TRUEE; a bioinformatic pipeline to define the functional microRNA targetome of Arabidopsis

Gigi Y. Wong and Anthony A. Millar[*] iD

Division of Plant Science, Research School of Biology, The Australian National University, Canberra ACT, 2601, Australia

SUMMARY

Central to plant microRNA (miRNA) biology is the identification of functional miRNA-target interactions (MTIs). However, the complementarity basis of bioinformatic target prediction results in mostly false positives, and the degree of complementarity does not equate with regulation. Here, we develop a bioinformatic workflow named TRUEE (Targets Ranked Using Experimental Evidence) that ranks MTIs on the extent to which they are subjected to miRNA-mediated cleavage. It sorts predicted targets into high (HE) and low evidence (LE) groupings based on the frequency and strength of miRNA-guided cleavage degradome signals across multiple degradome experiments. From this, each target is assigned a numerical value, termed a *Category Score*, ranking the extent to which it is subjected to miRNA-mediated cleavage. As a proof-of-concept, the 428 Arabidopsis miRNAs annotated in miRBase were processed through the TRUEE pipeline to determine the miRNA 'targetome'. The majority of high-ranking *Category Score* targets corresponded to highly conserved MTIs, validating the workflow. Very few *Arabidopsis-specific*, *Brassicaceae-specific*, or *Conserved-passenger* miRNAs had HE targets with high *Category Score*s. In total, only several hundred MTIs were found to have *Category Score*s characteristic of currently known physiologically significance MTIs. Although non-exhaustive, clearly the number of functional MTIs is much narrower than many studies claim. Therefore, using TRUEE to numerically rank targets directly on experimental evidence has given insights into the scope of the functional miRNA targetome of Arabidopsis.

Keywords: microRNA, microRNA-target interaction, bioinformatics, degradome, target prediction, Arabidopsis thaliana.

INTRODUCTION

MicroRNAs (miRNAs) are short non-coding RNAs of approximately 20–22 nt in length, which guide the RNA Induced Silencing Complex to repress target mRNAs via transcript cleavage and/or translational repression. Given that a high degree of complementarity between a plant miRNA-target pair is necessary for a strong repression (Liu et al., 2014; Schwab et al., 2005), numerous bioinformatic target prediction programs based on mismatch scoring schemas have been developed (Bonnet et al., 2010; Dai & Zhao, 2011; Sun et al., 2011). These scoring schemas consider the positions of mismatches, weightings for different mismatches (G:U pairs) and potential miRNA-binding site accessibility (Allen et al., 2005; Bonnet et al., 2010; Dai & Zhao, 2011; Mallory et al., 2004; Schwab et al., 2005; Sun et al., 2011). As further studies experimentally identified miRNA-target pairs with complementarity that would not be detected by these initial scoring schemas (Brousse et al., 2014; Zheng et al., 2012), this has justified relaxing complementarity requirements of the bioinformatic prediction of miRNA targets. For example, in an updated version of the most widely cited miRNA-target prediction tool, psRNATarget (version 2), the default parameter relating to complementarity (expectation score) was relaxed from 3 to 5 (Dai et al., 2018). Although this improved the prediction (or recall) of 143 of 147 experimentally validated Arabidopsis targets, there were almost 10 000 predicted targets in the bioinformatic output (Dai et al., 2018). Therefore, the output is overwhelmed with likely false positives.

It has also become evident that miRNA-target complementarity does not correlate with a functional miRNA-mediated regulatory outcome. For example, of a family of seven Arabidopsis *GAMYB-like* genes that contained analogous conserved miR159-binding sites, only two genes were found to be strongly miR159-regulated (Allen et al., 2007; Zheng et al., 2017). This, with the myriad of potential false positives, and the inability to rank targets on complementarity, highlights the limitations of identifying the cohort of

functional plant miRNA-target genes using bioinformatics alone, and the need to develop a miRNA-target prediction scoring schema independent of miRNA-target binding site complementarity.

Degradome sequencing has been used to complement bioinformatics approaches experimentally (Addo-Quaye et al., 2008; German et al., 2008). As miRNA guides the target cleavage precisely between the 10th and 11th nucleotide of the miRNA-binding site, sequencing and then mapping of the 5′ ends of degraded transcripts can accurately identify miRNA-guided cleavage products. Mapping of these degradome reads to individual transcripts form target-plots (T-plots), in which the relative abundance of reads mapping precisely to the cleavage site of a potential miRNA target (cleavage tag) can be compared with all other reads on the transcript (Addo-Quaye et al., 2008; German et al., 2008). Based on the frequency of the cleavage tag relative to the other reads in a transcript, these T-plots can then be placed into four categories [Category (Cat) 1–4], indicating the most confident (Cat 1) to least confidence (Cat 4) of a target being subjected to miRNA-guided cleavage (Addo-Quaye et al., 2008). Most canonical miRNA targets are Cat 1 targets (the cleavage tag being the most abundant read), and this is considered a hallmark of a validated target (German et al., 2008). There has now been extensive degradome analysis done in many plant species, and these data are available to determine which predicted miRNA targets have degradome signatures. For example, the Whole-degradome-based Plant MicroRNA-target Interaction Analysis Server (WPMIAS) makes data from numerous publicly available degradome libraries across diverse species easily accessible (Fei et al., 2020). However, detection of a degradome signal will be reliant on isolating RNA from a tissue in which both the miRNA and target mRNA are present, so any one single degradome library will only reflect miRNA-target interactions (MTIs) in these tissues, or in plants grown under those specific conditions. Moreover, degradome analysis only detects miRNA-mediated cleavage, but not other mechanisms, such as translational repression. Furthermore, as this is a biochemical signature, detection of a degradome signature does not necessarily equate to a MTI of physiological significance, nor can there be an arbitrary cut-off implying that any one particular degradome signature defines that gene as a 'real' miRNA target.

Adding to this uncertainty, is the identification of *bona fide* miRNAs themselves. Currently, miRBase is the go-to repository of experimentally identified miRNAs, with the latest release (v22) detailing 1000s of different miRNA sequences that have been reported across many diverse plant species (Kozomara et al., 2019). However, many publications have queried the quality and validity of these miRNA entries, which are mostly user-submitted, and have suggested the greater majority of entries are potentially false positives (Axtell & Meyers, 2018; Taylor et al., 2014).

Identifying high evidence (HE) miRNA targets for these miRNAs would help determine whether these miRNAs are genuine or potentially mis-annotated small RNAs (sRNAs).

This study develops a bioinformatic workflow that attempts to address the limitations outlined above. Long lists of putative targets from complementary-based predictions (psRNATarget), are filtered using an online server (WPMIAS) in which multiple degradome libraries can be searched for corresponding cleavage tags. The workflow then assesses the frequency and strength of the degradome signatures for each predicted target, which can then be arbitrarily sorted into HE and LE targets, as well as the non-arbitrarily ranking score based on the frequency and strength of degradome signatures for each predicted target. Using Arabidopsis as a proof-of-concept, this workflow was applied to gain a better understanding of the functional scope of a plant miRNome, by obtaining an accurate estimate of the total number of MTIs that have degradome signatures characteristic of known physiologically significant MTIs (i.e. MTIs that when manipulated can alter a trait). We call the collection of targets the 'miRNA targetome', which estimates the number of MTIs that have degradome characteristics of physiologically relevant MTIs.

RESULTS

Bioinformatic workflow to facilitate the identification of HE miRNA targets

A workflow was developed that sorts predicted miRNA targets into groups of either HE or low evidence (LE) targets, and then ranks the HE targets on the strength and frequency of their T-plots across degradome experiments. This workflow has been designated the 'Targets Ranked Using Experimental Evidence' (TRUEE), and combines miRBase to retrieve miRNA sequences (Kozomara et al., 2019), psRNATarget to predict miRNA targets (Dai et al., 2018), which are then subsequently used as input into WPMIAS (Fei et al., 2020) to retrieve all corresponding degradome data (Figure 1). Both psRNATarget and WPMIAS were chosen, as they are highly accessible via user-friendly webservers, and for psRNATarget, it is the most used and cited miRNA target prediction tool. Parameters are then implemented, filtering the degradome data to distinguish HE from LE targets of miRNA-mediated regulation, and then a simple formula for ranking HE targets. Below is the description of the input parameters and the rationale for developing the TRUEE workflow.

An experimentally validated set of Arabidopsis miRNA targets to benchmark TRUEE parameters

To develop this workflow, the input parameters were benchmarked against a compiled set of 106 experimentally validated sRNA targets from Arabidopsis based on the
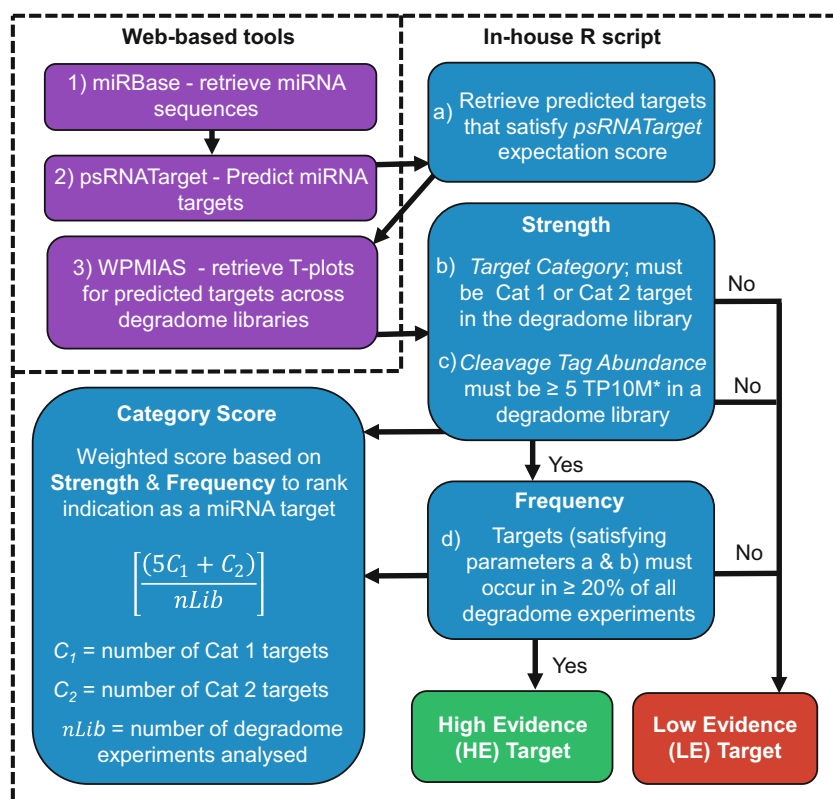
Figure 1. Workflow and parameters of TRUEE. Purple boxes indicate data retrieved from external web-based tools.

(a)–(d) Blue boxes indicate parameters, which were used to filter for HE targets and the *Category Score* (*Cat Score*) scoring schema. miRNAs were retrieved from miRBase (v22) (Kozomara et al., 2019). Potential miRNA target cleavage sites were then predicted using psRNATarget (Dai et al., 2018) and predicted targets with an expectation score ≤3.0 or ≤5.0 were used for further analysis. Degradome data for these cleavage sites were then retrieved using WPMIAS (Fei et al., 2020). *TP10M, Transcript Per 10 Million.

literature that we have termed the 'Validated Arabidopsis Target (VAT)' set. It is composed of targets of 28 miRNA families and one *trans*-acting siRNA (tasiRNA) family (the TAS3 phasing products, tasiARFs) and includes both widely and narrowly conserved MTIs (Table S1). To qualify as a validated target in this set, at least two independent lines of evidence from commonly used experimental approaches to identifying miRNA targets were required. This includes genetic evidence (altered mRNA/protein expression in *mirna* loss-of-function or miRNA overexpression plants, or expression of a miRNA-resistant target gene) or molecular evidence (degradome analysis, 5'-RACE cleavage assays or correlation of miRNA/target mRNA levels). The requirement of two independent lines of evidence to qualify for this list has resulted in a lower number of genes than other comparable lists in the literature (Dai et al., 2018; Folkes et al., 2012; Ma et al., 2018; Srivastava et al., 2014; Zheng et al., 2012).

Input parameters of TRUEE workflow

There are four parameters to consider: (a) *psRNATarget Expectation Score*; (b) *Cleavage Tag Abundance* (the number of degradome sequencing reads that coincide with the predicted cleavage site); (c) *Target Category* (corresponding to the Cat 1–4 categories of the T-Plots); and (d) *Library % Cut-off* [corresponding to the percentage of degradome libraries in which a predicted target occurs with the

defined (a), (b), and (c) parameters]. The optimal cut-offs for these parameters were determined via analysis of 61 Arabidopsis degradome libraries available on WPMIAS (Fei et al., 2020), from which identified targets were benchmarked against the VAT. The aim was to maximize the number of VAT targets identified, while minimizing additional targets that may represent either newly discovered targets or false positives (henceforth, potential targets).

(a) psRNATarget expectation score. The first parameter considered for TRUEE was the psRNATarget expectation score, a penalty score weighted on the number and position of mismatches between an miRNA and a predicted target gene (Dai et al., 2018). Using an expectation score too low will result in false negatives, while an expectation score too high will generate a multitude of false positives. The most recent version of psRNATarget (v2) has a default expectation score of 5, as some canonical target genes have expectation scores higher than 4 (Dai et al., 2018). As such, the expectation scores that were analysed ranged from 0 to 5. Using an expectation score of ≤5.0 predicted 2977 targets for the 29 miRNA/siRNA families, a >28-fold increase compared with the 106 targets of the VAT set. This predicted/validated target fold difference decreased with decreasing expectation score, although fewer of the VAT set were captured (Figure 2a). From the analysis, an expectation score of ≤3 appears optimal, resulting in a
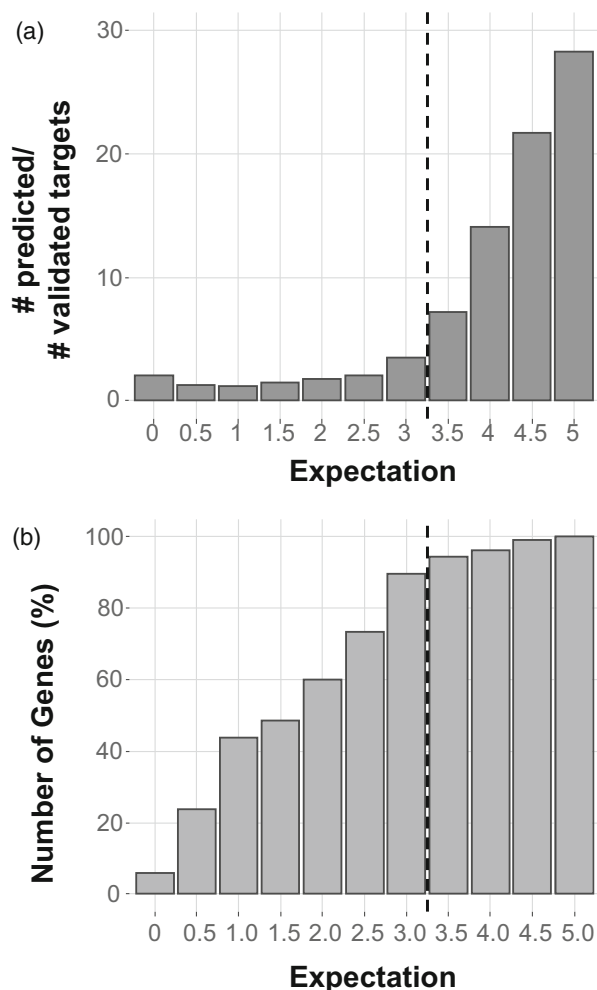
Figure 2. Determining the optimal psRNATarget Expectation Score cut-off. (a) Fold differences of the total number of targets predicted by psRNATarget over the number of targets in the VAT set identified at each expectation score cut-off. Black numbers above each bar are the total number of predicted targets/number of validated targets for each expectation score. (b) Cumulative percentage of the 106 targets of the VAT set that are retrieved at each expectation score cut-off. Red bar indicates the expectation score cut-off that was chosen for the TRUEE workflow. Total HE targets = Validated targets + potential targets.

relatively low-fold difference (3.5-fold), yet still included a large percentage of targets from the VAT set (89%) (Figure 2a). In comparison, using an expectation score any higher than 3 disproportionally increased the number of predicted targets captured (i.e. potential false positives), whereas an expectation score ≤2.5 failed to identify many of the VAT set (i.e. false negatives) (Figure 2b). For miRNAs with experimentally validated targets with expectation scores >3 (miR167, miR398, and miR408), the expectation score threshold was increased to ≤5.

Cleavage tag abundance. This parameter represents the number of cleavage tag reads for any given RNA, with the

greater the read, the more confidence for miRNA-mediated regulation. Therefore, targets with a low *Cleavage Tag Abundance* may represent fortuitous degradation events coinciding with the predicted cleavage site and thus represent a false positive. To determine an optimal value, TRUEE analysis was performed using a *Cleavage Tag Abundance* of ≥1, ≥5, and ≥10 when normalized to transcript per 10 million (TP10M), values that have been used in previous degradome studies (Jeong et al., 2013; Thody et al., 2020). This indicates that the degradome library for an RNA is only considered in analysis if the corresponding cleavage tag has at least 1, 5, or 10 TP10M, respectively. For this analysis, TRUEE was performed with variable *Library % Cut-offs* and *Target Categories*.

A *Cleavage Tag Abundance* of ≥1 TP10M identified the greatest number of the VAT set (Figure 3a–c). At a *Library % Cut-off* of 10%, nearly all of the VAT set was identified (97%). However, the number of potential targets was almost double the number of the VAT set (Figure 3a). Furthermore, across all *Library % Cut-offs*, the number of potential targets was many fold greater compared to when the *Cleavage Tag Abundance* was set to ≥5 and ≥10 TP10M (Figure 3a–c).

A *Cleavage Tag Abundance* of ≥5 TP10M appeared optimal. It identified a greater number of the VAT set compared with when setting of ≥10 TPM but had a greatly reduced number of potential targets compared with when the *Cleavage Tag Abundance* was set to ≥1 TPM (Figure 3a–c). Therefore, a *Cleavage Tag Abundance* of ≥5 appeared to minimize signals from potential random degradation, while maximizing identification of the VAT set.

Target category. For each predicted target RNA, the read-out of degradome analyses are T-plots. On WPMIAS, T-plots are placed into four *Target Categories* (1–4), with descending levels of confidence and so only inclusion of *Target Category* 1 and 2 targets are recommended and is set as the default (Fei et al., 2020). However, to identify the targets with greatest evidence, the stringency of TRUEE was increased by only including *Target Category* 1 targets. Results show that even at the lowest *Library % Cut-off* of 10%, only 75% of the VAT set were identified as HE targets (Figure 3d). This was 17 fewer targets compared with when using both *Target Category* 1 and 2 (Figure 3b). As only using *Target Category* 1 resulted in potentially many false negatives, for the third parameter, *Target Category* 1 and 2 were used to maximize the identification of the VAT set.

Library % Cut-off. The parameter, *Library % Cut-off*, assesses the frequency at which a predicted target satisfies the stated parameters in all degradome libraries analysed. The greater number of libraries a predicted target occurs in, the greater evidence it has as an miRNA target. As mentioned above, the *Library % Cut-offs* were 10%, 20%, 30%,
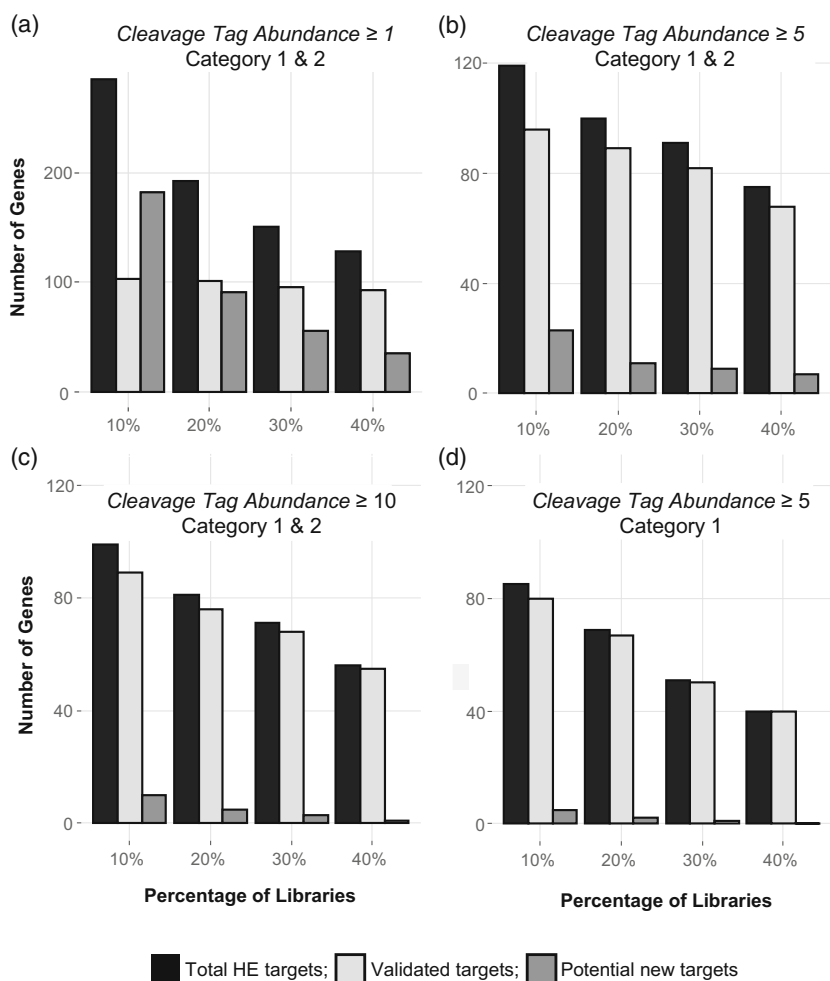
(a)

(b)

Figure 3. Number of genes defined as high-evidence (HE) targets, as determined by *Library % Cut-off*, *Cleavage Tag Abundance*, or *Target Category*.

'Total targets' indicate the total number of HE targets found by TRUEE. 'Validated targets' are the HE targets found in the VAT set. 'Potential targets' are HE targets that are not found in the VAT set. Note the differences in *y*-axis scales.

(c)

(d)

■ Total HE targets; □ Validated targets; ▨ Potential new targets

and 40%. Analysis was performed on the 61 Arabidopsis degradome libraries available on WPMIAS (Fei et al., 2020).

At a *Cleavage Tag Abundance* of ≥5, and a *Target Category* of 1 and 2, *Library % Cut-off* was assessed (Figure 3b). At the *Library % Cut-off* of 10%, 97 VATs and 22 new potential targets were identified. The number of VAT set and potential targets identified decreased to 90 and 10, respectively, at a 20% *Library % Cut-off*. These values continued to decrease with increasing *Library % Cut-off*. Based on this, a *Library % Cut-off* of 20% appears optimal, as most of the VAT set was identified as HE targets, with less than 50% of additional new potential targets compared with a *Library % Cut-off* of 10%.

In conclusion, using a *Library % Cut-off* of 20%, with a *Target Category* of 1 and 2, and a *Cleavage Tag Abundance* of ≥5 TP10M TRUEE maximizes the identification of VAT set targets, whilst minimizing potential targets.

Category Score, a simple scoring schema to rank HE targets

Within the HE targets identified from the above workflow, there will remain a large variation in the confidence and

extent to which the retrieved targets are being subjected to miRNA-mediated degradation. Therefore, ranking these HE targets based on the strength and frequency of the target across libraries will enable a clear indication of the confidence miRNA-mediated degradation for each target. As the *Target Category* approximates the extent of which miRNA-mediated cleavage contributes to RNA degradation of a target, a scoring schema was devised, which considers the number of libraries (frequency) a gene is found to be a Category 1 (C_1) or Category 2 (C_2) target with a *Cleavage Tag Abundance* of ≥5 (strength) (Figure 1). C_1 and C_2 were assigned the weighted values of 5 and 1, respectively. The heavier weighting for C_1 compared with C_2 targets was chosen considering the reduced confidence of the latter in reflecting miRNA-mediated degradation. The weighted number of libraries a gene was found to be a C_1 or C_2 target was then divided by the total number of libraries analysed (*nLib*). The category score (*Cat Score*) was calculated by:

$$Cat\ Score = \frac{(5C_1 + C_2)}{nLib}$$

This equation can give a maximum *Cat Score* = 5, which would mean the gene is a C_1 target in all degradome

libraries analysed. For such a scenario, both the miRNA and the target mRNA would need to be widely expressed to be detected in all degradome libraries.

Determining the *Cat Score* of the VAT set targets identified by TRUEE (Figure 3b) found that the *Cat Score* ranged from 4.15 to 0.12 (Table S1), enabling this ranking score to assess the extent of miRNA-mediated degradation rapidly for each HE target. Eight targets have a *Cat Score* ≥4, implying these MTIs are occurring strongly throughout Arabidopsis. Even within a family of miRNA targets, *Cat Score*s are highly variable. For instance, the *GROWTH REGUATORY FACTOR* (*GRF*) genes that are validated targets of miR396 have *Cat Score*s that vary from 4.02 (*GRF1*; At2g22840) to 0.12 (*GRF7*; At5g53660). Similarly, the *SQUAMOSA PROMOTER-BINDING PROTEIN LIKE* (*SPL*) genes that are validated targets of miR156 have *Cat Score*s that vary from 3.18 (*SPL13*; AT5G50570) to 0.33 (*SPL9*; AT2G42200), and the *TEOSINTE BRANCHED1, CYCLOIDEA*, and *PROLIFERATING CELL NUCLEAR ANTIGEN BINDING FACTOR* (*TCP*) genes that are validated targets of miR319 have *Cat Score*s that vary from 2.23 (*TCP4*; AT3G15030) to 0.28 (*TCP10*; AT2G31070). This enables clear identification of which paralogues with identical (or near identical) expectation scores are subjected to the strongest miRNA-mediated degradation. In addition, having a *Cat Score* ≥1 would indicate that the gene must be a C_1 target in at least one degradome

library. Of the 106 VAT set, 75 (70.8%) have a *Cat Score* of ≥1 (Table S1). This indicates that this cut-off will identify the majority of experimentally validated miRNA targets.

HE targets identified by TRUEE that are not in the VAT set

In the analysis above, TRUEE identified HE targets from Arabidopsis that were not present in the VAT set, and therefore may be new targets or false positives. These targets are analysed below in terms of their *Library % Cut-off*, their *Cat Score* and their highest *Target Category* (*Maximum Category*).

To maximize the potential of identifying new targets, the *Library % Cut-off* was lowered from 20% to 10%, resulting in the identification of a total of 22 new potential targets (Table 1). However, the 12 additional potential targets identified at the *Library % Cut-off* of 10%, all have a very low *Cat Score* (all but two were <0.5). This lends support to the justification of using the *Library % Cut-off* of 20% determined above. Of the 22 targets, only four had a *Cat Score* >1, and these were in 40% of libraries. Four of these targets showed evidence that was typical of canonical miRNA targets (Figure S1a-d). The highest ranked targets, *RNA PROCESSING FACTOR3* (*RPF3*) and *PENTATRICOPEPTIDE REPEAT1* (*PPR1*) are both family member homologues of genes in the VAT set with evidence of being miRNA targets, and therefore should have been included in the VAT

Table 1 Analysis of identified HE targets not present in the VAT set

miRNA	Target ID	Target description	Library % Cut-off				Max Cat	Cat S
			10	20	30	40		
miR161	AT1G62930	**RPF3, RNA Processing Factor 3**	X	X	X	X	1	2.311
miR161	AT1G06580	**PPR1, Pentatricopeptide Repeat 1**	X	X	X	X	1	1.180
miR167	AT5G58590	**RANBP1, RAN BINDING PROTEIN 1**	X	X	X	X	1	1.689
miR398	AT2G27530	**PGY1, PIGGYBACK 1**	X	X	X	X	1	1.246
miR398	AT1G03630	POR C,	X	X	X	X	2	0.492
miR408	AT3G01480	Cyclophilin 38	X	X	X	X	2	0.492
miR168	AT3G58030	**MUSE1**	X	X	X		1	0.852
miR408	AT1G68010	HPR, HYDROXYPYRUVATE REDUCTASE	X	X	X		2	0.328
miR395	AT1G50930	**Serine/threonine-kinase**	X	X			1	0.541
miR396	AT3G19400	Cysteine proteinases superfamily protein	X	X			1	0.393
miR161	AT1G64583	**Tetratricopeptide repeat (TPR)-like**	X				1	0.721
miR164	AT3G12977	**NAC (No Apical Meristem) domain**	X				1	0.525
miR167	AT1G51760	IAR3, IAA-Alanine Resistant 3,	X				1	0.295
miR167	AT5G10550	GTE2, Global Transcription Factor E2	X				2	0.148
miR172	AT3G05530	ATS6A.2, RPT5A, TRIPLE-A ATPASE 5A	X				2	0.131
miR396	AT1G48380	HYP7, HYPOCTYL 7, ROOT HAIRLESS 1	X				1	0.262
miR396	AT1G60140	TPS10, Trehalose Phosphate Synthase	X				1	0.295
miR398	AT4G24280	cpHsc70-1, chloroplast heat shock 70–1	X				2	0.164
miR398	AT5G14550	Beta-1,6-*N*-acetylglucosaminyltransferase	X				2	0.115
miR408	AT5G21930	PAA2, P-type ATPase of Arabidopsis 2	X				2	0.148
miR408	AT2G47900	TLP3, TUBBY LIKE PROTEIN 3	X				2	0.131
miR408	AT4G34230	CAD5, Cinnamyl Alcohol Dehydrogenase 5	X				2	0.131

Library % Cut-off threshold meet for each HE target is indicated by 'X'. Genes in bold type indicate HE targets that were found to possess T-plots comparable with those in the VAT set (Figure 4, S1). *Maximum Category* (*Max Cat*) indicates whether the highest T-plot Category found across degradome libraries is Cat 1 or Cat 2 and *Cat S* is *Category Score*. HE, high evidence.

set (Allen et al., 2004; Howell et al., 2007). However, no clear previous evidence exists for the miR167 target, *RNA BINDING PROTEIN 1* (*RANBP1*) or the miR398 target, *PIGGYBACK1* (*PGY1*), both of which had a *Maximum Category* of 1 with a high *Cleavage Tag Abundance* (Figure 4a,b). Both T-plots were comparable with that of previously validated miR167 target, *AUXIN RESPONSE FACTOR 6* (*ARF6*), or the miR398 target, *COPPER/ZINC SUPEROXIDE DISMUTASE 1* (*CSD1*) (Figure S1e–h), suggesting an analogous degree of miRNA-mediated regulation in this library. Neither the miRNA-binding site in *RANBP1* nor *PGY1* was conserved beyond the *Brassicaceae* family (Figure S2), and thus may explain why targets such as these have not been previously identified by bioinformatic tools that rely on conservation (Chorostecki et al., 2012; Ma et al., 2018).

In contrast, although *PROTOCHLOROPHYLLIDE OXIDOREDUCTASE C* (*POR C*) and *CYCLOPHILIN 38* (*CYP38*) were also found in more than 40% of libraries, they had comparatively lower *Cat Scores* (<0.5). In addition, their *Maximum Category* was 2, and subsequent investigation of their T-plots revealed *Cleavage Tag Abundances* to be comparable with other degradome reads mapping at many different nucleotide positions throughout the transcript (Figure 4c,

d). This suggests the occurrence of the high *Cleavage Tag Abundance* in a high percentage of degradome libraries may be due to RNA degradation pathways other than miRNA-mediated regulation.

Despite occurring in fewer libraries than *PORC1* and *CYP38*, four additional targets, *MUSE1*, *SERINE/ THREONINE-KINASE*, a *TPR* homologue, and a *NAC* homologue, have greater *Cat Scores* and their *Maximum Category* was 1. Again, both *TPR* and *NAC* are family members of genes previously found to be miRNA-regulated, but for *MUSE1* and *SERINE/THREONINE-KINASE* there is no known evidence for miRNA regulation, and both display T-plots characteristic of canonical targets (Figure 4e,f). This suggests that even at a *Library % Cut-off* of 10%, by considering targets with the highest *Cat Scores*, TRUEE is able to identify targets with T-Plots highly indicative of miRNA-mediated cleavage. Therefore, while also considering *Library % Cut-off* and *Maximum Category*, *Cat Score* enables the ranking of targets, which should be given priority for further investigation regarding potential miRNA regulation. In this regard, a *Library % Cut-off* of 10%, in addition to a *Cat Score* cut-off of ≥0.5, may be used as an alternate set of parameters to identify HE targets.
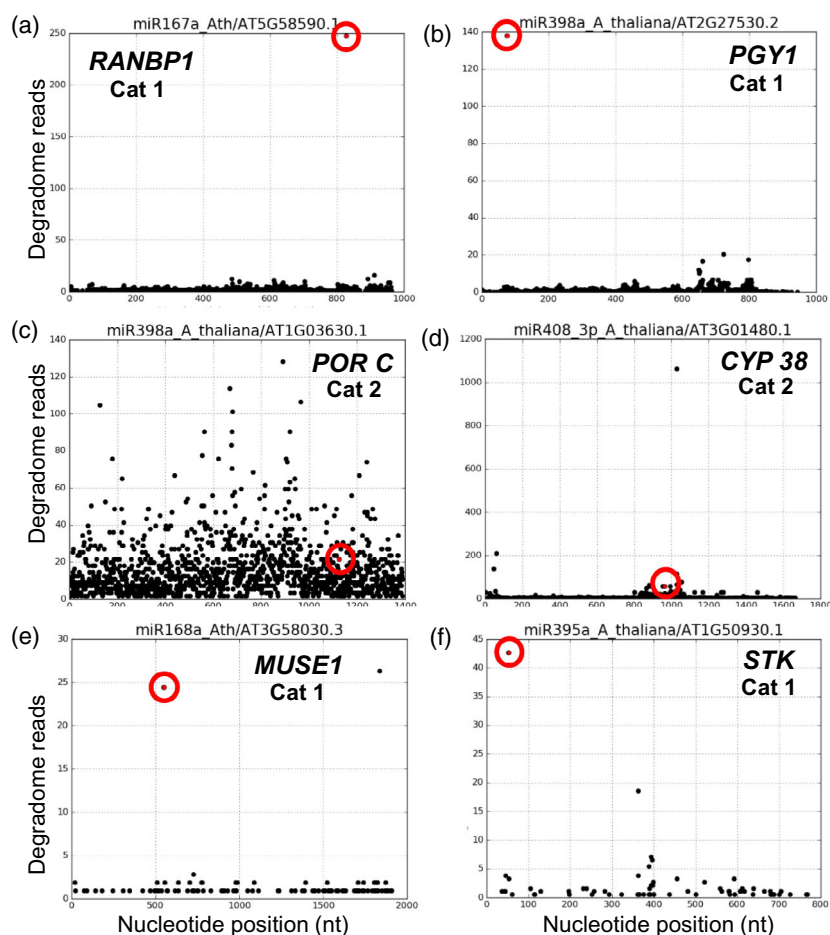

Figure 4. T-plots of high-evidence targets not from the VAT set.
T-plots of (a) *RNA BINDING PROTEIN 1* (*RANBP1*); (b) *PIGGYBACK1* (*PGY1*) that encodes a ribosomal protein L10aP; (c) *PROTOCHLOROPHYLLIDE OXIDOREDUCTASE C* (*POR C*); (d) *CYCLOPHILIN 38* (*CYP38*); (e) *MUSE1,* encodes a RING domain E3 ligase; (f) *VASCULAR-RELATED UNKNOWN PROTEIN 2*, which encodes a serine/threonine-kinase (*STK*). T-plot from the degradome library with the highest *Maximum Category* and highest *Cleavage Tag Abundance* was used for each miRNA target. Cleavage tag is circled in red. T-plot figures were adapted from WPMIAS (Fei et al., 2020).

Modification of TRUEE to consider narrow spatial and temporal expression

At a *Library % Cut-off* of 20%, only 16 of 106 of the VAT set were not identified by TRUEE (Table S1). Several of these are known canonical miRNA targets, most of which are only regulated under specific environmental/stress conditions and so are likely being overlooked by TRUEE due to insufficient degradome libraries under the specific environmental conditions that these MTIs occur. To overcome this, the analysis of select degradome libraries from a particular treatment or tissues may better detect these narrow spatial or temporal MTIs. For instance, narrowing TRUEE only to analyse root libraries finds large increases to the *Cat Score* of *SERINE/THREONINE-KINASE* (0.5–4.3), and a *NAC* homologue (At3g12977) (0.5–3.33), implying these MTIs occur preferentially in roots (Table 2). Therefore, by filtering which degradome libraries are analysed, TRUEE can allow the identification of more subtle MTIs, such as spatially specific MTIs.

Defining the Arabidopsis miRNA targetome

Most of the literature on Arabidopsis MTIs corresponds to the 29 miRNA and tasiRNA families whose targets compose the VAT set. However, this is only a small subset of Arabidopsis miRNAs, as there are 428 annotated miRNAs composing 231 families in Arabidopsis as reported in miR-Base v22 (Kozomara et al., 2019). Therefore, to gain a better understanding of the scope of MTIs in Arabidopsis, TRUEE was applied to this complete set of putative Arabidopsis miRNAs (Table S2). The analysis was performed on 34 Arabidopsis degradome libraries, which appeared to be the limit to which WPMIAS could process the 400 miRNAs. The initial analysis was performed at a *Library % Cut-off* of 10% to assist the identification of more subtle MTIs (henceforth, low stringency). The collection of HE targets identified by this analysis is defined as the 'miRNA targetome'.

Number of HE targets per miRNA family strongly correlates with miRNA conservation

Given the large numbers of miRNAs, they were first sorted into groups based on conservation (Table 3). These conservation-based groups were: (i) miRNAs that have only been identified in *Arabidopsis thaliana* (132 families; referred to as '*A. thaliana-specific*'); (ii) miRNAs conserved in at least one other species of the *Brassicaceae* (53 families; referred to as '*Brassicaceae-specific*'); these included many miRNAs that have only been found in *A. thaliana* and *Arabidopsis lyrata*; and (iii) miRNAs conserved across multiple clades of land plants (27 families; referred to as 'conserved'), as defined in Axtell & Meyers (2018).

Table 2 Additional TRUEE targets not in the VAT set from only analysing root-specific degradome libraries

miRNA	Target ID	Target description	Library % Cut-off				Max Cat	Cat S
			10	20	30	40		
miR161	AT1G06580	PPR1, Pentatricopeptide Repeat 1	X	X	X	X	1	4.167
miR161	AT1G62930	Tetratricopeptide repeat (TPR)-like superfamily protein	X	X	X	X	2	0.667
miR164	**AT3G12977**	**NAC (No Apical Meristem) domain**	**X**	**X**	**X**	**X**	**1**	**3.333**
miR172	AT3G05530	REGULATORY PARTICLE TRIPLE-A ATPASE 5A	X	X	X	X	2	0.5
miR395	**AT1G50930**	**Serine/threonine-kinase**	**X**	**X**	**X**	**X**	**1**	**4.333**
miR396	AT3G19400	Cysteine proteinases superfamily protein	X	X	X	X	2	0.5
miR398	AT2G27530	PGY1, PIGGYBACK 1	X	X	X	X	2	1
miR396	AT1G60140	TPS10, TREHALOSE PHOSPHATE SYNTHASE	X	X	X		2	0.333
miR398	AT4G26230	Ribosomal protein L31e family protein	X	X	X		2	0.333
miR408	AT4G34230	CINNAMYL ALCOHOL DEHYDROGENASE 5	X	X	X		2	0.333
miR857	AT5G36880	ACS, acetyl-CoA synthetase	X	X	X		2	0.333
miR159	AT2G21600	endoplasmatic reticulum retrieval protein 1B	X				2	0.167
miR159	AT3G08850	RAPTOR1B	X				2	0.167
miR161	AT1G64583	Tetratricopeptide repeat (TPR)-like protein	X				1	0.833
miR163	AT5G38100	SABATH family methyltransferase.	X				1	0.833
miR166	AT1G07810	RNA-binding (RRM/RBD/RNP motifs) protein	X				2	0.167
miR167	AT3G07810	RNA-binding (RRM/RBD/RNP motifs) protein	X				2	0.167
miR167	AT3G52190	Phosphate transporter traffic facilitator1	X				2	0.167
miR168	AT3G58030	MUSE1	X				1	0.833
miR398	AT1G75270	DHAR2, dehydroascorbate reductase 2	X				2	0.167
miR398	AT2G43900	Endonuclease/exonuclease/phosphatase	X				2	0.167
miR408	AT2G47900	TLP3, TUBBY LIKE PROTEIN 3	X				2	0.167

Library % Cut-off threshold meet for each high-evidence target is indicated by 'X'. Genes in bold type indicate high-evidence targets that were found to possess T-plots comparable with those in the VAT set. *Maximum Category* (*Max Cat*) indicates whether the highest T-plot Category found across degradome libraries is Cat 1 or Cat 2 and *Cat S* is *Category Score*.

Table 3 Low and high stringency miRNA targetome of Arabidopsis

miRNA group	miRNA[a] families	Predicted[b] targets	HE targets[c]		miRNA families with[d] HE targets	
			Low	High	Low	High
Conserved-guide	27	493	120	82	24	20
Conserved-passenger	19	478	27	6	10	4
Brassicaceae-specific	53	983	57	19	30	10
Arabidopsis thaliana-specific	132	1907	88	29	44	14
Total	231	3478	292	136	108	48

Abbreviation: HE, high evidence.
[a]Total number of miRNA family entries for *A. thaliana* on miRBase v22 (Kozomara et al., 2019).
[b]Number of predicted targets based on default settings of psRNATarget (Dai et al., 2018).
[c]Total number of HE targets identified using high and low stringency parameters in TRUEE.
[d]Number of miRNA families with HE targets using high and low stringency parameters in TRUEE.

Conserved miRNAs were further grouped into *Conserved-guide* (27 families) and *Conserved-passenger* (19 families) as there is evidence that the miRNA passenger strand (miRNA*) also have regulatory roles (Liu et al., 2017).

In total, 3478 targets were predicted for the 428 Arabidopsis miRNAs by psRNATarget (Table 3). Of these, TRUEE identified 292 as HE targets at a low stringency *Library % Cut-off* of 10% (Table 3). Therefore, the number of HE targets is at least an order of magnitude lower than the number of predicted targets. The *Conserved-guide* miRNA grouping had the greatest number of HE targets (41%), followed by the *A. thaliana-specific* (30%), *Brassicaceae-specific* (20%), and *Conserved-passenger* (9%) families. Therefore, HE targets of the *Conserved-guide* miRNA group contributes the most to the Arabidopsis targetome, despite this grouping having far fewer miRNA families than the *Brassicaceae-specific* or *A. thaliana-specific* groupings (Table 3).

Finally, TRUEE only identified HE targets in 108 of the 231 Arabidopsis miRNA families (Table 3). Whereas only 33% of *A. thaliana-specific* families had HE targets, the majority of families in the *Brassicaceae-specific* (30 of 53; 57%), *Conserved-passenger* (10 of 19; 53%), and *Conserved-guide* (24 of 27; 89%) groupings, had HE targets. Therefore, as the conservation of a miRNA family increased, the likelihood it had an HE target increased.

Upon analysing the distribution of HE targets by individual miRNA families, it was found that most *Conserved-guide* families had multiple HE targets (Figure 5a). In contrast, most *A. thaliana-specific* and *Brassicaceae-specific* families only had single HE targets, although a few of these families had many HE targets. The *Cat Scores* of the HE targets were determined for each conservation grouping (Figure 6a). It was found that the *Cat Scores* for HE targets from the *Conserved-guide* families were the most evenly distributed, ranging from 0.2 to 4.3. By contrast, the number of HE targets for *A. thaliana-specific* and *Brassicaceae-specific* families plateaued around a *Cat* *Score* of 0.75, and both had relatively few HE targets with a *Cat Score* >1 (Figure 6a). In particular, *Conserved-passenger* families had the fewest HE targets with a *Cat Score* ≥0.5, where none exceeded 0.7 (Figure 6a; Table S3). Therefore, most of the HE targets with high *Cat Scores* correspond to targets from the *Conserved-guide* families.

Most HE targets with the highest *Cat Scores* correspond to previously characterized MTIs

Next, the HE targets of *Conserved-guide* miRNA families were classified as either belonging to a conserved target family, or corresponding to being a non-conserved target (Table S4). Most of the HE targets (86%) from conserved target families had a *Cat Score* ≥0.5 (Figure 6b; Table S4). On the other hand, most non-conserved HE targets (77%) had a *Cat Score* <0.5 (Figure 6b). For non-conserved targets, the highest *Cat Score* was 2.6, whereas many conserved targets exceeded this value, with the highest *Cat Score* being 4.3.

Of the conserved HE targets that had *Cat Scores* ≥0.5, all but two were part of the VAT set (Table S1), indicating the vast majority of these MTIs have been previous characterized. Interestingly, the only two HE targets not part of the VAT set were both homologues of characterized targets, i.e. an *NAC* homologue (AT3G12977; miR164) and an *SBP-DOMAIN* homologue (AT5G50670; miR156). For non-conserved targets, the top two HE targets with the highest *Cat Scores*, *RELATED TO AP2 12* (*RAP2.12*; AT1G53910) and *CRY2-INTERACTING BHLH4* (*CIB4*; AT1G10120), were also part of the VAT set.

This was also true for the *Brassicaceae-specific* miRNA targets where 15 of the 19 of the HE targets with a *Cat Score* ≥0.5 were previously reported as miRNA targets in the literature (either part of the VAT, or otherwise) or were related to these targets (e.g. miR161:*PPR/TPR* family; miR163:*SAMT* family) (Table 4, Table S5). Furthermore, the *Brassicaceae-specific* HE targets with the highest *Cat Scores* also corresponded to the most highly studied MTIs, such as the miR161:*PPR/TPR* module and miR824:*AGL16*

Figure 5. Arabidopsis miRNA targetome.
High-evidence (HE) targets identified for all Arabidopsis miRNA families by conservation group at: (a) low stringency; (b) high stringency. Families are grouped by conservation so that pink indicates *Arabidopsis thaliana-specific* miRNA families, green indicates *Brassicaceae-specific* miRNA families, blue indicates *Conserved-guide* miRNA families, and purples indicates *Conserved-passenger* miRNA families. Each bar represents the number of HE targets per miRNA family when analysed by TRUEE.
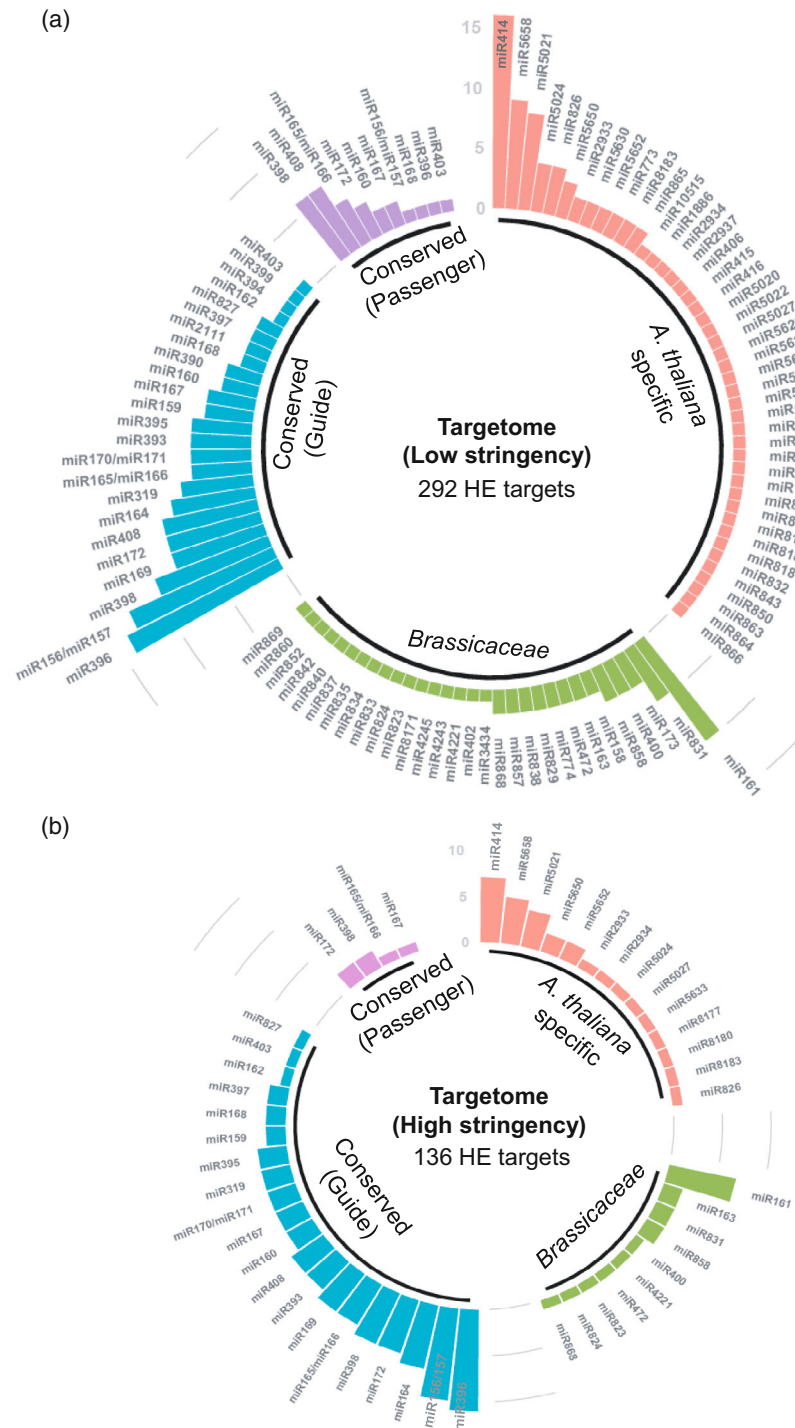
module (Howell et al., 2007; Kutter et al., 2007; Szaker et al., 2019). By contrast, only four of the 38 *Brassicaceae-specific* HE targets with a *Cat Score* <0.5 were part of the VAT set. Together, these results show that, for the *Conserved-guide* and *Brassicaceae-specific* miRNA groupings, most HE targets with the highest *Cat Score*s are well characterized miRNA targets, or are related to these

targets. This argues that the scope of functional MTIs in Arabidopsis has largely been identified.

Many HE targets of *Arabidopsis thaliana-specific* miRNAs are diverse genes with trinucleotide repeats

By contrast, most of the HE targets for the *A. thaliana-specific* families have not been previously described, and
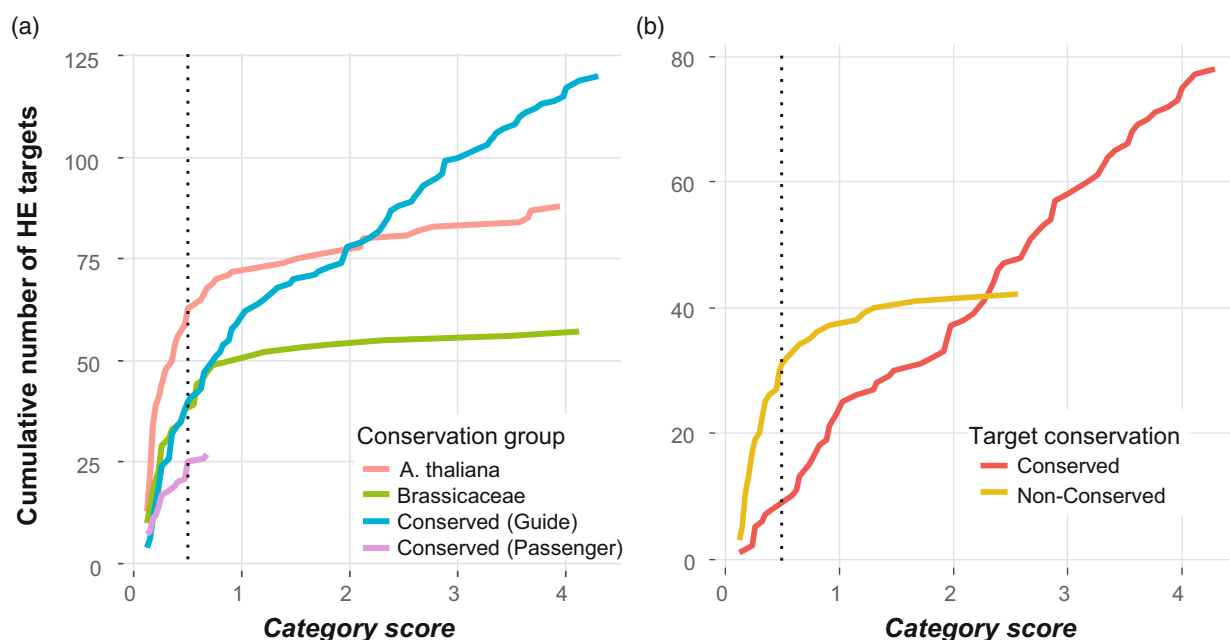
(a)

(b)

Figure 6. Distribution of high-evidence (HE) target *Cat Scores* that relate to conservation.
(a) Cumulative number of HE targets against *Cat Score* of the different miRNA conservation groups. Dotted line indicates a *Cat Score* cut-off of 0.5.
(b) Cumulative number of HE targets against *Cat Score* for conserved and non-conserved targets of the *Conserved-guide* miRNA families.

Table 4 HE targets of *Brassicaceae-specific* miRNA families with a *Cat Score* ≥0.5

miRNA	Target ID	*Cat Score*	Previously characterized	Gene symbol	Target description
miR161	AT5G41170	4.118	Yes[a]		PPR-like superfamily protein
miR824	AT3G57230	3.471	Yes[a]	AGL16	AGAMOUS-like 16
miR823	AT1G69770	2.294	Yes[a]	CMT3	CHROMOMETHYLASE 3
miR161	AT1G06580	1.794	Yes[a]		PPR superfamily protein
miR472	AT5G43740	1.529	Yes[b]		Disease resistance protein (CC-NBS-LRR class) family
miR163	AT1G66700	1.206	Yes[a]	PXMT1	*S*-adenosyl-L-methionine-dependent methyltransferase
miR4221	AT1G20500	1.059	No		AMP-dependent synthetase and ligase family protein
miR161	AT1G64583	1.059	Yes[a]		TPR-like superfamily protein
miR400	AT1G62720	0.735	Yes[c]		PPR-like superfamily protein
miR831	AT3G56020	0.735	No		Ribosomal protein L41 family
miR868	AT1G18270	0.676	No		Ketose-bisphosphate aldolase class-II family protein
miR831	AT3G08520	0.676	No		Ribosomal protein L41 family
miR858	AT2G47460	0.618	Yes[d]	MYB12	myb domain protein 12
miR161	AT1G62910	0.588	Yes[a]		PPR superfamily protein
miR161	AT1G62914	0.588	Yes[a]		PPR repeat-containing protein
miR161	AT1G62930	0.588	Yes[a]		TPR-like superfamily protein
miR161	AT1G63130	0.588	Yes[a]		TPR-like superfamily protein
miR163	AT3G44860	0.588	Yes[a]	FAMT	Farnesoic acid carboxyl-*O*-methyltransferase
miR858	AT4G26930	0.559	MYB12 related	MYB97	myb domain protein 97

[a]Part of or related to targets in the VAT set
[b]Boccara et al., 2014
[c]Park et al., 2014
[d]Sharma et al., 2016
Abbreviations: HE, high evidence; PPR, pentatricopeptide repeat; TPR, tetratricopeptide repeat.
List of HE targets with [a] indicating that the target is part of, or related to genes in the VAT set. [b, c, d] Indicate genes that are not in the VAT set but are supported in literature to have genetic and molecular evidence as miRNA targets.

none were present in the VAT set. Of the 29 HE targets with *Cat Scores* ≥0.5, 16 were targets of three miRNAs (miR414, miR5021, and miR5658), with some of these HE targets having very strong *Cat Scores* (Table 5). Curiously, all three miRNAs are mainly composed of repeated trinucleotide sequences, which was also characteristic of their binding sites in their HE targets. In addition, the HE targets of miR414, miR5021, and miR5658 did not appear to be related in identity, but had rather diverse mRNA targets containing these trinucleotide repeats.

A high stringency Arabidopsis miRNA targetome

Given the above analyses have shown the majority of MTIs with strong experimental evidence correspond to HE targets with *Library % Cut-off* of 10% and a *Cat Score* cut-off of ≥0.5, imposing these cut-offs appears justified in terms of capturing MTIs with the highest evidence in a bid to define a high stringency Arabidopsis targetome. Using these parameters, only 136 HE targets are identified, with the *Conserved-guide* HE targets now making up the majority of targets (60%), followed by *A. thaliana-specific* (21%), *Brassicaceae-specific* (14%), and *Conserved-passenger* (5%) families (Figure 5b). In

this high stringency targetome, the number of miRNA families with HE targets dropped to only 48 of the 231 miRNA families (21%), with the *A. thaliana-specific* (14 of 132; 11%), *Brassicaceae-specific* (10 of 53; 19%) and *Conserved-passenger* (4 of 19; 21%) groupings now all having a minority of miRNA families with HE targets. This reduction stems mainly from the exclusion of single HE target-miRNA interactions being filtered from this high stringency Arabidopsis targetome (Figure 5b). By contrast, the majority of *Conserved-guide* families still had HE targets (20 of 27; 74%). Hence, TRUEE is filtering out a set of targets that is in line with the long-standing notion that most functional MTIs are conserved (Axtell, 2008), rather than the possibility of promiscuous targeting of many mRNAs via a large and diverse miRNome (Brodersen & Voinnet, 2009).

DISCUSSION

A central question of plant miRNA biology is the identification of functionally important (physiologically relevant) MTIs. Here, TRUEE has been developed to filter and rank MTIs based on experimental evidence. This was then applied to Arabidopsis as a proof-of-concept to define an

Table 5 HE targets of *Arabidopsis thaliana*-specific miRNA families with a *Cat Score* ≥0.5

miRNA	Rep. miRNA	Target ID	Cat score	Gene symbol	Target description
miR414	Yes	AT5G55580	3.941		Mitochondrial transcription termination factor
miR5021	Yes	AT2G40520	3.676		Nucleotidyltransferase family protein
miR5021	Yes	AT5G24670	3.676		Cytidine/deoxycytidylate deaminase
miR5021	Yes	AT1G03190	3.647	UVH6	RAD3-like DNA-binding helicase protein
miR5021	Yes	AT3G23890	3.559	TOPII	Topoisomerase II
miR414	Yes	AT5G40340	2.765		Tudor/PWWP/MBT superfamily protein
miR8177		AT1G15710	2.618		Prephenate dehydrogenase family protein
miR5652		AT1G62670	2.529	RPF2	rna processing factor 2[a]
miR414	Yes	AT3G11810	2.118		(1 of 2) PTHR33133:SF7 - F26K24.10
miR414	Yes	AT5G55300	2.118	TOP1ALPHA	DNA topoisomerase I alpha
miR414	Yes	AT1G16150	2.088	WAKL4	Wall associated kinase-like 4
miR414	Yes	AT1G60220	1.853	ULP1D	UB-like protease 1D
miR5658	Yes	AT1G73710	1.706		Pentatricopeptide repeat (PPR) superfamily
miR5658	Yes	AT4G11600	1.5	GPX6	Glutathione peroxidase 6
miR5658	Yes	AT5G56860	1.382	GNC	GATA type zinc finger transcription factor
miR5633		AT2G35670	1.147	FIS2	VEFS-Box of polycomb protein
miR5652		AT5G16640	0.912		Pentatricopeptide repeat (PPR) superfamily[a]
miR5027		AT1G07610	0.882	MT1C	Metallothionein 1C
miR2933		AT4G32390	0.765		Nucleotide-sugar transporter family protein
miR5658	Yes	AT2G32310	0.735		CCT motif family protein
miR2934		AT5G03650	0.676	SBE2.2	Starch branching enzyme 2.2
miR8183		AT5G04220	0.676	SYTC	Calcium-dependent lipid-binding family
miR414	Yes	AT5G64830	0.676		PCD 2 C-terminal domain-containing protein
miR5658	Yes	AT4G20070	0.618	AAH	Allantoate amidohydrolase
miR5650		AT5G03240	0.618	UBQ3	Polyubiquitin 3
miR826		AT1G09730	0.5	ASP1	Arabidopsis sumo protease 1
miR5024		AT3G57290	0.5	EIF3E	Eukaryotic translation initiation factor 3
miR8180		AT4G29350	0.5	PFN2	Profilin 2
miR5650		AT5G20620	0.5	UBQ4	Ubiquitin 4

List of the HE targets, and indication of whether it is regulated by a miRNA with trinucleotide repeats (Rep. miRNA), *Cat Score*, and gene annotation. None of these targets are in the VAT set.
Abbreviations: HE, high evidence.

accurate estimation of the number of functional MTIs in a plant, termed the 'miRNA targetome'. Although non-exhaustive, the approach suggests Arabidopsis would have no more than 300 functionally MTIs, and likely, considerably fewer. In the context of this paper, functionally important refers to an MTI that if altered, would alter a plant trait (i.e. have a physiological impact).

TRUEE: a simple approach to rank MTIs independently of miRNA-target complementarity

We aimed to develop a simple bioinformatic approach based on currently available and widely utilized online tools. First, psRNATarget is the most widely used and cited plant miRNA target prediction tool that has been recently updated (Dai et al., 2018). It is a highly accessible, user-friendly webserver, and is compatible with WPMIAS (Fei et al., 2020). WPMIAS is also a highly accessible, user-friendly webserver and currently is the simplest tool to analyse multiple degradome libraries.

Unlike previous miRNA target prediction tools that are based on miRNA-target complementarity, the scoring schema of TRUEE is derived solely from degradome data. It is based on the strength and frequency of a target's T-plots across multiple degradome libraries, from which the *Cat Score* can be derived, a metric that directly relates to extent of miRNA-mediated cleavage. Like WPMIAS (Fei et al., 2020), *Target Categories* 3 and 4 were not considered strong enough evidence for miRNA-mediated cleavage (so are essentially given a weighted score of 0). This approach is justified in that using only *Target Categories* 1 and 2 was sufficient to identify the vast majority of the VAT set (Figure 3). *Target Category* 1 was given an arbitrary weighted value 5-fold greater than *Target Category* 2 plots given the much greater confidence that these signals are derived from miRNA-mediated regulation. This is illustrated in Figure 4, where it is unclear whether the *Target Category* 2 signals for *POR C* and *CYCLOPHILIN 38* is derived from miRNA-mediated cleavage or other degradation mechanisms.

Finally, if TRUEE is compared with data from the most recently published tool, TarHunter (Ma et al., 2018), it appears TRUEE is identifying less false positives. Using TarHunter in the ortho_mode (protein and nucleotide sequence at the target site is conserved) and most stringent number of mismatches, TarHunter identifies 59 targets for the conserved set of miRNAs in Arabidopsis (http://www.biosequencing.cn/TarHunter/ath.html). Of these, 17 (29%) are not present in the VAT set. Therefore, even at the highest stringency of TarHunter, it appears that TRUEE is identifying proportionally fewer false positives.

Limitations of TRUEE

First, given the presence of a degradome signal requires both the presence of the miRNA and transcription of the target mRNA, TRUEE will preferentially detect MTIs that are widespread, and potentially miss those MTIs that have a narrow temporal and spatial occurrence. Both the canonical nutrient-dependent miR399:*PHO2* and miR395:*SULTR2* MTIs had low *Cat Score*s (0.265, Table S4), as the majority of the degradome analyses have likely not been performed when conditions exist for these MTIs. To offset this, a selection of particular degradome libraries (conditions or tissues), potentially may help identify these narrow MTIs, as was demonstrated for the root MTIs. The current code (published on Open Science Framework) is customizable, so that the analysis of any subset(s) of degradome libraries is possible. As most degradome experiments are only a snapshot of miRNA-mediated activity at one particular developmental stage or growth condition, obviously the larger the number of degradome libraries analysed, the more comprehensive a picture will be of the miRNA targetome.

Secondly, TRUEE will not detect targets that are regulated solely by translational repression. However, this may be inconsequential, as nearly all canonical targets were identified using TRUEE, validating the use of this approach to detect the majority of miRNA targets. This is consistent with the observation that canonical targets that are known to undergo translational repression, are also cleaved by the miRNA (for review see Yu et al., 2017), implying there is no strong evidence that miRNA targets are solely regulated by translational repression.

Functional miRNA targetome of Arabidopsis

Currently, the functional scope of the plant miRNome remains contentious. As many studies claim that most miRNAs in a plant are lineage-specific (Cui et al., 2017), and that many of these MTIs are evolutionarily fluid (Smith et al., 2015), these notions align with the hypothesis that there are likely 100s of functional miRNAs and 1000s of MTIs. However, other researchers are more cautious, and question the validity of many of these species-specific miRNAs that have been annotated on databases such as miRBase (Axtell & Meyers, 2018) or argue that most non-conserved miRNAs are likely to be evolutionary transient with no functional targets (Axtell, 2008; Cuperus et al., 2011). In this study, by determining how many functional MTIs there are in a plant and the proportion of these that correspond to non-conserved miRNAs, we aimed to add weight to which hypothesis is more likely.

Our findings support the notion that only several 100 MTIs of functional importance are present in a plant (Li et al., 2014; Taylor et al., 2014). Although previously proposed, the value of reiterating this notion has merit in that many current studies assume there are 1000s of MTIs of functional importance as predicted by bioinformatics (Bülow et al., 2012; Dai et al., 2018; Fei et al., 2020; Kozomara et al., 2019; Lindow et al., 2007; Lindow & Krogh, 2005). Moreover, without the filters imposed by TRUEE, studies based on degradome data also claim 1000s

of targets (e.g. WPMIAS reports >10 000 MTIs in *Oryza sativa* from an analysis of 738 miRNAs; Fei et al., 2020). Our findings align with the view of Axtell & Meyers (2018), in that the prediction of 1000s of targets, followed by Gene Ontology or KEGG Ontology analysis to infer miRNA function is problematic (Eldem et al., 2012; Tiwari et al., 2020; Xu et al., 2020; Yaish et al., 2015; Yawichai et al., 2019), and likely has little relevance to miRNA function *in planta*. We advocate that using an approach such as TRUEE will enable rapid identification of which genes are being strongly regulated by miRNA, and therefore, what genetic targets would be best to modify in the bid to improving desired plant traits.

Our analyses support the idea that the majority of functional MTIs have already been identified in Arabidopsis. In the analysis of 34 Arabidopsis degradome libraries in WPMIAS (Fei et al., 2020), the known conserved canonical miRNA targets had the highest-ranking *Cat Score*s, indicating this metric was able to filter out and identify strong MTIs that have clear functional roles (Table S4). By contrast, there were very few uncharacterized MTIs that had a high *Cat Score*. This extended to the *Brassicaceae-specific* MTIs, where the highest *Cat Score*s were largely limited to previously documented MTIs, such as the well-studied miR824:*AGL16* and miR161:*PPR* modules (Howell et al., 2007; Kutter et al., 2007; Szaker et al., 2019).

It could be argued that only a subset of sRNAs were investigated, as the complex miRNome includes miRNA isoforms that arise through altered processing or modifications and that are predicted to confer altered specificity, and these were not included in the analysis. To investigate this possibility, we analysed the passenger strands (miRNA*s) of *Conserved-guide* miRNAs, as currently this class of alternative miRNA isoforms have the strongest evidence implicating them in functional MTIs (Du et al., 2017; Liu et al., 2017; Manavella et al., 2013; Zhang et al., 2011). However, only a few HE targets were identified for this *Conserved-passenger* grouping and all had low *Cat Score*s (<0.7). Moreover, previously reported functional miRNA*-target interactions, such as miR393* (Zhang et al., 2011) were not detected in the analysis. Again, it is possible that these classes of sRNAs have highly specific temporal and/or spatial expression, and so their MTIs are missed due to the absence of the corresponding degradome libraries, as TRUEE will be biased towards MTIs that are widespread. Nevertheless, despite the regulatory potential of the miRNA*s, none of their MTIs have *Cat Score*s characteristics of the known physiologically important MTIs.

For the majority of Arabidopsis miRNA entries in miR-Base, TRUEE either failed to identify a HE target (72% for the *Brassicaceae-specific* and 89% for the *A. thaliana-specific* groupings) or had a single target with a low *Cat Score*. This is consistent with the observation that most low confidence miRNA entries on miRBase corresponded to poorly expressed, evolutionarily young miRNAs that lack a functional target gene (Cuperus et al., 2011), and the annotation of many of these being *bona fide* miRNAs has been questioned (Axtell & Meyers, 2018; Taylor et al., 2017). It is consistent with the hypothesis of the existence within the plant cell of a large pool of diverse, evolutionarily young, and weakly expressed miRNAs from which new MTIs of functional significance may arise (Axtell, 2008; Axtell et al., 2007; Cuperus et al., 2011; Fahlgren et al., 2007; Rajagopalan et al., 2006). However, it has been hypothesized this is rare and that most young miRNAs remain targetless and undergo neutral drift until their sequences are no longer recognizable by DCL for processing (Axtell, 2008; Cuperus et al., 2011). Again, it may be argued that many young MTIs will not be identified by TRUEE because they have a narrow spatial and temporal expression. However, that any young MTI can be detected, such as miR824: *AGL16*, which are localized in stomatal complexes, suggests otherwise (Kutter et al., 2007).

Finally, the highest ranking HE targets of the *A. thaliana-specific* miRNAs, predominantly consisted of targets of three unrelated miRNAs that have trinucleotide repeats, miR414, miR5021, and miR5658. For each miRNA, their targets consisted of diverse genes with the common feature of trinucleotide repeats at their potential binding site. Trinucleotide repeat expansions are known to cause multiple human genetic diseases such as Huntington's disease and have been reported to cause sensitivity to high temperatures in the *A. thaliana* accession *Bur-0* (Bates et al., 2015; Tabib et al., 2016). Therefore, these *A. thaliana-specific* miRNAs may have a specialized role in silencing potentially deleterious genes with trinucleotide repeat expansions. However, these claims will need to be tested with experimental analyses.

CONCLUSIONS

TRUEE represents an approach to rank miRNA targets independently of complementarity, circumventing the limitation of that approach that has been a central feature of bioinformatic target prediction programs. We envision the approach can be applied to other species, once sufficient degradome analyses have been conducted. It will enable fast ranking of targets, and therefore, which genes to modify concerning the plant traits that miRNAs control.

EXPERIMENTAL PROCEDURES

Bioinformatics workflow

The parameters of TRUEE were developed via benchmarking the retrieval of the VAT set. The VAT set was assembled via systematically and manually reviewing the literature, requiring two independent lines of evidence from commonly used experimental approaches. The literature supporting the formation of the VAT set is found in Table S1.

Mature miRNA sequences were retrieved from miRBase v22 (Kozomara et al., 2019). Where multiple isomiRs were found, the isomiR with the highest abundance found on a plant next-generation sequencing database (https://mpss.danforthcenter.org) was used (Nakano et al., 2020). The most conserved tasiARF sequence as reported by Allen et al. (2005) was used in the analysis. For the Arabidopsis 'miRNA targetome', all 428 available mature miRNA sequences, which include isomiRs, were retrieved from miRBase v22 (Kozomara et al., 2019; note that tasiARFs were not analysed, as they are not on miRBase).

Sequences were used as input into psRNATarget v2, 2017 scoring schema (Dai et al., 2018). Default settings were used for analysis other than the expectation score, which was decreased to 3 for all sRNAs except miR167, miR398, and miR408. An expectation score of 5 was used for these miRNAs as their targets from the VAT set exceeds an expectation score of 3.

The resulting predicted targets were then analysed using WPMIAS (Fei et al., 2020). WPMIAS settings were: (i) analysis type – analysis > Advanced II > use psRNATarget predicted results directly; (ii) plant species – *A. thaliana*; cDNA libraries – Transcript, JGI genomic project, Phytozome 11, 167 TAIR10 (from psRNAtarget); (iii) offset from spliced position (nt) 0 (default), or 1 for miR162, miR396, and miR398, which can only be identified using an offset of 1 (Debernardi et al., 2012; Shao et al., 2015; Yamasaki et al., 2007); (iv) mismatches allowed for mapping degradome reads to references, i.e. 0 (default).

Degradome data retrieved from WPMIAS were then used as input and analysed using TRUEE to identify HE and LE targets as described in Figure 1. TRUEE was developed using an in-house R script. Analysed data from WPMIAS and R script for TRUEE is accessible on the Open Science Framework page for this project https://osf.io/k7rcs/. *Target Categories* as defined in WPMIAS were used in this study (Fei et al., 2020).

Data visualization

Multiple sequence alignments were performed using Multiple Alignment using Fast Fourier Transform (Katoh & Standley, 2013), and the resulting alignment visualized using Jalview (Waterhouse et al., 2009). T-plots of miRNA targets were adapted from WPMIAS (Fei et al., 2020). Figures determining the optimal expectation score (Figure 2), identifying the HE targets by TRUEE (Figure 3), and the Arabidopsis targetome (Figure 5) were generated using R package, ggplot2. Code and design for Figure 5 was by Holtz Yan and can be found at https://www.r-graph-gallery.com/297-circular-barplot-with-groups.html. All graphs were generated using the R package, ggplot2.

ACKNOWLEDGEMENTS

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. T-plots of HE targets not from the VAT set found at a Library % Cut-off of 40%.

Figure S2. Binding site conservation of HE targets is limited to the *Brassicaceae* family.

Table S1. The VAT set - previously validated miRNA and tasiRNA targets found in *Arabidopsis thaliana*.

Table S2. *A. thaliana* miRNAs retrieved from miRBase v22 and their conservation group.

Table S3. HE targets of *Conserved passenger* families with *Cat score*s.

Table S4. HE targets of *Conserved guide* families with *Cat score*s.

Table S5. HE targets of *Brassicaceae-specific* families with *Cat score*s.

Table S6. HE targets of *A. thaliana-specific* families with *Cat score*s.

REFERENCES

Addo-Quaye, C., Eshoo, T.W., Bartel, D.P. & Axtell, M.J. (2008) Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. *Current Biology*, **18**, 758–762. https://doi.org/10.1016/j.cub.2008.04.042.

Allen, E., Xie, Z., Gustafson, A.M. & Carrington, J.C. (2005) microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell*, **121**, 207–221. https://doi.org/10.1016/j.cell.2005.04.004.

Allen, E., Xie, Z., Gustafson, A.M., Sung, G.H., Spatafora, J.W. & Carrington, J.C. (2004) Evolution of microRNA genes by inverted duplication of target gene sequences *in Arabidopsis thaliana*. *Nature Genetics*, **36**, 1282–1290. https://doi.org/10.1038/ng1478.

Allen, R.S., Li, J., Stahle, M.I., Dubroué, A., Gubler, F. & Millar, A.A. (2007) Genetic analysis reveals functional redundancy and the major target genes of the Arabidopsis miR159 family. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 16371–16376. https://doi.org/10.1073/pnas.0707653104.

Axtell, M.J. (2008) Evolution of microRNAs and their targets: are all micro-RNAs biologically relevant? *Biochimica et Biophysica Acta*, **1779**, 725–734. https://doi.org/10.1016/j.bbagrm.2008.02.007.

Axtell, M.J. & Meyers, B.C. (2018) Revisiting criteria for plant microRNA annotation in the era of big data. *Plant Cell*, **30**, 272–284. https://doi.org/10.1105/tpc.17.00851.

Axtell, M.J., Snyder, J.A. & Bartel, D.P. (2007) Common functions for diverse small RNAs of land plants. *Plant Cell*, **19**, 1750–1769. https://doi.org/10.1105/tpc.107.05170.

Bates, G.P., Dorsey, R., Gusella, J.F., Hayden, M.R., Kay, C., Leavitt, B.R. et al. (2015) Huntington disease. *Nature Reviews. Disease Primers*, **1**, 15005. https://doi.org/10.1038/nrdp.2015.5.

Boccara, M., Sarazin, A., Thiébeauld, O., Jay, F., Voinnet, O., Navarro, L. et al. (2014) The Arabidopsis miR472-RDR6 silencing pathway modulates PAMP- and effector-triggered immunity through the post-transcriptional control of disease resistance genes. *PLoS Pathogens*, **10**, e1003883. https://doi.org/10.1371/journal.ppat.1003883.

Bonnet, E., He, Y., Billiau, K. & Van de Peer, Y. (2010) TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. *Bioinformatics*, **26**, 1566–1568. https://doi.org/10.1093/bioinformatics/btq233.

Brodersen, P. & Voinnet, O. (2009) Revisiting the principles of microRNA target recognition and mode of action. *Nature Reviews Molecular Cell Biology*, **10**, 141–148. https://doi.org/10.1038/nrm2619.

Brousse, C., Liu, Q., Beauclair, L., Deremetz, A., Axtell, M.J. & Bouché, N. (2014) A non-canonical plant microRNA target site. *Nucleic Acids Research*, **42**, 5270–5279. https://doi.org/10.1093/nar/gku157.

Bülow, L., Bolívar, J.C., Ruhe, J., Brill, Y. & Hehl, R. (2012) 'MicroRNA Targets', a new AthaMap web-tool for genome-wide identification of miRNA targets in *Arabidopsis thaliana*. *BioData Mining*, **5**, 7. https://doi.org/10.1186/1756-0381-5-7.

Chorostecki, U., Crosa, V.A., Lodeyro, A.F., Bologna, N.G., Martin, A.P., Carrillo, N. et al. (2012) Identification of new microRNA-regulated genes by conserved targeting in plant species. *Nucleic Acids Research*, **40**, 8893–8904. https://doi.org/10.1093/nar/gks625.

Cui, J., You, C. & Chen, X. (2017) The evolution of microRNAs in plants. *Current Opinion in Plant Biology*, **35**, 61–67. https://doi.org/10.1016/j.pbi.2016.11.006.

Cuperus, J.T., Fahlgren, N. & Carrington, J.C. (2011) Evolution and functional diversification of MIRNA genes. *Plant Cell*, 23, 431–442. https://doi.org/10.1105/tpc.110.082784.

Dai, X. & Zhao, P.X. (2011) PsRNATarget: a plant small RNA target analysis server. *Nucleic Acids Research*, 39(SUPPL. 2), 155–159. https://doi.org/10.1093/nar/gkr319.

Dai, X., Zhuang, Z. & Zhao, P.X. (2018) PsRNATarget: a plant small RNA target analysis server (2017 release). *Nucleic Acids Research*, 46(W1), W49–W54. https://doi.org/10.1093/nar/gky316.

Debernardi, J.M., Rodriguez, R.E., Mecchia, M.A. & Palatnik, J.F. (2012) Functional specialization of the plant miR396 regulatory network through distinct microRNA-target interactions. *PLoS Genetics*, 8, e1002419. https://doi.org/10.1371/journal.pgen.1002419.

Du, Q., Zhao, M., Gao, W., Sun, S. & Li, W.-X. (2017) microRNA/microRNA* complementarity is important for the regulation pattern of NFYA5 by miR169 under dehydration shock in Arabidopsis. *Plant Journal*, 91, 22–33. https://doi.org/10.1111/tpj.13540.

Eldem, V., Çelikkol Akçay, U., Ozhuner, E., Bakır, Y., Uranbey, S. & Unver, T. (2012) Genome-wide identification of miRNAs responsive to drought in peach (Prunus persica) by high-throughput deep sequencing. *PLoS One*, 7, e50298. https://doi.org/10.1371/journal.pone.0050298.

Fahlgren, N., Howell, M.D., Kasschau, K.D., Chapman, E.J., Sullivan, C.M., Cumbie, J.S. et al. (2007) High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. *PLoS One*, 2, e219. https://doi.org/10.1371/journal.pone.0000219.

Fei, Y., Mao, Y., Shen, C., Wang, R., Zhang, H. & Huang, J. (2020) WPMIAS: Whole-degradome-based Plant MicroRNA-Target Interaction Analysis Server. *Bioinformatics (Oxford, England)*, 36, 1937-1939. https://doi.org/10.1093/bioinformatics/btz820.

Folkes, L., Moxon, S., Woolfenden, H.C., Stocks, M.B., Szittya, G., Dalmay, T. et al. (2012) PAREsnip: a tool for rapid genome-wide discovery of small RNA/target interactions evidenced through degradome sequencing. *Nucleic Acids Research*, 40, e103. https://doi.org/10.1093/nar/gks277.

German, M.A., Pillay, M., Jeong, D.-H., Hetawal, A., Luo, S., Janardhanan, P. et al. (2008) Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nature Biotechnology*, 26, 941–946. https://doi.org/10.1038/nbt1417.

Howell, M.D., Fahlgren, N., Chapman, E.J., Cumbie, J.S., Sullivan, C.M., Givan, S.A. et al. (2007) Genome-wide analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 pathway in Arabidopsis reveals dependency on miRNA- and tasiRNA-directed targeting. *Plant Cell*, 19, 926–942. https://doi.org/10.1105/tpc.107.050062.

Jeong, D.-H., Schmidt, S.A., Rymarquis, L.A., Park, S., Ganssmann, M., German, M.A. et al. (2013) Parallel analysis of RNA ends enhances global investigation of microRNAs and target RNAs of Brachypodium distachyon. *Genome Biology*, 14, R145–R145. https://doi.org/10.1186/gb-2013-14-12-r145.

Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30, 772–780. https://doi.org/10.1093/molbev/mst010.

Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. (2019) miRBase: from microRNA sequences to function. *Nucleic Acids Research*, 47(D1), D155–D162. https://doi.org/10.1093/nar/gky1141.

Kutter, C., Schöb, H., Stadler, M., Meins, F. & Si-Ammour, A. (2007) MicroRNA-mediated regulation of stomatal development in Arabidopsis. *Plant Cell*, 19, 2417–2429. https://doi.org/10.1105/tpc.107.050377.

Li, J., Reichel, M., Li, Y. & Millar, A.A. (2014) The functional scope of plant microRNA-mediated silencing. *Trends in Plant Science*, 19, 750–756. https://doi.org/10.1016/j.tplants.2014.08.006.

Lindow, M., Jacobsen, A., Nygaard, S., Mang, Y. & Krogh, A. (2007) Intragenomic matching reveals a huge potential for miRNA-mediated regulation in plants. *PLoS Computational Biology*, 3, e238. https://doi.org/10.1371/journal.pcbi.0030238.

Lindow, M. & Krogh, A. (2005) Computational evidence for hundreds of non-conserved plant microRNAs. *BMC Genomics*, 6, 119. https://doi.org/10.1186/1471-2164-6-119.

Liu, Q., Wang, F. & Axtell, M.J. (2014) Analysis of complementarity requirements for plant MicroRNA targeting using a Nicotiana benthamiana quantitative transient assay. *Plant Cell*, 26, 741–753. https://doi.org/10.1105/tpc.113.120972.

Liu, W.W., Meng, J., Cui, J. & Luan, Y.S. (2017) Characterization and function of microRNA*s in plants. *Frontiers in Plant Science*, 8, 1–7. https://doi.org/10.3389/fpls.2017.02200.

Ma, X., Liu, C., Gu, L., Mo, B., Cao, X. & Chen, X. (2018) TarHunter, a tool for predicting conserved microRNA targets and target mimics in plants. *Bioinformatics*, 34, 1574–1576. https://doi.org/10.1093/bioinformatics/btx797.

Mallory, A.C., Reinhart, B.J., Jones-Rhoades, M.W., Tang, G., Zamore, P.D., Barton, M.K. et al. (2004) MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5′ region. *EMBO Journal*, 23, 3356–3364. https://doi.org/10.1038/sj.emboj.7600340.

Manavella, P.A., Koenig, D., Rubio-Somoza, I., Burbano, H.A., Becker, C. & Weigel, D. (2013) Tissue-specific silencing of Arabidopsis SU(VAR)3-9 HOMOLOG8 by miR171a. *Plant Physiology*, 161, 805–812. https://doi.org/10.1104/pp.112.207068.

Nakano, M., McCormick, K., Demirci, C., Demirci, F., Gurazada, S.G.R., Ramachandruni, D. et al. (2020) Next-generation sequence databases: RNA and genomic informatics resources for plants. *Plant Physiology*, 182, 136–146. https://doi.org/10.1104/pp.19.00957.

Park, Y.J., Lee, H.J., Kwak, K.J., Lee, K., Hong, S.W. & Kang, H. (2014) MicroRNA400-guided cleavage of pentatricopeptide repeat protein mRNAs renders Arabidopsis thaliana more susceptible to pathogenic bacteria and fungi. *Plant and Cell Physiology*, 55, 1660–1668. https://doi.org/10.1093/pcp/pcu096.

Rajagopalan, R., Vaucheret, H., Trejo, J. & Bartel, D.P. (2006) A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. *Genes and Development*, 20, 3407–3425. https://doi.org/10.1101/gad.1476406.

Schwab, R., Palatnik, J.F., Riester, M., Schommer, C., Schmid, M. & Weigel, D. (2005) Specific effects of microRNAs on the plant transcriptome. *Developmental Cell*, 8, 517–527. https://doi.org/10.1016/j.devcel.2005.01.018.

Shao, F., Qiu, D. & Lu, S. (2015) Comparative analysis of the dicer-like gene family reveals loss of miR162 target site in SmDCL1 from salvia miltiorrhiza. *Scientific Reports*, 5, 9891. https://doi.org/10.1038/srep09891.

Sharma, D., Tiwari, M., Pandey, A., Bhatia, C., Sharma, A. & Trivedi, P.K. (2016) MicroRNA858 is a potential regulator of phenylpropanoid pathway and plant development. *Plant Physiology*, 171, 944–959. https://doi.org/10.1104/pp.15.01831.

Smith, L.M., Burbano, H.A., Wang, X., Fitz, J., Wang, G., Ural-Blimke, Y. et al. (2015) Rapid divergence and high diversity of miRNAs and miRNA targets in the Camelineae. *Plant Journal*, 81, 597–610. https://doi.org/10.1111/tpj.12754.

Srivastava, P.K., Moturu, T.R., Pandey, P., Baldwin, I.T. & Pandey, S.P. (2014) A comparison of performance of plant miRNA target prediction tools and the characterization of features for genome-wide target prediction. *BMC Genomics*, 15, 348. https://doi.org/10.1186/1471-2164-15-348.

Sun, Y.H., Lu, S., Shi, R. & Chiang, V.L. (2011) Computational prediction of plant miRNA targets. *Methods in Molecular Biology*, 744, 175–186. https://doi.org/10.1007/978-1-61779-123-9_12.

Szaker, H.M., Darkó, É., Medzihradszky, A., Janda, T., Liu, H.C., Charng, Y.Y. et al. (2019) miR824/AGAMOUS-LIKE16 module integrates recurring environmental heat stress changes to fine-tune Poststress development. *Frontiers in Plant Science*, 10, 1–19. https://doi.org/10.3389/fpls.2019.01454.

Tabib, A., Vishwanathan, S., Seleznev, A., McKeown, P.C., Downing, T., Dent, C. et al. (2016) A polynucleotide repeat expansion causing temperature-sensitivity persists in Wild Irish accessions of Arabidopsis thaliana. *Frontiers in Plant Science*, 7, 1311. https://www.frontiersin.org/article/10.3389/fpls.2016.01311.

Taylor, R.S., Tarver, J.E., Foroozani, A. & Donoghue, P.C.J. (2017) Insights and perspectives MicroRNA annotation of plant genomes. Do it right or not at all. *BioEssays*, 1600113, 1–6. https://doi.org/10.1002/bies.201600113.

Taylor, R.S., Tarver, J.E., Hiscock, S.J. & Donoghue, P.C.J. (2014) Evolutionary history of plant microRNAs. *Trends in Plant Science*, 19, 175–182. https://doi.org/10.1016/j.tplants.2013.11.008.

Thody, J., Moulton, V. & Mohorianu, I. (2020) PAREameters: a tool for computational inference of plant miRNA–mRNA targeting rules using small RNA and degradome sequencing data. *Nucleic Acids Research*, 48, 2258–2270. https://doi.org/10.1093/nar/gkz1234.

Tiwari, J.K., Buckseth, T., Zinta, R., Saraswati, A., Singh, R.K., Rawat, S. *et al.* (2020) Genome-wide identification and characterization of microRNAs by small RNA sequencing for low nitrogen stress in potato. *PLoS One*, **15**, e0233076. https://doi.org/10.1371/journal.pone.0233076.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. & Barton, G.J. (2009) Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191. https://doi.org/10.1093/bioinformatics/btp033.

Xu, X., Chen, X., Chen, Y., Zhang, Q., Su, L., Chen, X. *et al.* (2020) Genome-wide identification of miRNAs and their targets during early somatic embryogenesis in *Dimocarpus longan* Lour. *Scientific Reports*, **10**, 4626. https://doi.org/10.1038/s41598-020-60946-y.

Yaish, M.W., Sunkar, R., Zheng, Y., Ji, B., Al-Yahyai, R. & Farooq, S.A. (2015) A genome-wide identification of the miRNAome in response to salinity stress in date palm (*Phoenix dactylifera* L.). *Frontiers in Plant Science*, **6**, 946. https://www.frontiersin.org/article/10.3389/fpls.2015.00946.

Yamasaki, H., Abdel-Ghany, S.E., Cohu, C.M., Kobayashi, Y., Shikanai, T. & Pilon, M. (2007) Regulation of copper homeostasis by micro-RNA in Arabidopsis. *Journal of Biological Chemistry*, **282**, 16369–16378. https://doi.org/10.1074/jbc.M700138200.

Yawichai, A., Kalapanulak, S., Thammarongtham, C. & Saithong, T. (2019) Genome-wide identification of putative MicroRNAs in cassava (Manihot esculenta Crantz) and their functional landscape in cellular regulation. *BioMed Research International*, **2019846**. https://doi.org/10.1155/2019/2019846.

Yu, Y., Jia, T. & Chen, X. (2017) The 'how' and 'where' of plant microRNAs. *New Phytologist*, **216**, 1002–1017. https://doi.org/10.1111/nph.14834.

Zhang, X., Zhao, H., Gao, S., Wang, W.C., Katiyar-Agarwal, S., Huang, H.D. *et al.* (2011) Arabidopsis Argonaute 2 regulates innate immunity via miRNA393(*)-mediated silencing of a Golgi-localized SNARE gene, MEMB12. *Molecular Cell*, **42**, 356–366. https://doi.org/10.1016/j.molcel.2011.04.010.

Zheng, Y., Li, Y.F., Sunkar, R. & Zhang, W. (2012) SeqTar: an effective method for identifying microRNA guided cleavage sites from degradome of polyadenylated transcripts in plants. *Nucleic Acids Research*, **40**, e28. https://doi.org/10.1093/nar/gkr1092.

Zheng, Z., Reichel, M., Deveson, I., Wong, G., Li, J. & Millar, A.A. (2017) Target RNA secondary structure is a major determinant of miR159 efficacy. *Plant Physiology*, **174**, 1764–1778. https://doi.org/10.1104/pp.16.01898.