This could have resulted from in vitro degradation prior to polyA enrichment. However, the most typical sources of such artifacts (spontaneous and RNaseA-catalyzed hydrolysis) leave fragments that have 5' hydroxyls and therefore are excluded from our 5'-monophosphate-dependent protocol. Thus, although more experiments are clearly needed, our data hint at a role for endonucleases during mRNA turnover in Arabidopsis.

Experimental Procedures

Degradome Library Construction and Gene-Specific 5'-RACE

Degradome libraries were constructed by ligation of polyA-enriched RNA samples to a custom RNA adaptor containing a 3' Mmel site, followed by reverse transcription (RT), second-strand synthesis, Mmel digestion, ligation of a 5' dsDNA adapter, gel-purification, and PCR amplification. Different RT and second-strand synthesis were used during development of the protocol, as noted (Supplemental Experimental Procedures). Amplified degradome tag libraries were then sequenced with a 454/Roche GS20 genome analyzer or a Solexa/Illumina genome analyzer. Gene-specific RLM 5'-RACE was performed as described [15]; oligo sequences are listed in the Table S2.

Data Analysis

Raw reads were processed to remove 5' and 3' adaptor sequences; tags with sizes of 20 or 21 nt (the sizes expected from Mmel cleavage) were retained. Tags that did not correspond to structural RNAs (rRNA, tRNA, snRNA, snoRNA) were then mapped to the sense polarity of annotated Arabidopsis transcripts (TAIR 7 release); the abundance of tags that matched more than one transcript was repeat normalized. A 35–36 nt extended signature was derived from each transcript-matched tag by adding 15 nt of upstream sequence; these were then aligned to a set of annotated mature Arabidopsis miRNAs [35], an expanded set of expressed 20–22-mers derived from miRNA hairpins, or a set of expressed 20–22-mers derived from known producers of secondary siRNAs (Supplemental Experimental Procedures). Alignments where the 5' tag position aligned with the tenth nucleotide of a miRNA were retained and scored as in [2]. Control alignments were also performed with cohorts of permuted miRNA queries controlled for di- and tri-nucleotide composition [36]. Targets were categorized as I, II, or III as described in the Supplemental Experimental Procedures.
Accession Numbers
The processed degradome tags have been deposited with NCBI GEO under the accession number GSE11007.

Supplemental Data
Additional Experimental Procedures, one figure, and three tables are available at http://www.current-biology.com/cgi/content/full/18/10/.

Acknowledgments
We thank Chanseok Shin and Sumeet Gupta for assistance with Solexa/Illumina sequencing. This work was supported by a grant from the National Science Foundation (0718501) to M.J.A. and a grant from the National Institutes of Health (GM067031) to D.P.B.

References