



## NRC Publications Archive Archives des publications du CNRC

### **Long noncoding miRNA gene represses wheat $\beta$ -diketone waxes**

Huang, D.; Feurtado, J. A.; Smith, M. A.; Flatman, L. K.; Koh, C.; Cutler, A. J.

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

For the publisher's version, please access the DOI link below. / Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

#### **Publisher's version / Version de l'éditeur:**

<https://doi.org/10.1073/pnas.1617483114>

*Proceedings of the National Academy of Sciences of the United States of America*, 114, 15, pp. E3149-E3158, 2017-03-28

#### **NRC Publications Record / Notice d'Archives des publications de CNRC:**

<https://nrc-publications.canada.ca/eng/view/object/?id=8bda481d-f701-4832-9e15-17b470f63242>

<https://publications-cnrc.canada.ca/fra/voir/objet/?id=8bda481d-f701-4832-9e15-17b470f63242>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at

<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site

<https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

#### **Questions?** Contact the NRC Publications Archive team at

PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

**Vous avez des questions?** Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.



# Long noncoding miRNA gene represses wheat $\beta$ -diketone waxes

Daiqing Huang<sup>a,1</sup>, J. Allan Feurtado<sup>a,1</sup>, Mark A. Smith<sup>a</sup>, Leah K. Flatman<sup>a</sup>, Chushin Koh<sup>a,2</sup>, and Adrian J. Cutler<sup>a,3</sup>

<sup>a</sup>Wheat Improvement Flagship Program, National Research Council of Canada, Saskatoon, Saskatchewan, SK S7N 0W9, Canada

Edited by David C. Baulcombe, University of Cambridge, Cambridge, United Kingdom, and approved March 7, 2017 (received for review November 9, 2016)

The cuticle of terrestrial plants functions as a protective barrier against many biotic and abiotic stresses. In wheat and other Triticeae,  $\beta$ -diketone waxes are major components of the epicuticular layer leading to the bluish-white glaucous trait in reproductive-age plants. Glaucousness in durum wheat is controlled by a metabolic gene cluster at the *WAX1* (*W1*) locus and a dominant suppressor *INHIBITOR OF WAX1* (*Iw1*) on chromosome 2B. The wheat D subgenome from progenitor *Aegilops tauschii* contains *W2* and *Iw2* paralogs on chromosome 2D. Here we identify the *Iw1* gene from durum wheat and demonstrate the unique regulatory mechanism by which *Iw1* acts to suppress a carboxylesterase-like protein gene, *W1-COE*, within the *W1* multigene locus. *Iw1* is a long noncoding RNA (lncRNA) containing an inverted repeat (IR) with >80% identity to *W1-COE*. The *Iw1* transcript forms a miRNA precursor-like long hairpin producing a 21-nt predominant miRNA, miRW1, and smaller numbers of related sRNAs associated with the nonglaucous phenotype. When *Iw1* was introduced into glaucous bread wheat, miRW1 accumulated, *W1-COE* and its paralog *W2-COE* were down-regulated, and the phenotype was nonglaucous and  $\beta$ -diketone-depleted. The IR region of *Iw1* has >94% identity to an IR region on chromosome 2 in *Ae. tauschii* that also produces miRW1 and lies within the marker-based location of *Iw2*. We propose the *Iw* loci arose from an inverted duplication of *W1-COE* and/or *W2-COE* in ancestral wheat to form evolutionarily young miRNA genes that act to repress the glaucous trait.

glaucous | inhibitor of wax | small RNA | long noncoding RNA | *WAX1*

**P**lant epicuticular waxes deposited on the outer surface of the plant cuticle produce a water-resistant layer that serves to reduce nonstomatal water loss and mitigate the effects of heat and UV radiation as well as pathogen and insect attacks (1). Grasses in the Triticeae tribe, subfamily Pooideae, which include the cultivated species barley (*Hordeum vulgare*;  $2n = 2x = 14$ ), rye (*Secale cereale*,  $2n = 2x = 14$ ), durum wheat (*Triticum durum*;  $2n = 4x = 28$ , AABB), and bread wheat (*Triticum aestivum*;  $2n = 6x = 42$ , AABBDD), have two predominant pathways for wax production: (i) an alcohol- and alkane-rich wax pathway and (ii) a pathway leading to  $\beta$ -diketones and derivatives including hydroxy- $\beta$ -diketones (2). The alcohol and alkane waxes are prevalent in earlier development and on leaves, whereas  $\beta$ -diketones dominate during the reproductive phase, particularly on leaf sheaths and flower heads (3, 4).  $\beta$ -Diketone wax is predominantly hentriacontane-14, 16-dione, which consists of a 31-carbon chain with carbonyl groups at C<sub>14</sub> and C<sub>16</sub>. In durum wheat, about 20% of the  $\beta$ -diketone is hydroxylated to form 25-hydroxy- $\beta$ -diketone, whereas in bread wheat hydroxylation is at C<sub>8</sub> or C<sub>9</sub> (5).  $\beta$ -Diketone wax deposition manifests visibly as glaucousness, a bluish-white coloration on stems, leaves, and flower heads. However, the relationship between a glaucous appearance and the total amount of cuticular wax can be inconsistent, especially during the later stages of wheat reproductive development (6, 7). Nonetheless,  $\beta$ -diketones are essential for the appearance of glaucousness and associated wax morphology (3).  $\beta$ -Diketone wax deposition and the development of glaucousness lead to a greater reflectance of incident light. Reduced light absorption can lower tissue temperatures, thereby reducing transpirational water loss, and also may reduce

photosynthesis under nonsaturating illumination (1). In wheat, a glaucous appearance has been shown to be associated with stabilizing grain yield, particularly in growth environments that are water-limited and prone to heat stress (6, 8, 9). Thus, because of the protective nature of the waxy  $\beta$ -diketone layer, the glaucous appearance has generally been selected for during the breeding of cultivated durum and bread wheat varieties (3, 10). In contrast, a nonglaucous (NG) state is prevalent in the uncultivated relatives of wheat, including progenitor species wild emmer (*Triticum dicoccoides*,  $2n = 4x = 28$ , AABB) and *Aegilops tauschii* ( $2n = 2x = 14$ , DD) (11, 12). As such, these species have been used in the development of NG wheat varieties and for studies on the characterization of the genes and genetic loci involved in wax deposition, in particular *W1/W2* and *Iw1/Iw2* (13, 14).

The complex evolution of durum and bread wheat as multilevel genome mosaics means each of the three subgenomes in wheat (A, B, and D) has the potential to contribute to the inheritance of glaucousness (15). However, only the B and D subgenomes contain major glaucousness loci, and the A genome progenitor *Triticum urartu* does not contain any appreciable  $\beta$ -diketones and is NG (5, 16). Within the B subgenome, *WAX1* (*W1*) and *INHIBITOR OF WAX1* (*Iw1*) have been mapped very close to each other on the distal end of 2BS. Conversely, in the D subgenome, *W2* and *Iw2* have been mapped far apart, with *W2* at the proximal end and *Iw2* at the distal end of 2DS (16). As the name suggests, the *Iw* loci

## Significance

Higher plants have waxy surface layers that prevent uncontrolled water loss. Many wheat cultivars accumulate diketone epicuticular waxes in reproductive-age plants that produce a glaucous appearance. We identify *INHIBITOR OF WAX1* (*Iw1*), a dominant glaucous repressor, as a young miRNA gene (MIRNA) that produces an miRNA, miRW1, which targets the transcript of the biosynthetic gene *WAX1-CARBOXYLESTERASE* (*W1-COE*) for degradation. The high sequence similarity between the *Iw1* hairpin sequence and *W1-COE* suggests that this MIRNA gene arose from an inverted duplication of its target. The cleavage specificity of miRW1 for its target gene defines the unique role of a young MIRNA gene in the regulation of an important agricultural trait related to stress tolerance.

Author contributions: D.H., J.A.F., and A.J.C. designed research; D.H., J.A.F., M.A.S., and L.K.F. performed research; D.H., J.A.F., M.A.S., and C.K. analyzed data; and D.H., J.A.F., and A.J.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The small RNA and RNA-sequencing data have been submitted to the Sequence Read Archive (SRA) at the National Center for Biotechnology Information (NCBI) [accession nos. SAMN05725181–SAMN05725246 (mRNAs and sRNA in *Triticum durum* and *Triticum aestivum*)]. The *Iw1* full-length cDNA sequence was submitted to the NCBI expressed sequence tags database (dbEST) (accession no. KX823910).

<sup>1</sup>D.H. and J.A.F. contributed equally to this work.

<sup>2</sup>Present address: Global Institute for Food Security, University of Saskatchewan, Saskatoon, SK S7N 4J8, Canada.

<sup>3</sup>To whom correspondence should be addressed. Email: adrian.cutler@nrc-cnrc.gc.ca.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1617483114/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1617483114/-DCSupplemental).

provide epistatic dominant inhibition over the *W* loci. In recent years, several reports have furthered the characterization of these loci, including fine-mapping of *Iw1* (11), demonstrating *Iw1* suppression of  $\beta$ -diketone wax accumulation (17), fine mapping of *Iw2* in *Ae. tauschii* (12), comparative mapping of *Iw1* and *Iw2* in hexaploid wheat (18), determining the impact of *W* and *Iw* loci on glaucousness and cuticle permeability (3), and fine mapping of *W1* in hexaploid wheat (19). The synthesis and chemistry of diketone waxes has been studied extensively in barley, chiefly through the characterization of loss-of-function mutants of three tightly linked loci collectively referred to as “*Cer-cqu*” (2, 20, 21). The *Cer-cqu* operon, as it has been described, is associated with the *cer-c*, *-q*, and *-u* complementation groups and corresponding mutants *glossy sheath 6* (*gsh6*), *gsh1*, and *gsh8*, respectively, and have been mapped close to the terminus of the short arm of chromosome 2H (22). A recent study identified the CER gene cluster including *GSH1* (*Cer-Q*) as encoding a lipase/carboxylesterase, *GSH6* (*Cer-C*), as a chalcone synthase-like polyketide synthase and *GSH8* (*Cer-U*) as a cytochrome P450-type hydroxylase (23). More recently, the wheat *W1* locus was identified as a gene cluster that is collinear to the barley CER gene cluster (24) and includes orthologs of *Cer-Q*, *Cer-C*, and *Cer-U* (23), which we define as *W1-COE* (carboxylesterase), *W1-PKS* (polyketide synthase), and *W1-CYP* (cytochrome P450 hydroxylase), respectively. However, it is not known which of the genes at the *W1* locus are regulated by *Iw1* and therefore are responsible for the presence or absence of diketone waxes and the glaucous phenotype. More importantly, the inhibitor genes *Iw1* and *Iw2* have not been identified, and their mechanism of action is unknown.

Long noncoding RNAs (lncRNAs) are a large and diverse class of RNA transcripts with a length of more than 200 nt that do not encode proteins. lncRNAs are emerging as important regulators in a wide range of essential biological processes. In humans, lncRNAs represent more than 68% of the transcriptome, and 79% of the lncRNAs were previously unannotated (25). Our current knowledge of their biological functions is limited, and lncRNA research in plants lags behind lncRNA research in animals. To date, very few lncRNAs have been characterized in detail (26). Some lncRNAs can be precursors of small RNAs (sRNAs) including miRNAs, which are a class of sRNA ranging from 20–24 nt in length that regulate numerous pathways and biological processes (27). miRNAs play significant roles in posttranscriptional gene regulation through base pairing with specific target sequences in their complementary mRNA targets, leading to transcript degradation (28, 29). The mechanism of action of miRNAs implies that they typically act as genetically dominant-negative regulators. Therefore, because the *Iw* loci are dominant repressors of the glaucous trait, we investigated the possible involvement of sRNAs in the regulation of wax accumulation. Using near-isogenic lines (NILs) from durum wheat that differed in glaucousness (30, 31), we compared the sRNAs in each of the isogenic pairs and identified a set of related miRNAs associated with repression of  $\beta$ -diketone deposition. We show that the *Iw1* locus is a miRNA gene (MIRNA) that encodes a miRNA precursor and represses  $\beta$ -diketone deposition via miRNA-mediated cleavage of *W1-COE* transcripts.

## Results

To investigate the genetic basis for glaucousness in wheat, we characterized four pairs of NILs of durum wheat, AG1, AG2, AE3, and D051, defined by the presence or absence of the glaucous trait (30, 31). These lines were produced by back-crossing a NG cultivar to a glaucous parent and then maintaining heterozygosity for glaucousness in the F4 and later generations (*SI Appendix, Table S1A*) (31). Glaucous lines showed bluish-white coloration from the booting stage in the stems, leaves, and floral tissues, whereas NG lines were green and glossy (Fig. 1A). The cuticular wax content and composition from leaf sheaths of all four NIL

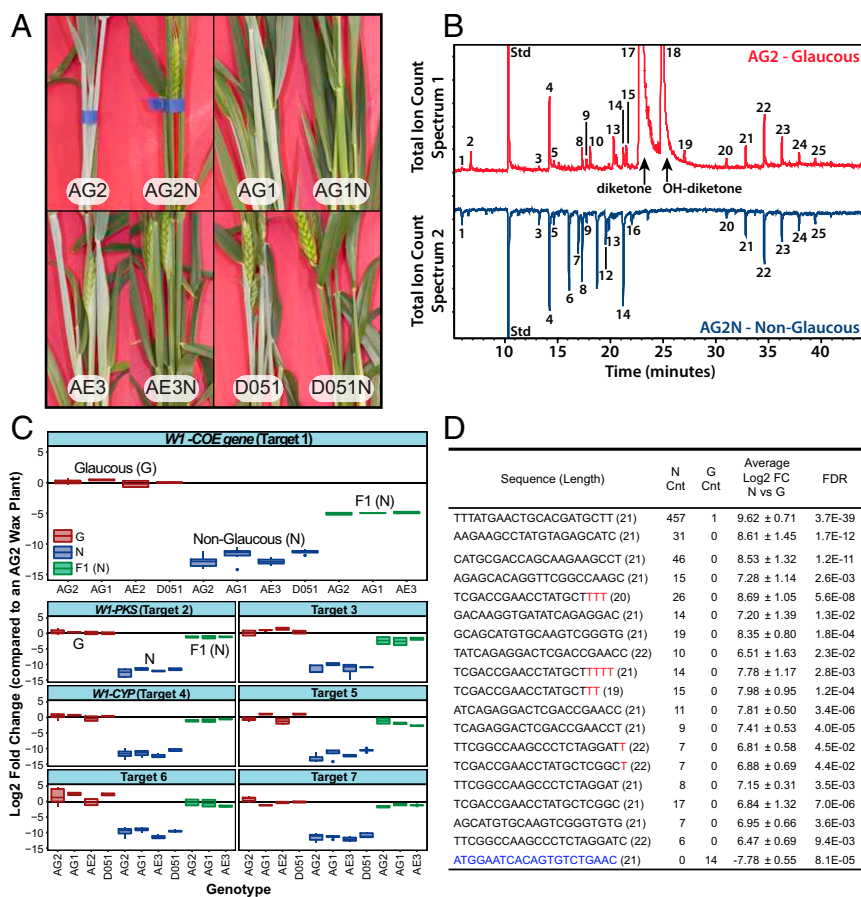
pairs was analyzed by GC-MS. Glaucous lines contained primarily  $\beta$ -diketone (hentriacontane-14, 16-dione) and 25-hydroxy- $\beta$ -diketone (25-hydroxyhentriacontane-14, 16-dione) (5, 32), whereas the NG lines contained no detectable diketone waxes (Fig. 1B and *SI Appendix, Table S1*). Importantly, crosses between three pairs of NILs, AG1, AG2, and AE3 (glaucous  $\times$  NG), resulted in F1 plants that were 100% NG (*SI Appendix, Fig. S1*), confirming the dominance of the NG trait, as observed previously (30, 31).

## Transcripts Associated with Wax Production in the Durum NILs.

Transcripts in NIL pairs were compared to identify differences that were consistently associated with the loss of diketone wax production and glaucousness. Potential wax-related genes that were strongly down-regulated in NG lines were first identified by mapping reads to the National Center for Biotechnology Information (NCBI) unigene set and determining significant differences based on count data (edgeR,  $P \leq 0.05$ ). We found 16 unigenes that were commonly down-regulated in all four glaucous/NG NIL comparisons (*Dataset S1*). Consistent with previous mapping studies locating *W1* and *Iw1* on 2BS (16), blasting the 16 unigenes against the International Wheat Genome Sequencing Consortium (IWGSC) wheat survey sequences from AABB genomes revealed that most of the contigs were located on 2BS scaffolds (*Dataset S1*). Through further bioinformatics analyses, we defined these 16 unigenes into seven potential target genes (*Dataset S1*). Three of these target genes, targets 1, 2, and 4, were from the *W1* locus gene cluster: *W1-COE*, *W1-PKS*, and *W1-CYP*. To confirm the significance of the differentially expressed genes from the NCBI unigene reference, we also used the IWGSC transcript set (v1 from EnsemblPlants) as a reference for analysis with the addition of the unannotated *W1-COE* sequence. Using this reference, the RNA sequencing (RNA-seq) data were reanalyzed using the pseudoalignment program kallisto and the Bioconductor package DESeq2 (adjusted  $P$  value  $\leq 0.05$ ) (33, 34). Similar differentially expressed target genes were obtained and included the *W1* gene cluster: *W1-COE*, *W1-CYP* (*Traes\_2BS\_163390FC4*), target 6 (*Traes\_2BS\_D6F1011EA*), and *W1-PKS* (*Traes\_2BS\_9E10D26DB*) on 2BS and transcripts with high homology to *W1-PKS*: *Traes\_3B\_FC275A64D*, *Traes\_4BS\_AB8E1AD32*, and *Traes\_6BS\_C400F1983* (*Dataset S2*). With respect to expression level, all seven of the potential target genes showed virtually no expression in NG NILs with an average down-regulation of more than 2,800-fold (Fig. 1C). In F1 heterozygous lines, *W1-COE* expression was still down-regulated by 31.3-fold on average, but the expression of all other targets, including *W1-PKS* and *W1-CYP*, recovered to a large extent, showing down-regulation between 1.7- and 4.7-fold (Fig. 1C), suggesting that *W1-COE* is most likely the gene controlling the glaucous trait in these durum NILs.

To establish whether *W1-COE* is indeed involved in diketone wax production, we used virus-induced gene silencing (VIGS) to block its expression transiently in wheat. Fragments of *W1-COE* were integrated into a VIGS system using the barley stripe mosaic virus (35) and were applied to the leaves of glaucous AG2 plants at the tillering stage, before visible glaucousness was apparent. Then the development of glaucousness was monitored for 4–6 wk. *W1-COE* fragments all produced large reductions in visible glaucousness relative to waxy controls (*SI Appendix, Fig. S2*) and in total diketone wax accumulation in leaf sheaths (*SI Appendix, Figs. S3A and S4*). Control infections with a *PHYTOENE DESATURASE* (*PDS*) fragment produced a slight reduction in wax content which may be attributed both to the general effects of viral infection and to the reduction in pigment accumulation resulting from the inhibition of *PDS*. The levels of *W1-COE* expression in VIGS-treated plants were measured by quantitative PCR (qPCR), which showed that all four of the tested fragments reduced the expression of the gene (*SI Appendix, Fig. S3B*). In addition, there was a linear correlation between the amount of  $\beta$ -diketone wax and the expression of *W1-COE* (*SI Appendix, Fig. S3C*). The VIGS





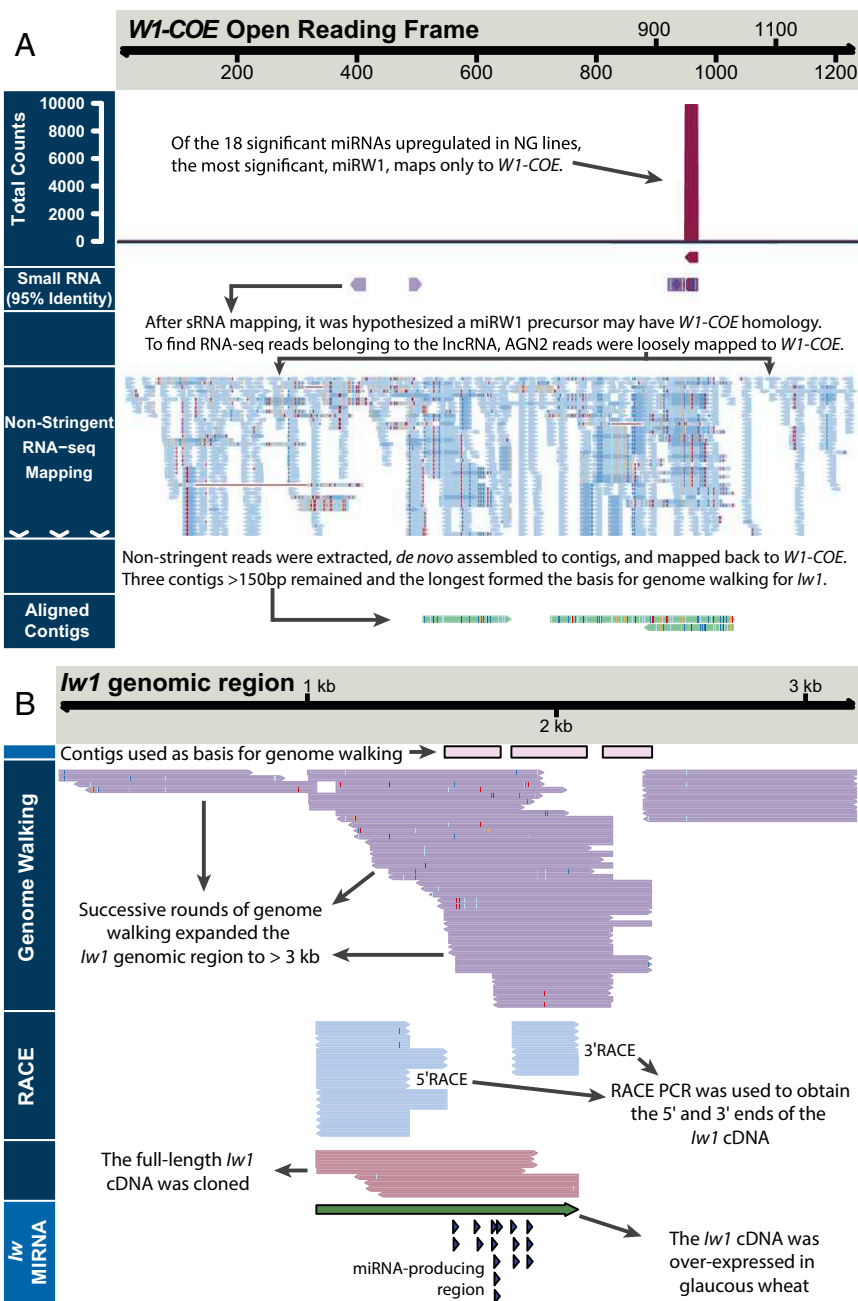
**Fig. 1.** Characterization of durum wheat NILs differing in the presence or absence of the glaucous trait. (A) Phenotypes of the four NIL pairs, AG2/AG2N, AG1/AG1N, AE3/AE3N, and D051/D051N. Each pair consists of a glaucous line and a corresponding NG (N) isoline. (B) GC-MS analysis of the surface waxes confirms that the glaucous appearance is caused by the presence of  $\beta$ -diketones. (C) qPCR analysis of seven target genes present on chromosome 2B5 including the *W1* cluster genes *W1-COE*, *W1-PKS*, and *W1-CYP*. Target genes were identified through differential expression analysis using RNA-seq (Dataset S1). (D) Differentially expressed sRNAs common to all four NIL comparisons described in A. Black sequence color indicates a perfect match to the putative *W1* sequence; red indicates probable sRNA U/A tailing modifications; blue-colored sRNA up-regulated in glaucous lines does not map to *W1*. Total read counts (Cnt) were normalized to 10 million total reads; FDR represents the highest significance of an isogenic pair; the significance threshold was at an adjusted *P* value (FDR)  $\leq 0.05$ .

result further confirmed the hypothesis that *W1-COE* has a primary role in regulating diketone wax production in the durum wheat NILs and is in agreement with recent reports identifying the barley *Cer-cqu* and wheat *W1* gene clusters (23, 24).

**sRNAs Associated with the NG State Show Targeting Specificity for *W1-COE*.** Differential expression analysis of sRNAs of 19–28 nt in length (edgeR adjusted *P* value  $\leq 0.05$ ) revealed a series of 19- to 22-nt sequences in NG lines that were almost completely absent in glaucous lines (Fig. 1D and Dataset S3A). These sRNA sequences up-regulated in NG lines could not be perfectly mapped to the IWGSC wheat genome survey sequences. However, the most abundant sRNA, 21 nt in length with 9,403 total reads and 457 average reads per 10 million (Fig. 1D), mapped to *W1-COE* with one mismatch (Dataset S3A and B). From the sRNA reads, five other 19- to 22-nt sequences also mapped to *W1-COE* with one mismatch or less (Dataset S3B). Because the most abundant sRNA was complementary to a specific sequence in *W1-COE*, we designated the sRNA sequence as “microRNA specific to *W1-COE*” (miRW1) (Fig. 2). As mentioned, expression of miRW1 and other related sRNAs was almost absent in glaucous lines but was present in NG lines, including the F1 heterozygous progeny of crosses between the glaucous and NG NILs (Fig. 3A and Dataset S3A and B). Because miRW1 had no sequence homolog other than *W1-*

*COE* sequences in the wheat NCBI unigene set or coding sequences within the IWGSC survey sequence (Dataset S4A), we considered the possibility that miRW1 was derived from *W1-COE*. However, RNA structure-prediction software indicated that the *W1-COE* transcript could not fold to form a hairpin loop structure characteristic of miRNA precursors. Furthermore, the concept of *W1-COE* as a miRNA precursor is inconsistent with the inverse correlation between *W1-COE* and miRW1 expression (i.e., *W1-COE* being expressed in glaucous lines and miRW1 being preferentially expressed in NG lines). Therefore, the more likely explanation is that miRW1 is produced from an unknown precursor gene and targets *W1-COE* for suppression.

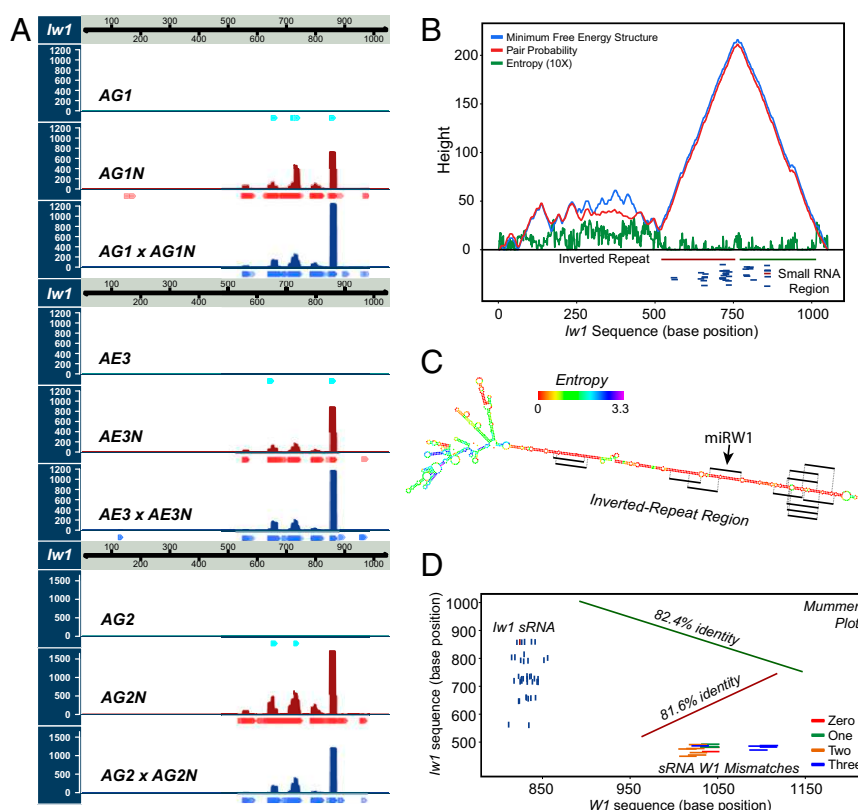
**The miRW1 Precursor Contains a Hairpin-Forming Inverted Repeat with Homology to *W1-COE*.** We hypothesized that the putative miRW1 precursor could have weak homology to *W1-COE* based on evidence from the literature indicating that miRNAs and their targets can have sequence similarities that extend beyond the sequence of the miRNA itself (36, 37). Because 8 of the 18 differentially expressed miRNAs had homology to *W1-COE* (with three mismatches or fewer) (Dataset S3B), we surmised that the NG lines may contain RNA-seq reads from a precursor sequence with homology to *W1-COE*. Thus, all RNA-seq reads from NG lines, in which the putative miRNA precursor, but not *W1-COE*, would be



**Fig. 2.** Identification and cloning of *lw1* from NG durum wheat NILs. (A) Schematic representation of the *lw1* cloning strategy. The premise of the experiment was to identify the *lw1* sequence based on its suspected loose homology to the suspected regulatory target *W1-COE*. (B) After identification of potential *lw1* contigs, genome walking was performed to isolate a 3-kb genomic DNA fragment from NG NILs. Using 5' and 3' RACE PCR, a 1-kb cDNA was obtained that included a region where differential sRNA mapped (Fig. 1D and Dataset S3). Nucleotide polymorphisms within sequencing reads and contigs are represented by colored bars.

expressed were pooled and aligned against *WI-COE* with low stringency (requiring a contiguous aligned read length of >20% to *WI-COE* with >80% homology). The mapped sequences were extracted and collected for de novo assembly. Three contigs of greater than 150 nt were obtained (Fig. 2A). Several of the differentially expressed sRNA sequences, including the most abundant miRW1 sequence, could be perfectly mapped to two of these contigs, suggesting that one or more of the contigs was part of the miRW1 precursor (Fig. 2B). These contigs were the starting point for a series of genome-walking experiments that allowed us to obtain a putative 3,207-nt genomic sequence fragment (Fig. 2B). Because many *MIRNA* genes have a 5' cap structure and 3'

polyadenylation (38, 39), we performed 5' and 3' RACE-PCR from primers designed around the location where the sRNAs mapped within the precursor fragment obtained from genome walking (Fig. 2B and *SI Appendix, Table S4*). Through RACE, two primary miRNA sequences differing by three bases at the 5' end were obtained, the longest of which was 1,051 nt. Sequence analysis revealed that the miRW1 precursor is a lncRNA that contains an inverted repeat (IR) from nucleotides 520–756 and from nucleotides 770–1014 with high base-pairing probability (Fig. 3B). Structure prediction indicated with high confidence that the lncRNA could fold into a long hairpin loop structure with the IR forming the stem (Fig. 3C). Analysis of the similarity between the



**Fig. 3.** Characterization of the *lwl* lncRNA. (A) Expression of sRNA mapping to *lwl* in three pairs of NILs and in the F1 heterozygous generation resulting from genetic crosses between each pair of isolines. (B) *lwl* contains an IR with high base-pairing probability. The mountain plot demonstrates that the 3' end of *lwl* (~500 bp) has a high probability of forming a hairpin structure in the region of an IR. An sRNA-producing region is identified within this hairpin IR. (C) The *lwl* hairpin structure. The IR region has the highest base-pairing probability consistent with low entropy values. Significant sRNAs with perfect homology are shown. (D) A mummer plot displays the relationship of *lwl* and its potential target gene *W1-COE*. The IR region shows greater than 80% identity with portions of the *W1* gene. sRNA mapping to *lwl* (perfect match) and to *W1-COE* (up to three mismatches) are also displayed.

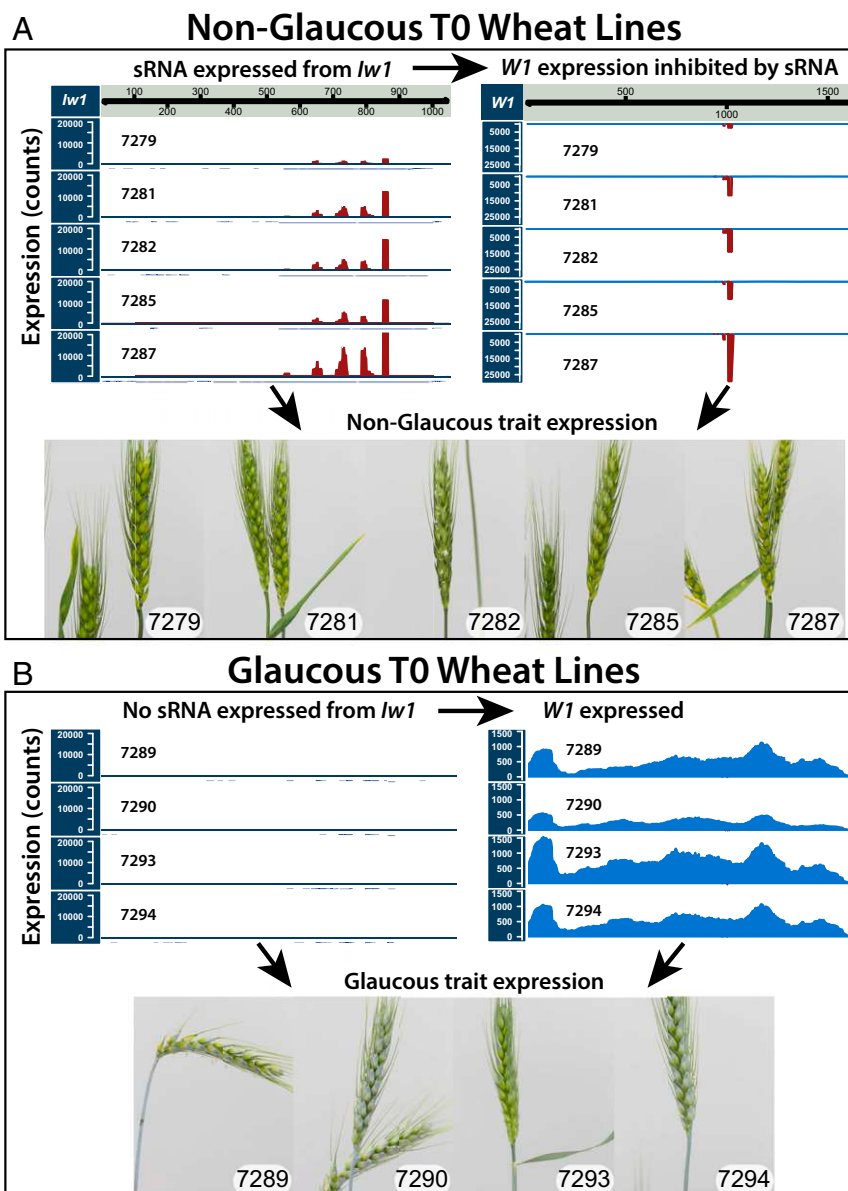
lncRNA and *W1-COE* showed that each repeat of the *MIRW1* hairpin shares ~82% identity with the *W1-COE* sequence (Fig. 3D), suggesting that the IR of the lncRNA originated from an inverted duplication of *W1-COE*, the mechanism proposed by Allen et al. (40). Of the 18 sRNAs significantly up-regulated in NG lines, 13, including the miRW1 sequence, mapped perfectly to the foldback region of the lncRNA, and all the sRNAs up-regulated in NG lines can be mapped to the lncRNA if sRNA tailing is considered (Fig. 3C and Dataset S4B). Tailing involves the nontemplated addition of bases to the 3' end of sRNAs through adenylation or uridylation (41–43). Taking these findings together, we conclude that the miRW1 precursor forms a long hairpin structure that is processed to produce miRW1 and other sRNAs that specifically target *W1-COE*.

**Expression of the miRW1 Precursor in Glaucous Wheat Creates an NG Phenotype Through Repression of *W1-COE* and *W2-COE*.** Introduction of the 1,051-nt miRW1 precursor driven by the maize ubiquitin1 promoter into the bread wheat cultivar Bobwhite resulted in an obvious NG appearance in 20 of 29 T0 plants (Fig. 4 and Dataset S5A). Analysis of diketone waxes in T0 transgenic lines also revealed that the NG trait was the result of the absence of  $\beta$ -diketones (SI Appendix, Fig. S5). The NG phenotype and the absence of diketone waxes was heritable and carried over to the T1 generation (SI Appendix, Figs. S6 and S7).

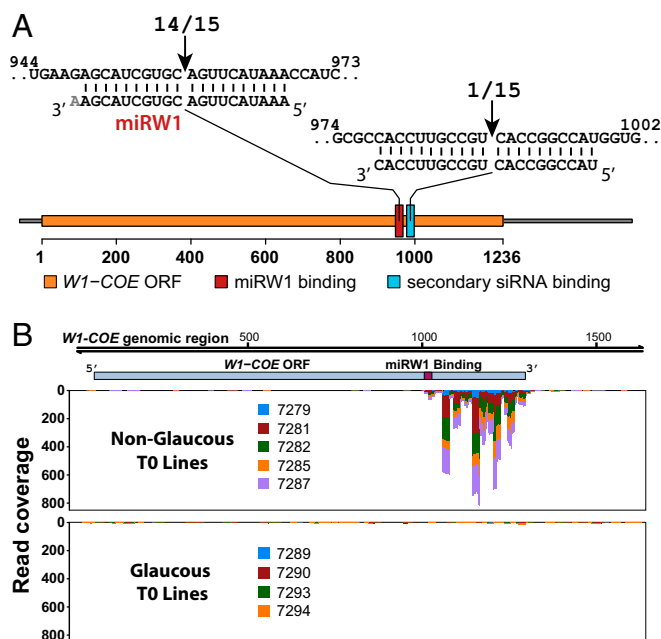
RNA-seq analyses of five NG and four glaucous T0 plants were carried out to determine both differentially expressed genes and sRNAs. Similar to the analysis of the wax NILs, we used two approaches: mapping to the NCBI unigene set and pseudoalignment

to the IWGSC v1 transcript set. Four unigenes in the NG T0 lines were significantly down-regulated (edgeR adjusted  $P$  value  $\leq 0.05$ ); all were fragments of *W1-COE* or its paralog on chromosome 2DS, *W2-COE* (SI Appendix, Table S2). Following pseudoalignment with kallisto, the only differentially expressed transcript was *W1-COE* (DESeq2 adjusted  $P$  value  $\leq 0.05$ ) (SI Appendix, Table S3). Of the sRNAs, 222 were differentially up-regulated in the NG T0 lines (edgeR adjusted  $P$  value  $\leq 0.05$ ) (Dataset S5B), and 208 of the 222 could be mapped to the miRW1 precursor when tailing was considered (Dataset S4C). Other significantly up-regulated sRNAs mapped perfectly to *W1-COE* and its paralog on 2DS (*W2-COE*) (Dataset S4D). The numbers of sRNAs mapping to the miRW1 precursor and to the targets *W1-COE* and *W2-COE* are consistent with the observed reductions in *W1-COE* and *W2-COE* transcript levels in NG phenotypes (Fig. 4 and SI Appendix, Fig. S8).

To validate that *W1-COE* is the target of miRNA-guided cleavage in the transgenic lines, we performed a 5' RACE assay to map possible cleavage sites within *W1-COE* (Fig. 5A). The principal position of cleavage within *W1-COE* was within the miRW1-binding site located between nucleotides 10 and 11 and was present in 14 of 15 cloned sequences. Furthermore, in NG transgenic lines with active cleavage of *W1-COE*, we detected the presence of secondary siRNA in the 3' cleavage fragment (Fig. 5B). The majority of these siRNAs were 21 nt in length and were positively related to both the abundance of miRW1 and its monouridylated form (SI Appendix, Fig. S9A and B). Secondary siRNAs also were apparent in the 3' region adjacent to the miRW1-binding site in *W2-COE* (SI Appendix, Fig. S9C). The presence of secondary siRNAs also was detected in the heterozygous







**Fig. 5.** Validation of the cleavage of the *W1-COE* mRNA target by miRW1. (A) *W1-COE* cleavage sites identified by 5' RACE. Sequences of 14 of 15 clones showed cleavage between nucleotides 10 and 11 of the miRW1-binding site. One cleavage site was located within a minor sRNA-binding site between the 10th and 11th nucleotide. (B) Mapping of secondary siRNAs to *W1-COE* in *Iw1* T0 transgenic lines. (Upper) sRNAs with perfect homology to *W1-COE* mapped to a region 3' of the miRW1-binding site in NG lines. (Lower) Almost none did so in glaucous lines.

*Iw2*, we propose that two miRW1 precursors, *Iw1* and *Iw2*, are present on chromosomes 2BS and 2DS, respectively. The IR and sRNA-producing regions of *Iw1* and *Iw2* are highly homologous (94%), and both are energetically favored to form a hairpin structure (Fig. 6C). Additional evidence of an *Iw2* hairpin-forming RNA comes from sRNA sequencing experiments downloaded at the NCBI short-read archive from Li et al. (46). Similar to *Iw1*, the *Iw2* region also produces a series of sRNAs, including the most dominant sequence miRW1 (Fig. 6D).

## Discussion

We have identified and validated the regulatory function of *Iw1* and confirmed the role of a key gene, *W1-COE*, within the *W1* locus. *Iw1* and its homolog *Iw2* are young MIRNA genes with long hairpin precursors which ultimately suppress  $\beta$ -diketone wax production. *Iw1* and *Iw2* produce miRW1, which specifically targets and represses the expression of the putative carboxylesterase genes that are necessary for the production of  $\beta$ -diketone waxes in wheat (24). Identification of the *Iw* loci represents a major step forward in our regulatory understanding of the glaucous trait in wheat and related species, from both a functional and an evolutionary standpoint.

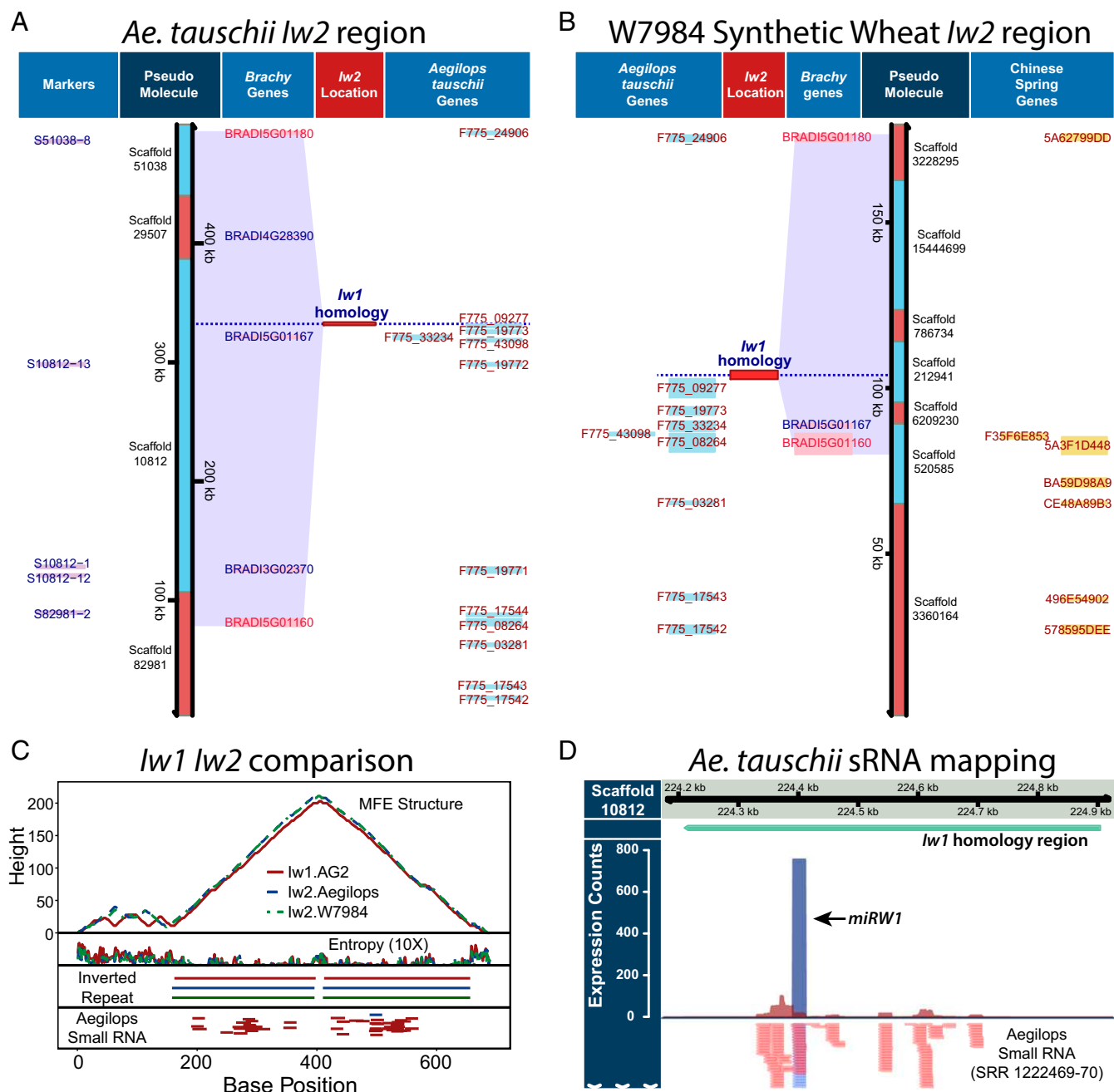
miRNAs have a significant regulatory role in plants and target a wide range of transcripts for degradation and therefore are inherently dominant-negative genetic factors (28, 41). Many evolutionarily conserved miRNA families play critical roles in plant development and adaptation to diverse environments. There also are many nonconserved, evolutionarily recent miRNAs and their corresponding targets that are present only within a few closely related species or appear to be unique to specific species (47–49). Wheat miRNA sequencing, identification, profiling, and characterization have been reported extensively (50–56). However, neither miRW1 nor its precursor has been reported, possibly because of the atypical characteristics of *Iw1* and *Iw2*.

In fact, the *Iw1* sequence does not exist in any available wheat genome reference or sequencing database such as the NCBI. The identification of *Iw1* and *Iw2* as long noncoding, hairpin-forming, sRNA-producing RNAs with IRs similar to their target sequence places them among the few functional lncRNAs described in monocots, the most notable previous example being the maize *Mu killer* locus (57). *Mu killer* arose from an inverted duplication of a sequence similar to its target, the MuDR transposon; however *Mu killer* acts via the production of siRNAs and an epigenetic mechanism (58, 59) instead of the miRNA-based silencing mechanism of the *Iw* genes. lncRNA-mediated gene regulation is emerging as a common regulatory mechanism in plants. A variety of lncRNA-mediated regulation mechanisms have been unraveled, including target mimicry, transcription interference, PRC2-associated histone methylation, and DNA methylation (26). However, although the number of known plant lncRNAs is expanding, the great majority have no known function (60–65). In wheat, the lncRNA landscape has been profiled during fungal responses and heat stress, but the characterization of function is deficient (66, 67).

As described here, *Iw1* and *Iw2* serve as miRNA precursors and repress target gene expression through a miRNA-mediated mechanism. Several lines of evidence indicate that *Iw1* and *Iw2* are evolutionarily young MIRNA genes that arose by inverted duplication of their target gene (39, 49). First, the foldback region of *Iw1* has extended similarity (>80%) with the target *W1-COE* beyond that of the miRW1 region. Second, the *Iw1* primary transcript (1,051 nt) is much longer than typical miRNA primary sequences; 98% of miRNA precursor lengths are <336 nt with a mean of 146 nt (68). Third, the foldback regions of *Iw1* and *Iw2* are hairpin structures >200 nt that resemble a dsRNA and do not resemble the typical short structure of miRNA hairpins. In *Arabidopsis*, Ben Amor et al. (27) identified nine ncRNAs corresponding to miRNA, trans-acting siRNA, and 24-nt siRNA precursors, including a young MIRNA gene *MIR869A*. The transcript of *MIR869A* is processed by DCL4, because its secondary structure is closer to that of dsRNA than to that of a typical, short miRNA precursor processed by DCL1 (69). The example of *MIR869a* might indicate the evolutionary path of the *Iw* genes, with younger dsRNA-forming MIRNA genes evolving through the production of miRNA-like siRNA, because the *Iw1* and *Iw2* hairpin precursors also produce other sRNAs in addition to the predominant 21-nt miRW1. Fourth, we show evidence that miRW1, the predominant miRNA, is primarily responsible for the cleavage of the *W1-COE* transcript. Moreover, cleavage between nucleotides 10 and 11 of miRW1 is consistent with the principle hallmark of miRNA-guided degradation (40, 70, 71). The presence of secondary siRNAs mapping to the 3' cleavage fragment in the NG *Iw1* overexpression and F1 heterozygous crosses of NIL pairs provides additional evidence of miRNA-directed degradation of *W1-COE* and *W2-COE*. In plants there are two models, “one-hit” and “two-hit,” for secondary siRNA production. In the one-hit model, the trigger can be binding of 22-nt miRNAs, and in the two-hit model the trigger can be two neighboring miRNA target sites on the same mRNA (72, 73). Both models are possible triggers for the secondary siRNAs arising from the cleavage of *W1-COE* (and *W2-COE*). A 22-nt monouridylated form of miRW1 supports the one-hit model; alternatively, additional miRNAs arising from *Iw1* with *W1-COE* as a potential binding target support the two-hit model. As mentioned above, one curious aspect of the discovery of secondary siRNA arising from *W1-COE* is the absence of secondary siRNA in the homozygous NILs, and this difference leads into a discussion of whether *W1* genes are present in NG lines and cultivars.

One feature of the results leading to the identification of *W1* and *Iw1* is that in homozygous NG NILs other genes in addition to *W1-COE* are strongly down-regulated (Fig. 1C), notably *W1-PKS* and *W1-CYP* in the *W1* gene cluster on chromosome 2BS, which





**Fig. 6.** *lw2* in the D subgenome is a paralog of *lw1* from the B subgenome. (A and B) The IR, sRNA-producing sequence of *lw1* maps to the *lw2* region in *Ae. tauschii* and the W7984 synthetic hexaploid wheat (SHW). The location of the *lw1* homology region is consistent with *Ae. tauschii* S10812 markers and the *Brachypodium distachyon* (Brachy) genes BRADI5G01180 and BRADI5G01160. (C) Comparison of *lw1* and the homologous *Ae. tauschii* and W7984 SHW sequences reveals a high similarity including ~95% sequence identity between *lw1* and the D-genome sequences, the presence of IRs, and high base-pairing probabilities in the minimum free energy (MFE) structure. (D) sRNA libraries from spikes of *Ae. tauschii* revealed a similar pattern of sRNAs in the 689-bp *lw1* homology region on *Ae. tauschii* scaffold 10812 with the 21-bp sequence miRW1 predominating.

also are involved in the  $\beta$ -diketone and OH- $\beta$ -diketone synthesis pathways (24). However, *lw1* dominantly regulates glaucousness through miRW1-promoted cleavage and mRNA degradation of *W1-COE*, and overexpression of *lw1* in bread wheat down-regulated only *W1-COE* and its paralog *W2-COE* (Fig. 4 and *SI Appendix*, Tables S2 and S3). Moreover, in NG F1 heterozygous lines, which manifest *lw1* dominance, *W1-COE* was the key down-regulated gene related to diketone wax synthesis (Fig. 1C). These results show, first, that *lw1*-mediated repression of *W1-COE* causes loss of the glaucous phenotype and, second, that there is

another mechanism that down-regulates multiple genes at the *W1* locus in NG homozygotes. Interestingly, Hen-Avivi et al. (24) provide evidence that *W1-COE*, *W1-PKS*, and *W1-CYP* are missing from the *W1/lw1* genomic interval in the glossy, *lw1*-containing wild emmer accession TTD140 but are present in the *W1* metabolic gene cluster found in the glaucous cultivar Zavitan. The idea that the *W1* genes are missing or have moved to a transcriptionally inactive part of the genome in the NG genotype is interesting and is consistent with our observations of very strong down-regulation of the *W1* cluster and the lack of secondary siRNAs from *W1-COE*

in the homozygous NILS. However, the relationship between the *W1* gene cluster and *Iw* needs to be explored further by sequencing more glaucous and NG cultivars.

The presence of *Iw* in selected species within the Triticeae tribe allows us to propose an approximate evolutionary origin of *Iw*. Barley, which contains diketone wax but in which there are no reports of a dominant wax inhibitor gene, diverged from wheat 8–12 Mya, suggesting that the inverted duplication event that created *Iw* occurred after this date (74–77). The inverted duplication may have been a single event in an ancestral wheat genome lineage or separate later events resulting in convergent evolution in B (*Iw1*) and D (*Iw2*) genome species. A single inverted duplication of *W1-COE* in an ancestral B genome is plausible, based on the evidence presented by Marcussen et al. (15), who suggest that a hybridization event between A and B lineages occurred ~5.5 Mya and led to the origin of the D genome lineage. The time of *Iw* creation at <12 Mya is comparable to the creation of young MIRNA genes in the *Arabidopsis* genus at <20 Mya or in *Arabidopsis thaliana* itself at <10 Mya (49, 78–81). In contrast, more ancient conserved MIRNA genes (e.g., miR156) predate the separation of the monocots and dicots at ~150 Mya (82).

In summary, the specific and unique interaction between *Iw* and miRW1 with *W-COE* represents a mechanism for dominant gene repression and provides a basis for genome-wide identification of other nonconserved lncRNA functions or atypical MIRNA genes. Furthermore, the identification of the *Iw* genes as a major regulatory mechanism governing *W-COE* expression

and  $\beta$ -diketone deposition suggests the possibility of precise gene editing or marker-based manipulation of glaucousness for better adaptation to specific conditions and environments.

## Materials and Methods

Details of sample preparation, experimental procedures, and data analysis with associated references can be found in *SI Appendix, Materials and Methods*.

The sRNA and RNA-seq data have been submitted to the Sequence Read Archive (SRA) at the NCBI with the accession numbers SAMN05725181–SAMN05725246 (mRNAs and sRNAs in *Triticum durum* and *Triticum aestivum*). The *Iw1* full-length cDNA sequence was submitted to dbEST (NCBI). The accession number is KX823910.

**ACKNOWLEDGMENTS.** We thank Drs. Jitao Zou of the National Research Council of Canada (NRC), Andrew Sharpe (Global Institute for Food Security, University of Saskatchewan), and Weiren Wu (Fujian Agriculture and Forestry University) for helpful comments during the course of the project and in the preparation of this paper; Dr. Ron Knox (Agriculture and Agri-Food Canada) for providing the NILs used in this research; Mr. Joe Hammerlindl and Mr. Allan Kolenovsky of the NRC-Saskatoon Plant Transformation Service Facility for wheat transformation and selection of transformants; Dr. Shawn Clark and Mr. Enwu Liu for suggestions regarding VIGS experiments; Mr. Darwin Reed for optimizing GC-MS conditions for wax analyses and for locating wax standards synthesized by the late Dr. Pat Tulloch and colleagues; the NRC-Saskatoon Genomics Service Facility for DNA, RNA, and sRNA sequencing; Mr. Dustin Cram for bioinformatics assistance; and Drs. Assaf Distelfeld (Tel-Aviv University) and Cristobal Uauy (John Innes Centre) for providing access to additional genome sequence data. Funding for this project was provided by the NRC through the Canadian Wheat Alliance. This paper is NRCC No. 56262.

- Shepherd T, Wynne Griffiths D (2006) The effects of stress on plant cuticular waxes. *New Phytol* 171:469–499.
- Von Wettstein-Knowles P (2012) *Plant Waxes* (John Wiley & Sons, Ltd, Chichester, UK).
- Zhang Z, Wang W, Li W (2013) Genetic interactions underlying the biosynthesis and inhibition of  $\beta$ -diketones in wheat and their impact on glaucousness and cuticle permeability. *PLoS One* 8:e54129.
- Wang Y, et al. (2015) Developmental changes in composition and morphology of cuticular waxes on leaves and spikes of glossy and glaucous wheat (*Triticum aestivum* L.). *PLoS One* 10:e0141239.
- Tulloch AP, Baum BR, Hoffman LL (1980) A survey of epicuticular waxes among genera of Triticeae. 2. Chemistry. *Can J Bot* 58:2602–2615.
- Clarke JM, McCaig TN, DePauw RM (1993) Relationship of glaucousness and epicuticular wax quantity of wheat. *Can J Plant Sci* 73:961–967.
- Johnson DA, Richards RA, Turner NC (1983) Yield, water relations, gas exchange, and surface reflectance of near-isogenic lines differing in glaucousness. *Crop Sci* 23:318–325.
- Richards RA, Rawson HM, Johnson DA (1986) Glaucousness in wheat: Its development and effect on water-use efficiency, gas exchange and photosynthetic tissue temperature. *Aust J Plant Physiol* 13:465–473.
- Monneveux P, et al. (2004) Relationships between grain yield, flag leaf morphology, carbon isotope discrimination and ash content in irrigated wheat. *J Agron Crop Sci* 190:395–401.
- Tsunewaki K (1966) Comparative gene analysis of common wheat and its ancestral species. II. Waxiness, growth habit and awnness. *Jpn J Bot* 19:175–254.
- Xu Z, Yuan C, Wang J, Fu D, Wu J (2015) Mapping the glaucousness suppressor *Iw1* from wild emmer wheat 'PI 481521'. *Crop J* 3:37–45.
- Nishijima R, Iehisa JCM, Matsuoka Y, Takumi S (2014) The cuticular wax inhibitor locus *Iw2* in wild diploid wheat *Aegilops tauschii*: Phenotypic survey, genetic analysis, and implications for the evolution of common wheat. *BMC Plant Biol* 14:246.
- Simmonds JR, et al. (2008) Mapping of a gene (*Vir*) for a non-glaucous, viridescent phenotype in bread wheat derived from *Triticum dicoccoides*, and its association with yield variation. *Euphytica* 159:333–341.
- Liu Q, et al. (2007) Molecular mapping of a dominant non-glaucousness gene from synthetic hexaploid wheat (*Triticum aestivum* L.). *Euphytica* 155:71–78.
- Marcussen T, et al.; International Wheat Genome Sequencing Consortium (2014) Ancient hybridizations among the ancestral genomes of bread wheat. *Science* 345:1250092.
- Tsunewaki K, Ebana K (1999) Production of near-isogenic lines of common wheat for glaucousness and genetic basis of this trait clarified by their use. *Genes Genet Syst* 74:33–41.
- Adamski NM, et al. (2013) The inhibitor of wax 1 locus (*Iw1*) prevents formation of  $\beta$ - and OH- $\beta$ -diketones in wheat cuticular waxes and maps to a sub-cM interval on chromosome arm 2BS. *Plant J* 74:989–1002.
- Wu H, et al. (2013) Comparative high-resolution mapping of the wax inhibitors *Iw1* and *Iw2* in hexaploid wheat. *PLoS One* 8:e84691.
- Lu P, et al. (2015) Comparative fine mapping of the *Wax 1* (*W1*) locus in hexaploid wheat. *Theor Appl Genet* 128:1595–1603.
- von Wettstein-Knowles P, Sogaard B (1980) The *cer-cqu* region in barley: Gene cluster or multifunctional gene. *Carlsberg Res Commun* 45:125–141.
- von Wettstein-Knowles P (1995) Biosynthesis and genetics of waxes. *Waxes: Chemistry, Molecular Biology and Functions*, ed RJ Hamilton (Oily, Alloway, Ayr, Scotland), pp 91–130.
- Anonymous (1996) Description of glossy sheath 1. *Barley Genet Newsl* 26:292.
- Schneider LM, et al. (2016) The *Cer-cqu* gene cluster determines three key players in a  $\beta$ -diketone synthase polyketide pathway synthesizing aliphatics in epicuticular waxes. *J Exp Bot* 67:2715–2730.
- Hen-Avivi S, et al. (2016) A metabolic gene cluster in the wheat W1 and the barley *Cer-cqu* loci determines  $\beta$ -diketone biosynthesis and glaucousness. *Plant Cell* 28:1440–1460.
- Iyer MK, et al. (2015) The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet* 47:199–208.
- Zhang J, Mujahid H, Hou Y, Nallamilli BR, Peng Z (2013) Plant long ncRNAs: A new frontier for gene regulatory control. *Am J Plant Sci* 4:1038–1045.
- Ben Amor B, et al. (2009) Novel long non-protein coding RNAs involved in *Arabidopsis* differentiation and stress responses. *Genome Res* 19:57–69.
- Ghildiyal M, Zamore PD (2009) Small silencing RNAs: An expanding universe. *Nat Rev Genet* 10:94–108.
- Bartel DP (2009) MicroRNAs: Target recognition and regulatory functions. *Cell* 136:215–233.
- Clarke JM, McCaig TN, DePauw RM (1994) Inheritance of glaucousness and epicuticular wax in durum wheat. *Crop Sci* 34:327–330.
- Clarke JM, et al. (1995) Registration of seven pairs of durum wheat genetic stocks near-isogenic for glaucousness. *Crop Sci* 35:1241.
- Bianchi G, Figini ML (1986) Epicuticular waxes of glaucous and nonglaucous durum wheat lines. *J Agric Food Chem* 34:429–433.
- Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 34:525–527.
- Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550.
- Yuan C, et al. (2011) A high throughput barley stripe mosaic virus vector for virus induced gene silencing in monocots and dicots. *PLoS One* 6:e24648.
- Axtell MJ, Bowman JL (2008) Evolution of plant microRNAs and their targets. *Trends Plant Sci* 13:343–349.
- Nozawa M, Miura S, Nei M (2012) Origins and evolution of microRNA genes in plant species. *Genome Biol Evol* 4:230–239.
- Xie Z, et al. (2005) Expression of *Arabidopsis* MIRNA genes. *Plant Physiol* 138:2145–2154.
- Bologna NG, et al. (2013) Multiple RNA recognition patterns during microRNA biogenesis in plants. *Genome Res* 23:1675–1689.
- Allen E, et al. (2004) Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nat Genet* 36:1282–1290.
- Borges F, Martienssen RA (2015) The expanding world of small RNAs in plants. *Nat Rev Mol Cell Biol* 16:727–741.
- Chou MT, et al. (2015) Tailor: A computational framework for detecting non-templated tailing of small silencing RNAs. *Nucleic Acids Res* 43:e109.
- Zhai J, et al. (2013) Plant microRNAs display differential 3' truncation and tailing modifications that are ARGONAUTE1 dependent and conserved across species. *Plant Cell* 25:2417–2428.

44. Jia J, et al.; International Wheat Genome Sequencing Consortium (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* 496:91–95.
45. Chapman JA, et al. (2015) A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biol* 16:26.
46. Li A, et al. (2014) mRNA and small RNA transcriptomes reveal insights into dynamic homeolog regulation of allopolyploid heterosis in nascent hexaploid wheat. *Plant Cell* 26:1878–1900.
47. Cuperus JT, Fahlgren N, Carrington JC (2011) Evolution and functional diversification of MIRNA genes. *Plant Cell* 23:431–442.
48. Taylor RS, Tarver JE, Hiscock SJ, Donoghue PC (2014) Evolutionary history of plant microRNAs. *Trends Plant Sci* 19:175–182.
49. Fahlgren N, et al. (2010) MicroRNA gene evolution in *Arabidopsis lyrata* and *Arabidopsis thaliana*. *Plant Cell* 22:1074–1089.
50. Sun F, et al. (2014) Whole-genome discovery of miRNAs and their targets in wheat (*Triticum aestivum* L.). *BMC Plant Biol* 14:142.
51. Pandey R, Joshi G, Bhardwaj AR, Agarwal M, Katiyar-Agarwal S (2014) A comprehensive genome-wide study on tissue-specific and abiotic stress-specific miRNAs in *Triticum aestivum*. *PLoS One* 9:e95800.
52. Han R, et al. (2014) Identification and characterization of microRNAs in the flag leaf and developing seed of wheat (*Triticum aestivum* L.). *BMC Genomics* 15:289.
53. Budak H, Khan Z, Kantar M (2015) History and current status of wheat miRNAs using next-generation sequencing and their roles in development and stress. *Brief Funct Genomics* 14:189–98.
54. Yao Y, Sun Q (2012) Exploration of small non coding RNAs in wheat (*Triticum aestivum* L.). *Plant Mol Biol* 80:67–73.
55. Meng F, et al. (2013) Development-associated microRNAs in grains of wheat (*Triticum aestivum* L.). *BMC Plant Biol* 13:140.
56. Agharbaoui Z, et al. (2015) An integrative approach to identify hexaploid wheat miRNAome associated with development and tolerance to abiotic stress. *BMC Genomics* 16:339.
57. Slotkin RK, Freeling M, Lisch D (2003) *Mu* killer causes the heritable inactivation of the Mutator family of transposable elements in *Zea mays*. *Genetics* 165:781–797.
58. Slotkin RK, Freeling M, Lisch D (2005) Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. *Nat Genet* 37:641–644.
59. Tricker PJ (2015) Transgenerational inheritance or resetting of stress-induced epigenetic modifications: Two sides of the same coin. *Front Plant Sci* 6:699.
60. Bazin J, Bailey-Serres J (2015) Emerging roles of long non-coding RNA in root developmental plasticity and regulation of phosphate homeostasis. *Front Plant Sci* 6:400.
61. Heo JB, Lee YS, Sung S (2013) Epigenetic regulation by long noncoding RNAs in plants. *Chromosome Res* 21:685–693.
62. Wang H, et al. (2015) Analysis of non-coding transcriptome in rice and maize uncovers roles of conserved lncRNAs associated with agriculture traits. *Plant J* 84:404–416.
63. Zhang YC, et al. (2014) Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol* 15:512.
64. Li L, et al. (2014) Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol* 15:R40.
65. Wang H, et al. (2014) Genome-wide identification of long noncoding natural anti-sense transcripts and their responses to light in *Arabidopsis*. *Genome Res* 24:444–453.
66. Xin M, et al. (2011) Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing. *BMC Plant Biol* 11:61.
67. Zhang H, et al. (2016) Genome-wide identification and functional prediction of novel and fungi-responsive lincRNAs in *Triticum aestivum*. *BMC Genomics* 17:238.
68. Thakur V, et al. (2011) Characterization of statistical features for plant microRNA prediction. *BMC Genomics* 12:108.
69. Rajagopalan R, Vaucheret H, Trejo J, Bartel DP (2006) A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes Dev* 20:3407–3425.
70. Kasschau KD, et al. (2003) P1/HC-Pro, a viral suppressor of RNA silencing, interferes with *Arabidopsis* development and miRNA function. *Dev Cell* 4:205–217.
71. Huntzinger E, Izaurralde E (2011) Gene silencing by microRNAs: Contributions of translational repression and mRNA decay. *Nat Rev Genet* 12:99–110.
72. Axtell MJ, Jan C, Rajagopalan R, Bartel DP (2006) A two-hit trigger for siRNA biogenesis in plants. *Cell* 127:565–577.
73. Chen HM, et al. (2010) 22-Nucleotide RNAs trigger secondary siRNA biogenesis in plants. *Proc Natl Acad Sci USA* 107:15269–15274.
74. Huang S, et al. (2002) Phylogenetic analysis of the acetyl-CoA carboxylase and 3-phosphoglycerate kinase loci in wheat and other grasses. *Plant Mol Biol* 48:805–820.
75. Dvorak J, Akhunov ED (2005) Tempos of gene locus deletions and duplications and their relationship to recombination rate during diploid and polyploid evolution in the *Aegilops-Triticum* alliance. *Genetics* 171:323–332.
76. Chalupska D, et al. (2008) Acc homoeoloci and the evolution of wheat genomes. *Proc Natl Acad Sci USA* 105:9691–9696.
77. Middleton CP, et al. (2014) Sequencing of chloroplast genomes from wheat, barley, rye and their relatives provides a detailed insight into the evolution of the Triticeae tribe. *PLoS One* 9:e85761.
78. Koch MA, Haubold B, Mitchell-Olds T (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Mol Biol Evol* 17:1483–1498.
79. Wright SI, Lauga B, Charlesworth D (2002) Rates and patterns of molecular evolution in inbred and outbred *Arabidopsis*. *Mol Biol Evol* 19:1407–1420.
80. Ossowski S, et al. (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* 327:92–94.
81. Smith LM, et al. (2015) Rapid divergence and high diversity of miRNAs and miRNA targets in the Camelineae. *Plant J* 81:597–610.
82. Floyd SK, Bowman JL (2004) Gene regulation: Ancient microRNA target sequences in plants. *Nature* 428:485–486.



## **A novel long-noncoding MIRNA represses wheat $\beta$ -diketone waxes**

Huang et al. 2017

1. SI Materials and Methods	Pages 2-7
2. SI References	Pages 8-10
3. Supplementary Figures	Pages 11-19
4. Supplementary Tables	Pages 19-22

## SI Materials and Methods

### ***Plant materials and growth conditions***

Four pairs of glaucous (G) and non-glaucous (N) near-isogenic lines (NILs), AE3 versus AE3N, AG1 versus AG1N, AG2 versus AG2N, and D051 versus D051N, were obtained from the Swift Current Research and Development Centre of Agriculture and Agri-Food Canada (1, 2, 3). G X N crosses were made for 3 of the near-isogenic lines to produce heterozygous F1 plants for AE3, AG1 and AG2. The bread wheat cultivar Bobwhite was used in the production of transgenic plants. *Nicotiana benthamiana* were used as host plants in the VIGS procedure documented below.

Growth of wheat and *N. benthamiana* plants was carried out in 4-inch-square or 6-inch-round pots containing Sunshine mix #8 (Sun Gro Horticulture) mixed with a slow-release 14-14-14 fertilizer. Plants were maintained in a growth cabinet or sunroom with a 16 h day/8 h night, light intensity of 600-800  $\mu\text{mol m}^{-2} \text{s}^{-1}$ , and day/night temperatures of 22°C/18°C.

### ***Wax extraction for GC-MS Profiling of Wax Composition***

Cuticular wax was extracted from flag leaf sheaths at the late booting / early heading stage by submerging tissues in a glass tube containing 10 mL of HPLC grade chloroform (Fisher Scientific) and 10 ng of tetracosane (Sigma-Aldrich) as an internal standard and vortexing manually for 1 min. The tissue was rinsed with an additional 5 mL of chloroform, and the two extracts were combined. The wax extract was dried under a nitrogen stream and resuspended in 250  $\mu\text{L}$  of toluene. Wax components were separated and identified by GC-MS using an Agilent 6890N GC equipped with a 15 m MXT-1 capillary column ([www.restek.com](http://www.restek.com)) and an Agilent 5973N mass selective detector. Samples were injected in split mode with hydrogen as carrier and an initial column temperature of 125°C. To enable separation of multiple components ranging from long chain fatty acids ( $>\text{C}_{14}$ ) to wax esters ( $>\text{C}_{52}$ ), the oven temperature was increased at 5°C/min to a maximum of 350°C and held at that temperature for 5 minutes. Components were identified by comparison of retention time and mass spectra to standards. To confirm the position of hydroxyl groups in hydroxy- $\beta$ -diketones, 50  $\mu\text{L}$  aliquots of each sample were transferred to fresh glass GC vial inserts ([www.chromspec.com](http://www.chromspec.com)) and mixed with an equal volume of 50% N,O-Bis(trimethylsilyl)acetamide (BSA) in pyridine. After 30 minutes at room temperature, samples were analysed by GC-MS using the conditions described above.

### ***DNA and RNA isolation***

Leaf sheaths in 100-150 mg amounts were ground to a fine powder in liquid N<sub>2</sub> prior to extraction. Total genomic DNA was isolated from the NIL wheat pairs using the GenElute Plant Genomic DNA Miniprep Kit (Sigma) following the manufacturer's instructions, and used subsequently for PCR, genome-walking, and cloning. RNAs were extracted and small RNAs enriched from all samples with the mirVana miRNA isolation kit in combination with Plant RNA Isolation Aid according to the manufacturer's protocols (ThermoFisher Scientific). Purification

included a phenol step and used 100-150 mg of material. Samples for RNA extraction were typically the leaf sheaths between the flag leaf and the penultimate leaf unless otherwise noted. Contaminating DNAs were removed using the TURBO DNA-free Kit (ThermoFisher Scientific) prior to RNA-seq, qPCR, cDNA RACE and full-length cDNA cloning. The quantity of DNA, RNA and small RNA were measured spectrophotometrically on a Synergy H1 plate reader with Take3 adapter (Biotek). RNA quality was analysed with the Agilent 2100 BioAnalyzer with the RNA 6000 Nano kit (large RNA) according to the manufacturer's instructions (Agilent Technologies).

### ***Library preparation for DNA- and RNA-sequencing***

Stranded RNA sequencing libraries for the NILs, using 5-6 µg total RNA as starting material, were prepared using the Ribo-Zero rRNA Removal Kit for Plants and the ScriptSeq v2 RNA-Seq Library Preparation Kit with ScriptSeq Index PCR primers according to the manufacturer's instructions (Epicentre, an Illumina Company). For all subsequent RNA library preparations, the TruSeq RNA Library Preparation Kit v2 was used according to manufacturer's instructions (Illumina). For Small RNA, the TruSeq Small RNA Library Preparation kit was used according to the manufacturer's instructions (Illumina). RNA and library quality control was accomplished using the RNA 6000 Pico kit and High Sensitivity DNA and DNA 1000 kits (Agilent Technologies).

### ***Gene expression and small RNA processing and abundance calculations***

Paired-end sequencing with 101 or 151 cycles was performed on a HiSeq2500 (Illumina). The sequencing reads were assigned to corresponding sample libraries based on specific barcode sequences added to each sample during library preparation. Two workflows were used for the analyses of high-throughput RNA sequencing data from the NILs and *lw1* over-expression (*lw*-OE) lines. Analyses based on mapping to the NCBI Unigene set were accomplished with the CLC Genomics workbench (Qiagen Bioinformatics). Following removal of low quality sequence and adaptor trimming, reads were mapped to the NCBI *Triticum aestivum* Unigene Set (Build # 63, <http://www.ncbi.nlm.nih.gov/UniGene/UGOrg.cgi?TAXID=4565>), consisting of 178,464 sequences. The expression levels of the unigenes were calculated as unique read counts per unigene using the CLC genomic workbench RNA-seq analysis module. A second workflow incorporated transcripts produced from the International Wheat Genome Sequencing Consortium (IWGSC) Chinese Spring Wheat Chromosome Survey Sequence (v1). 'Triticum\_aestivum.IWGSC1+popseq.29.cdna.all' was downloaded from Ensembl Plants (4) and was subsequently modified into two separate references: ABD-W1-COE, consisting of AABBD genome transcripts with the addition of the *W1*-COE cDNA and AB-W1-COE, consisting of AABBD genome transcripts with addition of the *W1*-COE cDNA. The *W1*-COE cDNA sequence added to these references was from Chinese Spring wheat. Following removal of low quality reads and trimming of adaptors with trimmomatic 0.35 (5), kallisto 0.42.4 was used to pseudo-align the reads to the either the AB-W1-COE or ABD-W1-COE reference transcripts for NILs or *lw*-OE lines, respectively, to produce an abundance file containing, transcript lengths, estimated counts and transcripts per million (6).



Small RNA sequencing, consisting of 51 cycles, was performed on a HiSeq2500 (Illumina). The analysis workflow for small RNA was accomplished using CLC Genomics Workbench (Qiagen) and included filtering for quality and adaptor trimming, filtering the reads to between 19-28 nt and a count greater than 1, and assembly into tabular read counts files containing reads per small RNA sequence per sample.

For qPCR expression measurements of potential target genes in the NILs including the F1 generation, 1.6 µg of RNA was used with the SuperScript III First-Strand Synthesis SuperMix for qRT-PCR (ThermoFisher Scientific) according to the manufacturer's protocol except that the synthesis reaction occurred in 40 µL at 50°C for 60 min. cDNA was diluted 12.5-fold in water and 8 µL of this dilution was used for qRT-PCR expression analysis. For qPCR, the Power SYBR Green Master Mix (ThermoFisher Scientific) was used in a total reaction volume of 20 µL and 0.2 µM primer concentration. Cycle threshold (CT) values were collected on an ABI StepOne Plus using the StepOne v2.3 Software and relative quantitation was calculated based on using the  $2^{-\Delta\Delta CT}$  method (7). The gene, Ta.46201, Cell Division Control protein 48 homolog E-like, was selected as a reference based on Paolacci et al. (8) and Giménez et al. (9). Measurement of *W1-COE* (target1) expression following VIGS treatments utilized the procedure described above except that the High-Capacity RNA-to-cDNA (ThermoFisher Scientific) was used to synthesize cDNA from 1 µg of RNA following the manufacturer's instructions.

### **Genome walking and RACE-PCR**

Cloning of an expanded genomic region containing *lw1* was accomplished using the Universal GenomeWalker 2.0 kit following the manufacturer's protocol (Clontech laboratories). AG1N and D051N genomic DNA were digested with four separate restriction enzymes (EcoRV, DraI, PvuII, and StuI) and subsequently ligated to the GenomeWalker adaptors. GenomeWalker DNA was then used as a template for primary PCR amplification. Gene specific primers (GSP) were designed from the contigs assembled from the RNA-seq reads which loosely mapped to *W1-COE* (target1). Nested PCR was performed using outer and inner adapter primers provided in the kit and nested primers specific to the identified *lw* sequence (Fig. 6, primers listed *SI Appendix*, Table S4). A series of PCR products, representing the genome walking steps, were cloned into the pCR-Blunt II TOPO vector (ThermoFisher Scientific), sequenced sequentially, and assembled to eventually to obtain a 3knt fragment (Fig. 2).

A rapid amplification of cDNA ends (RACE) approach was used to clone the *lw1* cDNA using the GeneRacer kit following the manufacturer's protocol (ThermoFisher Scientific). Briefly, two to 3 µg of total RNA from AG1N and AG2N leaf sheaths was dephosphorylated and de-capped prior to ligation of the GeneRacer RNA oligo to full-length mRNA. The RNA was reverse transcribed using SuperScript III and PCR products were cloned using the Zero-Blunt TOPO PCR Cloning Kit. Five- and 3'-ends were obtained using nested PCR with GeneRacer 5' and 3' primers, including nested versions, together with 5' and 3' targeted GSP that were designed from the *lw1* template obtained from genome walking (Fig. 2, *SI Appendix* Table S4). Following analysis of the RACE PCR fragments, the full-length *lw1* cDNA was cloned into the Gateway-

cloning vector pDONR221 using NT Clonase II (ThermoFisher Scientific) for use in subsequent analyses (see *Biolistic-mediated transformation of wheat*). To map the miRNA cleavage sites, a modified RACE procedure was performed with the GeneRacer kit (ThermoFisher Scientific). Briefly, 1.2 µg of RNA from transgenic lines 7279 and 7282 were ligated to the GeneRacer RNA oligo without the dephosphorylation and de-capping steps to capture cleaved and incomplete cDNA products. Following a phenol-chloroform and precipitation purification step, RNA was reverse transcribed using the GeneRacer oligoDT primer. Five-prime-ends were obtained using nested PCR with the GeneRacer 3' primers and GSP (*SI Appendix*, Table S4) and the resultant cDNA products were cloned using the Zero-Blunt TOPO PCR Cloning Kit (ThermoFisher Scientific). PCR amplification for genome walking, RACE, and full-length cDNA cloning was accomplished using the high-fidelity-enzyme PfuUltra II Fusion HS DNA polymerase (Agilent Technologies).

### ***Statistical analyses, data mining, and data visualizations***

The R statistical computing language, running within the RStudio integrated development environment, was used for significance testing, data analysis and production of graphs and figures; this included integration of the Bioconductor software framework for many of the tasks as outlined below (10, 11, 12).

Two approaches were used to test for differential gene expression associated with the glaucous-non-glaucous states. Initial experiments using unique counts per NCBI unigene (CLC Genomics analysis pipeline) used a non-adjusted p-value of  $\leq 0.05$  as the level of significance to test for differentially-expressed genes (DEGs). DEG significance testing was accomplished using the Bioconductor package edgeR using the suggested 'classic approach' parameters for pairwise comparisons (13). Second follow-up analyses using IWGSC transcripts (trimmomatic-kallisto analysis pipeline) used an adjusted p-value of  $\leq 0.05$ . Transcript abundances from kallisto output were imported into R using the R Bioconductor package tximport (14) and readr package (15). The standard workflow of the Bioconductor package DESeq2 was followed for calculation of the DEGs (16) including p-value adjustment using the Benjamini & Hochberg method (17).

Testing for differential small RNA expression was accomplished using the Bioconductor package edgeR (13). To reduce the interference of low counts in significance testing using the edgeR pipeline, sequences were included in the significance analysis if present at greater than 1 count per million (cpm) in at least one-quarter of the samples. DEG significance testing was accomplished using the R Bioconductor package edgeR using the suggested 'classic approach' parameters for pairwise comparisons except that upper-quartile normalization was used (13, 18, 19).

The following R packages also contributed to the analysis of data: data.table (20); dplyr (21); stringr (22); tidyr (23). In particular, dplyr was used heavily within the R environment to arrange data and select common elements between dataframes (e.g. common DEGs between NIL pairs). Graphs were produced using the package ggplot2 and ggrepel (24, 25).

For some datasets, preliminary data arrangement and summary occurred in Microsoft Excel. Most sequence-based data analyses occurred in CLC Genomics Workbench (Qiagen). For sequence-based figures, fasta, bam, or gff files were exported from CLC Genomics Workbench and were produced using the Bioconductor package GViz (26). Mapping of the miRW1 cleavage site was produced with the Bioconductor package trackViewer (27). Additional Bioconductor packages which were used in data analysis and figure preparation included: Biostrings and Rsamtools (28, 29). Bam files were converted to bedGraphs, containing coverage data, for plotting in GViz with the Samtools 1.3 and BEDTools 2.25.0 software packages (30, 31). Mapping of the *lw* location in the D genome was accomplished using the *Aegilops tauschii* genome sequences from Jia et al. (32) and the “Sequencing the *Aegilops tauschii* Genome” Project based at the University of California, Davis (<http://aegilops.wheat.ucdavis.edu/ATGSP/>), the W7984 synthetic wheat genome sequence (33), and sequence resources available from Ensembl Plants (4) (<http://plants.ensembl.org/index.html>). Marker sequences in the *lw2* region from *Ae. tauschii* were obtained from Nishijima et al. (34). The software package MUMMER 3.23, including NUCMER, was used to compare the *lw* genomic regions of *Ae. tauschii* and *T. aestivum* W7984 (35) (*SI Appendix*, Fig. S8). The GMAP software (version 2014-12-29) was used to map markers and sequences of *Ae. tauschii*, *T. aestivum*, and *Brachypodium distachyon* to the *lw* genomic region (36). PDF figure files were arranged and edited in Adobe Illustrator, Adobe Photoshop, and Adobe InDesign for final production.

### ***Small RNA mapping and lw1 sequence analysis***

Due to the presence and observation of small RNA trimming and tailing, small RNA mapping was accomplished using the software Tailor 1.1, which has the ability to detect and account for tailing during alignment and mapping to the genome (37). Significantly differentially expressed small RNAs from the NIL comparisons and *lw*-OE were mapped to *lw1* to determine how many short sequences arose from *lw1* and the extent of tailing. In addition, the same set of small RNAs was also mapped to the IWGSC AABBDD Chinese Spring Wheat genome (v1). To avoid issues during indexing and mapping, individual contigs were concatenated into non-ordered pseudo-chromosomes with 100xN between each contig for each chromosome arm. The software packages Samtools 1.3 (30) and BEDOPS 2.4.16 (38) were used to determine the closest transcript (feature) to the mapping locations, within the genome, of the significant small RNAs. In addition, small RNA reads with perfect match to the *lw1* transcript were mapped with CLC Genomics Workbench (Qiagen) and to the *W1-COE* ORF with up to 3 base mismatches using the Burrows-Wheeler Aligner (BWA v0.7.10-r789) (39).

The *lw1* cDNA and homologous regions from *Ae. tauschii* and *T. aestivum* W7984 were analyzed with the RNAfold program and mountain.pl and replot.pl scripts of the ViennaRNA v2.2.5 package (40). To compare the *lw1* cDNA with the *W1-COE* ORF sequences, we used the NUCMER program from the MUMMER 3.23 package (35). The *lw1* cDNA was analyzed for inverted repeats using the einverted program written by Richard Durbin (Sanger Institute, Cambridge, UK) and Peter Rice (European Bioinformatics Institute, Cambridge, UK); einverted



is included in the Jembooss package v6.6.0-1(41). The final mummer and mountain plots were created using ggplot2 (24) with data wrangling in dplyr (21) and tidyr (23).

### ***Virus-Induced Gene Silencing (VIGS)***

The BSMV-VIGS system developed by Yuan et al. (42), comprising three T-DNA binary plasmids, pCaBS- $\alpha$ , pCaBS- $\beta$ , and pCa- $\gamma$ bLIC, was used with modifications. The pCa- $\gamma$ bLIC vector was modified by introducing ccdB gene for efficient selection of recombinant clones. The resulting vector pCa- $\gamma$ bLIC-ccdB was used to clone a series of target gene fragments by a ligation independent cloning (LIC) strategy (43). The primers for target gene fragments are listed in *SI Appendix*, Table S4. The target gene fragments were PCR amplified from wax line AG2 genomic DNA using PfuUltra II Fusion HS DNA Polymerase (Agilent Technologies). The PCR fragments were gel purified with QIAquick Gel Extraction buffer and MinElute column (Qiagen). The purified PCR fragments were cloned into the pCa- $\gamma$ bLIC-ccdB vector by using a LIC strategy. The resulting clones were sequencing verified. The plasmid DNAs of pCaBS- $\alpha$ , pCaBS- $\beta$ , and pCa- $\gamma$ bLIC with target gene fragments were transformed into *Agrobacterium tumefaciens* strain c58. *A. tumefaciens* strains harboring pCa- $\gamma$ bLIC vector with PDS (42) and GFP fragment were used as controls. Agroinfiltration into 3- to 4-week-old *N. benthamiana* plants and viral inoculation of wheat and barley at tillering stages were carried out as described previously (42).

### ***Biolistic-mediated transformation of wheat***

The full length of *lw1* cDNA in pDONR221 was transferred into the pANIC 5E vector (44) using LR Clonase II (ThermoFisher Scientific) following the manufacturer's protocol. The pANIC 5E vector contains the *ZmUbi1* promoter (maize ubiquitin 1 promoter and intron) to drive the expression of the wheat *lw1* cDNA and a selectable marker, bar gene, which was placed under the transcriptional control of the rice actin 1 gene (*OsAct1*) promoter.

Macrocarriers (0.6  $\mu$ M gold particles) were prepared and coated with plasmid DNA according to the protocol by Jordan (45). Immature embryos from the bread wheat cultivar Bobwhite with 2-3 mm in size were isolated and arranged in the center of a petri dish of Osmotic Medium with the scutellum facing upwards. The embryos were bombarded with plasmid DNA coated gold particles at 650 psi from rupture disks using the Biolistic PDS-1000/He Particle Delivering System (Bio-Rad, USA) according to manufacturer's instructions. The media and the culture of explants after bombardment and the selection of transformants were prepared and performed as described by Jordan (45) with modifications.

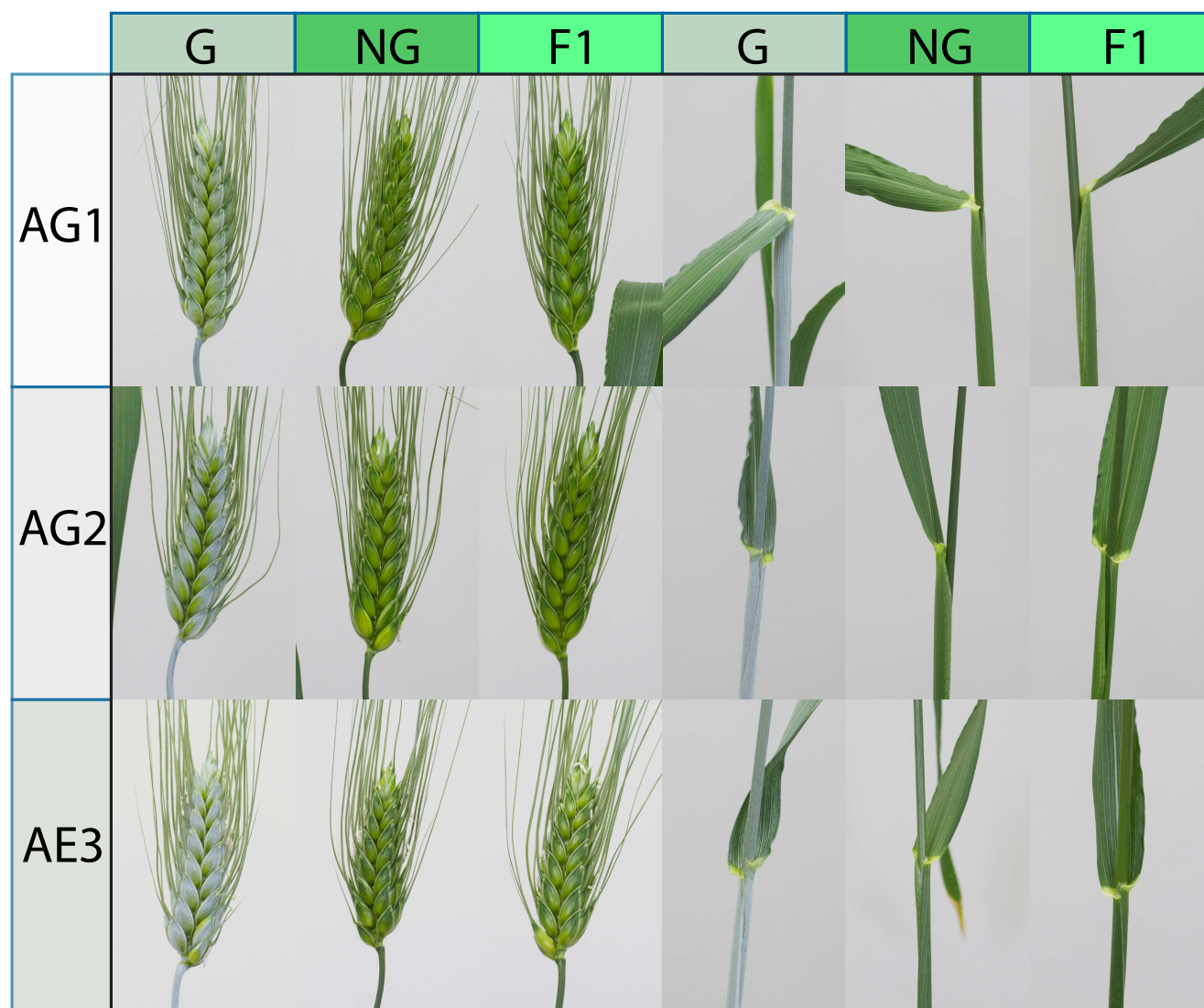
## SI References

1. Clarke JM, McCaig TN, DePauw RM (1993) Relationship of glaucousness and epicuticular wax quantity of wheat. *Can J Plant Sci* 73(4):961-967.
2. Clarke JM, McCaig TN, DePauw RM (1994) Inheritance of glaucousness and epicuticular wax in durum wheat. *Crop Sci* 34(2):327-330.
3. Clarke JM, et al. (1995) Registration of seven pairs of durum wheat genetic stocks near-isogenic for glaucousness. *Crop Sci* 35(4):1241.
4. Kersey PJ, et al. (2016) Ensembl Genomes 2016: more genomes, more complexity. *Nucleic Acids Res* 44(D1): D574-580.
5. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.
6. Bray N, Pimentel H, Melsted P, Pachter L (2016) Near-optimal RNA-Seq quantification. *Nat Biotechnol* 34(5):525-527.
7. Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29(9):e45.
8. Paolacci AR, Tanzarella OA, Porceddu E, Ciaffi M (2009) Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. *BMC Mol Biol* 10:11.
9. Giménez, MJ, Pistón F, Atienza SG (2011) Identification of suitable reference genes for normalization of qPCR data in comparative transcriptomics analyses in the Triticeae. *Planta* 233(1):163-173.
10. R Core Team (2015) R: A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria <https://www.R-project.org/>
11. RStudio Team (2015) RStudio: Integrated Development for R RStudio, Inc, Boston, MA <http://www.rstudio.com/>
12. Huber W, et al. (2015) Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods* 12(2):115-121.
13. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139-140.
14. Soneson C, Love MI, Robinson MD (2015) Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* 4:1521.
15. Wickham H, Francois R (2015) readr: Read Tabular Data. R package version 0.22 <https://CRAN.R-project.org/package=readr>
16. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):550.
17. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *JR Statist Soc B* 57:289–300.

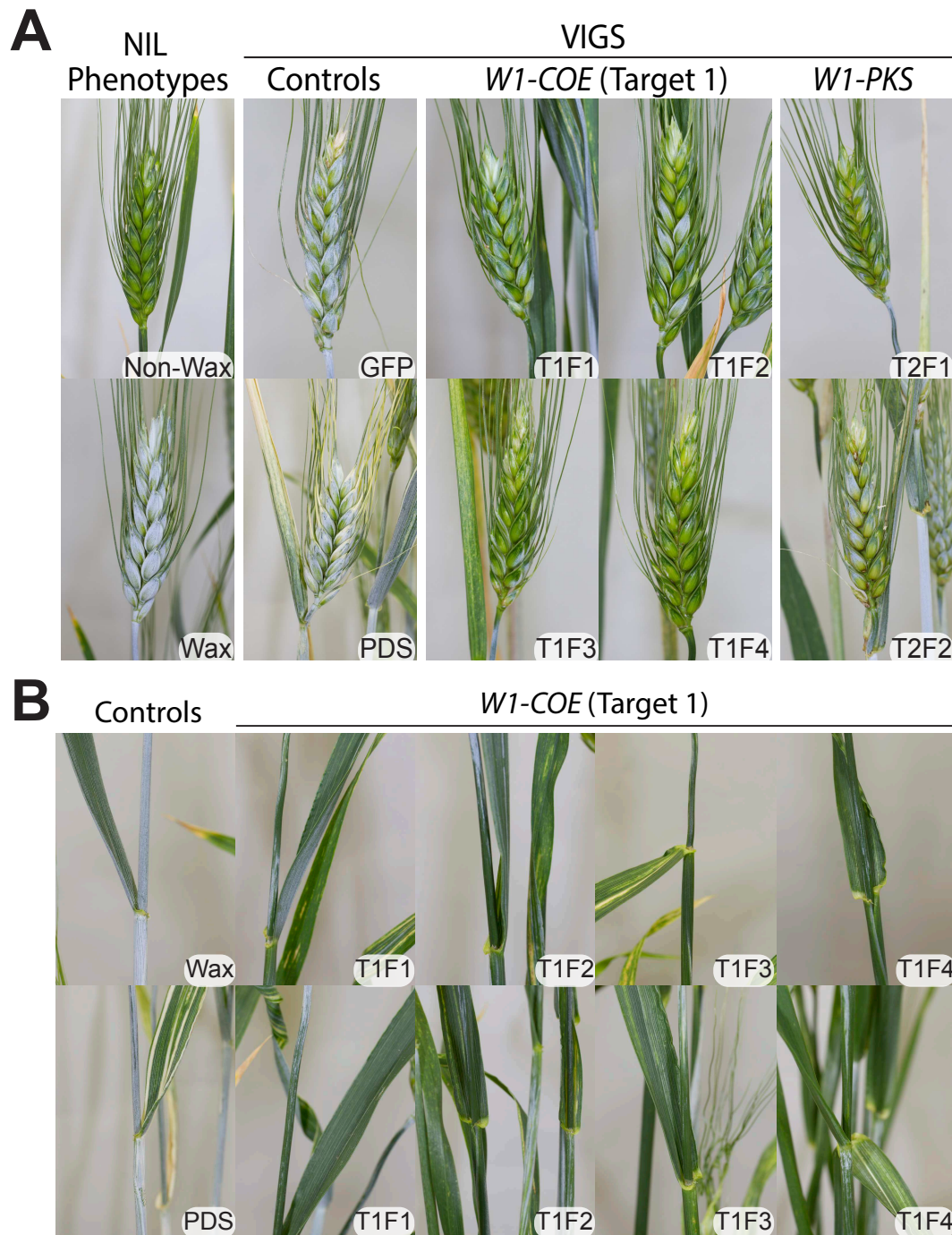
18. Bullard JH, Purdom E, Hansen KD, Dudoit S (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* 11:94.
19. Tam S, Tsao MS, McPherson JD (2015) Optimization of miRNA-seq data preprocessing. *Brief Bioinform* 16(6):950-963.
20. Dowle M, et al. (2015) Data.table: Extension of Data.frame. R package version 196 <https://CRANR-projectorg/package=datatable>
21. Wickham H, Francois R (2015) dplyr: A Grammar of Data Manipulation. R package version 043 <https://CRANR-projectorg/package=dplyr>
22. Wickham H (2015) stringr: Simple, Consistent Wrappers for Common String Operations. R package version 100 <https://CRANR-projectorg/package=stringr>
23. Wickham H (2016) tidyr: Easily Tidy Data with spread() and gather() Function. R package version 041 <https://CRANR-projectorg/package=tidyr>
24. Wickham H (2009) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 213pp
25. Slowikowski K(2016) ggrepel: Repulsive Text and Label Geoms for 'ggplot2'. R package version 0.6.5. <https://CRAN.R-project.org/package=ggrepel>
26. Hahne F, et al. (2016) Gviz: Plotting data and annotation information along genomic coordinates. R package version 1144, <https://www.bioconductor.org/packages/release/bioc/html/Gviz.html>
27. Ou J, Wang Y, Zhu LJ (2016) trackViewer: A bioconductor package with minimalist design for drawing elegant tracks or lollipop plot. R package version 1.10.
28. Morgan M, Pagès H, Obenchain V, Hayden N (2016) Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. R package version 1220, <http://bioconductor.org/packages/release/bioc/html/Rsamtools.html>
29. Pages H, Aboyoun P, Gentleman R, DebRoy S (2016) Biostrings: String objects representing biological sequences, and matching algorithms R package version 2384, <https://www.bioconductor.org/packages/release/bioc/html/Biostrings.html>
30. Li H, et al. (2009) The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25(16):2078-2079.
31. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841-842.
32. Jia J, et al. (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* 496(7443):91-95.
33. Chapman JA, et al. (2015) A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biol* 16:26.
34. Nishijima R, Iehisa JCM, Matsuoka Y, Takumi S (2014) The cuticular wax inhibitor locus *lw2* in wild diploid wheat *Aegilops tauschii*: phenotypic survey, genetic analysis, and



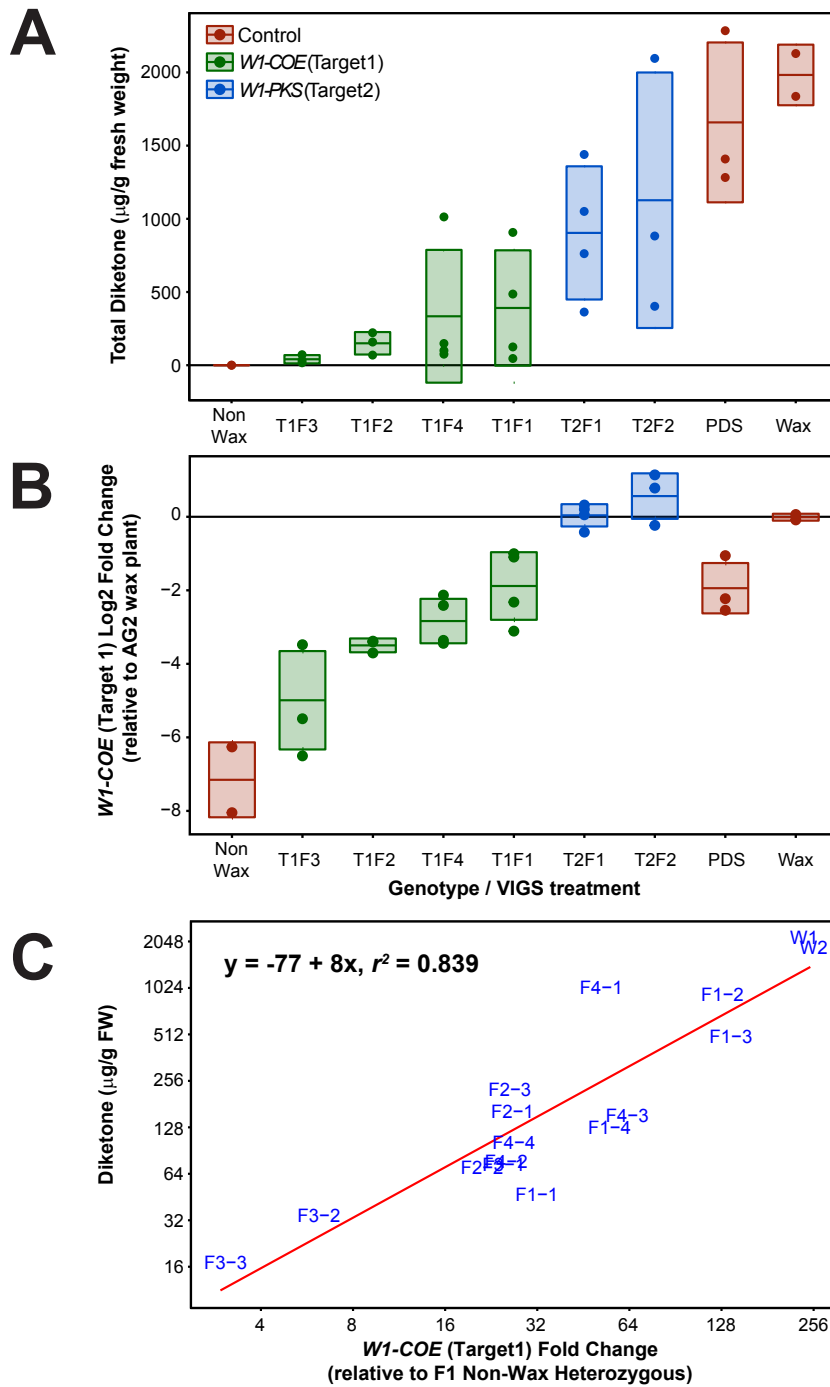
- implications for the evolution of common wheat. *BMC Plant Biol* 14:246.
35. Kurtz S, et al. (2004) Versatile and open software for comparing large genomes. *Genome Biol* 5(2):R12.
  36. Wu TD, Watanabe CK (2005) GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21(9):1859-1875.
  37. Chou MT, et al. (2015) Tailor: a computational framework for detecting non-templated tailing of small silencing RNAs. *Nucleic Acids Res* 43(17): e109.
  38. Neph S, et al. (2012) BEDOPS: high-performance genomic feature operations. *Bioinformatics* 28(14):1919-20.
  39. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14):1754-1760.
  40. Lorenz R, et al. (2011) ViennaRNA Package 2.0. *Algorithms Mol Biol* 6:26.
  41. Carver T, Bleasby A (2003) The design of Jemboss: a graphical user interface to EMBOSS. *Bioinformatics* 19(14):1837-1843.
  42. Yuan C, et al. (2011) A high throughput barley stripe mosaic virus vector for virus induced gene silencing in monocots and dicots. *PLoS ONE* 6(10):e26468.
  43. Haun RS, Servent IM, Moss J (1992) Rapid, reliable ligation-independent cloning of PCR products using modified plasmid vectors. *BioTechniques* 13(4):515-518.
  44. Mann DG, et al. (2012) Gateway-compatible vectors for high-throughput gene functional analysis in switchgrass (*Panicum virgatum* L.) and other monocot species. *Plant Biotechnol J* 10(2):226-236.
  45. Jordan MC (2000) Green fluorescent protein as a visual marker for wheat transformation. *Plant Cell Rep* 19(11):1069-1075.



**Fig. S1.** Dominance of the non-glaucous (NG) trait. Crosses between glaucous (G) and non-glaucous (NG) isolines of AG1, AG2, and AE3 resulted in F1 heterozygous plants that were non-glaucous. The three columns of the left show representative heads from each line and the three right hand columns show representative stems and flag leaves. Glaucous tissue has a blue-white coloration.

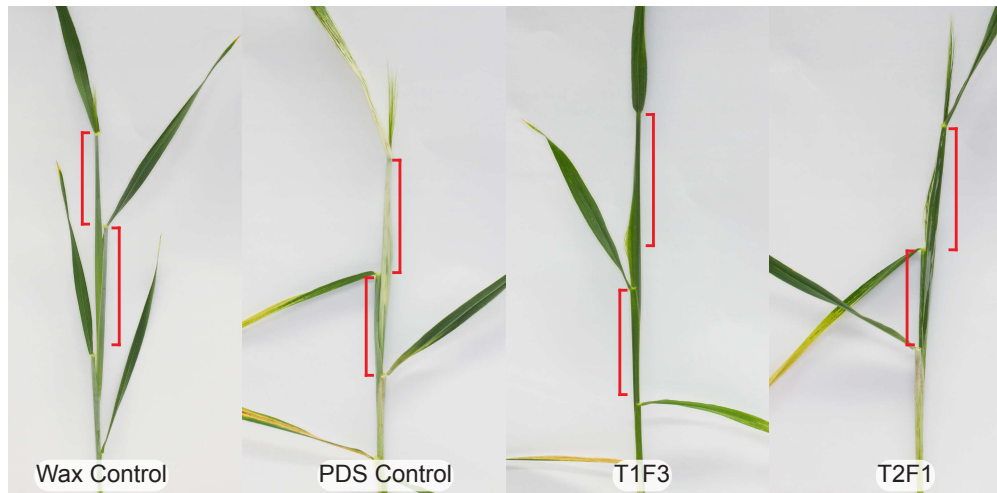


**Fig. S2.** Reduction in glaucousness resulting from Virus Induced Gene Silencing (VIGS). (A) VIGS using DNA fragments (F1, F2 etc.) of target genes 1 (T1F1-T1F4) and 2 (T2F1-T2F2), representing *W1-COE* and *W1-PKS* respectively, results in decreased glaucousness on wheat heads. Controls include non-VIGS-treated glaucous and non-glaucous AG2 NILs and VIGS-treated plants using DNA fragments targeting green fluorescent protein (GFP) or Phytoene desaturase (PDS). Knockdown of the PDS gene results in chlorotic regions on leaves, stems, heads, and awns. (B) VIGS using DNA fragments of target gene 1 (*W1-COE*) results in decreased glaucousness on wheat stems. Wax and PDS controls are described in (A). NG phenotypes began to appear 3 weeks after VIGS treatment and were monitored for 4-5 weeks.

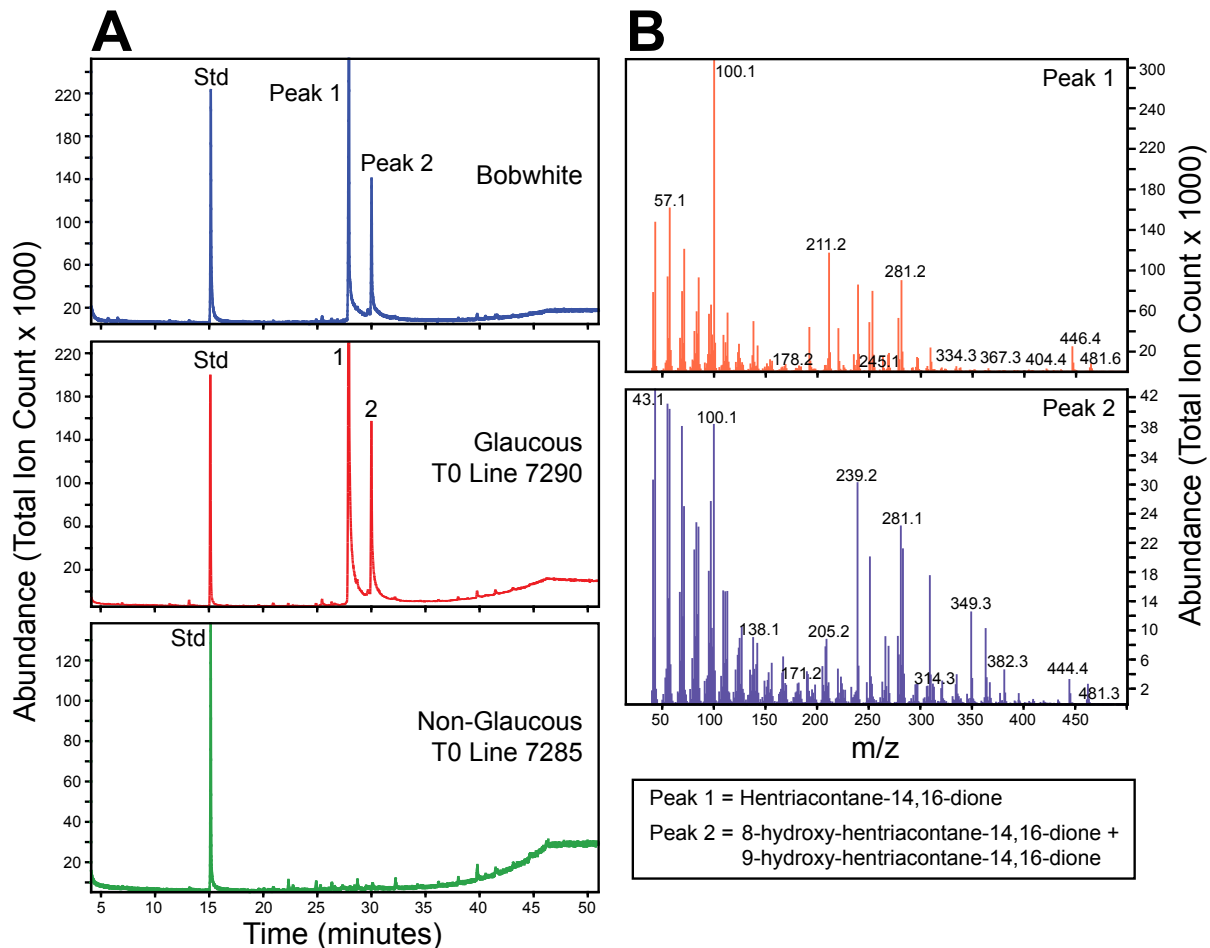


**Fig. S3.** Reduction of diketones in leaf sheaths correlates with reductions in *W1-COE* (Target 1) gene expression during VIGS suppression in AG2 NILs. (A) Total diketone, including hentriacontane-14,16-dione ( $\beta$ -diketone) and 25-hydroxy- $\beta$ -diketone, are reduced after VIGS treatments of target 1 and 2 genes. For example, leaf sheaths of target 1 fragment 3 (T1F3) plants are almost completely depleted of diketones. (B) Expression of *W1-COE*, as measured by qRT-PCR, is reduced in AG2 glaucous plants treated with VIGS fragments for *W1-COE*. (C) Diketone amount correlates with the fold change in *W1-COE* expression in the VIGS treated plants. F represents fragment of *W1-COE* used for VIGS, W1/W2 represent glaucous control plants. Selection of leaf sheaths for analysis is shown in Fig. S4. Non-Wax in (A-B) is an F1 heterozygous plant. Boxplots in (A-B) represent the mean  $\pm$  standard deviation with data points indicated.

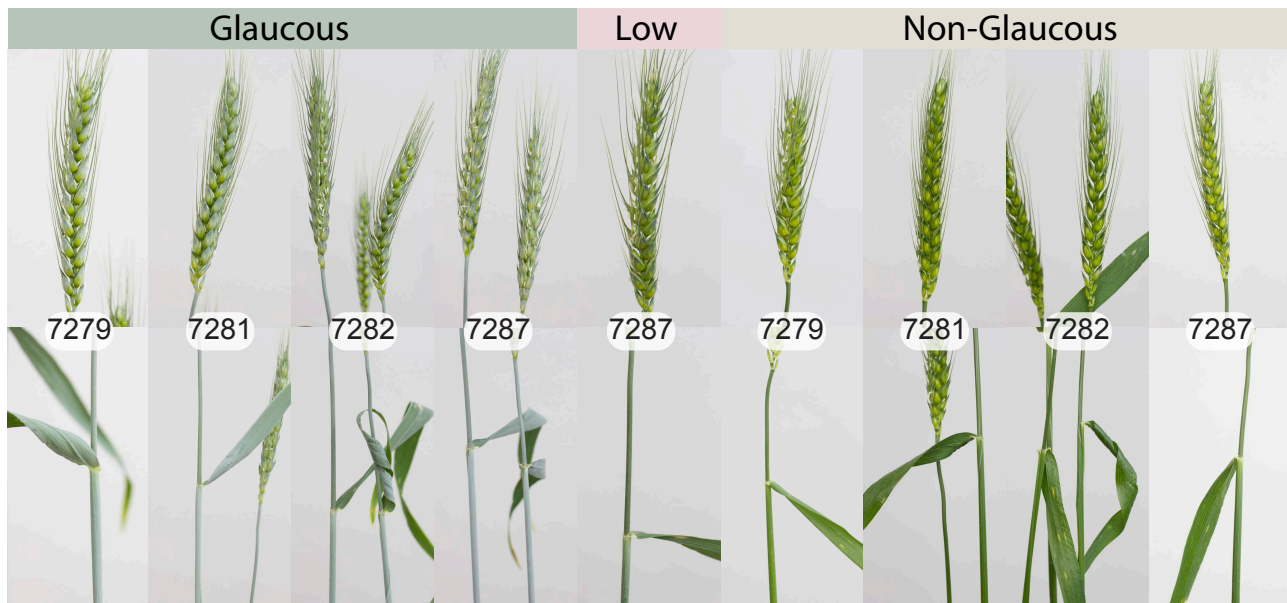




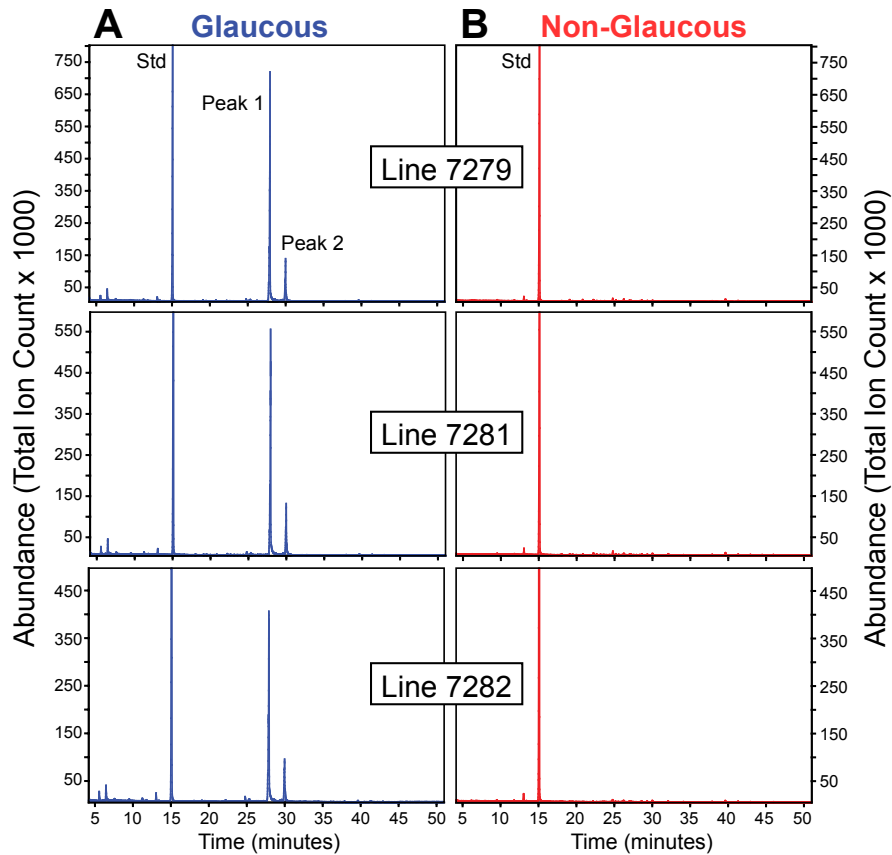
**Fig. S4.** Leaf sheaths from the flag leaf and penultimate leaf (shown in red) were used for wax and gene expression analyses. Examples for the wax and PDS controls and target 1 fragment 3 and target 2 fragment 1 are depicted as examples.



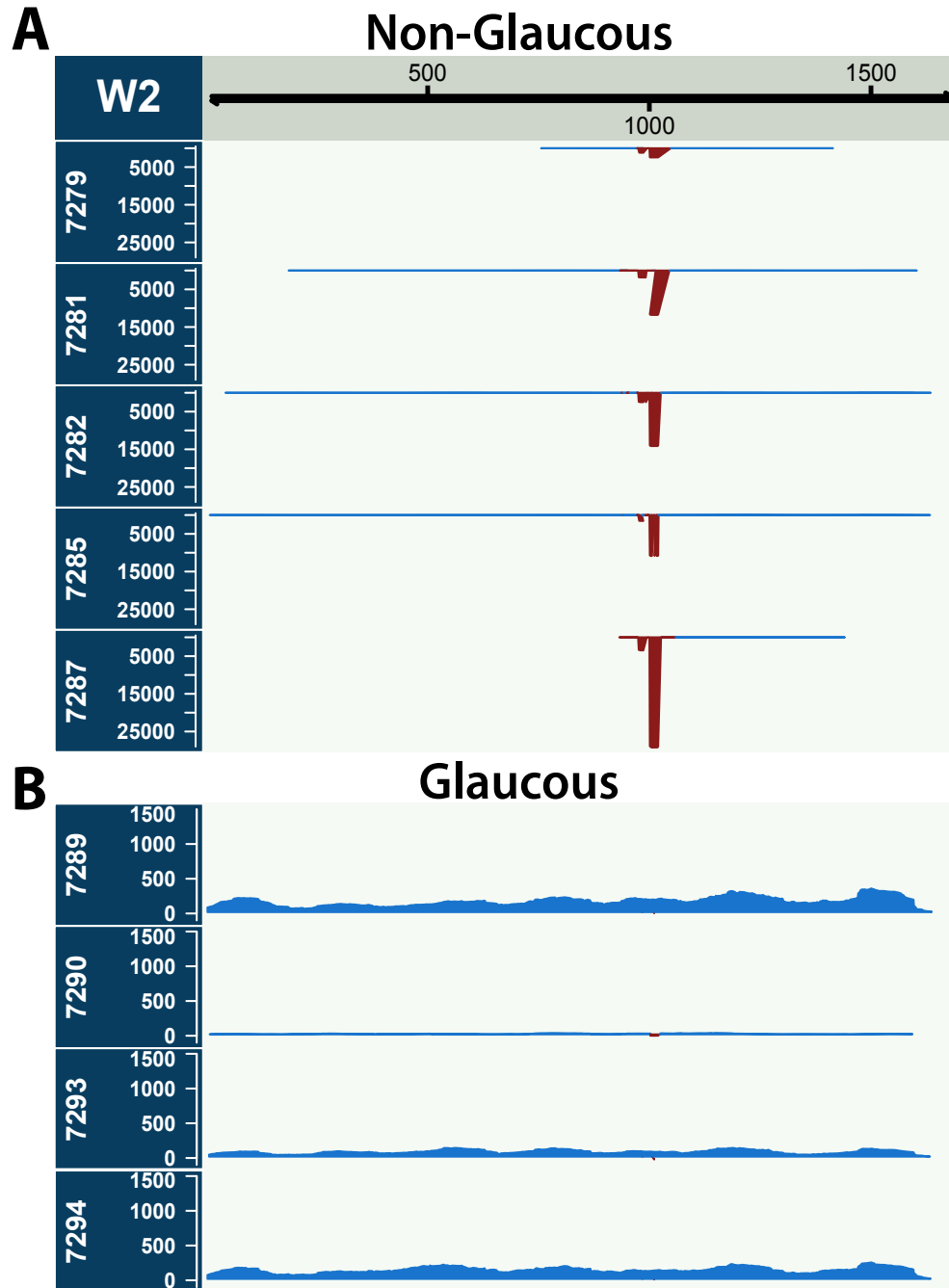
**Fig. S5.** Characterization of diketone waxes in flag leaf sheaths of *lw1* over-expressing (OE) lines. (A) Total Ion Count (TIC) spectra of cuticular wax from leaf sheath of untransformed spring wheat cultivar Bobwhite and primary *lw1* OE transformants (T0 generation), as shown in Figure 4. Std =  $C_{24}$  standard added. (B) Mass spectra of the primary cuticular wax components, Hentriacontane-14,16-dione (Peak 1) and a mixture of 8-hydroxy-hentriacontane-14,16-dione and 9-hydroxy-hentriacontane-14,16-dione (Peak 2).



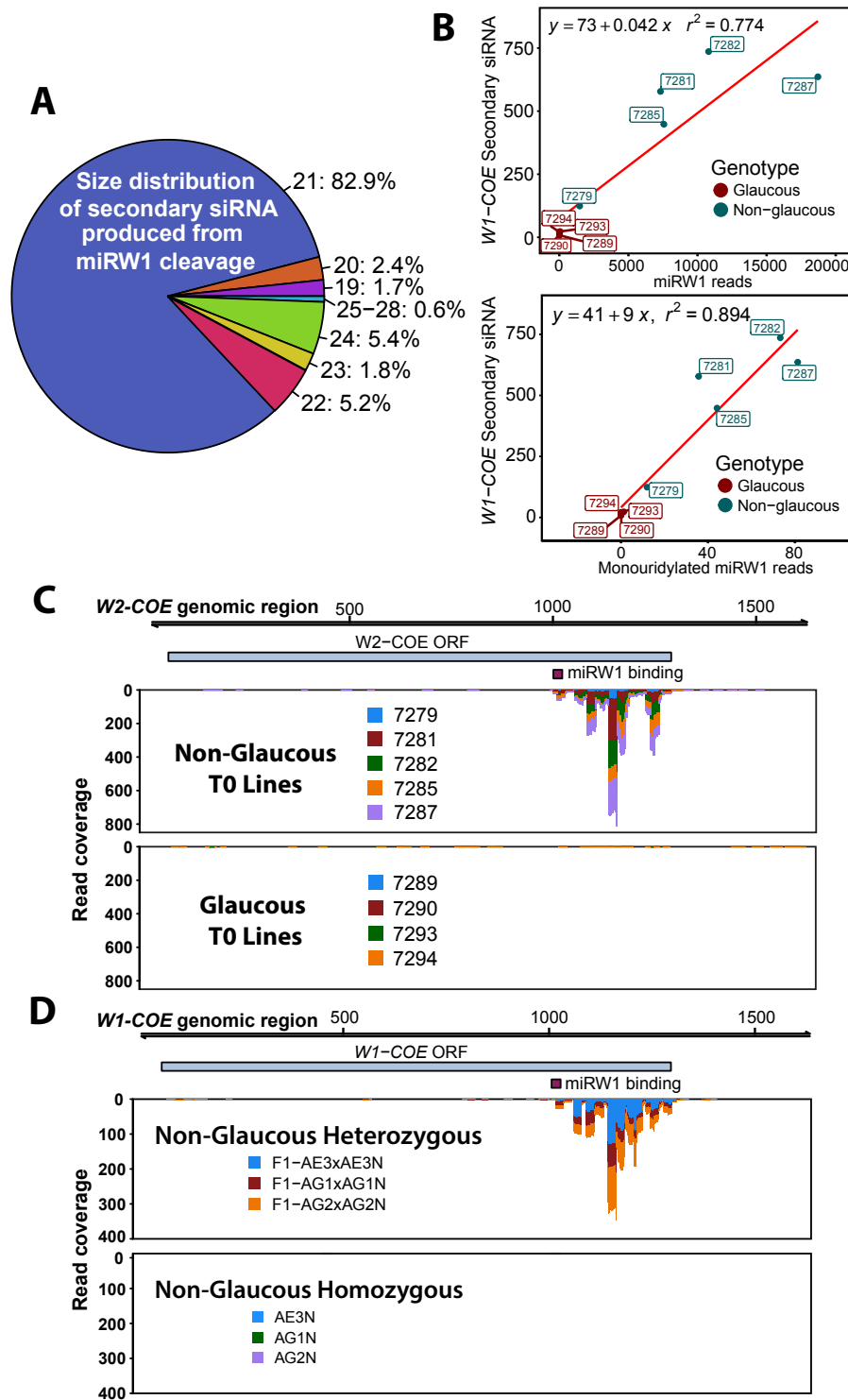
**Fig. S6.** Segregation of the glaucous trait in *lw1* over-expression lines in the T1 generation. In addition to clearly glaucous and non-glaucous plants, a few showed an intermediate low glaucous phenotype; line 7287 is shown as an example. Heads after anthesis and stems surrounding the flag leaf are shown.



**Fig. S7.** Total Ion Count (TIC) spectra of cuticular wax from flag leaf sheaths of glaucous (A) and non-glaucous (B) T1 plants of lines 7279, 7281 and 7282 segregating for *lw1* over-expression. Std = C<sub>24</sub> standard added. Peak 1 = Hentriacontane-14,16-dione (C<sub>31</sub>-14,16-dione); Peak 2 = 8-hydroxy-hentriacontane-14,16-dione and 9-hydroxy-hentriacontane-14,16-dione (8-OH-C<sub>31</sub>-14,16-dione and 9-OH-C<sub>31</sub>-14,16-dione).



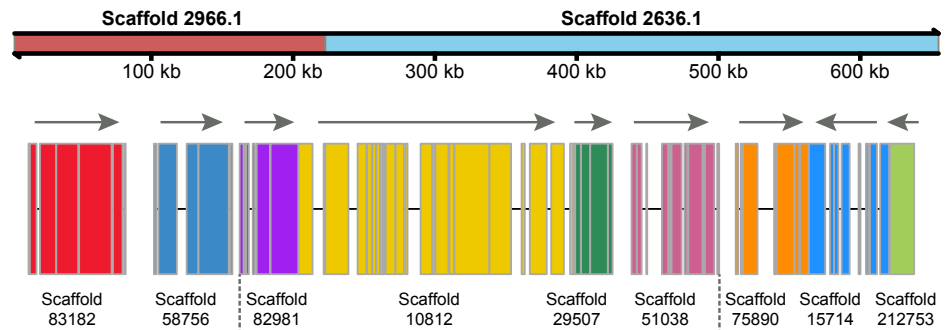
**Fig. S8** Mapping of RNA and small RNA sequencing reads to *W2-COE* in T0 transgenic lines over-expressing *lw1* (miRW1 precursor). (A) Non-glaucous lines displayed expression of sRNA (shown in red) derived from *lw1* and repression of *W2-COE* (blue). (B) In contrast, glaucous lines showed almost no expression of sRNA from *lw1* and transcription of *W2-COE*. Expression of *W2-COE* and its higher expressed paralog *W1-COE* (Fig. 4) would allow  $\beta$ -diketone synthesis and deposition to occur and thus the presence of a glaucous state.



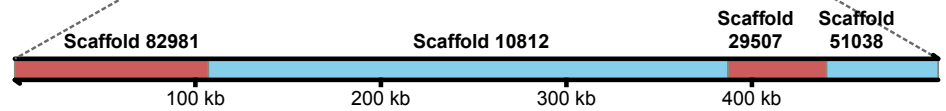
**Fig. S9** Production of secondary siRNAs from *W1-COE* and *W2-COE*. (A) Small RNA size distribution of secondary siRNA arising from *W1-COE*. (B) The relationship between *W1-COE* secondary siRNA and miRW1 or monouridylated miRW1 reads. Reads are normalized per 10 million. (C) Secondary siRNA mapping to *W2-COE*. (D) Secondary siRNA mapping to *W1-COE*. The siRNAs mapped are perfect matches to either *W1-COE* or *W2-COE* and do not map to *lw1*. (A-C) are data from miRW1 precursor (*lw1*) over-expression lines while (D) are from the near-isogenic lines (NILs) for glaucousness.



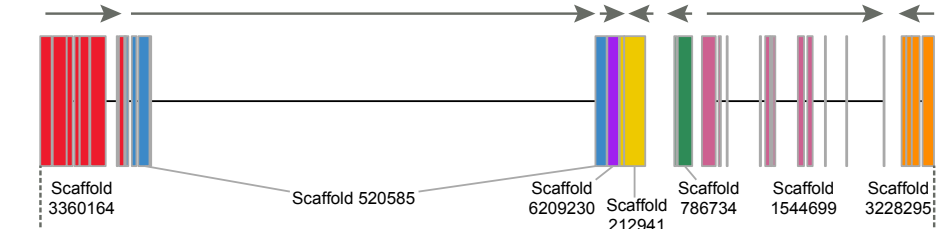
1) Compare *Aegilops tauschii* draft genomes: Map Jia et al. 2013 to UC Davis scaffolds.



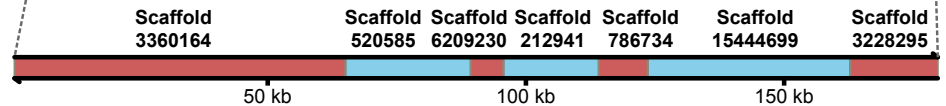
2) Since UC Davis scaffolds missing key part of Scaffold 82981, create pseudomolecule with Jia et al. scaffolds in *lw2* region.



3) Map W7984 synthetic wheat scaffolds to the *Aegilops* pseudomolecule.



4) Create pseudomolecule of *lw2* region in W7984 Synthetic wheat.



**Fig. S10.** Ordering of *Aegilops tauschii* and synthetic hexaploid wheat W7984 scaffolds to create the pseudomolecules used for mapping the *lw1* sequence to the location of *lw2* (Figure 5A-B).

**Table S1A. Parental lines used by Clark et al. (30, 31) in the development of four pairs of glaucous (G) and non-glaucous (NG) near-isogenic durum wheat lines as shown in Figure 1A.**

Genotype	Cross
AG1	Valgerardo (G) / Tibula (NG)
AG2	Valitalico (G) / Trinakria (NG)
AE3	Valitalico (G) / Trinakria (NG)
D051	Safra Man (PI283154) (NG) / DT369 (PI546362) (G)

**Table S1B. Identification of peaks listed in GC-MS spectrum in Figure 1B.**

Peak ID	Wax Component	Qualitative Presence / Amount	
		Glaucous	Non-Glaucous
1	18:3 Fatty acid (Linolenic acid)	Low	Y
2	Unknown	Y	No
3	C24-aldehyde	Low	Y
4	C24-primary alcohol	Y	Y
5	C27-alkane	Y	Y
6	C26-aldehyde	Low	Y
7	C26-primary alcohol	Low	Y
8	C29-alkane	Y	Y
9	Unknown	Y	Y
10	C29-alkan-2-ol	Y	No
11	C28-aldehyde	Low	Y
12	C28-primary alcohol	Low	Y
13	C29-12,14-dione	Y	No
14	C30-aldehyde	Y	Y
15	C29-ketone	Y	No
16	Unknown	Low	Y
17	C31-14,16-dione	Y	No
18	25-OH-C31-14,16-dione	Y	No
19	C37-alkan-2-ol	Y	No
20	C40 wax ester	Y	Y
21	C42 wax ester	Y	Y
22	C44 wax ester	Y	Y
23	C46 wax ester	Y	Y
24	C48 wax ester	Y	Y
25	C50 wax ester	Y	Y

**Table S2. Significant differentially-expressed (DE) unigenes between glaucous and non-glaucous *lw1* over-expressing transgenic plants.**

Unigene ID	<i>Ta#S61776684</i>	<i>Ta#S65655630</i>	<i>Ta#S65662797</i>	<i>Ta#S65693199</i>	<i>Ta#S16192919</i>
logFC	-5.48	-4.86	-4.86	-4.82	-3.26
logCPM	-1.85	0.81	4.06	2.10	3.96
PValue	1.35E-07	4.44E-17	2.20E-22	4.31E-17	5.43E-08
FDR	4.83E-03	2.64E-12	3.92E-17	2.64E-12	2.42E-03
Feature/ Scaffold	Traes4AL 4FE4DB56B.1	IWGSC-2BS scaff5155034	Traes2DS A38359A00	IWGSC-2BS scaff5155034	IWGSC-2BS scaff5155034
hit	7	2	3	3	5
E-Value	0	0	4.71E-134	1.44E-100	8.78E-85
Homolog	<i>W1-COE*</i>	<i>W1-COE</i>	<i>W2-COE</i>	<i>W1-COE</i>	<i>W1-COE</i>

Notes: Count analysis for unigenes was performed in CLC genomic workbench and significance analysis of counts was performed in edgeR. All DE unigenes were related to *W1-COE* or *W2-COE* sequences except for *Ta#S61776684* which had weak homology\* to *W1-COE*.

**Table S3. Significant differentially-expressed genes (DEGs) between glaucous and non-glaucous *lw1* overexpressing transgenic plants.**

DESeq2 Results	gene_ID	Traes_2BS_W1-COE
	baseMean	1668.54
	log2 FoldChange	-3.11
	lfcSE	0.25
	stat	-12.42
	pvalue	1.96E-35
	padjusted	1.59E-30
Kallisto Transcripts per Million (TPM)	Line 7279 (NG)	3.82
	Line 7281 (NG)	4.24
	Line 7282 (NG)	5.63
	Line 7285 (NG)	6.19
	Line 7287 (NG)	3.55
	Line 7289.2 (G)	93.62
	Line 7290 (G)	54.15
	Line 7293 (G)	96.37
	Line 7294 (G)	103.87
Average TPM	Glaucous Lines	87.0
	Non-Glaucous Lines	4.7

Note: Count analysis was performed in Kallisto and significance analysis in DESeq2. *W1-COE* was the only significant gene found. G represents a glaucous phenotype and NG a non-glaucous phenotype.

**Table S4. Primers used in this study.**

Name	Sequence	Purpose
T1S1	AAGGAAGTTTAAATTTATGAACTGCACGATGCTC	Cloning for VIGS
T1S2	AACCACCACCACCGTAGTTTGTGTGCAGCATGC	Cloning for VIGS
T1S3	AAGGAAGTTTAAATCCCGTAGCCAAATATGTCCTT	Cloning for VIGS
T1S4	AACCACCACCACCGTTGGATAAGCTGTGGCCGTTTCGT	Cloning for VIGS
T2S1	AAGGAAGTTTAAAGTTACTGCACCGAGACCAT	Cloning for VIGS
T2S2	AACCACCACCACCGTCTTCTTTGCTGCTGCTGC	Cloning for VIGS
T2S3	AAGGAAGTTTAAACCCTCATCACTCATGCGT	Cloning for VIGS
T2S4	AACCACCACCACCGTAGTGAAGCCTGGTCCAAAT	Cloning for VIGS
attB1_T1F1	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGCCTG- CAAACAAGACTTACCCCT	ORF cloning
attB2_T1R1	GGGGACCACTTTGTACAAGAAAGCTGGGTCTAGAAA- CAGTTCCTCATCACGAAT	ORF cloning
attB1_T2F1	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGCAGG- CAGCTCACCGAAGGTTAG	ORF cloning
attB2_T2R1	GGGGACCACTTTGTACAAGAAAGCTGGGTC- TATTTTTTCTTGAGAGCGCCGGTTG	ORF cloning
NG-M1F1	CCGTCTACATGCAACCAGCAAGAAGCT	Genome walking
NG-M1F2	AAGCATCGTGCAGTTCATAAATTATCACGCCA	Genome walking
NG-M1F3	CAGATCAATCCTTTGGACGAGGAGATCGT	Genome walking
NG-M1F4	CCTGGCGGAGAAGGACTTGGACAGAAA	Genome walking
NG-M1F5	CGAGGGGTTGGTCATTGTGCATCCATT	Genome walking
NG-M1F6	GACACTTAGCTCGCCAATACACAGACCT	Genome walking
NG-M1R1	CAGGTTTCGGCCAAGCCCTCTAGGAT	Genome walking
NG-M1R2	GTGGCGTGATAATTTATGAACTGCACGATGCT	Genome walking
NG-M1R3	ATGGATGCACAATGACCAACCCCTCGA	Genome walking
NG-M2F1	CTATCCTTGTGAGACCAGTTGTGCCGGT	Genome walking
NG-M2F2	CAAGAAGCCTATGTAGAGCATCCTGCGGT	Genome walking
NG-M2F3	AGGTGATATCAGAGGACTCGACCGAACC	Genome walking
NG-M2R1	GGTTCGGTCGAGTCCTCTGATATCACCT	Genome walking
NG-M2R2	CAGGATGCTCTACATAGGCTTCTTGCT	Genome walking
NG-M1R4	GGTGTGCACAGTGTGCTACATGATGTT	Genome walking
NG-M1R5	CCGGAGGTCTGTGTATTGGCGAGCTAA	Genome walking
NG-M1R6	GATGTTGCCACCAAGGAGGGGAACGT	Genome walking
NG-M1F7	CGGACCTCTGATTCTTCTCCGGTGGCT	Genome walking
NG-M1F8	CTTCTCCGGTGGCTGGTTTGGGTATTTA	Genome walking
NG-M1F9	CACCTTTCTCTGCTTTGTTGTAAGGCGGA	Genome walking
NG-M1F10	CGGCAATGTGATCTAGATTATTGACTCTAGT	Genome walking
NG-M1R7	ACTAGAGTCAATAATCTAGATCACATTGCCG	Genome walking
NG-M1R8	TAAATACCCAAACCAGCCACCGGAGAAGA	Genome walking
NG-M1F11	CGACAAAGAAGATGTCCTGAGTTGGTCA	Genome walking



**Table S4 Continued. Primers used in this study.**

Name	Sequence	Purpose
NG-M1F12	CGGAATAAACTATCACCCGGGTAAAGCT	Genome walking
NG-M1R9	GCAACCACATTAGCTTTACCCGGGTGAT	Genome walking
NG-M1R10	CCGATGTCATAATCCTTGACCAACTCAGGA	Genome walking
SC350	TCTCCAGAAGAAGATGCAGGA	sequencing VIGS constructs
BS10	GGTGCTTGATGCTTTGGATAAGG	sequencing VIGS constructs
BS32	TGGTCTTCCCTTGGGGGAC	sequencing VIGS constructs
M2F2	CAAGAAGCCTATGTAGAGCATCCTGCGGT	RACE3'end
M1R1	CAGGTTTCGGCCAAGCCCTCTAGGAT	RACE3'end
M1R2	GTGGCGTGATAATTTATGAACTGCACGATGCT	RACE3'end
M2R1	GGTTCGGTCGAGTCCTCTGATATCACCT	RACE 5'end
M2R2	CAGGATGCTCTACATAGGCTTCTTGCT	RACE 5'end
RACE5_6	CAAGGATAGTACAAGGACAGAGCCCA	RACE 5'end
RACE5_7	AGAACGCACTCTCCCTTTAGAACTGG	RACE 5'end
Gpri_F1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATCA-GAGGTCCGCCTTACAACAAAGCA	Full-length lw1 cloning (Gateway)
Gpri_R1	GGGGACCACTTTGTACAAGAAAGCTGGGTATGACG-GTAAGCAGGTGTGGTGGTAGTT	Full-length lw1 cloning (Gateway)
pri_F1	AATCAGAGGTCCGCCTTACAACAAAGCA	Full-length lw1 cloning
pri_R1	TGACGGTAAGCAGGTGTGGTGGTAGTT	Full-length lw1 cloning
T1_q1	AGTAGGCCATTTAAGGACATATTTG	qPCR target1
T1_q2	CTCATCACGAATCTTACCGAAGTT	qPCR target1
T2_q1	GAACAAGGTGTCCCAAGAAGAGTA	qPCR target1
T2_q2	TGCCCCGTTAGACCACATATTATCT	qPCR target2
CDC48_q3	GCTTATCTACATCCCTCTTCCTGA	qPCR reference gene
CDC48_q4	GACCTCATCCACCTCATCCTC	qPCR reference gene
T3_q9	ATCATGTTTCATCATCCTCCAAG	qPCR target3
T3_q10	CGAAGAAGTCCATATTTCTCTCC	qPCR target3
T4_q3	ATAGCGACCATCATCAAGAGTTTC	qPCR target4
T4_q4	AGTTTATAGATGTAGCGGGAGGA	qPCR target4
T5_q5	ATGGAGAGGAGGAAGATGAATAC	qPCR target5
T5_q6	CTGGGCGAGTTTGTAGAAGAT	qPCR target5
T6_q1	AAGTTTCTCCTTCTCCCTGTCC	qPCR target6
T6_q2	TTGTTCCGTATGTGAATTATACCCA	qPCR target6
T7_q3	ACTCGTTAGTCGGTCAAATG	qPCR target7
T7_q4	CACATATTACAGGGATCGGAAGA	qPCR target7
CRACE1	GAAACAGTTCCTCATCACGAATCTT	cleavage 5'RACE
CRACE2	CAGTTCCTCATCACGAATCTTAC	cleavage 5'RACE