

The emergence, evolution, and diversification of the **miR390-TAS3-ARF pathway in land plants**

Rui Xia^{1,2,3*}, Jing Xu^{2,3}, Blake C. Meyers^{3, 4*}

1. State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, South China Agricultural University, Guangzhou, 510642, China
2. College of Horticulture, South China Agricultural University, Guangzhou 510642, China.
3. Donald Danforth Plant Science Center, St. Louis, MO 63132, USA.
4. University of Missouri – Columbia, Division of Plant Sciences, 52 Agriculture Lab, Columbia, MO 65211, USA.

*Corresponding author email addresses: bmeyers@danforthcenter.org; rxia@scau.edu.cn

1 **Abstract**

2 In plants, miR390 directs the production of tasiRNAs from *TRANS-ACTING SIRNA 3 (TAS3)* transcripts to
3 regulate *AUXIN RESPONSIVE FACTOR (ARF)* genes, transcription factors critical for auxin signaling; these
4 tasiRNAs are known as tasiARFs. This pathway is highly conserved, with the *TAS3* as the only one
5 noncoding gene present almost ubiquitously in land plants. To understand the evolution of this miR390-
6 *TAS3-ARF* pathway, we characterized homologs of these three genes from thousands of plant species,
7 from bryophytes to angiosperms. Both miR390 and *TAS3* are present and functional in liverworts,
8 confirming their ancestral role to regulate *ARFs* in land plants. We found the lower-stem region of
9 *MIR390* genes, critical for accurate DCL1 (DICER-LIKE 1) processing, is conserved in sequence in seed
10 plants. We propose a model for the transition of functional tasiRNA sequences in *TAS3* genes occurred
11 at the emergence of vascular plants, in which the two miR390 target sites of *TAS3* genes showed distinct
12 pairing patterns in different plant lineages. Based on the cleavability of miR390 target sites and the
13 distance between target site and tasiARF we inferred a potential bidirectional processing mechanism
14 exists for some *TAS3* genes. We also demonstrated a tight mutual selection between tasiARF and its
15 target genes, and characterized unusual aspects and diversity of regulatory components of this pathway.
16 Taken together, these data illuminate the evolutionary path of the *miR390-TAS3-ARF* pathway in land
17 plants, and demonstrate the significant variation that occurs in the production of phasiRNAs in plants,
18 even in the functionally important and archetypal *miR390-TAS3-ARF* regulatory circuit.

19

20 **Introduction**

21 In plants, small RNAs (sRNAs) play crucial regulatory functions in growth and development,
22 resistance to abiotic and biotic stresses, and reproduction (Chen 2009; Axtell 2013; Bartel 2009). Based
23 on features such as their biogenesis and function, sRNAs are classified into two major groups,
24 microRNAs (miRNAs) and small interfering RNAs (siRNAs). miRNAs are generated from precursor mRNAs
25 that fold back to form double-stranded stem-loop structures, while siRNAs are produced from double-
26 stranded RNAs (dsRNAs) biosynthesized secondarily by RNA-dependent RNA polymerase (RDR) (Axtell
27 2013). *Trans-acting* small interfering RNAs (tasiRNAs) are a special type of small RNAs found only in
28 plants, so far. Precursor genes of tasiRNAs (*TAS* genes) are sliced in a miRNA-directed event, and the
29 cleaved fragment is made double-stranded by RDR6; the resulting dsRNA is chopped by DICER-LIKE 4
30 (DCL4) into 21-nt siRNAs that map back to the precursors in a head-to-tail arrangement initiating from
31 the miRNA cleavage site (Allen et al. 2005; Yoshikawa et al. 2005).

32 Among plant *TAS* genes, the most well-studied is *TAS3*; its transcript bears two target sites of
33 miR390, generating tasiRNAs via the so-called “two-hit” mechanism (Axtell et al. 2006). The conserved,
34 resulting tasiRNA is known as “tasiARF” as it targets auxin responsive factor (*ARF*) genes (Allen et al.
35 2005; Axtell et al. 2006). To date, there are two kinds of *TAS3* genes described in plants; one contains a
36 single, centrally-located tasiARF, while the other generates two tasiARFs denoted as *TAS3*-short (*TAS3S*)
37 and *TAS3*-long (*TAS3L*), respectively (Xia et al. 2015b). In *TAS3L*, only the 3' miR390 target site is
38 cleavable and this sets the phase of tasiRNA production, giving rise to the two in-phase tasiARFs (Allen
39 et al. 2005; Axtell et al. 2006). The 5' target site of *TAS3L* is usually non-cleavable because of the
40 presence of a central mismatch (10th position) in the pairing of miR390 and target site (Axtell et al. 2006).
41 It serves as an important binding site of ARGONAUTE 7 (AGO7), a specialized protein partner of miR390
42 (Montgomery et al. 2008a). In contrast, both target sites of *TAS3S* are cleavable, and both can
43 potentially initiate tasiRNA generation (Howell et al. 2007; Xia et al. 2012, 2015b). The single tasiARF of
44 *TAS3S* is in near-perfect phasing to both miR390 sites as there is only 2-nt difference between the phase
45 registers set by the two target sites (Xia et al. 2012, 2015b).

46 Auxin, a plant hormone, regulates seemingly every aspect of plant growth and development. The
47 small class of ARF transcription factors can either activate or repress expression of downstream auxin-
48 regulated genes through protein–protein interactions with auxin/indole-3-acetic acid (Aux/IAA) family
49 members (Guilfoyle and Hagen 2007). Plant genomes contain ~10 to 30 *ARF* genes; for example, there
50 are 23 members in the model plant *Arabidopsis*. ARFs are classified into three clades: ARF5/6/7/8 (Clade
51 A), ARF1/2/3/4/9 (Clade B), and ARF10/16/17 (Clade C) (Finet et al. 2013). The *TAS3*-derived tasiARF
52 specifically targets *ARF* genes of Clade B (ARF2/3/4). This miR390-*TAS3*-*ARF* pathway is of critical
53 function in the regulation of plant growth and development, including leaf morphology, developmental
54 timing and patterning, and lateral root growth (Garcia et al. 2006; Fahlgren et al. 2006; Adenot et al.
55 2006; Marin et al. 2010; Hunter et al. 2006). It was recently found that ARF3, with the transcription
56 factor INDEHISCENT (IND), comprises an alternative auxin-sensing mechanism (Simonini et al. 2016).
57 Loss-of-function mutants of AGO7, the specialized AGO partner of miR390, show varying degrees of
58 growth and developmental disorders due to the malfunction of the tasiARF pathway (Yifhar et al. 2012;
59 Zhou et al. 2013; Dotto et al. 2014). For example, maize *ago7* (*leafbladeless1*) plants have thread-like
60 leaves lacking top/bottom polarity (Dotto et al. 2014), and *Medicago ago7* (*lobed leaflet1*) mutant plants
61 displayed lobed leaf margins and extra lateral leaflets (Zhou et al. 2013).

62 All three components of the pathway, miR390, *TAS3*, and *ARFs*, are present in the oldest land
63 plants, liverworts (Krasnikova et al. 2013; Finet et al. 2013). Interestingly, in bryophytes, the *TAS3* genes

64 are different from those found in flowering plants. Although bryophyte *TAS3* genes also have two
65 miR390 target, they generate tasiRNAs targeting not only *ARF* genes but also *AP2* genes (described, for
66 example, in the moss *Physcomitrella patens*) (Axtell et al. 2007). Moreover, the bryophyte *ARF*-targeting
67 tasiRNA is a different sequence compared to the tasiARF in flowering plants (Allen et al. 2005; Axtell et
68 al. 2007). How and when this transition in *TAS3* gene composition occurred in the evolution of land
69 plants is fascinating but unknown. We recently characterized ~20 *TAS3* genes in the gymnosperm
70 Norway spruce (*Picea abies*), demonstrating diverse features of these genes distinct from those
71 characterized in flowering plants (Xia et al. 2015a). In this study, we aimed to understand the
72 evolutionary history of and critical changes in the miR390-TAS3-ARF pathway for the major lineages of
73 land plants. We used > 150 plant genomes and the large dataset from the 1000 Plant Transcriptomes
74 (1KP) project, in combination of additional sequencing data and computational approaches; these
75 resources identified hundreds of *MIR390* genes and thousands of *TAS3* and *ARF* genes, across numerous
76 plant species. From these data, we elucidated with high-resolution the dynamic nature of the
77 evolutionary route of the miR390-TAS3-ARF pathway, revealing new regulatory features of the three
78 critical components of the pathway.

79

80 **Results**

81 ***Gene identification from plant genomic and transcriptomic data***

82 miR390-TAS3-ARF comprise a regulatory pathway highly conserved in plants. To maximize the
83 possibility of characterizing the full diversity of the three main components of this pathway (miR390,
84 *TAS3*, and *ARF* genes), we collected 159 sequenced plant genomes, ranging from liverworts to
85 angiosperms, plus the 1KP data (Matasci et al. 2014). For the identification of *TAS3* genes, only genomic
86 loci from sequenced genomes or transcripts (from 1KP data) containing at least one miR390 target site
87 and one tasiRNA targeting *ARF* gene were considered valid for our analysis. Using bioinformatics tools
88 and customized scripts (see Methods), we identified 374 *MIR390* genes from 163 plant species, 1923
89 *TAS3* genes from 792 species, and 2912 *ARF* genes (targets of tasiRNAs) from 934 species. We were
90 unable to identify homologs of *MIR390* or *TAS3* genes in five algal genomes, consistent with the earlier
91 conclusion that the miR390-TAS3-ARF pathway originated in land plants (Krasnikova et al. 2013).

92 To evaluate the evolutionary changes of three components of the pathway, we classified all the
93 plant species into one of seven groups (liverworts, mosses, monilophytes or ferns, gymnosperms, basal
94 angiosperms, monocots, eudicots); each group was considered independently in our subsequent

95 assessments. Monocots and eudicots accounted for the two largest groups of species and yielded the
96 vast majority of *MIR390* genes; many fewer were identified in the liverwort, monilophyte, and basal
97 angiosperm groups (Fig. S1A). Similarly, most of the *TAS3* genes identified were from angiosperms,
98 although there were many from gymnosperms as well (Fig. S1A).

99 We next examined variation in the length and GC content of *MIR390* and *TAS3*. Flanking
100 sequences of 50 bp (5' of the miR390 and 3' of the miR390* for *MIR390*; 5' of the 5' miR390 target site
101 and 3' of the 3' target site for *TAS3* genes) were included for these analyses. The length of the *MIR390*
102 genes ranged from approximately 150 bp to 250 bp, with the *MIR390* copies in monocots significantly
103 larger than those in gymnosperms (2.00E-05, t-test) and eudicots (4.00E-07, t-test) (Fig. S1B, at left). The
104 GC content of *MIR390* genes was similar among different plant groups, with the exception of the
105 eudicots in which they had a substantially lower GC content (1.40E-09, t-test) (Fig. S1C, at left). For the
106 *TAS3* genes, their length was noticeably shorter in monilophytes, while significantly longer in
107 gymnosperms than in monocots and eudicots (2.00E-16, t-test); in gymnosperms, there is an apparently
108 bimodal distribution of lengths, possibly reflecting that there are two major types of *TAS3* genes of
109 different length (Fig. S1B, at right). The GC content of the *TAS3* genes of two groups, the mosses and
110 eudicots, was exceptional: the moss *TAS3* genes were of higher GC content (2.00E-16, t-test), while the
111 eudicot *TAS3* genes had a much lower GC content (2.00E-16, t-test, Fig. S1C, at right).

112 In *Arabidopsis*, the proper execution of the miR390-TAS3-ARF pathway requires that miR390 is
113 loaded into a specific and highly selective AGO partner protein, AGO7 (Montgomery et al. 2008a). AGO7
114 is an indispensable component of the pathway, and thus we also investigated the evolutionary history of
115 AGO7. To complement our recent survey of AGO proteins that mainly focused on flowering plants
116 (Zhang et al. 2015), our analyses here focused on AGO proteins from non-flowering plants. We identified
117 237 AGO protein sequences with \geq 800 amino acids, and these were used for the construction of a
118 phylogenetic tree in combination with AGO proteins from three representative angiosperms: *Amborella*
119 *trichopoda*, *Oryza sativa*, and *Arabidopsis thaliana*. As previously documented (Vaucheret 2008; Mallory
120 and Vaucheret 2010; Zhang et al. 2015), AGO proteins clustered into three major clades, AGO1/5/10,
121 AGO2/3/7, and AGO4/6/8/9 (Fig. S2A). The AGO2/3/7 clade consisted of members all from vascular
122 plants, except two moss AGO proteins (4_Pp3c17_350V3.1 from *Physcomitrella patens* and
123 4_Sphflax0148s0007.1 from *Sphagnum fallax*) (Fig. S2B); we interpreted this as an indication that the
124 ancestor of the AGO2/3/7 clade likely separated from the AGO1/5/10 clade in mosses. Also, AGO7 was
125 apparently not specified until the emergence of gymnosperms, as only gymnosperm and *Amborella*
126 AGOs joined the eudicot AGO7 copies to form a subclade (Fig. S2B). These results suggest that the

127 specific partner AGO of miR390, AGO7, emerged much later than the miRNA and the pathway, possibly
128 to enable unique functions of the miR390-TAS3-ARF pathway in seed and flowering plants.
129

130 ***The lower stem region of MIR390 is under strong selection for conservation***

131 miR390 is one of the most ancient miRNAs, well conserved in land plants. During the course of
132 evolution, *MIRNA* genes (i.e. the precursor mRNAs) are relatively labile, typically displaying conservation
133 only in the sequences of the miRNA and miRNA* in the foldback region (Jones-Rhoades et al. 2006;
134 Fahlgren et al. 2010; Ma et al. 2010). Indeed, in our analysis, the sequences of miR390 and miR390*
135 were extremely conserved in land plants, as shown in the sequence alignment in Fig. 1A. Interestingly, in
136 addition to the miR390/miR390* region, we identified another two regions of relatively high
137 conservation in the precursors (Fig. 1A); these are the sequences forming the lower stem of the *MIR390*
138 stem-loop structure (Fig. 1B). They displayed a substantially greater consensus, especially for the seed
139 plants, than any other regions of the precursors except the miR390/miR390* duplex (Fig. 1A).

140 In plants, the release of a miRNA/miRNA* duplex relies on two sequential cuts by DCL1 in the
141 *MIRNA* stem-loop precursors. These two sequential cuts are directional, either base-to-loop or loop-to-
142 base. For base-to-loop processing, the first cut is defined by the distance from the miRNA/miRNA*
143 duplex to a large loop at the base; the distance is usually ~15 nt (Werner et al. 2010; Song et al. 2010;
144 Mateos et al. 2010). miR390 is one such base-to-loop-processed miRNA (Bologna et al. 2013). The
145 conservation in the *MIR390* lower stem, exemplified in Fig. 1B, is likely to maintain the consistent
146 distance of ~15 nt to ensure the accuracy of the first cut by DCL1 of the *MIR390* stem-loop precursor.
147

148 ***TAS3 originated to regulate ARF genes***

149 *TAS3* genes in bryophytes were firstly characterized in the moss *Physcomitrella patens*, consisting
150 in that genome of a small family of six genes (Axtell et al. 2007; Arif et al. 2012). Many *TAS3* genes were
151 subsequently described in mosses (Krasnikova et al. 2013). All known moss *TAS3* genes have similar
152 sequence components: two miR390 target sites, a tasiRNA targeting *AP2* genes (tasiAP2), and a tasiRNA
153 targeting *ARF* genes (tasiARF)(Axtell et al. 2007; Krasnikova et al. 2013). A *TAS3* gene was also identified
154 in a liverwort *Marchantia polymorpha*, representing the most ancient extant lineage of land plants
155 (Krasnikova et al. 2013). However, this *TAS3* gene was described to produce only a single tasiRNA of
156 sequence similar to the moss tasiAP2. We found five *TAS3* genes from liverwort species in addition to
157 that of *Marchantia polymorpha*. Sequence alignment of these six liverwort genes revealed the presence

158 of another conserved region, aside from the two miR390 target sites and the previously-described
159 tasiAP2, that could also produce an siRNA (Fig. 2A). Analyses of public sRNA data from *Marchantia*
160 *polymorpha* showed that a highly abundant tasiRNA was produced from the anti-sense strand of this
161 siRNA site. This tasiRNA (hereafter, “tasiARF-a1”) was predicted to target an ARF gene in *M. polymorpha*,
162 with the cleavage of the target site confirmed by PARE analysis (Fig. 2A). While the previously-described
163 tasiAP2 site is highly conserved, we were unable to validate its target interaction in *M. polymorpha* in
164 which we attempted by combining whole genome target analysis with sRNA and PARE data. This is likely
165 for several reasons: first, the corresponding tasiRNA was produced of low abundance; second, no AP2
166 homolog was predicted as a target of the tasiRNA even using relaxed prediction criteria (alignment score
167 ≤ 7); third, after checking *M. polymorpha* homologs of moss AP2 genes that are validated targets of
168 moss tasiARF-a1, we found no tasiAP2 target sites (data not shown). Therefore, the miR390-TAS3
169 machinery likely originated to regulate ARF genes, and not AP2 genes, unlike previous reports.

170 For the bryophytes, we identified a large number of *TAS3* genes including 67 genes from 36 moss
171 species), in addition to the six liverwort *TAS3* copies described above. For the 67 genes, we built a
172 multiple sequence alignment; from this, conserved sequence motifs, including the two miR390 target
173 sites, and both tasiAP2 and tasiARF, were detected as previously reported and as observed in the
174 mosses (Fig. S3A). In addition, we identified another tasiRNA site which is conserved only in a subset of
175 the moss *TAS3* genes. Target predictions indicated that this tasiRNA may target ARF genes as well, and it
176 is conserved in only a few members of the *TAS3* family, for instance, three *TAS3* genes of *P. patens*
177 (*a/d/f*) encode this tasiRNA sequence, but *TAS3b/c/e* lack it (Fig. 2B and Fig. S3A). In contrast to the
178 previously-identified tasiARF (tasiARF-a2, on the 3' end) which was produced in the antisense strand and
179 in phase with the 3' miR390 target site, the newly identified tasiARF-a3 is located in the sense strand
180 and in phase with the 5' miR390 target site (Fig. 2B). These three tasiARFs in liverworts (tasiARF-a1) and
181 mosses (tasiARF-a2, -a3) have no sequence similarity, originate from either strand of *TAS3* genes, and
182 target different regions of ARF genes, consistent with independent origins.

183 The distribution of tasiARF-a2 and -a3 in moss *TAS3* genes is consistent with distinct evolutionary
184 paths for these genes. To infer the possible evolutionary paths of *TAS3* in bryophytes, we constructed a
185 phylogenetic tree using their *TAS3* genes. The phylogenetic tree (Fig. S3B) yielded three major classes:
186 class I contained all six liverwort *TAS3* genes (tasiAP2 and tasiARF-a1); class II included moss *TAS3* genes
187 containing tasiAP2 and tasiARF-a2; and class III comprised moss *TAS3s* with tasiAP2, tasiARF-a2, and
188 tasiARF-a3. Intriguingly, class II is closer to liverwort *TAS3* genes (class I), indicating that class III *TAS3s*

189 likely evolved after the appearance of the class II *TAS3* genes, which raises an interesting question of the
190 origin of tasiARF-a3.

191

192 ***The evolutionary path of TAS3 genes in land plants***

193 *TAS3* genes found in seed plants are different from the bryophyte *TAS3*s. As summarized in Fig. 3A,
194 two types of *TAS3* genes, *TAS3L* with two tandem tasiARFs and *TAS3S* with one tasiARF, have been
195 previously characterized in gymnosperms and many angiosperms. Despite a similar arrangement of two
196 miR390 target sites, the near-identical tasiARFs in *TAS3L* and *TAS3S* are distinct from moss tasiARFs
197 (tasiARF-a1/a2/a3) in bryophyte *TAS3* genes, in terms of sequence, position and strand (Fig. 3A). These
198 differences suggest a significant change occurred during *TAS3* evolution in land plants. To better
199 understand when this change happened, we cataloged *TAS3* genes with tasiARFs; we found two *TAS3*
200 genes from a lycophyte *Phylloglossum drummondii*, one with two tasiARFs (*Pdr-TAS3L*) and the other
201 with a single tasiARF (*Pdr-TAS3S*) (Fig. 3B). The cDNA sequence of *Pdr-TAS3S* was too short to include the
202 5' miR390 target site. We generated sRNA sequencing data which confirmed the phased generation of
203 tasiARFs from *Pdr-TAS3L* (Fig. 3C). Both tasiARFs were predicted to target two *ARF* genes, found among
204 the cDNA sequences from the same species (Fig. 3C). Therefore, we believe that this transformation of
205 *TAS3* genes, and particularly the tasiARF transition, occurred after mosses and before or in lycophytes,
206 perhaps with the emergence of vascular plants. Thus, we named all the *TAS3* genes producing these
207 characteristic tasiARFs (i.e. not tasiARF-a1/a2/a3) as “vascular *TAS3*” genes.

208 We next asked how the transition of *TAS3* genes happened. In other words, how was this
209 signature tasiARF sequence generated in vascular plants. We compared the tasiARF sequence to
210 available cDNA or genome sequences. We found the tasiARF sequence shared substantial sequence
211 similarity to a region partially overlapped with the 5' miR390 target site from the cognate *TAS3* gene in
212 some species, as exemplified in a few *TAS3* genes shown in Fig. 3D. In a *TAS3* gene of the liverwort
213 *Marchantia polymorph* (*Mpo-TAS3*), the tasiARF sequence has 15 nucleotides of identity with the 5'
214 miR390 target sequence, with an overlap of 11 nt. This sequence similarity is even greater in *TAS3* genes
215 in a monotypic gymnosperm, *Welwitschia mirabilis* (*Wmi-TAS3*), and the basal angiosperm *Amborella*
216 *trichopoda* (*Atrich-TAS3*) (Fig. 3D). This finding of sequence similarity is consistent with a hypothesis that
217 the tasiARF was derived from the 5' miR390 target site from *TAS3*.

218 We previously reported that the genome of the gymnosperm Norway spruce includes a large
219 number of *TAS3* genes, of which many have non-canonical sequence features (Xia et al. 2015a). We
220 extended this observation to other plant species, finding *TAS3* genes with varied motif structures in our

221 large dataset (Fig. S4). For example, some have two 5' or 3' target sites due to short sequence
222 duplications; some have two or three non-adjacent tasiARFs. We propose a model, consistent with these
223 extant *TAS3* arrangements, for the tasiARF transition from bryophyte *TAS3* genes to vascular *TAS3* genes
224 (Fig. 3E). In the first step, the 5' miR390 target site of a bryophyte *TAS3* gene was duplicated through
225 segmental duplication, as evidenced in a couple of gymnosperm *TAS3* genes. Next, the miR390 target
226 site in the middle evolved into a tasiARF and was retained because of its essential function, yielding the
227 short *TAS3* gene (*TAS3S*); after this, two tasiARFs in a single *TAS3* gene resulted from the duplication of
228 tasiARF. Finally, the gap between the two tasiARFs was lost, forming a tandem repeat of tasiARFs,
229 yielding the long *TAS3* gene (*TAS3L*) present in vascular plants. This series of steps is consistent with the
230 *TAS3* variants present in plant genomes (Fig. S4).

231

232 ***Distinct pairing patterns of two miR390 target sites***

233 *TAS3* genes usually comprise a small gene family in plants. For instance, in bryophytes, only one
234 *TAS3* gene was identified in *M. polymorpha*, and six *TAS3* copies in *P. patens*. For comparison, there are
235 three *TAS3* copies in Arabidopsis, five in rice, and nine in maize – all vascular plants. Comparing across
236 the 157 vascular plants with full-genome sequences that we utilized, we found that this size of the *TAS3*
237 gene family is maintained across angiosperms, with most having fewer than ten *TAS3* genes and a mean
238 of four genes (Fig. S5). This is in a sharp contrast to gymnosperms in which the *TAS3* family is
239 substantially larger. The five gymnosperm species surveyed have at least 28 copies of *TAS3* genes, with
240 the *Pinus taeda* encoding as many as 71 *TAS3* copies. Another noticeable feature of the vascular *TAS3*
241 genes is that almost all of the species have both variants of *TAS3* genes (*TAS3L* and *TAS3S*) (Fig. S5), from
242 which we infer that these two types of *TAS3* genes likely have non-redundant functions.

243 We next evaluated how essential sequence motifs of *TAS3* genes changed in vascular plants. We
244 identified 3684 target sites of miR390 in 1847 vascular *TAS3* copies, including 1793 5' sites and 1891 3'
245 sites. These 5' and 3' miR390 sites showed different patterns of pairing with miR390, of which sequence
246 is highly conserved (Fig. 4A). In general, the majority of the 5' sites encode a central, 10th-position
247 mismatch, while the last four nucleotides of the pairing (18th to 21st, relative to the 5' end of miR390) are
248 always mismatched in the 3' target site (Fig. 4A). More specifically, the middle region (8th to 12th
249 nucleotides) of the 5' target site are of greater nucleotide diversity, with the 10th position generally
250 unpaired and the 11th position is predominantly a G:U pair. In contrast, the 5' five nucleotides (17th to
251 21st, relative to the 5' end of miR390) of the 3' target sites vary substantially in sequence, with the last
252 four (18th to 21st) always unpaired with miR390. Noticeable is that the final nucleotide of the 3' site (1st

253 relative to miR390) is not well conserved at all, maintained as a mismatch with miR390, unlike the 5' site
254 (Fig. 4A).

255 To assess the history of diversification of the pairing between miR390 and its target sites in *TAS3*
256 genes, we grouped all identified miR390 target sites according to the seven species lineages of land
257 plants described above, and we generated similar plots to represent miR390-*TAS3* pairing. We observed
258 substantial variation in pairing in the 5' site, especially for the middle region (8th to 12th positions) (Fig.
259 4B). Interestingly, the position most important for AGO-mediated slicing, the 10th position (of miR390)
260 was always matched in bryophytes, yet in later-diverged species, the mismatch at this position appeared
261 and seemed preferentially retained, as the proportion of mismatches gradually increased over plant
262 evolution. This was particularly noticeable in the basal angiosperms and monocots in which there were
263 almost no matched interactions at this position. For the 11th position of the 5' site, G:U pairing
264 predominated in all the lineages in spite of a substantial portion of perfect G:C pairing at the 10th
265 position observed in Monilophytes (Fig. 4B). Regarding the 3' site, its main features do not vary much
266 among the groups, including the 5' end mismatch region, the perfect match in the middle, and the high
267 proportion of mismatches for the final nucleotide (except for the Monilophytes) (Fig. 4B).

268

269 ***Evolutionary dynamic distances between tasiARFs and miR390 target sites***

270 The tasiARF is another functionally essential component of the pathway of our investigation. To
271 correctly generate the tasiARF, this siRNA needs to be in phase with a miR390 target site; in other words,
272 the distance from the cleavage site of miR390 target site to the end of the tasiARF must be a multiple of
273 21 nucleotides. Therefore, we calculated the distances and evaluated their evolutionary changes from
274 both 5' and 3' miR390 target sites to the tasiARF ends. Given that the tasiARF in vascular plants is
275 distinct from tasiARF-a1/a2/a3 found in bryophytes, which themselves vary substantially, and given the
276 large number of *TAS3* genes identified for vascular plants, we performed distance analyses only for
277 vascular *TAS3* genes.

278 Overall, there was substantial variation in the tasiARF distances (5'-site → tasiARF and tasiARF →
279 3'-site) in all lineages of vascular plants, with the exception of the eudicots, in which the tasiARF → 3'-
280 site distance of *TAS3L* and the 5'-site → tasiARF distance of *TAS3S* were highly consistent in length (Fig.
281 5A and B). For *TAS3L*, both distances were significantly shorter in the Monilophytes, but the
282 gymnosperms had a much longer 5'-site → tasiARF region compared with other lineages (Fig. 5A).

283 Monilophyte *TAS3S* also had a shorter 5'-site → tasiARF region, but the tasiARF → 3'-site distance was
284 more or less similar to those of other lineages (Fig. 5B).

285 Next, we assessed the distance from the tasiARFs to a miR390 cleavage site in terms of the phase
286 cycles of phased siRNAs, to determine which site was the trigger. The tasiARFs of *TAS3L* are mostly out-
287 of-phase with the 5' target site, with the exception of those from gymnosperms in which the tasiARFs
288 are consistently positioned at the 9th cycles according to the cleavage site of the 5' site (Fig. 5C left). In
289 contrast, the *TAS3L* tasiARFs are consistently in phase to the 3' site; in other words, the distances of the
290 3'-site → tasiARF were almost uniformly a multiple of 21 nucleotides, despite considerable length
291 variation in some groups (Fig. 5C). For *TAS3S*, its tasiARF is largely not in phase to the 5' site, except in
292 the eudicots, which had a consistent 5'-site → tasiARF distance of approximately three cycles, or 65
293 nucleotides. As with *TAS3L*, although variation in the length was observed for the 3'site → tasiARF
294 region, the distance was almost uniformly phased as well, i.e. a multiple of 21 nucleotides (Fig. 5D).
295 These results indicated that the 3' site is the main trigger site of tasiARF generation in both *TAS3L* and
296 *TAS3S*, but the gymnosperm *TAS3L* and eudicot *TAS3S* likely also generate tasiARFs triggered by the 5'
297 miR390 target site.
298

299 ***The cleavability of the 5' site and its in-phase tasiARF are selected coordinately***

300 The non-cleavable feature of 5' miR390 target site is functionally important for its role as a binding
301 site of the miR390-RISC complex, and this non-cleavability results from the presence of a mismatch at
302 the 10th position of the target site pairing (Montgomery et al. 2008b; Axtell et al. 2006). As
303 aforementioned, our analysis of the middle region of the miR390:target-site pairing of the 5' site (Fig. 4B)
304 demonstrated that, consistent with previous studies, the 10th position mismatch is indeed conserved in
305 the majority of the *TAS3* genes in vascular plants. However, we also observed that a not-insignificant
306 fraction of interactions of the 10th position of miR390 with *TAS3* are perfectly paired, especially in
307 monilophytes, gymnosperms and eudicots (Fig. 4B). Given the finding that the tasiARF in gymnosperm
308 *TAS3L* and eudicot *TAS3S* copies are mostly in phase with the 5' site as well, it is conceivable that the
309 portion of matched 10th position is contributed by the 5' sites capable of setting the phase of the tasiARF.
310 To check this possibility, we separated the 5' sites of *TAS3L* from those of *TAS3S*, and focused our
311 analyses on the middle region (8th to 12th positions, relative to the miRNA), as shown in Fig. 5E and F.
312 Although the general pattern was similar for *TAS3L* and *TAS3S*, i.e. a predominant 10th position
313 mismatch in most lineages and preferential 11th position G:U pairing, we found a few dissimilarities

314 between *TAS3L* and *TAS3S* in the pairing at these positions. Most noticeable was the level of perfect
315 matches at the 10th position for *TAS3S* compared to the majority of mismatches in *TAS3L* at the same
316 position (Fig. 5E). We then asked whether those 5' sites in phase to tasiARF were more likely to display a
317 10th position perfect match or not. When we divided the 10th position into two groups, the matched
318 group (with a "U" matching the 10th position "A" of miR390), and the mismatched group ("A", "C", or
319 "G"), and we calculated the proportion of in-phase target sites and out-of-phase target sites, separately.
320 The matched group had a much higher proportion of in-phase sites in eudicots (Fig. 5G), suggesting that
321 the in-phase and cleavable 5' site was coordinately selected during *TAS3* evolution in eudicots.
322

323 ***Strong mutual selection between tasiARF and its target site in ARF genes***

324 The miR390-TAS3-ARF pathway exerts its function via the silencing of a subgroup of *ARF* genes,
325 *ARF2/3/4* in Arabidopsis (Allen et al. 2005). In Arabidopsis, the *ARF* genes are classified into three clades
326 Clade A (*ARF5/6/7/8*); Clade B (*ARF1/2/3/4/9*), and Clade C (*ARF10/16/17*) (Finet et al., 2012). The
327 vascular tasiARFs target *ARF2/3/4* belong to Clade B, the ancestor of which likely emerged in liverworts
328 (Finet et al. 2013). Typically, *ARF2* has a single target site for the tasiARF, and *ARF3/4* have two target
329 sites (Allen et al. 2005; Axtell et al. 2006). As described above, the *ARF* genes in Clade B of bryophytes
330 are regulated by tasiARF-a1 to -a3, thereafter in evolution, this group was targeted by the tasiARF that
331 emerged in vascular plants. However, we found that some Clade B genes from mosses (for example,
332 from *P. patens*) bear analogous target site sequence of the vascular tasiARF, suggesting that this target
333 site predates the emergence of the tasiARF of vascular *TAS3* genes (Fig. S6). Combining these data with
334 the *ARF* evolution history illustrated in Finet et al. (2012), we summarized the likely path of
335 diversification of tasiARF target sites during the evolution of the Clade B *ARF* genes (Fig. 6A). The
336 interaction pattern of tasiARF with *ARF* genes was likely formed in lycophytes with only one target site.
337 In Monilophytes, genes in Clade B are targeted at a single site in most species, but a few species display
338 dual target sites. Thereafter, in evolutionary terms, this dual targeting was maintained in the subclade,
339 and likely eventually gave rise to the *ARF3/4* genes, while the single targeting was selectively retained in
340 the *ARF2* subclade, but lost in the *ARF1/9* group (Fig. 6A).

341 The target sites of vascular tasiARF were located in the middle region between two functional
342 domains (ARF and AUX/IAA) of *ARF2/3/4* genes (Fig. 6B). We recently reported that the miR482/2118
343 family displays significant sequence variation at positions matching the 3rd nucleotide of codons at the
344 miRNA target site, implying a strong selection from the functionally important P-loop motif of NB-LRR
345 proteins that shapes miRNA-target pairing (Zhang et al. 2016). In contrast, the tasiARF sequence is of

346 much lower sequence divergence, and it did not show a pattern like the miR482/2118 family, indicating
347 the selection on tasiARF pairing is distinct from the miR482/2118 case. Similarly, the tasiARF target sites,
348 unlike the miR390 target sites in *TAS3* genes which are of considerable diversity, are less divergent in
349 sequence, and consistently encode the amino acid sequence K/RVLQGQE (Fig. 6B). We also assessed
350 nucleotide diversity of the *ARF2/3/4* genes and we found that the three functional domains were, as
351 expected, of relatively low nucleotide diversity. However, the tasiARF target sites (one in *ARF2* and two
352 in *ARF3/4*) showed substantially lower nucleotide diversity, even compared to the encoded, conserved
353 functional domain, indicating a strong selection on them during evolution (Fig. 6C). Given the fact that
354 tasiARF sequences in *TAS3* genes are also extremely conserved in vascular plants, we hypothesize that
355 there is strong mutual selection between tasiARF in *TAS3* genes and its target sites in *ARF* genes.
356

357 **New regulatory mechanism of *TAS3* genes**

358 When we were annotating *TAS3* genes, we observed several other previously-undescribed
359 features of the *TAS3* gene family. First, we found a few vascular *TAS3* genes that display transcript
360 isoforms generated by alternative splicing. A good example is the *TAS3c* locus in maize, in which many
361 alternative splicing sites giving rise to numerous transcript isoforms (Fig. 7A). These isoforms selectively
362 spliced out the three essential components of the *TAS3* gene; for instance, the 5' site is missing for splice
363 variant T4, the 3' site is missing for T1 and T2, both target sites missing for T5, T8 and T9, and all of the
364 target sites and tasiARFs missing for T7, T10 and T13 (Fig. 7A). While we currently have no evidence of
365 functional relevance for these variants, it's possible that alternative splicing can serve as another layer
366 of regulation to fine-tune the activity of *TAS3* genes and subsequent tasiARF production. For example,
367 there is evidence that small ORFs encoded by *TAS* genes play functional roles (Yoshikawa et al., 2016),
368 and these splice variants could mediate ribosome loading, stalling, or peptide production, independent
369 of tasiRNA biogenesis.

370 Another new feature that we observed is an abnormal pairing interaction of miR390 with a few
371 target sites in *TAS3* genes; this pairing displays large bulges in the seed sequence region (2nd to 13th
372 positions) (Fig. 7B). The first example is the *TAS3-1* gene in soybean (*Glycine max*). We can infer based
373 on the register of the siRNAs that its 3' target site is cleaved to set the phasing, despite a predicted 4-nt
374 bulge present between the 6th and 7th positions of miR390 (Fig. 7B). Another three cases were predicted
375 to generate an 8- or 3-nt bulge (Fig. 7B). This type of abnormal, bulge-containing miRNA-target
376 interaction was recently reported and validated in Arabidopsis for miR398 (Brousse et al. 2014), the only

377 other known case of this type. Our results suggest that this is pairing is not unique to miR398, and large
378 asymmetrical bulges in miRNA:target pairing are at least allowable in plants.

379

380 Discussion

381 The presence of miR390 and *TAS3* was tracked back to liverworts (Lin et al. 2016), while the ARF
382 domain encoded by *ARF* genes likely first appeared in the land plants (Finet et al. 2013). We
383 demonstrated that *TAS3* in liverworts produces tasiRNAs to target *ARF* genes, suggesting this was the
384 earliest function of *TAS3*, a key function maintained throughout land plants. We also observed in
385 liverworts the conservation of another *TAS3*-derived tasiRNA that, in mosses, targets *AP2* genes
386 (referred to as tasiAP2), but we were unable to confirm this function in liverworts. It is possible that this
387 tasiRNA in liverwort *TAS3* genes emerged before the appearance of tasiAP2 target sites in *AP2* genes, or
388 this tasiRNA has an unidentifiable function or target.

389 Although the role of *TAS3* in regulating *ARF* genes is conserved across land plants, the bryophyte
390 *TAS3* genes are structurally different from those in vascular plants (Axtell et al. 2006). In other words,
391 tasiARFs are different in sequence between bryophytes and vascular plants. In our model for tasiARF
392 evolution, the tasiARF was derived from the duplication of the 5' target site of miR390, and the short
393 *TAS3* variant (*TAS3S*) is the ancestor of the long *TAS3* (*TAS3L*). We identified the vascular *TAS3* in a
394 lycophyte, indicating the transition of tasiRNA sequences is likely associated with the development of
395 vascular tissue in plants, as lycophytes were among the first vascular plants on earth. Measured across
396 the vascular plants, there are nearly always two types of *TAS3* genes (*TAS3S* and *TAS3L*) present in each
397 plant genome and totaling approximately four members in most species. The deep conservation of
398 these structures suggests they are not functionally redundant. Future work could address why, perhaps
399 by selective deletion of the two types using CRISPR/Cas9. Another striking observation was that the
400 *TAS3* copy number is significantly expanded in conifers, reminiscent of the expansion of *NB-LRR*-
401 targeting miRNAs (Xia et al., 2015). Despite evidence of whole genome duplications in spruce (Li et al.
402 2015), the >10-fold higher copy number in conifers relative to angiosperms is extraordinary. Perhaps
403 functional differences in tasiRNA movement in gymnosperms required added copies of *TAS3* genes, but
404 these copies were made redundant and lost by angiosperm-specific evolutionary adaptations, possibly
405 improved tasiRNA mobility.

406 In plants, miRNA/miRNA duplexes are released by two sequential cuts of their hairpin precursors
407 by DCL1; these cuts can occur either base-to-loop or loop-to-base (Bologna et al. 2009, 2013). For other
408 miRNAs, it was observed that a conserved length of the basal stem region ensures accurate cuts made

409 by DCL1 (Werner et al. 2010; Song et al. 2010; Mateos et al. 2010). We found that miR390 is processed
410 in a base-to-loop direction, with the first cut by DCL1 occurring at a position ~15 nt from a basal
411 unpaired region (> 4 nt). While this basal region (the unpaired region to the site of the first cut) is a
412 length consistent with other base-to-loop processed miRNAs, by comparison across the seed plants, we
413 found that the sequence is also relatively conserved, indicating that selection can maintain bases in the
414 hairpin other than the miRNA/miRNA*. The conservation of this paired region was recently described for
415 many plant miRNAs (Chorostecki et al. *in preparation*).

416 One of the major differences between two main mechanisms of tasiRNA/phasiRNA biogenesis
417 (“one-hit” and “two-hit” models) is the direction of tasiRNA production. In the “two-hit” model,
418 tasiRNAs are produced in a 3' to 5' direction, in contrast to the predominant 5' to 3' Dicer processing (i.e.
419 the “one-hit” model). miR390-TAS3 is the quintessential “two-hit” locus, yet its 3' to 5' processing is
420 distinctive and rare. Evolutionary analyses of miR390-TAS3 pairing revealed two distinct patterns of
421 pairing of the two target sites: (i) a conserved mismatched region mainly caused by the 10th position of
422 the 5' site (previously known – see below), and (ii) an open, unpaired region in the 3' end of the 3' target
423 site (from our study). The wide conservation in vascular plants of these features implies functional
424 relevance. Studies in *Arabidopsis* have shown that the non-cleavability of the 5' site, caused by the
425 central mismatch (10th position), is essential, mediating miR390 binding via AGO7 (Rajeswaran and
426 Pooggin 2012). Changing the 10th mismatch into a perfect match compromises tasiRNA biogenesis
427 (Axtell et al. 2006; Montgomery et al. 2008a). However, a substantial portion of *TAS3* genes, especially
428 the *TAS3S* subset, having a cleavable 5' site (A:U pair at the 10th position), and many of these sites,
429 particularly in eudicots, trigger tasiARF production. This indicates that the non-cleavability of the 5' site
430 is helpful but not necessary for tasiARF production. Another notable feature of the 5' site pairing is the
431 predominant G:U pairing at the 11th position; this preferential wobble pairing might be helpful for
432 maintaining the non-cleavability of the 5' site, which is believed to be mainly caused by the 10th position
433 mismatch. In contrast, the pairing of the 3' site has a consistently matched middle region, but an open,
434 unpaired 3' end region. The paired middle region could ensure the cleavage of the 3' site, make it the
435 typical trigger site for secondary tasiRNAs. The 3' end open region may direct the 3' to 5' production of
436 *TAS3* tasiRNAs. Perhaps after cleavage, the 3' end open region makes the cleaved mRNA end more
437 accessible to RDR6 to facilitate downstream tasiRNA production.

438 Besides the cleavability of the target site, the distance between the miR390 target site and tasiARF
439 also appears to be a determinant for phasiRNA biogenesis. tasiARF production requires distances in
440 multiples of 21 nucleotides from the cleavage site (“in register”). We showed that the distance of the 3'

441 side of *TAS3* is more consistently a multiple of 21 nt despite considerable length variation; the 3' site
442 also displayed fewer 10th position mismatches (i.e. better cleavability). However, we noted several
443 exceptions. The *TAS3L* in gymnosperms and the *TAS3S* in eudicots had a highly consistent distance on
444 the 5' side, in approximate phase with tasiARF, and the cleavability of the 5' site is coordinately selected
445 with the in-phase distance to the tasiARF in eudicots, suggesting that the 5' site in those *TAS3* genes is
446 likely to serve as a trigger site of tasiARF production as well. Therefore, our results suggest that some
447 *TAS3* loci in vascular plants are likely bi-directionally processed, consistent with the observation of the
448 original bidirectional processing of functional tasiRNAs in bryophytes (Axtell et al. 2006). For instance,
449 the two target sites of *TAS3* in *P. patens* are both cleavable, and tasiARF-a2 is in phase with the 3' site
450 while tasiARF-a3 is in phase with the 5' site. This bidirectional processing thus yields additional questions
451 about this “two-hit” mechanism. How is the activity of the two sites coordinated? Does cleavage occur
452 simultaneously at both sites or one site at a time?

453 miRNAs, tasiRNAs, or other type of sRNAs and their targets genes are a pair of partners,
454 functioning via their interactions, based on sequence complementarity. Few studies have deeply
455 investigated this sRNA:target partnership over evolutionary time. In the case of the widely conserved
456 miR482/2118 family , we described selection from target protein-coding genes to miRNAs; in other
457 words, the essential function of the P-loop encoded in *NB-LRR* genes, targeted by miR482/2118, is most
458 important, as miRNA variation matches a degenerate nucleotide change at the third position of each
459 codon in the target (Zhang et al. 2016). In the current study, we detected a distinct pattern of selection
460 between tasiARF and target site in *ARF* genes in vascular plants. Both are depleted of variation,
461 indicating a strong mutual selection. The tasiARF target site sequences in *ARF* genes show no periodical
462 variation (at the third position), indicating the target site sequence is not under strong selection at the
463 amino acid level, in accordance with the location of the tasiARF target site between two encoded
464 domains of ARF proteins, the ARF domain and AUX/IAA domain. The target site in the middle region is of
465 less functional importance at the protein level. This is in contrast to the location of miR482/2118 target
466 site in a functionally critical domain (Zhang et al. 2016). However, the sequence variation (nucleotide
467 diversity) of the tasiARF target sites in *ARF* genes is dramatically less than other gene regions, even the
468 conserved functional protein domains, suggesting that the tasiARF target site is under a selective force
469 stronger than selection than that of the encoded protein domains. Combined with the fact that tasiARF
470 sequences in vascular *TAS3* copies demonstrate substantially less sequence variation, we believe that
471 there is a robust selective connection between tasiARF and its target site in *ARF* genes, which permit
472 little sequence variation in either component over evolutionary time.

473

474 **Methods**

475 ***Genome sequences and 1KP data***

476 Genome sequences of 159 species were retrieved from either the Phytozome or NCBI. The
477 assembled transcriptome data of the 1000 Plant Transcriptome Project (“1KP”) was kindly shared by the
478 Wang lab at the University of Alberta, Canada (Matasci et al. 2014).

479

480 ***NGS data and analyses***

481 RNA of *Phylloglossum drummondii* was extracted using PureLink Plant RNA Reagent. A sRNA
482 library was constructed using the Illumina TruSeq sRNA kit, and sequenced on the Illumina HiSeq
483 platform at the University of Delaware. The sRNA data was deposited in NCBI GEO (Gene Expression
484 Omnibus) under the accession number GSE90706.

485 sRNA and PARE data of *Marchantia polymorpha* were retrieved from NCBI Short Read Archive
486 (SRA) under accession numbers SRR2179617 and SRR2179371, respectively (Lin et al. 2016). sRNA reads
487 were mapped to reference genome or transcripts by Bowtie (Langmead et al. 2009), and PARE data was
488 analyzed using Cleveland 2.0 (Addo-Quaye et al. 2009).

489 Paired-end RNA-seq data for maize were downloaded from NCBI SRA under accession numbers
490 SRR1213570 and SRR1213571 (Wang et al. 2015). RNA-seq reads were mapped to maize genome using
491 STAR v2.4.2a (Dobin et al. 2013) and transcripts of the *ZmTAS3c* locus were annotated using Cufflinks
492 v2.2.1 (Trapnell et al. 2012). The mapped bam files of two libraries were merged and viewed using
493 Integrative Genomics Viewer v2.3.59 (Robinson et al. 2011).

494

495 ***Homologous gene identification***

496 For the identification of *MIR390* genes, mature sequences of miR390 were retrieved from
497 miRBase, and used to search for homologous sequence using FASTA36 allowing two mismatches. After
498 that, \pm 500 bp sequence were excerpted for each homologous sequence from reference sequences and
499 used for the evaluation of secondary structure. Only those genomic loci or transcripts with a good stem
500 loop structure (\leq 4 nt mismatches and \leq 1 nt bulge) and with the mature miRNA in the 5' arm were
501 regarded as good *MIR390* genes.

502 For the identification of *TAS3* genes, < 500 bp genomic loci (for genomes) or EST sequences (for
503 transcriptome data) with evidence of at least two components of the two miR390 target sites and one
504 tasiRNA (tasiARF for vascular plants, tasiAP2 or tasiARF-a2 for bryophytes) were considered as *TAS3*
505 candidates. Their identity as a *TAS3* gene was further assessed by manual sequence comparisons. The
506 tool MEME (Bailey et al. 2009) was also used to profile the signature sequence motif of *TAS3* genes.

507 To identify tasiARF-targeted *ARF* genes, firstly Arabidopsis and subsequently rice *ARF* proteins
508 were used as bait sequences to identify *ARF* homologous genes, using either TBLASTN for annotated
509 genomes or 1KP transcriptome data or genBlast (She et al. 2011) for unannotated genomes. Secondly,
510 TargetFinder (<https://github.com/carringtonlab/TargetFinder>) was used to identify tasiARF-targeted *ARF*
511 genes. Thirdly, *ARF3/4* and *ARF2* genes were distinguished by the number of target sites as *ARF3/4*
512 genes have two tasiARF target sites and *ARF2* genes have only one target site. AGO proteins were
513 identified using BLASTP for selected annotated genomes or TBLASTN for 1KP data using Arabidopsis and
514 rice AGO proteins as bait sequences. Only full-length AGO protein sequences from sequenced genomes
515 and AGO sequences with ≥ 800 amino acids from the 1KP data were chosen for subsequent phylogenetic
516 tree construction.

517

518 ***Multiple alignment and tree construction***

519 Amino acid sequences of Argonautes (≥ 800 amino acids), annotated from transcripts and
520 genomes, were aligned using MUSCLE v3.8.31 with default parameters (Edgar 2004). The regions poorly
521 aligned were trimmed using trimAl v1.4 (Capella-Gutiérrez et al. 2009), and the trimmed alignments
522 were used for construction of a maximum likelihood (ML) tree using RAxML v8.1.1 under the GTRCAT
523 model (Stamatakis 2014). For a tree of bryophyte *TAS3* genes (Fig. S3B), the nucleotide sequences of
524 those genes were aligned and edited similarly, and the ML tree was made using RAxML under the
525 PROTGAMMAUTO model. For each tree, 100 replicates were conducted to generate bootstrap values.
526 The trees were viewed using Dendroscope v3.5.7 (Huson and Scornavacca 2012).

527 Jalview was used for the viewing of alignment results (Waterhouse et al. 2009). The R package was
528 used to make violin plots and conduct statistical analyses. Sequence logos of sRNA and target sites were
529 generated using Weblogo (Crooks et al. 2004). To calculate the nucleotide diversity (π) of *ARF* genes,
530 the amino acid sequences of *ARFs* were generated by translation of the genes, aligned using MUSCLE,
531 then the protein sequence alignment was used to generate the alignment of nucleotide sequences using
532 PAL2NAL (Suyama et al. 2006). Subsequently, poorly aligned regions, those with $<30\%$ nucleotide

533 coverage, were removed, and finally the nucleotide diversity (π) at a single nucleotide level was
534 calculated using a 20 nt sliding window.

535

536

537 **Acknowledgements**

538 We thank members of the Meyers lab for helpful discussions and input. This study was supported
539 by US National Science Foundation, Division of Integrative Organismal Systems award #1257869 and the
540 Chinese Thousand Young Talents Program. We are grateful to Dennis Stevenson and Ryan Lister for
541 assistance in obtaining *Phylloglossum drummondii* material. We are also grateful to Gane Ka-Shu Wang
542 for assistance with access to the 1KP data.

543 **References**

- 544 Addo-Quaye C, Miller W, Axtell MJ. 2009. CleaveLand: a pipeline for using degradome data to find
545 cleaved small RNA targets. *Bioinformatics* **25**: 130–131.
- 546 Adenot X, Elmayan T, Lauressergues D, Boutet S, Bouché N, Gasciolli V, Vaucheret H. 2006. DRB4-
547 dependent TAS3 trans-acting siRNAs control leaf morphology through AGO7. *Curr Biol* **16**: 927–932.
- 548 Allen E, Xie Z, Gustafson AM, Carrington JC. 2005. microRNA-directed phasing during trans-acting siRNA
549 biogenesis in plants. *Cell* **121**: 207–221.
- 550 Arif MA, Fattash I, Ma Z, Cho SH, Beike AK, Reski R, Axtell MJ, Frank W. 2012. DICER-LIKE3 activity in
551 *physcomitrella patens* DICER-LIKE4 mutants causes severe developmental dysfunction and sterility.
552 *Mol Plant* **5**: 1281–1294.
- 553 Axtell MJ. 2013. Classification and comparison of small RNAs from plants. *Annu Rev Plant Biol* **64**: 137–
554 159.
- 555 Axtell MJ, Jan C, Rajagopalan R, Bartel DP. 2006. A two-hit trigger for siRNA biogenesis in plants. *Cell* **127**:
556 565–577.
- 557 Axtell MJ, Snyder JA, Bartel DP. 2007. Common functions for diverse small RNAs of land plants. *Plant Cell*
558 **19**: 1750–1769.
- 559 Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME Suite:
560 Tools for motif discovery and searching. *Nucleic Acids Res* **37**: 202–208.
- 561 Bartel DP. 2009. MicroRNAs: target recognition and regulatory functions. *Cell* **136**: 215–233.
- 562 Bologna NG, Mateos JL, Bresso EG, Palatnik JF. 2009. A loop-to-base processing mechanism underlies
563 the biogenesis of plant microRNAs miR319 and miR159. *EMBO J* **28**: 3646–3656.
- 564 Bologna NG, Schapire AL, Zhai J, Bologna G, Schapire AL, Zhai J, Chorostecki U, Boisbouvier J, Meyers BC,
565 Palatnik JF. 2013. Multiple RNA recognition patterns during microRNA biogenesis in plants.
566 *Genome Res* **23**: 1675–1689.
- 567 Brousse C, Liu Q, Beauclair L, Deremetz A, Axtell MJ, Bouché N. 2014. A non-canonical plant microRNA
568 target site. *Nucleic Acids Res* **42**: 5270–5279.
- 569 Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: A tool for automated alignment
570 trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.
- 571 Chen X. 2009. Small RNAs and their roles in plant development. *Annu Rev Cell Dev Biol* **25**: 21–44.
- 572 Crooks G, Hon G, Chandonia J, Brenner S. 2004. WebLogo: a sequence logo generator. *Genome Res* **14**:
573 1188–1190.
- 574 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013.

- 575 STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **1–7**.
- 576 Dotto MC, Petsch KA, Aukerman MJ, Beatty M, Hammell M, Timmermans MCP. 2014. Genome-wide
577 analysis of *leafbladeless1*-regulated and phased small RNAs underscores the importance of the
578 *TAS3* ta-siRNA pathway to maize development. *PLoS Genet* **10**: e1004826.
- 579 Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic
580 Acids Res* **32**: 1792–1797.
- 581 Fahlgren N, Jogdeo S, Kasschau KD, Sullivan CM, Chapman EJ, Laubinger S, Smith LM, Dasenko M, Givan
582 SA, Weigel D, et al. 2010. MicroRNA gene evolution in *Arabidopsis lyrata* and *Arabidopsis thaliana*.
583 *Plant Cell* **22**: 1074–89.
- 584 Fahlgren N, Montgomery TA, Howell MD, Allen E, Dvorak SK, Alexander AL, Carrington JC. 2006.
585 Regulation of *AUXIN RESPONSE FACTOR3* by *TAS3* ta-siRNA affects developmental timing and
586 patterning in *Arabidopsis*. *Curr Biol* **16**: 939–944.
- 587 Finet C, Berne-Dedieu A, Scutt CP, Marlétaz F. 2013. Evolution of the *ARF* gene family in land plants: Old
588 domains, new tricks. *Mol Biol Evol* **30**: 45–56.
- 589 Garcia D, Collier SA, Byrne ME, Martienssen RA. 2006. Specification of leaf polarity in *Arabidopsis* via the
590 trans-acting siRNA pathway. *Curr Biol* **16**: 933–938.
- 591 Guilfoyle TJ, Hagen G. 2007. Auxin response factors. *Curr Opin Plant Biol* **10**: 453–460.
- 592 Howell MD, Fahlgren N, Chapman EJ, Cumbie JS, Sullivan CM, Givan S a, Kasschau KD, Carrington JC.
593 2007. Genome-wide analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 pathway in
594 *Arabidopsis* reveals dependency on miRNA- and tasiRNA-directed targeting. *Plant Cell* **19**: 926–942.
- 595 Hunter C, Willmann MR, Wu G, Yoshikawa M, de la Luz Gutiérrez-Nava M, Poethig SR. 2006. Trans-acting
596 siRNA-mediated repression of *ETTIN* and *ARF4* regulates heteroblasty in *Arabidopsis*. *Development*
597 **133**: 2973–2981.
- 598 Huson DH, Scornavacca C. 2012. Dendroscope 3: An interactive tool for rooted phylogenetic trees and
599 networks. *Syst Biol* **61**: 1061–1067.
- 600 Jones-Rhoades MW, Bartel DP, Bartel B. 2006. MicroRNAs and their regulatory roles in plants. *Annu Rev
601 Plant Biol* **57**: 19–53.
- 602 Krasnikova MS, Goryunov D V, Troitsky A V, Solovyev AG, Ozerova L V, Morozov SY. 2013. Peculiar
603 evolutionary history of miR390-guided *TAS3-like* genes in land plants. *Sci World J* **2013**.
- 604 Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short
605 DNA sequences to the human genome. *Genome Biol* **10**: R25.
- 606 Li Z, Baniaga AE, Sessa EB, Scascitelli M, Graham SW, Rieseberg LH, Barker MS. 2015. Early genome

- 607 duplications in conifers and other seed plants. *Sci Adv* **1**: e1501084.
- 608 Lin PC, Lu CW, Shen BN, Lee GZ, Bowman JL, Arteaga-Vazquez MA, Daisy Liu LY, Hong SF, Lo CF, Su GM,
609 et al. 2016. Identification of miRNAs and their targets in the liverwort *Marchantia polymorpha* by
610 integrating RNA-Seq and degradome analyses. *Plant Cell Physiol* **57**: 339–358.
- 611 Ma Z, Coruh C, Axtell MJ. 2010. *Arabidopsis lyrata* small RNAs: transient MIRNA and small interfering
612 RNA loci within the *Arabidopsis* genus. *Plant Cell* **22**: 1090–103.
- 613 Mallory A, Vaucheret H. 2010. Form, function, and regulation of ARGONAUTE proteins. *Plant Cell* **22**:
614 3879–3889.
- 615 Marin E, Jouannet V, Herz A, Lokerse AS, Weijers D, Vaucheret H, Nussaume L, Crespi MD, Maizel A.
616 2010. miR390, *Arabidopsis TAS3* tasiRNAs, and their AUXIN RESPONSE FACTOR targets define an
617 autoregulatory network quantitatively regulating lateral root growth. *Plant Cell* **22**: 1104–1117.
- 618 Matasci N, Hung L-H, Yan Z, Carpenter EJ, Wickett NJ, Mirarab S, Nguyen N, Warnow T, Ayyampalayam S,
619 Barker MS, et al. 2014. Data access for the 1,000 Plants (1KP) project. *Gigascience* **3**: 17.
- 620 Mateos JL, Bologna NG, Chorostecki U, Palatnik JF. 2010. Identification of microRNA processing
621 determinants by random mutagenesis of *Arabidopsis MIR172a* precursor. *Curr Biol* **20**: 49–54.
- 622 Montgomery T a., Howell MD, Cuperus JT, Li D, Hansen JE, Alexander AL, Chapman EJ, Fahlgren N, Allen
623 E, Carrington JC. 2008a. Specificity of ARGONAUTE7-miR390 interaction and dual functionality in
624 TAS3 trans-acting siRNA formation. *Cell* **133**: 128–141.
- 625 Montgomery T a, Yoo SJ, Fahlgren N, Gilbert SD, Howell MD, Sullivan CM, Alexander A, Nguyen G, Allen E,
626 Ahn JH, et al. 2008b. AGO1-miR173 complex initiates phased siRNA formation in plants. *Proc Natl
627 Acad Sci U S A* **105**: 20055–20062.
- 628 Rajeswaran R, Pooggin MM. 2012. RDR6-mediated synthesis of complementary RNA is terminated by
629 miRNA stably bound to template RNA. *Nucleic Acids Res* **40**: 594–599.
- 630 Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011.
631 Integrative Genomics Viewer. *Nat Biotechnol* **29**: 24–26.
- 632 She R, Chu JSC, Uyar B, Wang J, Wang K, Chen N. 2011. genBlastG: Using BLAST searches to build
633 homologous gene models. *Bioinformatics* **27**: 2141–2143.
- 634 Simonini S, Deb J, Moubayidin L, Stephenson P, Valluru M, Freire-rios A, Sorefan K, Weijers D,
635 Østergaard L. 2016. A noncanonical auxin-sensing mechanism is required for organ morphogenesis
636 in *Arabidopsis*. **30**: 2286–2296.
- 637 Song L, Axtell MJ, Fedoroff N V. 2010. RNA secondary structural determinants of miRNA precursor
638 processing in *Arabidopsis*. *Curr Biol* **20**: 37–41.

- 639 Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large
640 phylogenies. *Bioinformatics* **30**: 1312–1313.
- 641 Suyama M, Torrents D, Bork P. 2006. PAL2NAL: Robust conversion of protein sequence alignments into
642 the corresponding codon alignments. *Nucleic Acids Res* **34**: 609–612.
- 643 Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012.
644 Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and
645 Cufflinks. *Nat Protoc* **7**: 562–78.
- 646 Vaucheret H. 2008. Plant ARGONAUTES. *Trends Plant Sci* **13**: 350–358.
- 647 Wang H, Niu QW, Wu HW, Liu J, Ye J, Yu N, Chua NH. 2015. Analysis of non-coding transcriptome in rice
648 and maize uncovers roles of conserved lncRNAs associated with agriculture traits. *Plant J* **84**: 404–
649 416.
- 650 Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2-A multiple
651 sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189–1191.
- 652 Werner S, Wollmann H, Schneeberger K, Weigel D. 2010. Structure determinants for accurate processing
653 of miR172a in *Arabidopsis thaliana*. *Curr Biol* **20**: 42–48.
- 654 Xia R, Xu J, Arikit S, Meyers BC. 2015a. Extensive families of miRNAs and PHAS loci in norway spruce
655 demonstrate the origins of complex phasiRNA networks in seed plants. *Mol Biol Evol* **32**: 2905–
656 2918.
- 657 Xia R, Ye S, Liu Z, Meyers BC, Liu Z. 2015b. Novel and recently evolved microRNA clusters regulate
658 expansive F-BOX gene networks through phased small interfering RNAs in wild diploid strawberry.
659 *Plant Physiol* **169**: 594–610.
- 660 Xia R, Zhu H, An Y, Beers EP, Liu Z. 2012. Apple miRNAs and tasiRNAs with novel regulatory networks.
661 *Genome Biol* **13**: R47.
- 662 Yifhar T, Pekker I, Peled D, Friedlander G, Pistunov A, Sabban M, Wachsman G, Alvarez JP, Amsellem Z,
663 Eshed Y. 2012. Failure of the tomato trans-acting short interfering RNA program to regulate AUXIN
664 RESPONSE FACTOR3 and ARF4 underlies the wiry leaf syndrome. *Plant Cell* **24**: 3575–3589.
- 665 Yoshikawa M, Peragine A, Park MY, Poethig RS. 2005. A pathway for the biogenesis of trans-acting
666 siRNAs in *Arabidopsis*. *Genes Dev* **19**: 2164–2175.
- 667 Zhang H, Xia R, Meyers BC, Walbot V. 2015. Evolution, functions, and mysteries of plant ARGONAUTE
668 proteins. *Curr Opin Plant Biol* **27**: 84–90.
- 669 Zhang Y, Xia R, Kuang H, Meyers BC. 2016. The diversification of plant NBS-LRR defense genes directs the
670 evolution of microRNAs that target them. *Mol Biol Evol* **33**: 2692–2705.

671 Zhou C, Han L, Fu C, Wen J, Cheng X, Nakashima J, Ma J, Tang Y, Tan Y, Tadege M, et al. 2013. The trans-
672 acting short interfering RNA3 pathway and NO APICAL MERISTEM antagonistically regulate leaf
673 margin development and lateral organ separation, as revealed by analysis of an *argonaute7/lobed*
674 *leaflet1* mutant in *Medicago truncatula*. *Plant Cell* **25**: 4845–4862.
675

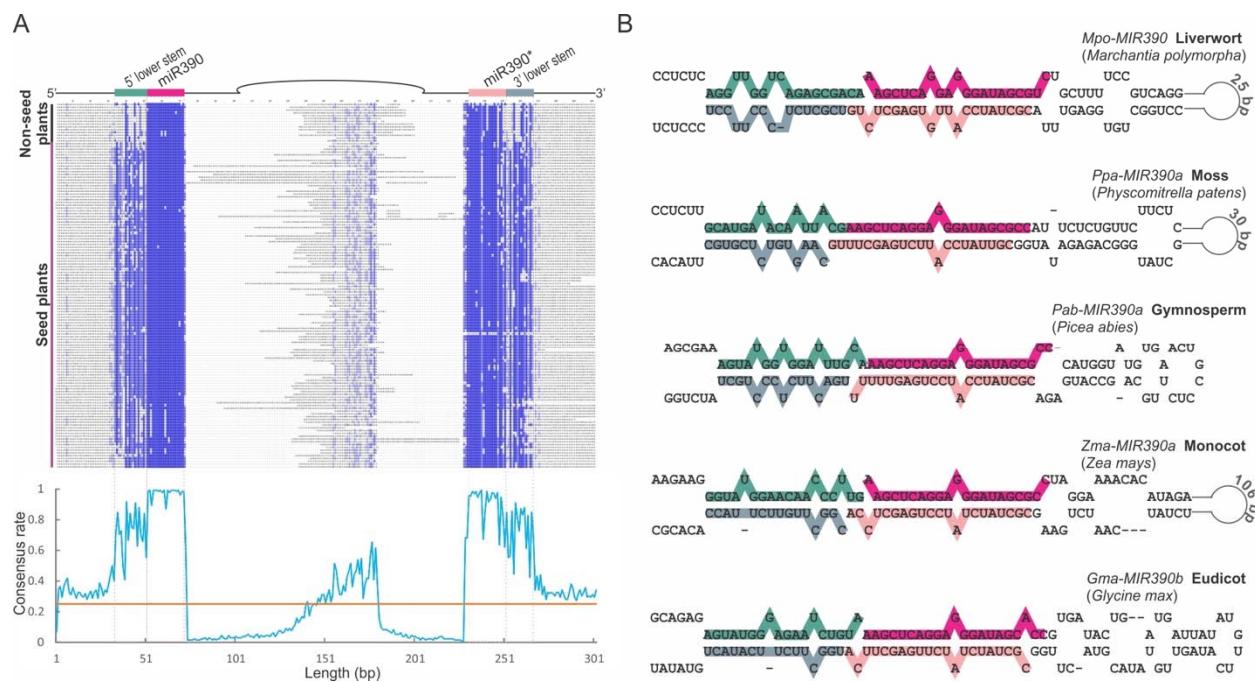


Figure 1. The lower stem region of *MIR390* is conserved in land plants.

(A) Nucleotide sequence alignment of *MIR390* precursor genes (± 50 bp before/after the miR390/miR390* region) with different sequence regions denoted above. The consensus rate of diversity of each position in the alignment is shown in the plot below with the orange line indicating the 25% level, since in a sequence randomized by neutral evolution, each nucleotide (A/U/C/G) would comprise 25% of each position. (B) Examples of stem-loop structures of *MIR390* precursor transcripts. The miRNA and lower-stem regions are indicated according to the colors shown in the top of panel A.

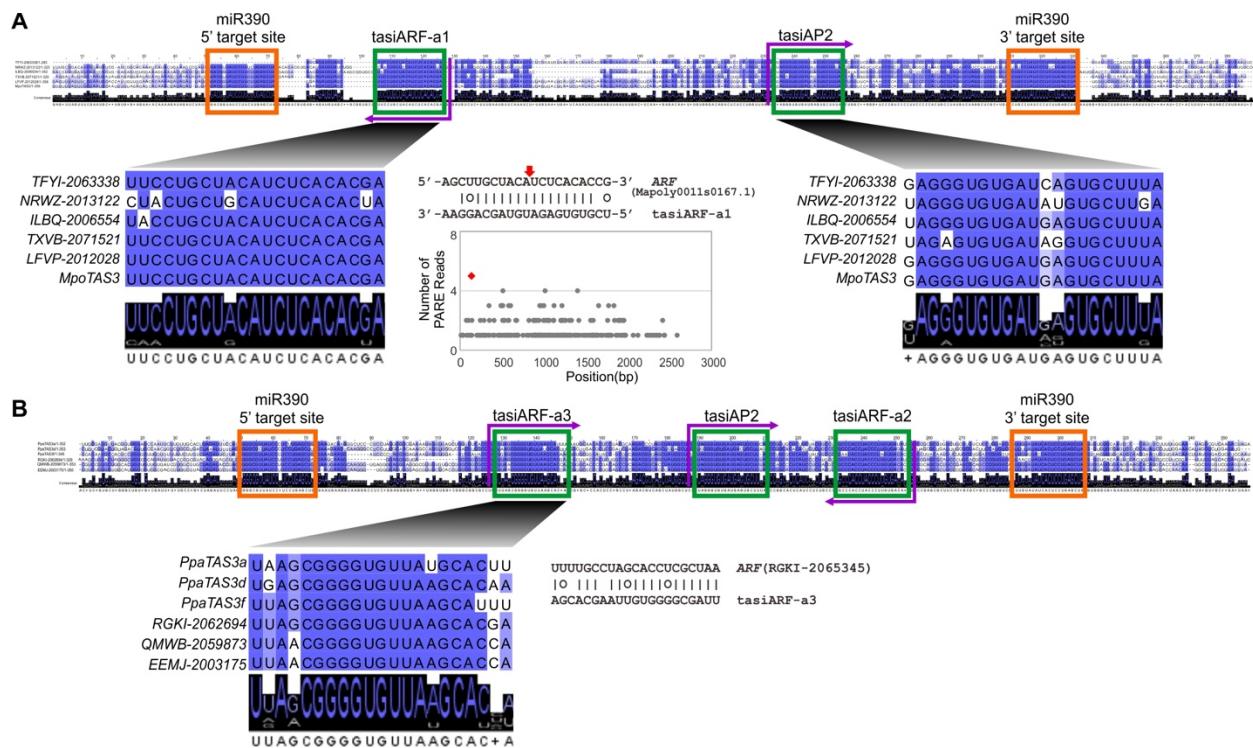


Figure 2. TAS3 originated to regulate ARF genes in land plants.

(A) Conserved motifs in TAS3 transcripts from liverworts. Purple arrows indicate the encoded strand of tasiRNAs; the left-pointing arrow indicates that the functional tasiRNA is located on the anti-sense strand, and the right-pointing arrow indicates that the tasiRNA is on the sense strand. tasiARF-a1 encoded in liverwort TAS3 genes targets ARF genes with the tasiRNA:target pairing (cleavage site marked with a red arrow) and validating, experimentally-derived PARE data shown in the middle. The red dot in the plot of PARE data marks the cleavage site directed by tasiARF-a1. (B) Conserved motifs in a few representative TAS3 genes in mosses. Besides tasiARF-a2, previously reported to target ARF genes, another tasiRNA, denoted as tasiARF-a3, was predicted to target ARF genes. The tasiARF-a2 and tasiARF-a3 sequences are encoded in the anti-sense and sense strands of TAS3 transcripts, respectively.

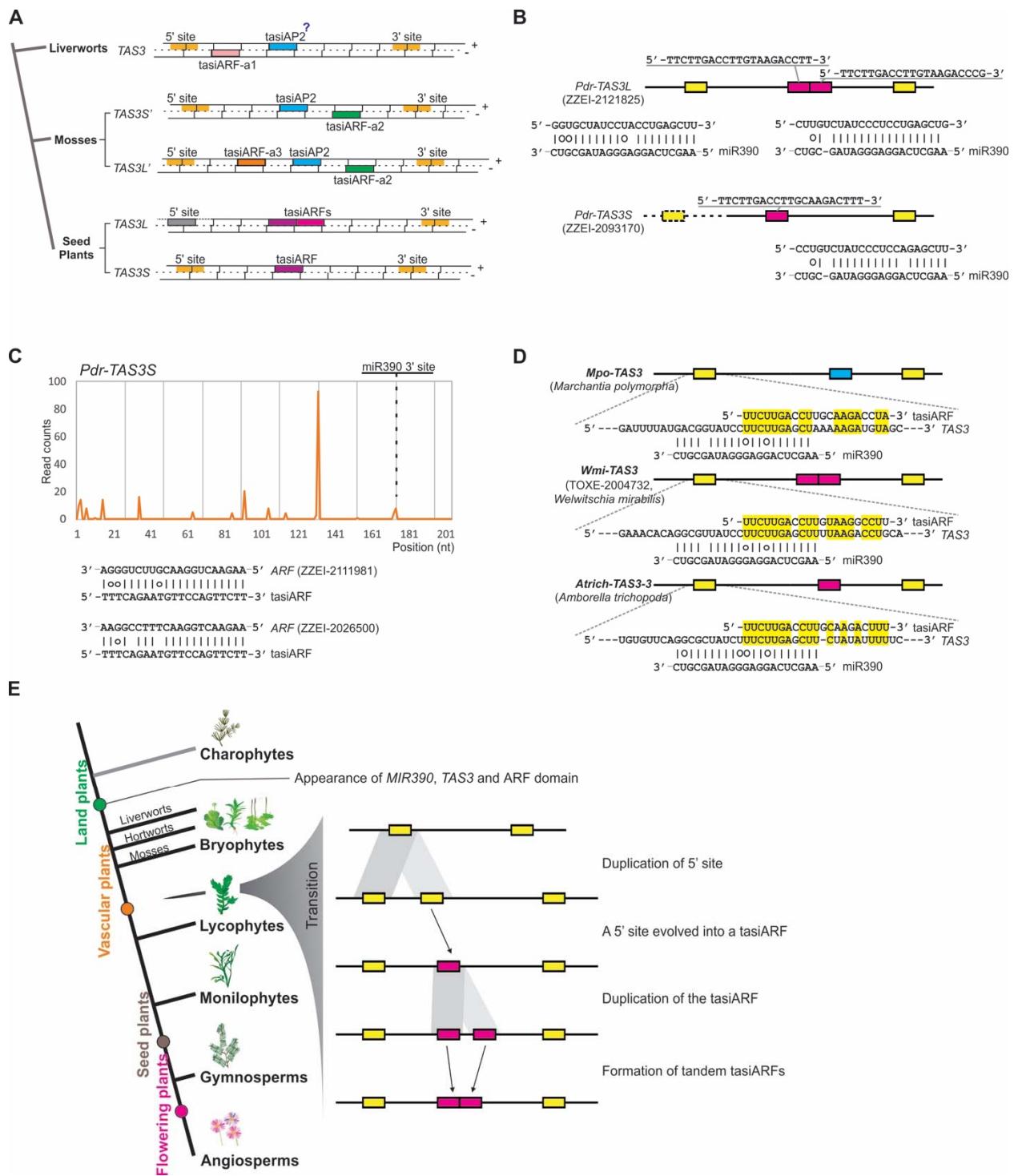


Figure 3. The inferred evolutionary progression of TAS3 genes in land plants.

(A) A summary of TAS3 gene structures observed in land plants. Colored bars denote different features, as indicated; the grey 5' miR390 site is not cleaved. The question mark "?" denotes that the function of tasiAP2 (targeting AP2 genes) could not be validated in liverworts. (B) Two TAS3 gene structures found in the lycophyte species, *Phylloglossum drummodii*. (C) TAS3

transcripts produce tasiARFs to regulate *ARF* genes in *Phylloglossum drummodii*. (D) tasiARF shows sequence similarity to the region partially covering the 5' miR390 target site of cognate *TAS3* genes. Three representative *TAS3* genes from different species are displayed here. Identical nucleotides between tasiARF and the region partially covering the 5' miR390 target site are highlighted in yellow. (E) An evolutionary model for the divergence of *TAS3* genes in land plants. The tasiARF sequence originated from the duplication of the 5' miR390 target site and the *TAS3S* genes (with a single tasiARF) might be the ancestor of the *TAS3L* genes (with two tandem tasiARFs).

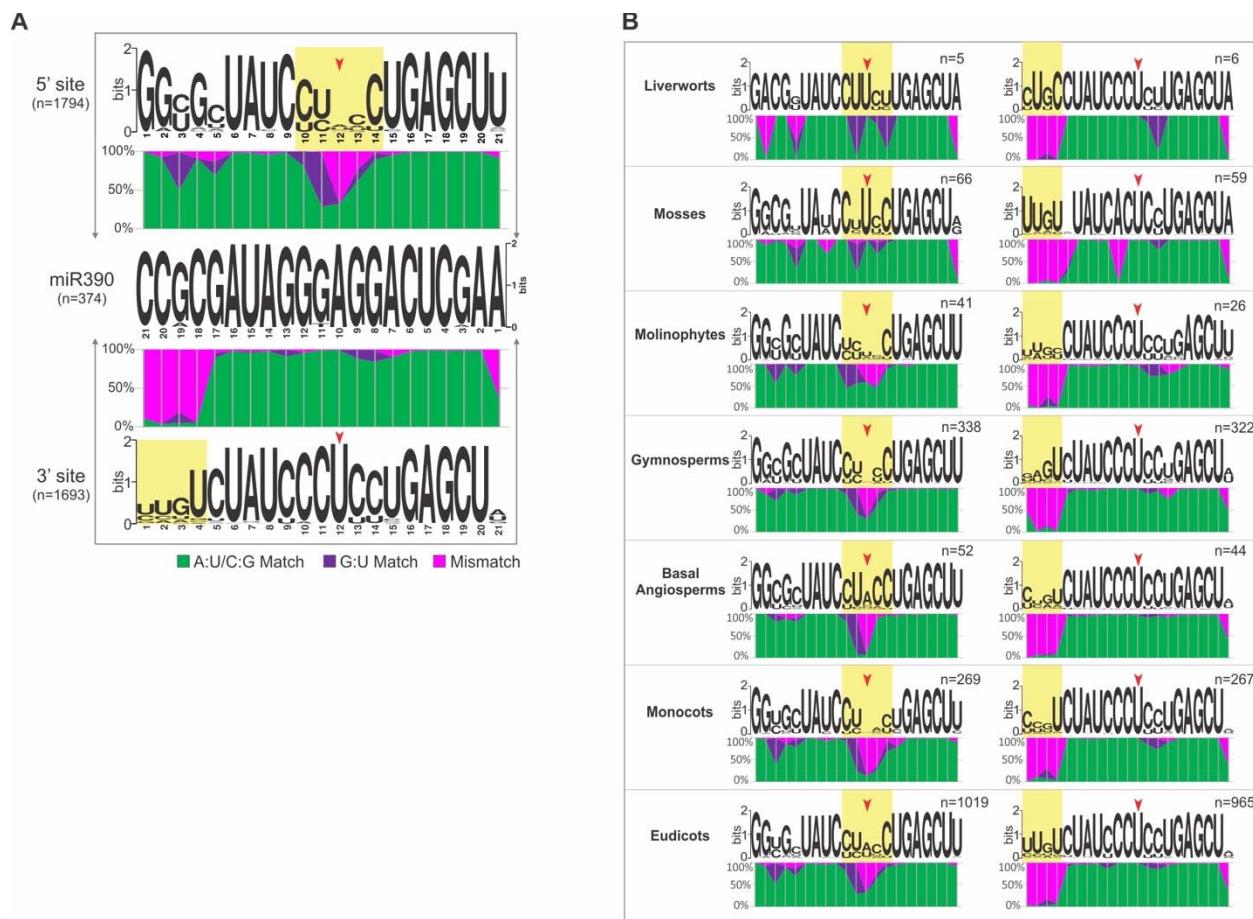


Figure 4. Pairing features and evolutionary variation of the two target sites of miR390 in *TAS3* genes.

(A) Distinct pairing patterns of the two miR390 target sites in *TAS3* genes. Sequence logos were generated using WebLogo. Different nucleotide pairings at each position in the target site (compared to the highly conserved miR390 sequence in the middle) are indicated by different colors, with A:U/C:G matches denoted in green, G:U matches in purple, and all mismatches in pink. The red arrow marks the 10th position, relative to the 5' end of miR390. The yellow shading indicates regions of substantially imperfect pairing. The upper graph shows the 5' target site of *TAS3*, the lower graph shows the 3' target site; the number of sequences analyzed is indicated for each panel. (B) Variation in the pairing of the two miR390 target sites in *TAS3* genes in different species or lineages of land plants. The images are as described for panel A, but the left graph shows the analysis of the 5' target sites of *TAS3*, and the right graph shows the analysis of the 3' target sites.

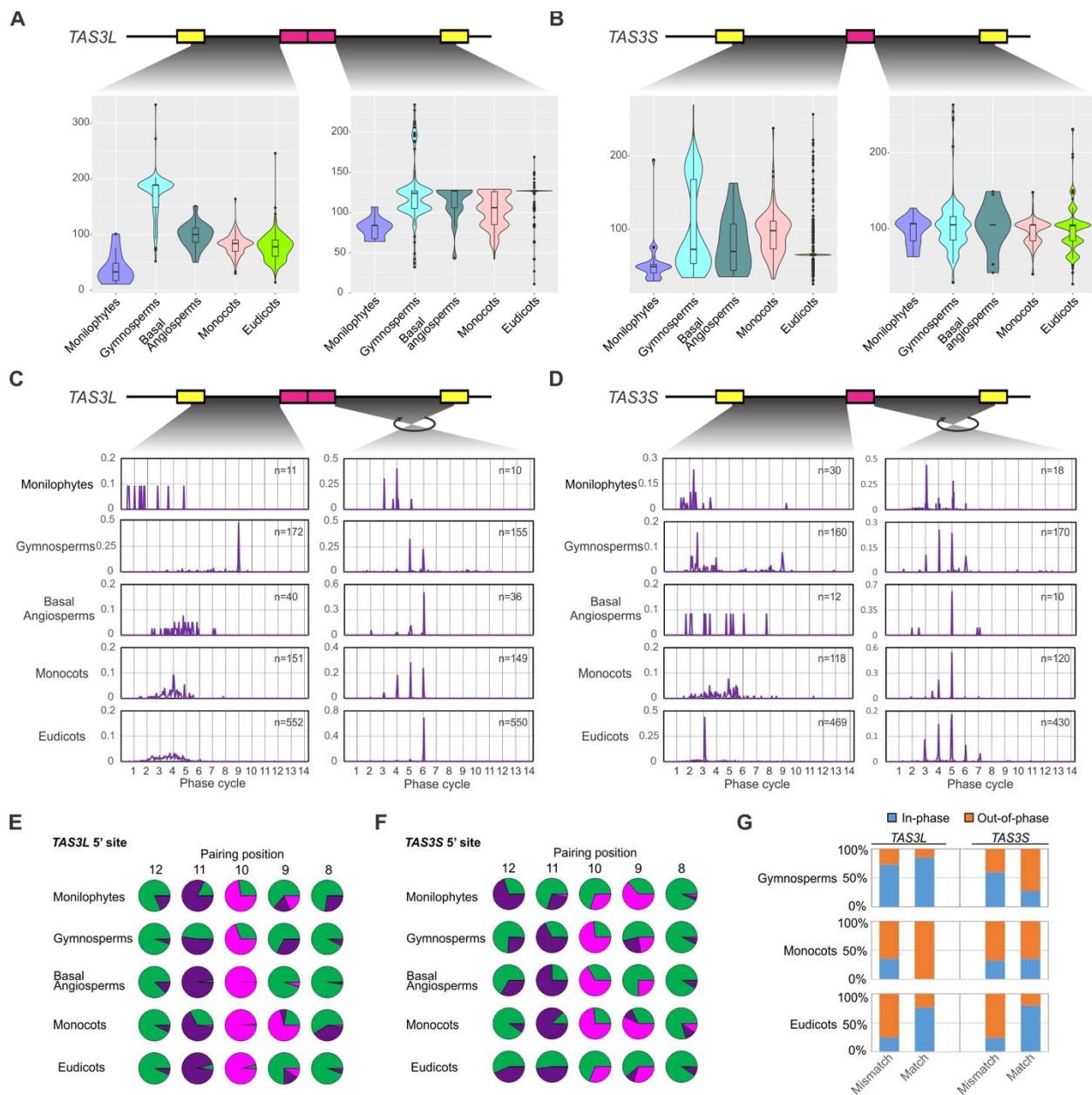


Figure 5. The distances between the two target sites of miR390 and the central tasiARF are under strong selection.

Panels (A) and (B) display the variation of the distances between two miR390 target sites and tasiARF of *TAS3L* (panel A) or *TAS3S* (panel B) genes in different lineages of vascular plants. In both panels, the lower graphs contain violin plots for each lineage representing the distribution of these distances; internal boxes represent the median as a heavy line surrounded by a box defining the upper and lower quartiles. Panels (C) and (D) display the distribution of the distances between two miR390 target sites and tasiARF of *TAS3L* (C) or *TAS3S* (D) genes in different lineages of vascular plants. The Y-axis is the percentage of *TAS3* genes with distances occurring within a given position (the X-axis). The 21-nt phased positions (phase “cycles”) are marked as grey gridlines. Panels (E) and (F) display the variation in pairing of the 8th to 12th

nucleotide positions (relative to the 5' end of miR390) of the 5' target site of *TAS3L* (E) and *TAS3S* (F). The type of miR390-TAS3 pairing observed at different nucleotide positions, relative to the 5' end of miR390, with A:U/C:G matches denoted in green, G:U matches in purple, and all mismatches in pink. (G) Ratio of the 5' miR390 target sites in phase or out-of-phase to the tasiARF in terms of different nucleotide pairing at the 10th position (match: U; mismatches: A, C, G as the 10th nucleotide of miR390 is "A").

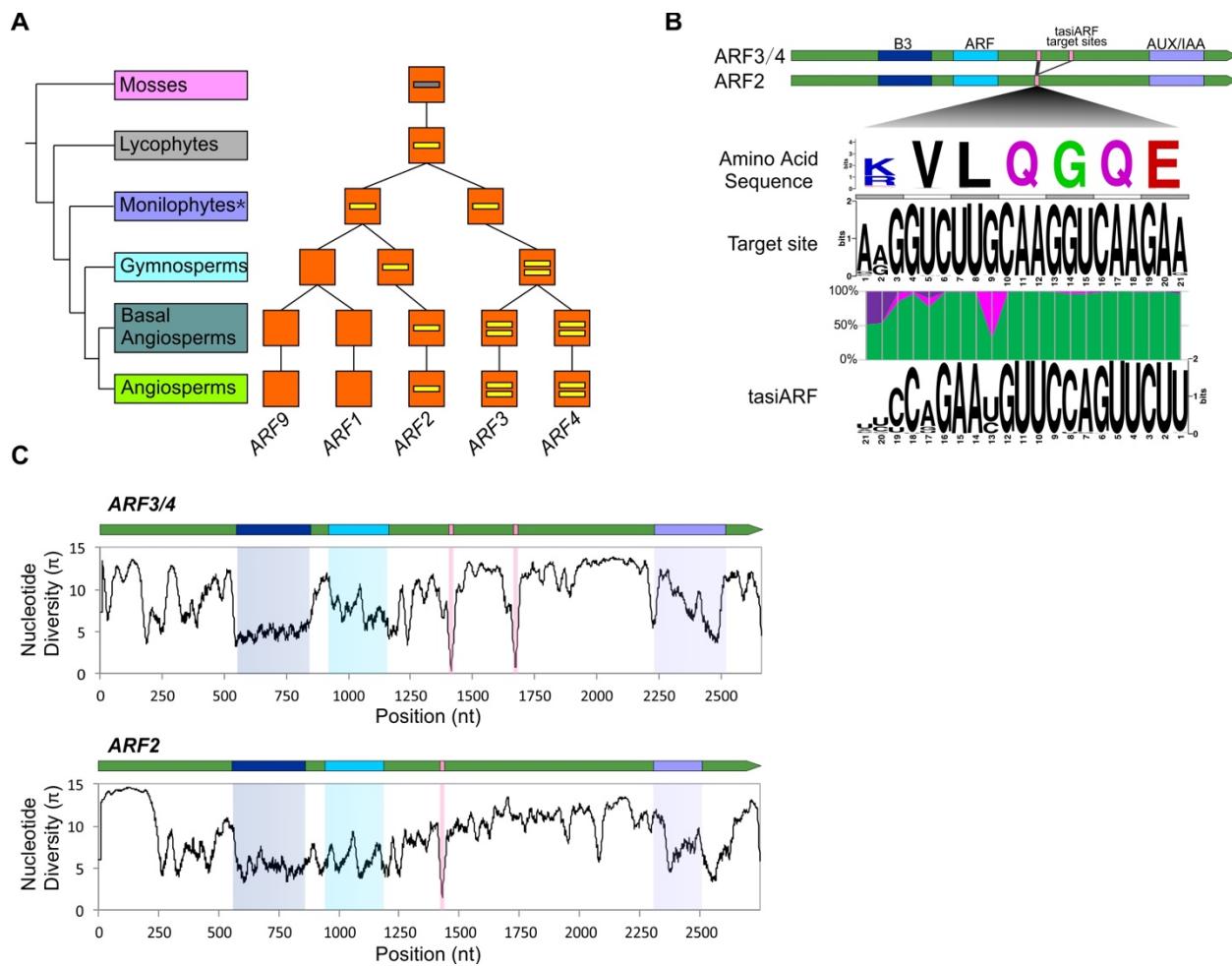


Figure 6. Evolutionary diversification of tasiARF target sites in ARF genes.

(A) Evolution of the number of tasiARF target sites in plant ARF genes. The evolutionary route of ARF genes was adapted from Finet *et al.* (2012). The number of short yellow lines in orange boxes denote the number of tasiARF target sites. The grey line means that there are potential tasiARF target sites in ARF genes in mosses. In monilophytes (marked with a “*”), some ARF3/4 homologous genes have already evolved two tasiARF target sites. (B) Sequence features of the target site of tasiARF in ARF genes and their encoded proteins. Gene structures of tasiARF-targeted ARF2/3/4 are displayed on the top, including the encoded protein motifs, with the tasiARF target site indicated as pink bars. The target site encodes a short peptide with a consensus sequence of K/RVLQGQE, as indicated with the encoding sequence. Pairing between tasiARF and its target site is color-coded with A:U/C:G matches denoted in green, G:U matches in purple, and all mismatches in pink. (C) Distribution of nucleotide diversity along tasiARF-targeted ARF2/3/4 genes, with the encoded functional domains and tasiARF target site marked in colors according to those in panel B.

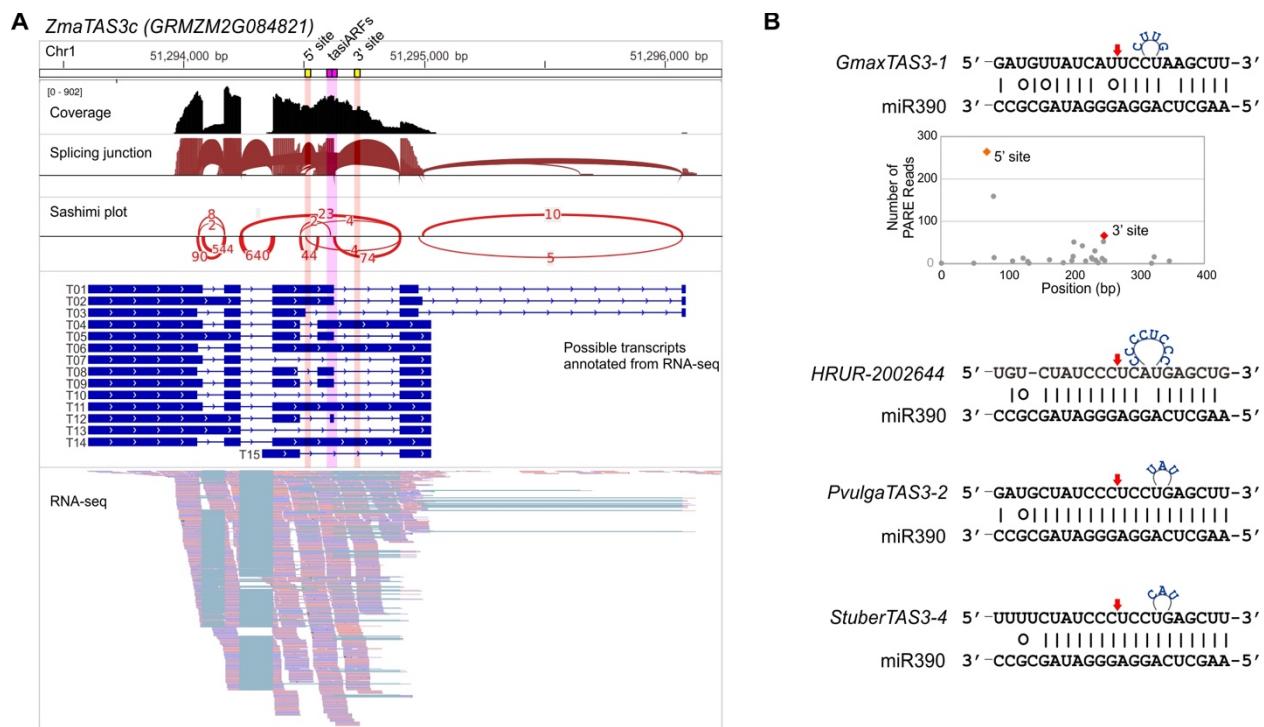


Figure 7. New regulatory features found for the miR390-TAS3-ARF pathway.

(A) Alternative splicing affects the structure of *TAS3* transcripts in maize. tasiARF (pink) and the two target sites of miR390 (yellow) are marked at the top, and they are alternatively present/absent in different splicing variants (T1...T15). The “Sashimi plot” displays the frequency with which different splice ends were joined in the RNA-seq data (at bottom). (B) miR390-interactions predict a large bulge in the so-called “seed” region of the miRNA:target interaction, for the 3’ target site of several *TAS3* genes; the example at the top is from soybean. Soybean PARE data were consistent with the successful cleavage of this 3’ target site of miR390 in *GmaxTAS3-1*. Three other cases from different species are displayed below this, from *Utricularia* sp. (*HRUR-2002644*), common bean (*Phaseolus vulgaris*, *PvulgaTAS3-2*), and potato (*Solanum tuberosum*, *StuberTAS3-4*).