Supporting Information (SI) Appendix

A novel long-noncoding MIRNA represses wheat β-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 $\mu mol\, m^2\, s^{-1}$, 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 μL 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) 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 ($>C_{14}$) to wax esters ($>C_{52}$), 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-β-diketones, 50 μL aliquots of each sample were transferred to fresh glass GC vial inserts (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 N2 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 AABBDD genome transcripts with the addition of the *W1-COE* cDNA and AB-W1-COE, consisting of AABB 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 <= 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 <= 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 Jemboss 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-α, pCaBS-β, and pCa-γbLIC, was used with modifications. The pCa-γbLIC vector was modified by introducing ccdB gene for efficient selection of recombinant clones. The resulting vector pCa-γbLIC-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-γbLIC-ccdB vector by using a LIC strategy. The resulting clones were sequencing verified. The plasmid DNAs of pCaBS-α, pCaBS-β, and pCa-γbLIC with target gene fragments were transformed into *Agrobacterium tumefaciens* strain c58. *A. tumefaciens* strains harboring pCa-γbLIC 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 *Iw1* 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 *Iw1* 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 μM 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://wwwR-projectorg/
11. RStudio Team (2015) RStudio: Integrated Development for R RStudio, Inc, Boston, MA http://wwwrstudiocom/
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 022 https://CRANR-projectorg/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://wwwbioconductororg/packages/release/bioc/html/Gvizhtml
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://bioconductororg/packages/release/bioc/html/Rsamtoolshtml
29. Pages H, Aboyoun P, Gentleman R, DebRoy S (2016) Biostrings: String objects representing biological sequences, and matching algorithms R package version 2384, https://wwwbioconductororg/packages/release/bioc/html/Biostringshtml
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 *Iw2* 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.

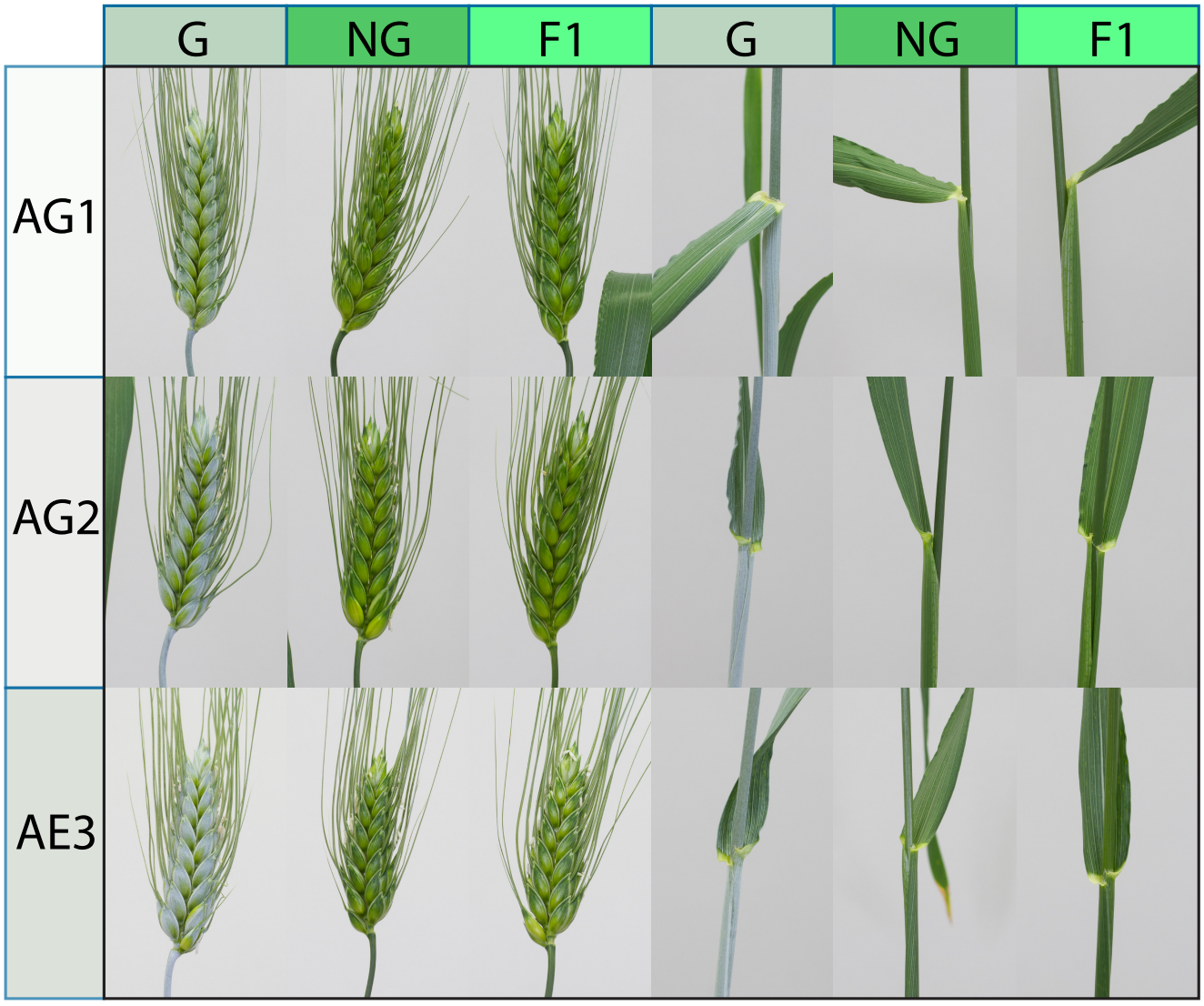

Fig. S1. Dominance of the non-glaucous (NG) trait. Crosses between glaucous (G) and non-glaucous (NG) isolines of AG1, AG2, and AE3 resulted in F1 heterozygous plants that were non-glaucous. The three columns of the left show representative heads from each line and the three right hand columns show representative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

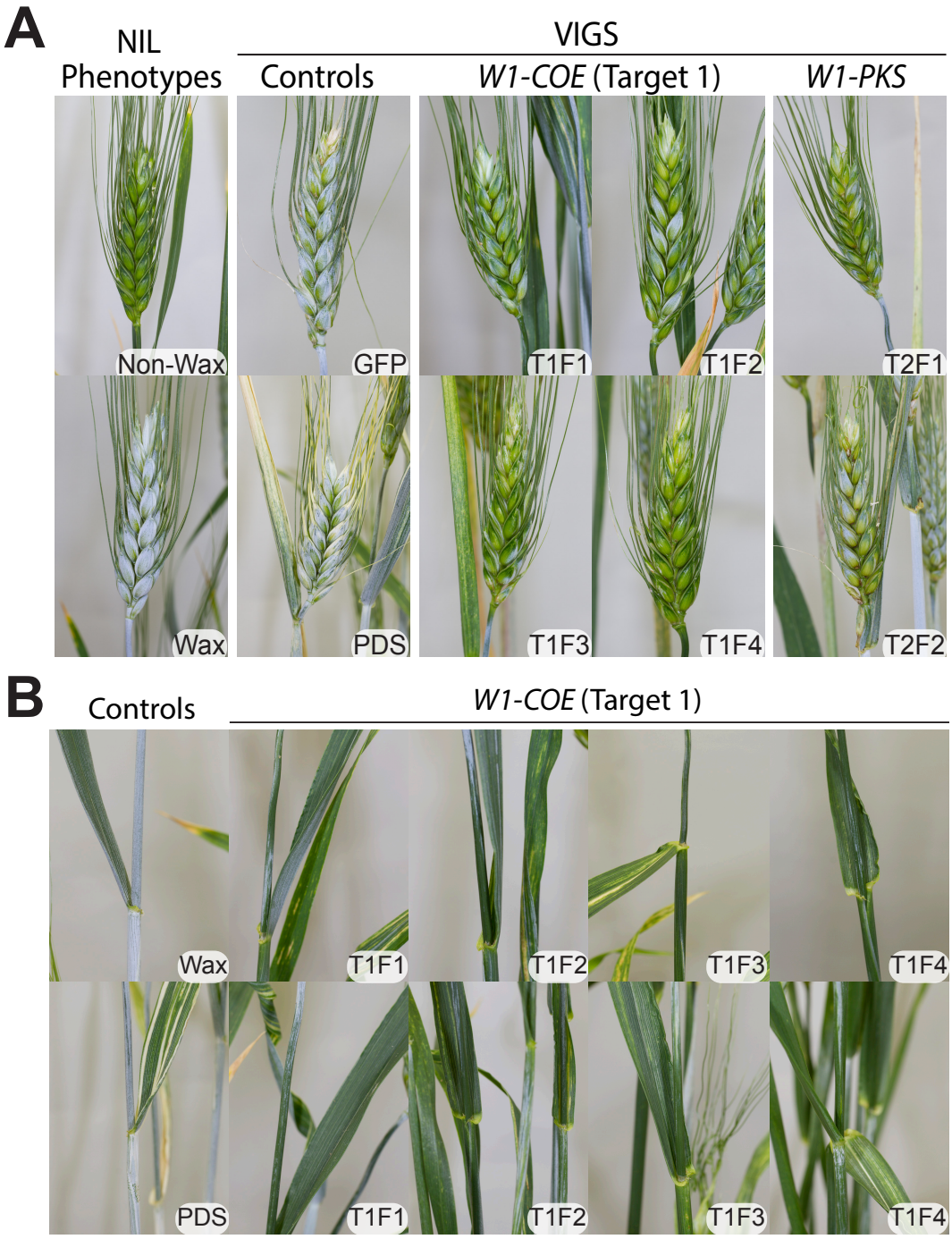

Fig. S2. Reduction in glaucousness resulting from Virus Induced Gene Silencing (VIGS). (*A*) VIGS using DNA fragments (F1, F2 etc.) of target genes 1 (T1F1-T1F4) and 2 (T2F1-T2F2), representing *W1-COE* and *W1-PKS* respectively, results in decreased glaucousness on wheat heads. Controls include non-VIGS-treated 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.

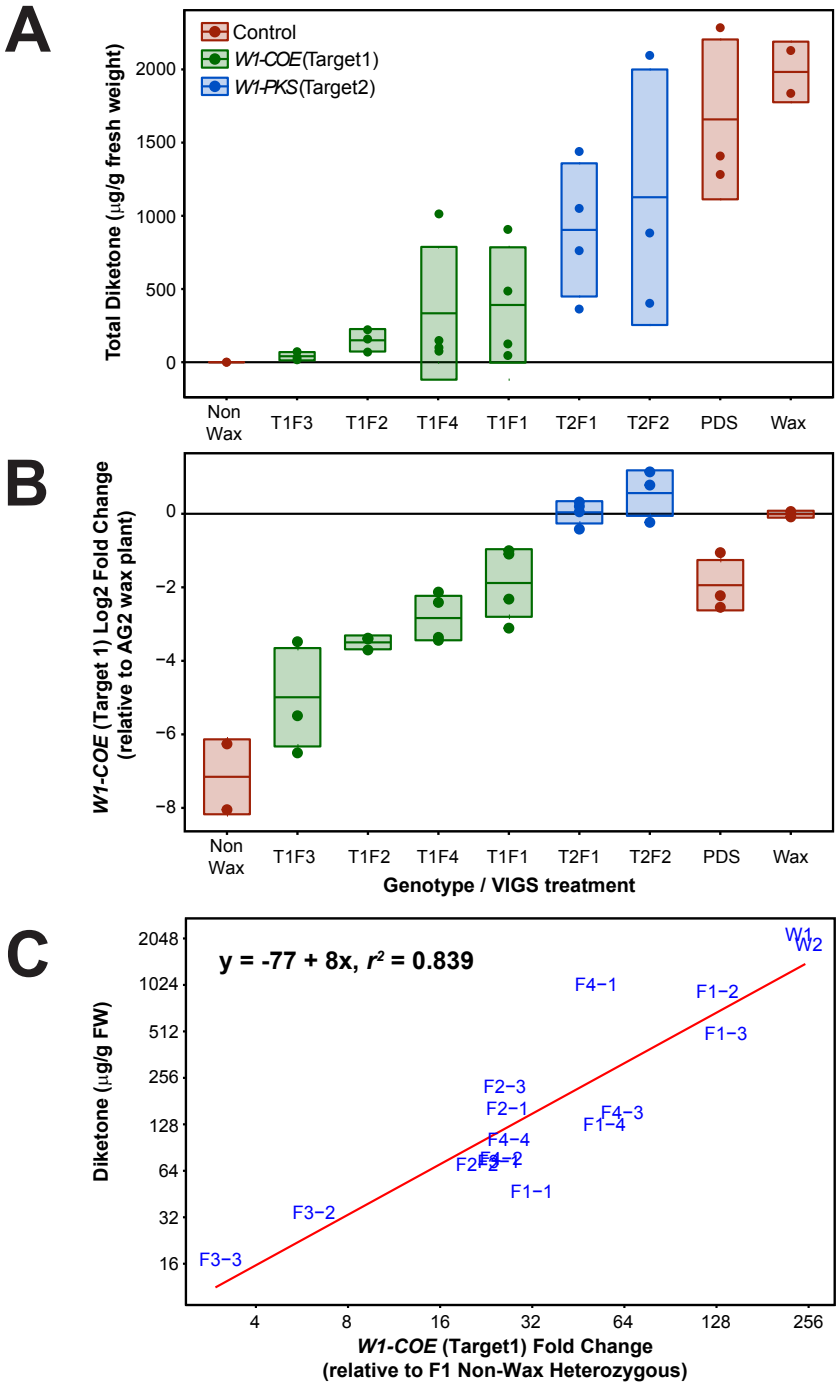

Fig. S3. Reduction of diketones in leaf sheaths correlates with reductions in *W1-COE* (Target 1) gene expression during VIGS suppression in AG2 NILs. (*A*) Total diketone, including hentriacontane-14,16-dione (β-diketone) and 25-hydroxy-β-diketone, are reduced after VIGS treatments of target 1 and 2 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 fragements 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 ± standard deviation with data points indicated.

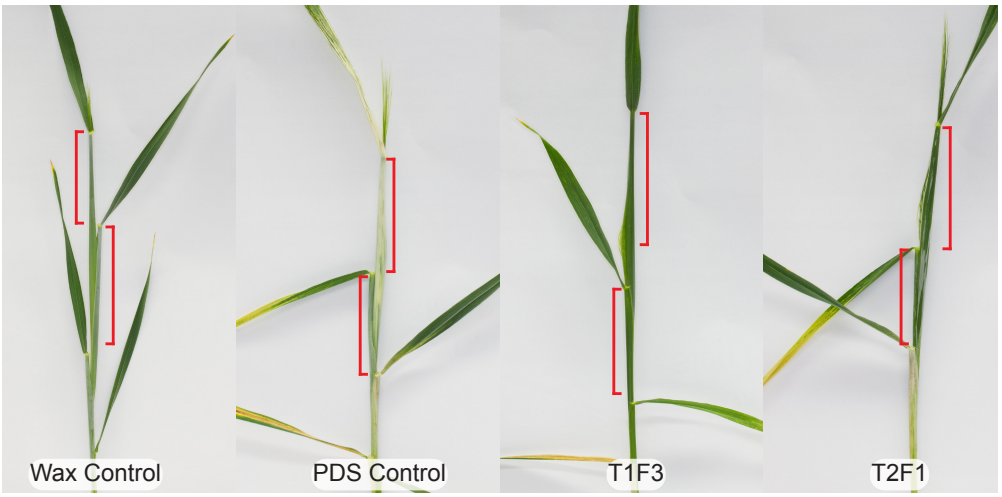

Fig. S4. Leaf sheaths from the flag leaf and penultimate leaf (shown in red) were used for wax and gene expression analyses. Examples for the wax and PDS controls and target 1 fragment 3 and target 2 fragment 1 are depicted as examples.

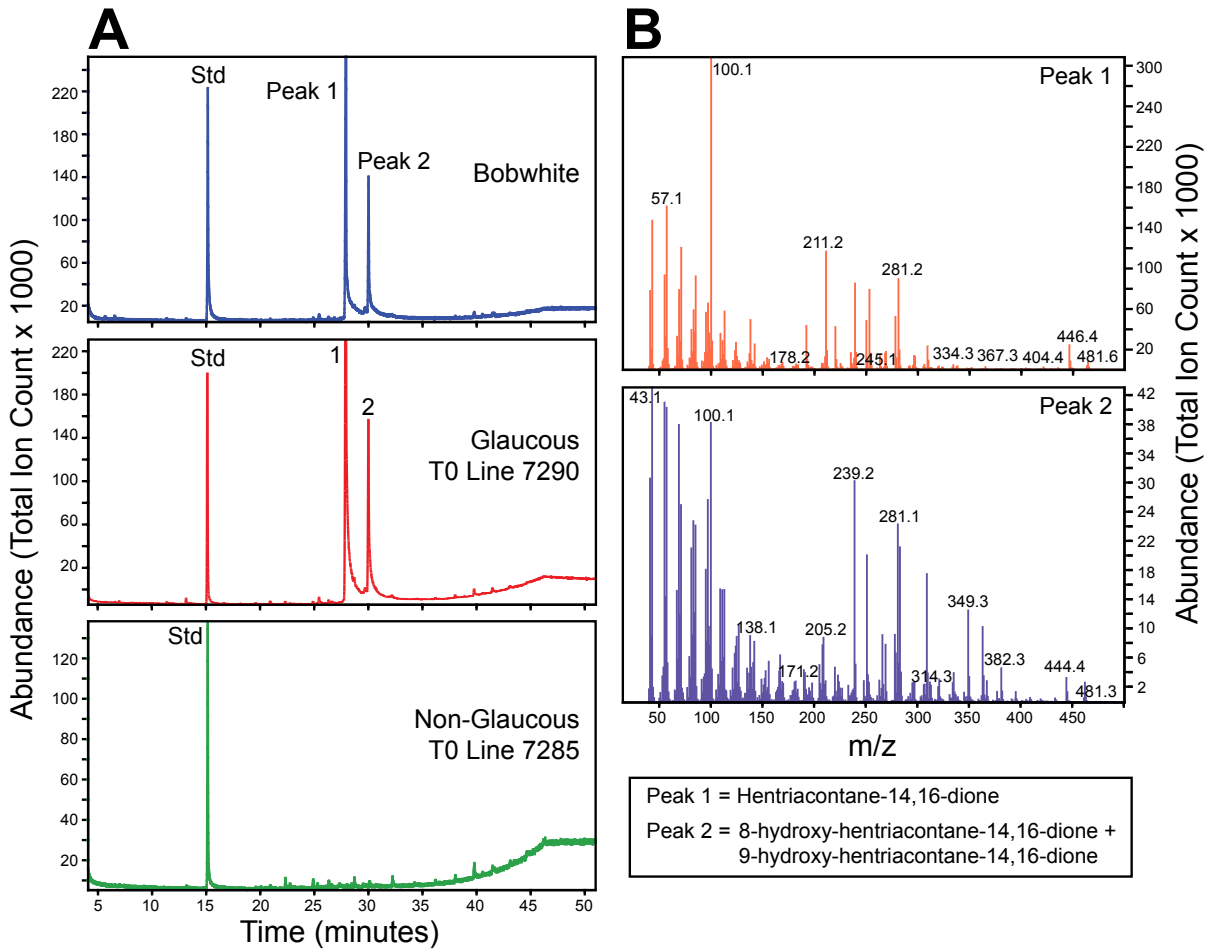

Fig. S5. Characterization of diketone waxes in flag leaf sheaths of *lw1* over-expressing (OE) lines. (*A*) Total Ion Count (TIC) spectra of cuticular wax from leaf sheath of untransformed spring wheat cultivar Bobwhite and primary *lw1* OE transformants (T0 generation), as shown in Figure 4. Std = C_{24} 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).

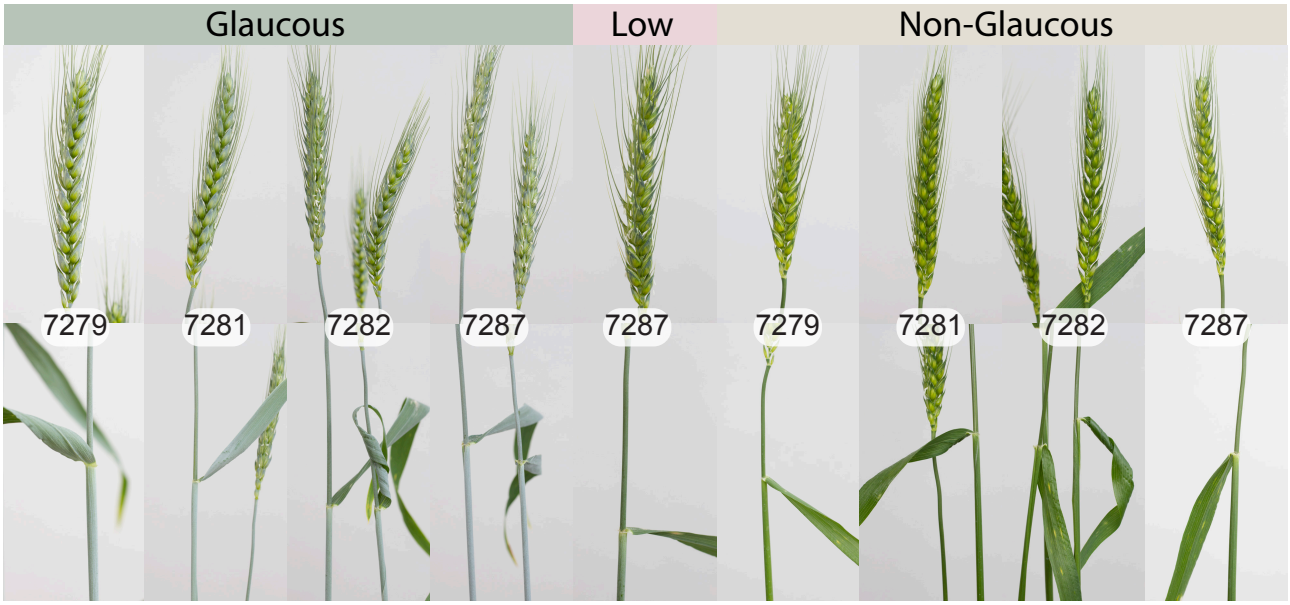

Fig. S6. Segregation of the glaucous trait in *Iw1* over-expression lines in the T1 generation. In addition to clearly glaucous and non-glaucous plants, a few showed an intermediate low glaucous phenotype; line 7287 is shown as an example. Heads after anthesis and stems surrounding the flag leaf are shown.

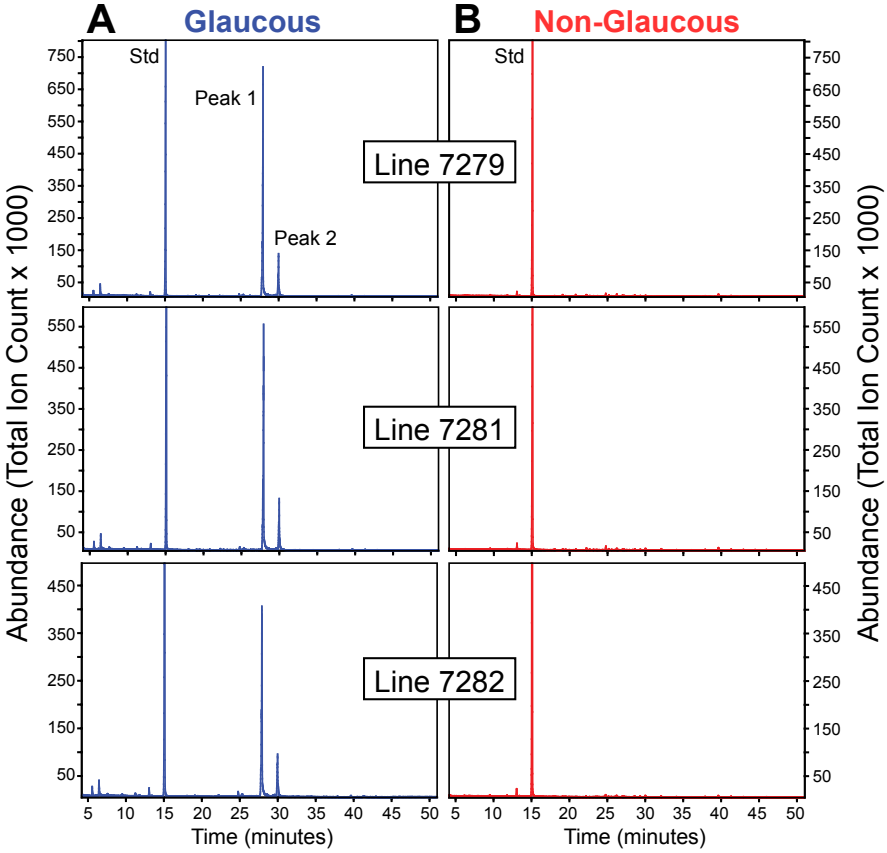

Fig. S7. Total Ion Count (TIC) spectra of cuticular wax from flag leaf sheaths of glaucous (*A*) and non-glaucous (*B*) T1 plants of lines 7279, 7281 and 7282 segregating for *Iw1* over-expression. Std = C_{24} standard added. Peak 1 = Hentriacontane-14,16-dione (C31-14,16-dione); Peak 2 = 8-hydroxy-hentriacontane-14,16-dione and 9-hydroxy-hentriacontane-14,16-dione (8-OH-C31-14,16-dione and 9-OH-C31-14,16-dione).

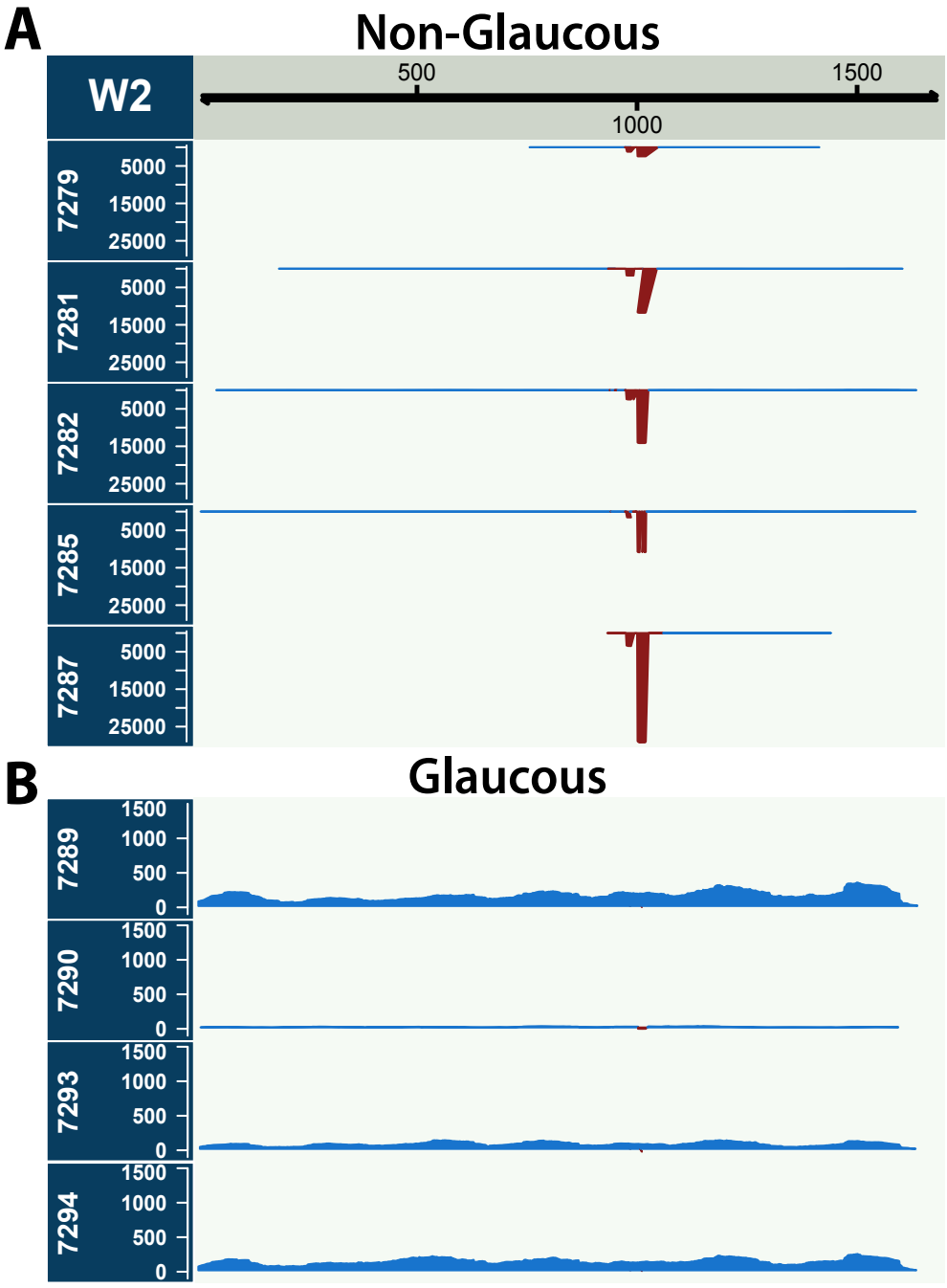

Fig. S8 Mapping of RNA and small RNA sequencing reads to *W2-COE* in T0 transgenic lines over-expressing *Iw1* (miRW1 precursor). (*A*) Non-glaucous lines displayed expression of sRNA (shown in red) derived from *Iw1* and repression of *W2-COE* (blue). (B) In contrast, glaucous lines showed almost no expression of sRNA from *Iw1* and transcription of *W2-COE*. Expression of *W2-COE* and its higher expressed paralog *W1-COE* (Fig. 4) would allow β-diketone synthesis and deposition to occur and thus the presence of a glaucous state.

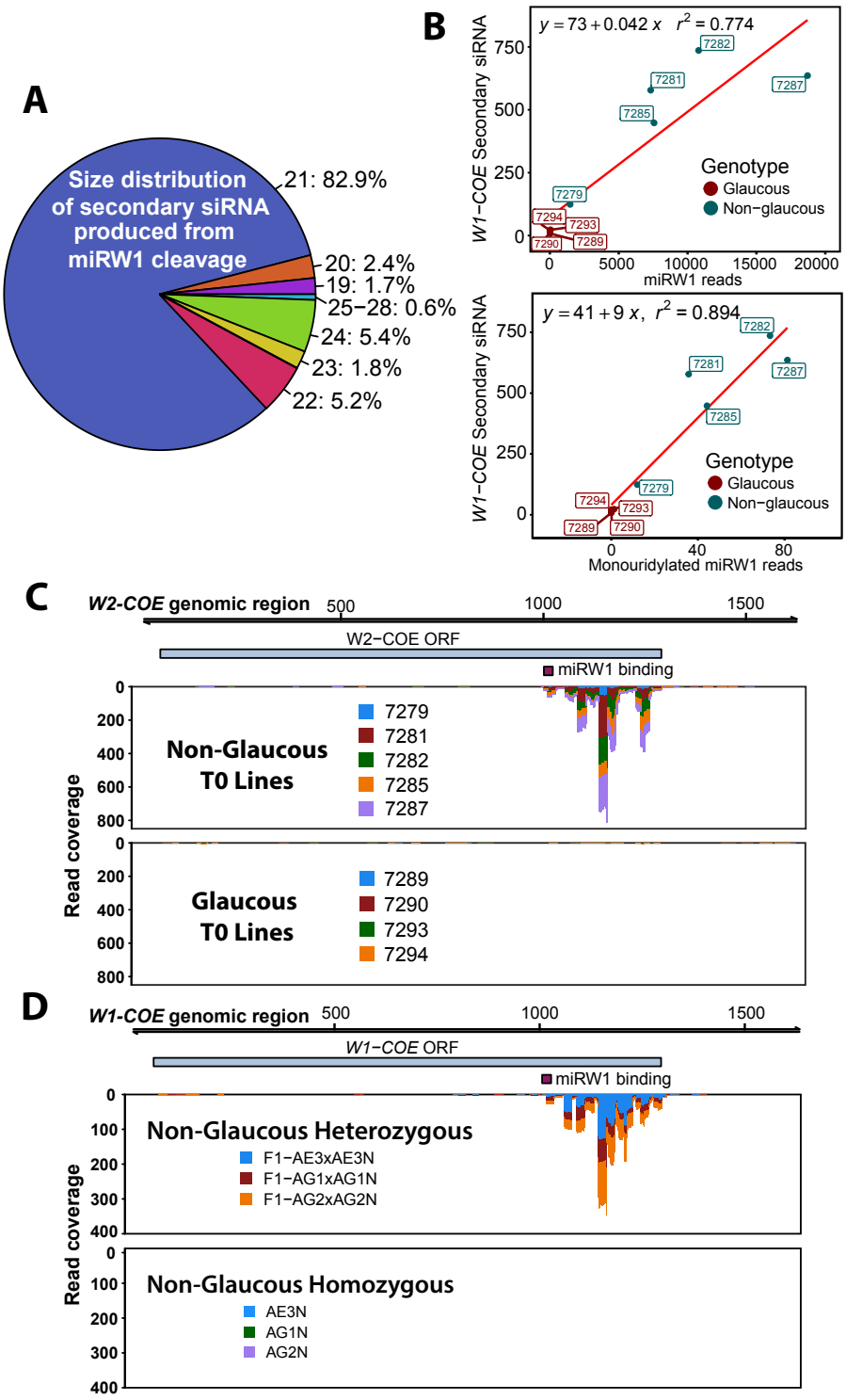

Fig. S9 Production of secondary siRNAs from *W1-COE* and *W2-COE*. (*A*) Small RNA size distribution of secondary siRNA arising from *W1-COE.* (*B*) The relationship between *W1-COE* secondary siRNA and miRW1 or monouridylated miRW1 reads. Reads are normalized per 10 million. (*C*) Secondary siRNA mapping 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 *Iw1*. (*A-C*) are data from miRW1 percursor (*Iw1*) over-expression lines while (*D*) are from the near-isogenic lines (NILs) for glaucousness.

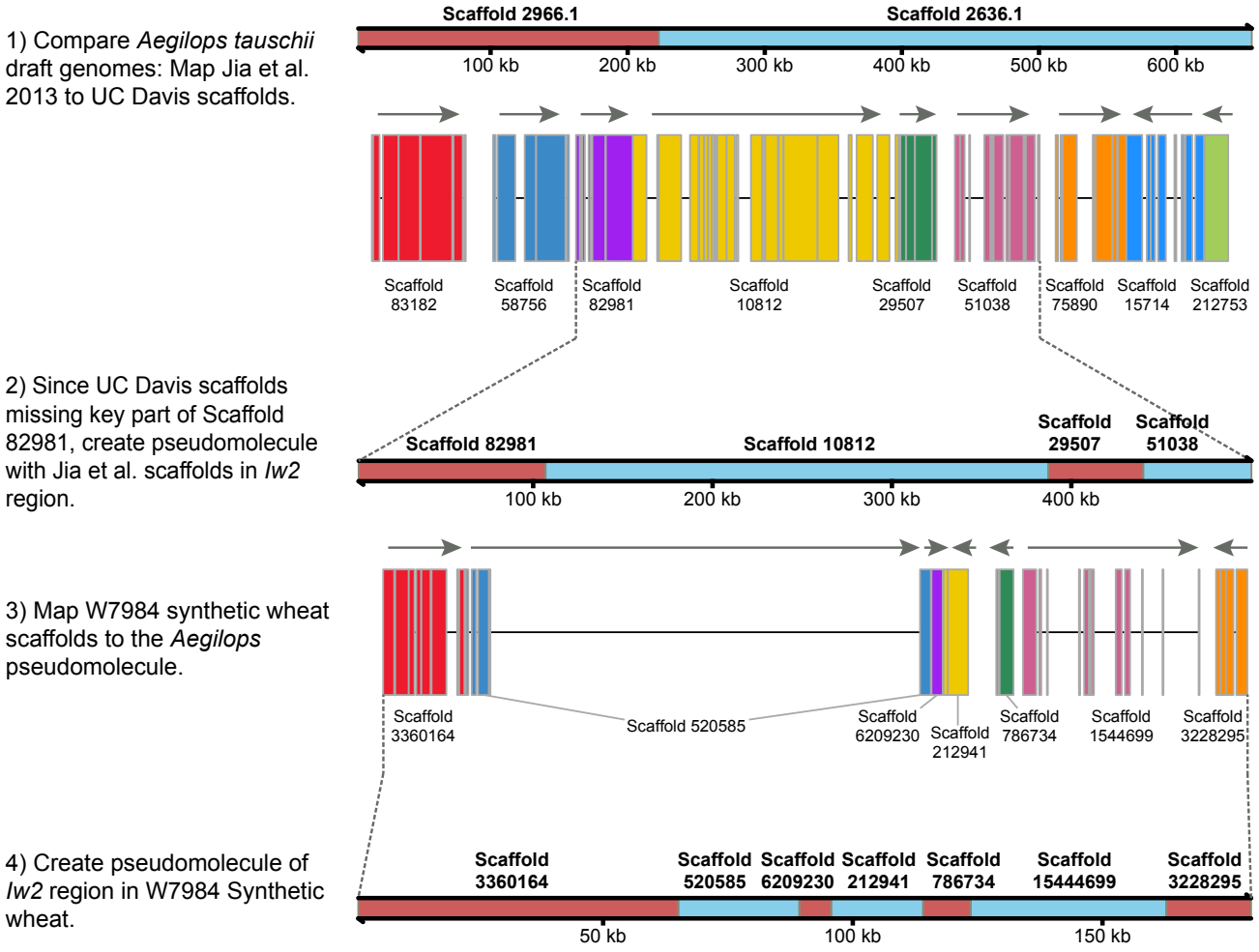

Fig. S10. Ordering of *Aegilops tauschii* and synthetic hexaploid wheat W7984 scaffolds to create the pseudomolecules used for mapping the *Iw1* sequence to the location of *Iw2* (Figure 5*A*-*B*).

Table S1*A*. 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 1*A*.

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 S1*B*. Identification of peaks listed in GC-MS spectrum in Figure 1*B*.

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-adehyde	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	*Ta#S61776684*	*Ta#S65655630*	*Ta#S65662797*	*Ta#S65693199*	*Ta#S16192919*
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	*W1-COE**	*W1-COE*	*W2-COE*	*W1-COE*	*W1-COE*

Notes: Count analysis for unigenes was performed in CLC genomic workbench and significnace 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.

	gene_ID	**Traes_2BS_W1-COE**
DESeq2 Results	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	AAGGAAGTTTAATTTATGAACTGCACGATGCTC	Cloning for VIGS
T1S2	AACCACCACCACCGTAGTTTGTGTGCAGCATGC	Cloning for VIGS
T1S3	AAGGAAGTTTAAATCCCGTAGCCAAATATGTCCTT	Cloning for VIGS
T1S4	AACCACCACCACCGTTGGATAAGCTGTGGCCGTTCGT	Cloning for VIGS
T2S1	AAGGAAGTTTAAGTTACTGCACCGAGACCAT	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	CAGGTTCGGCCAAGCCCTCTAGGAT	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	CAGGTTCGGCCAAGCCCTCTAGGAT	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 Iw1 cloning (Gateway)
Gpri_R1	GGGGACCACTTTGTACAAGAAAGCTGGGTATGACG-GTAAGCAGGTGTGGTGGTAGTT	Full-length Iw1 cloning (Gateway)
pri_F1	AATCAGAGGTCCGCCTTACAACAAAGCA	Full-length Iw1 cloning
pri_R1	TGACGGTAAGCAGGTGTGGTGGTAGTT	Full-length Iw1 cloning
T1_q1	AGTAGGCCATTTAAGGACATATTTG	qPCR target1
T1_q2	CTCATCACGAATCTTACCGAAGTT	qPCR target1
T2_q1	GAACAAGGTGTCCCAAGAAGAGTA	qPCR target1
T2_q2	TGCCCGTTAGACCACATATTATCT	qPCR target2
CDC48_q3	GCTTATCTACATCCCTCTTCCTGA	qPCR reference gene
CDC48_q4	GACCTCATCCACCTCATCCTC	qPCR reference gene
T3_q9	ATCATGTTCATCATCCTCCAAG	qPCR target3
T3_q10	CGAAGAAGTCCATATTTCCTCC	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