PNAS
www.pnas.org

Supplementary Information for:

A co-opted steroid synthesis gene, maintained in sorghum but not maize, is associated with a divergence in leaf wax chemistry.

Author names and affiliations:

Lucas Busta[1,2,5,6*], Elizabeth Schmitz[1,2,6], Dylan K. Kosma[3], James C. Schnable[2,4], and Edgar B. Cahoon[1,2*]

[1]Department of Biochemistry, University of Nebraska – Lincoln, Lincoln, Nebraska, 68588, USA

[2]Center for Plant Science Innovation, University of Nebraska – Lincoln, Lincoln, Nebraska, 68588, USA

[3]Department of Biochemistry and Molecular Biology, University of Nevada, Reno, Nevada 89557 USA

[4]Department of Agronomy and Horticulture, University of Nebraska – Lincoln, Lincoln, NE 68583 USA

[5]Current address: Department of Chemistry and Biochemistry, University of Minnesota Duluth – Duluth, MN, 55812 USA

[6]Co-first authors

*Corresponding authors Lucas Busta and Edgar B. Cahoon

Corresponding Author Email: lucasbusta1@gmail.com

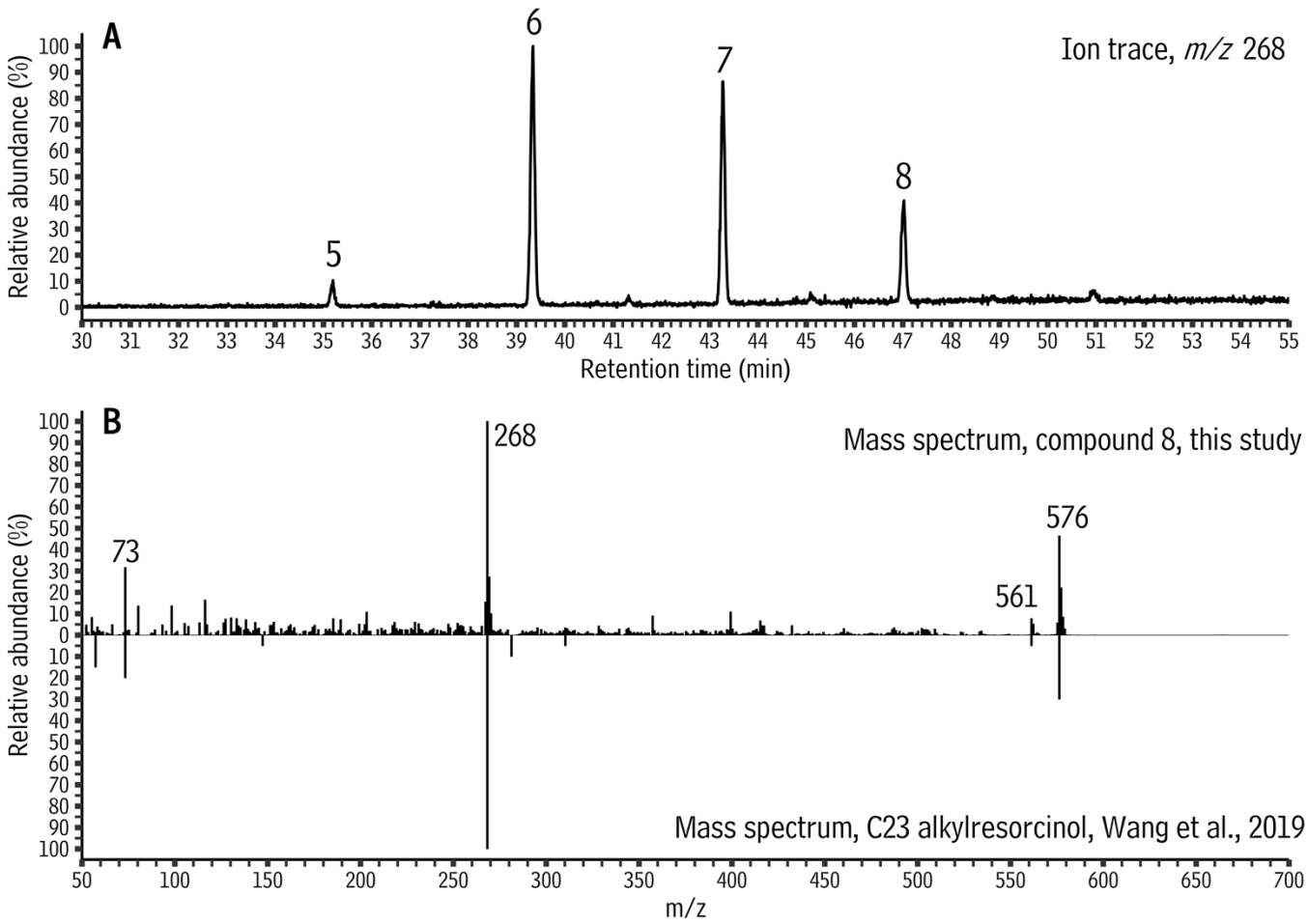
ORCID: 0000 0002 0102 9986

This PDF file includes:

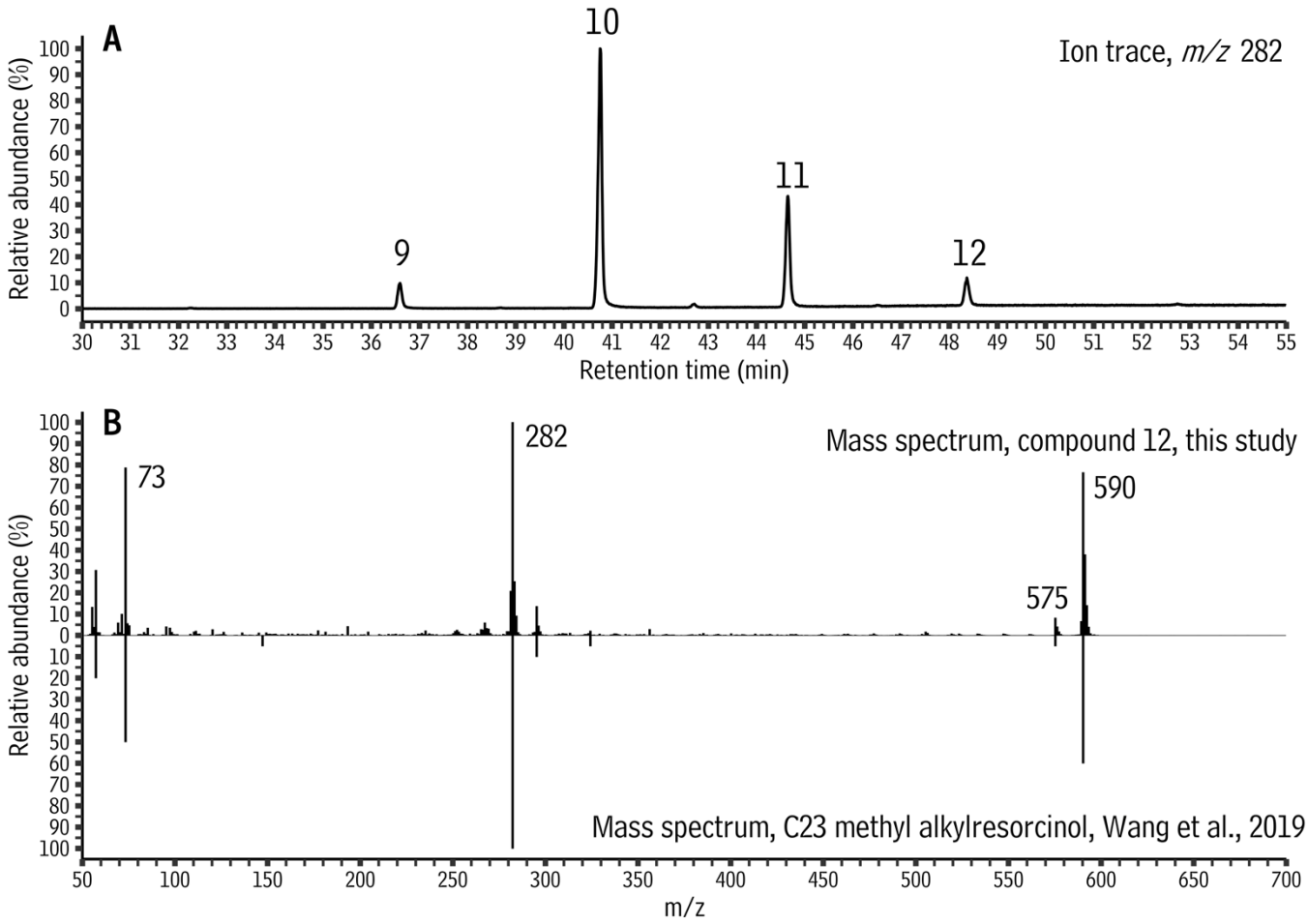
Figures S1 to S15

Other supplementary materials for this manuscript include the following:

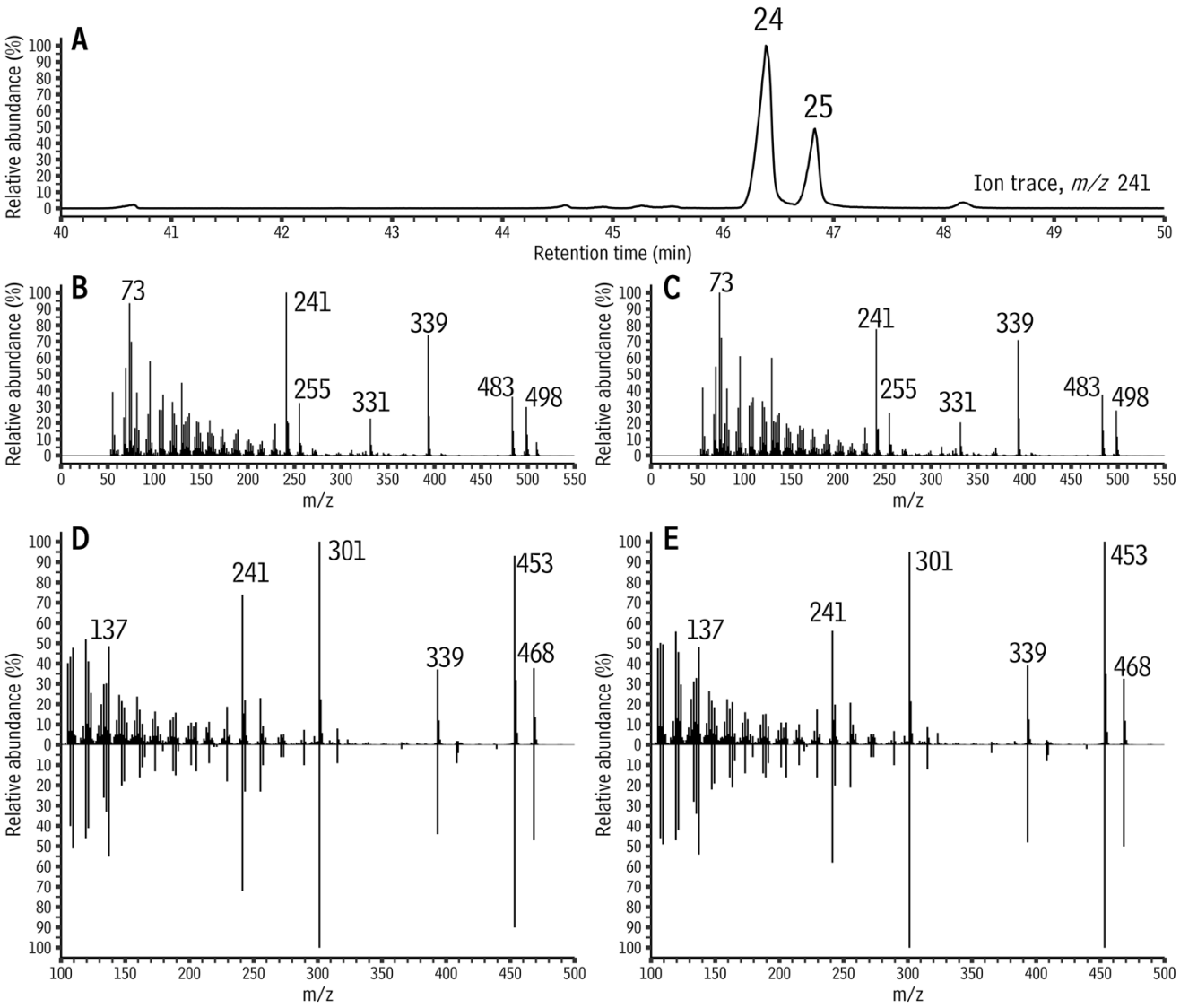
Supplementary Datasets S1 to S5



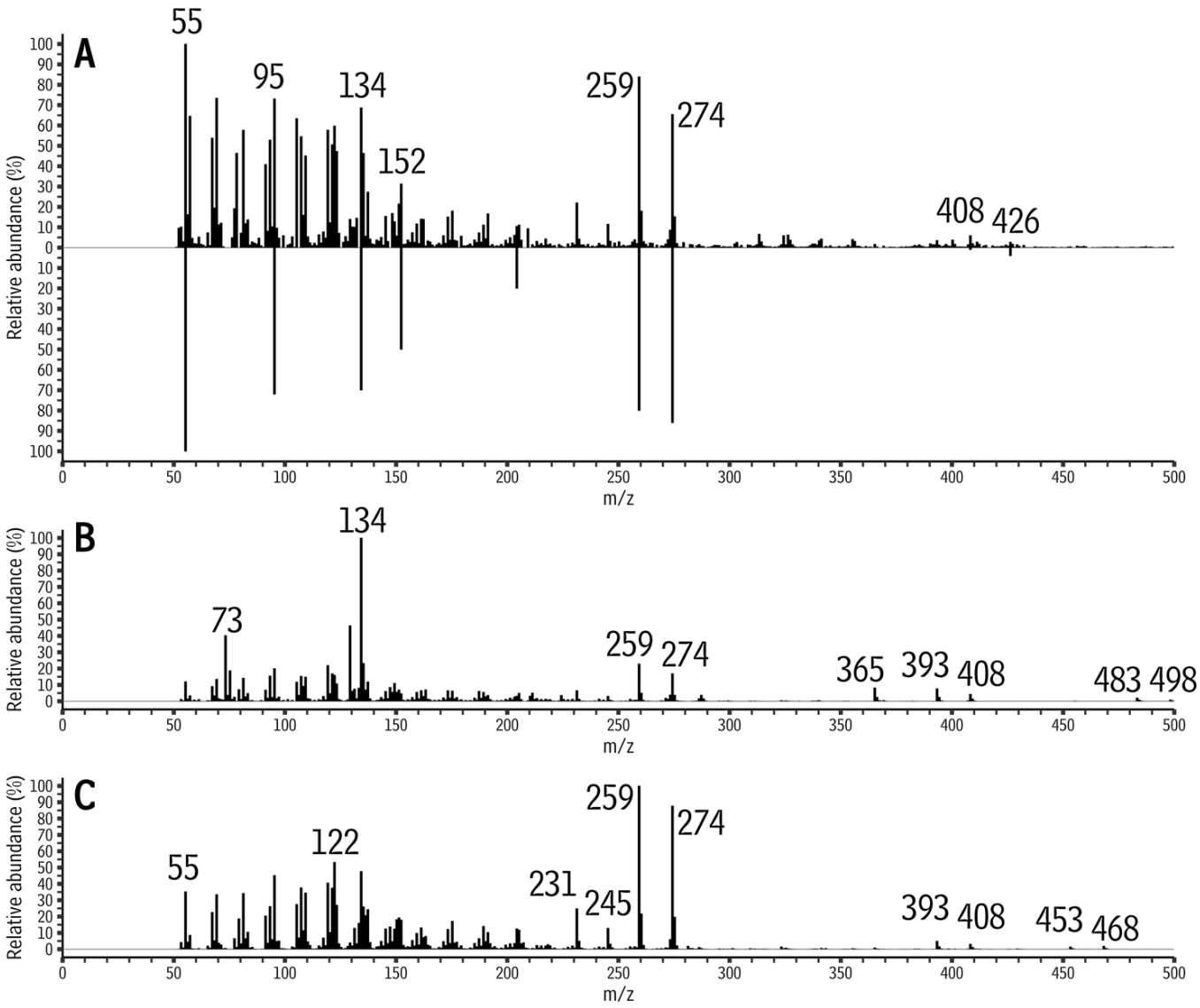
Supplemental Figure S1: A. Gas chromatographic trace of *m/z* 268 from TLC fraction R_f 0.21. **B.** Top: mass spectrum of compound 8 from A after trimethylsilyl derivatization, bottom: published mass spectrum of a 23:0 alkylresorcinol after trimethylsilyl derivatization.



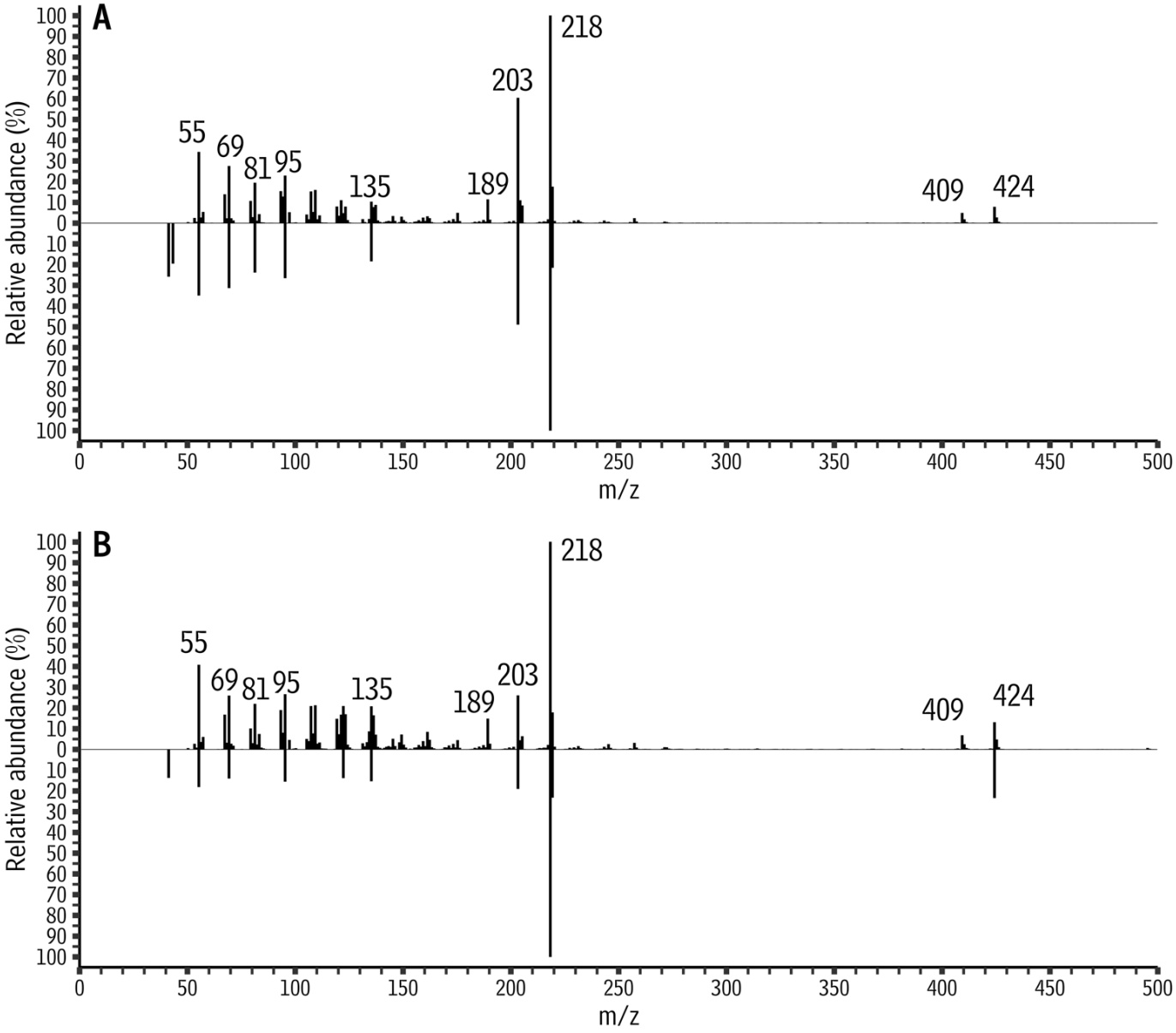
Supplemental Figure S2: A. Gas chromatographic trace of *m/z* 282 from TLC fraction R_f 0.36. **B.** Top: mass spectrum of compound 12 from A after trimethylsilyl derivatization, bottom: published mass spectrum of a 23:0 methyl alkylresorcinol after trimethylsilyl derivatization.



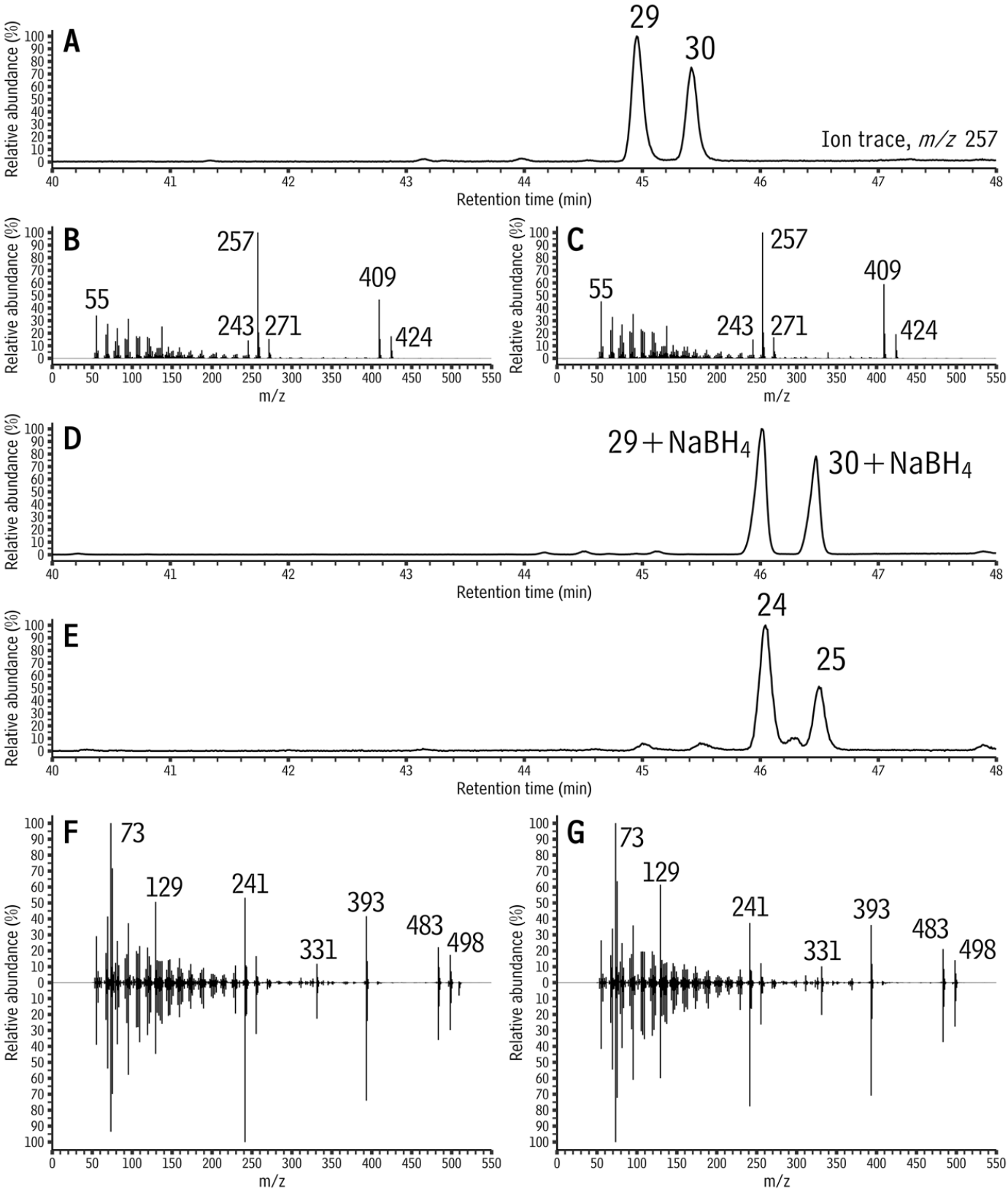
Supplemental Figure S3: A. Gas chromatographic trace of *m/z* 241 from TLC fraction R_f 0.55. **B.** Mass spectrum of compound 22 from A after trimethylsilyl derivatization. **C.** Mass spectrum of compound 23 from A after trimethylsilyl derivatization. **D.** Top: mass spectrum of compound 22 after acetylation, bottom: published mass spectrum of fernenol after acetylation. **E.** Top: mass spectrum of compound 23 after acetylation, bottom: published mass spectrum of isoarborinol after acetylation.



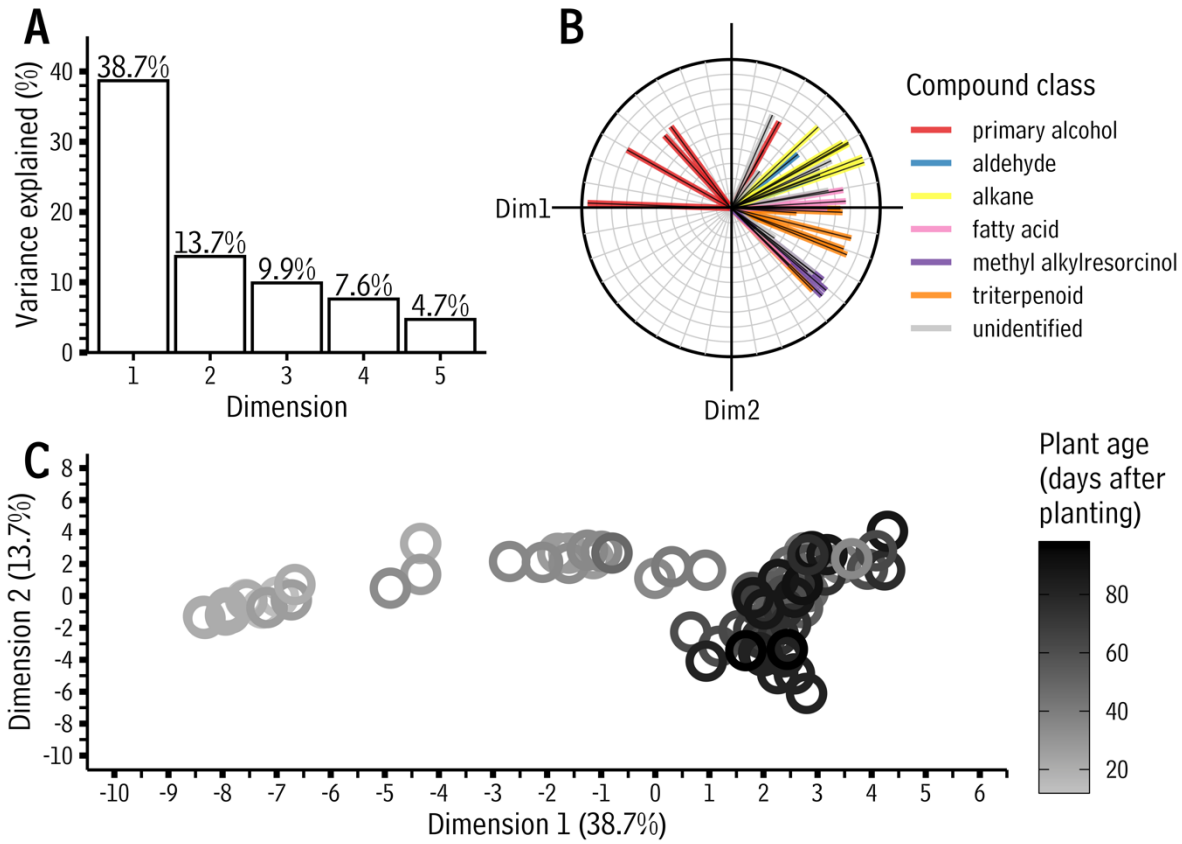
Supplementary Figure S4: A. Top: mass spectrum of compound 26 from TLC fraction R_f 0.70, bottom: published mass spectrum of simiarenol, a previously reported sorghum leaf wax compound. **B.** Mass spectrum of compound 26 after trimethylsilyl derivatization. **C.** Mass spectrum of compound 26 after acetylation.



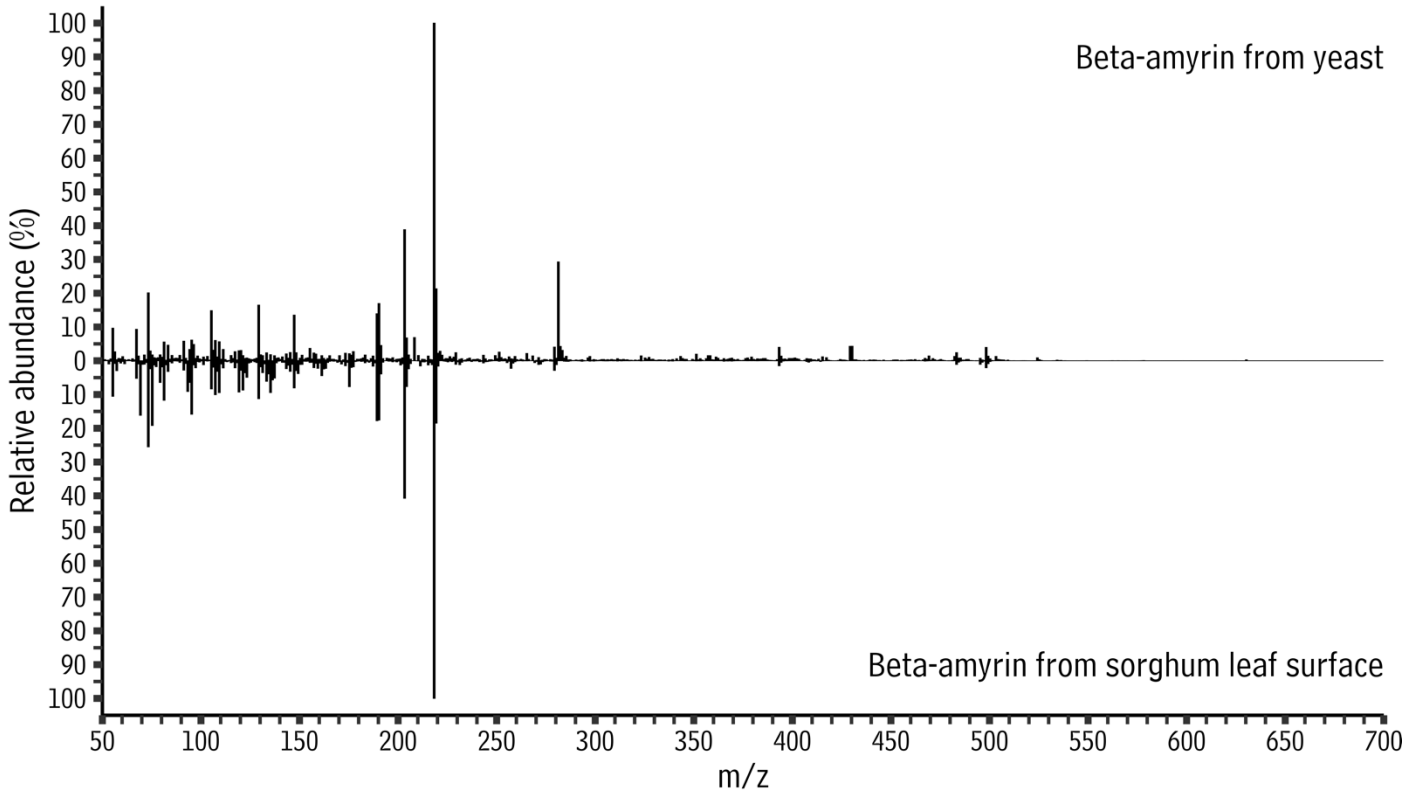
Supplementary Figure S5: A. Top: mass spectrum of compound 27 after trimethylsilyl derivatization, bottom: published mass spectrum of β-amyrone. **B.** Top: mass spectrum of compound 28 after trimethylsilyl derivatization, bottom: published mass spectrum of α-amyrone.



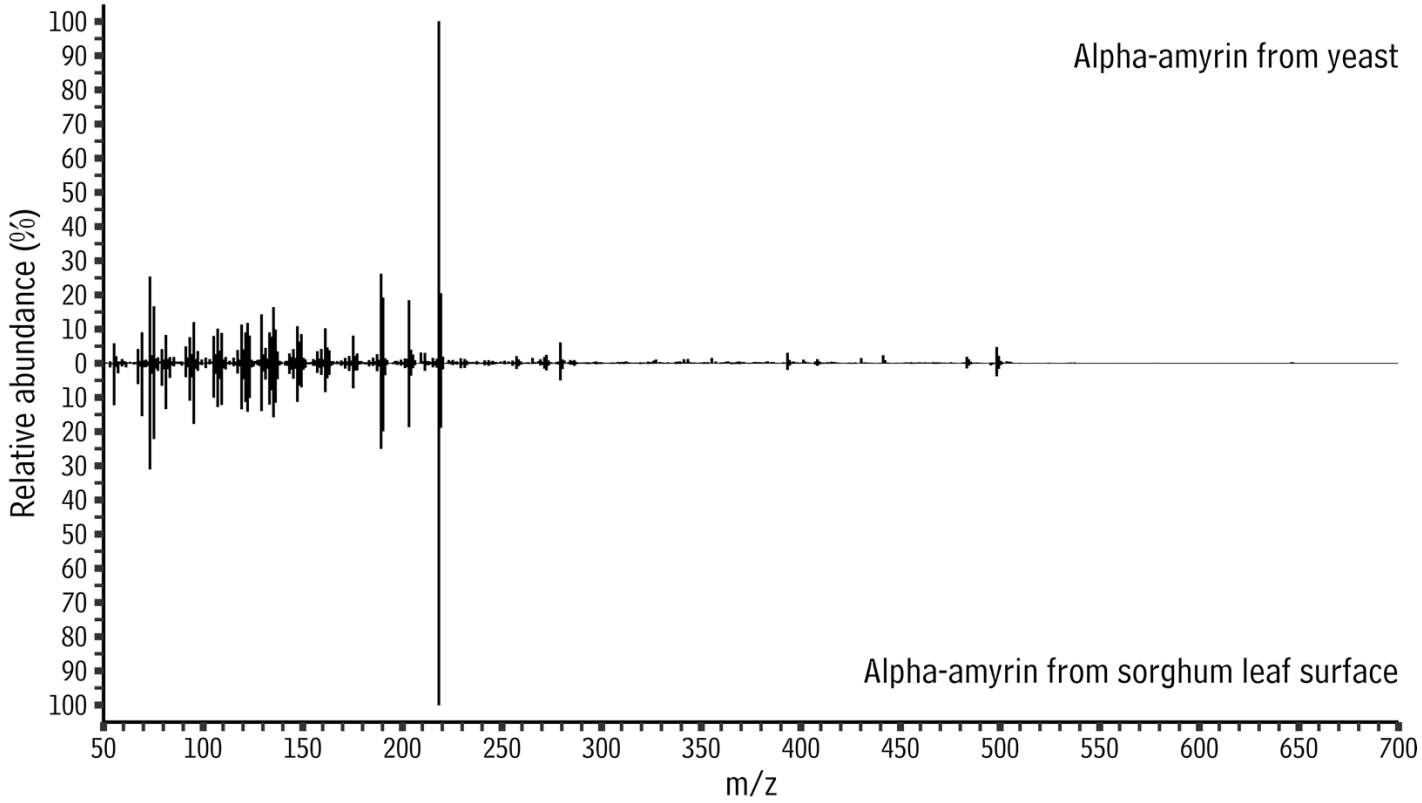
Supplementary Figure S6: A. Gas chromatographic trace of *m/z* 257 from TLC fraction R_f 0.80. **B.** Mass spectrum of compound 29 after trimethylsilyl derivatization. **C.** Mass spectrum of compound 30 after trimethylsilyl derivatization. **D.** Gas chromatographic trace of *m/z* 241 from TLC fraction R_f 0.80 after treatment with sodium borohydride. **E.** Gas chromatographic trace of *m/z* 241 from TLC fraction R_f 0.55. **F.** Top: mass spectrum of compound 29 after treatment with sodium borohydride and trimethylsilyl derivatization, bottom: mass spectrum of fernenone. **G.** Top: mass spectrum of compound 30 after treatment with sodium borohydride and trimethylsilyl derivatization, bottom: mass spectrum of isoarborinone.



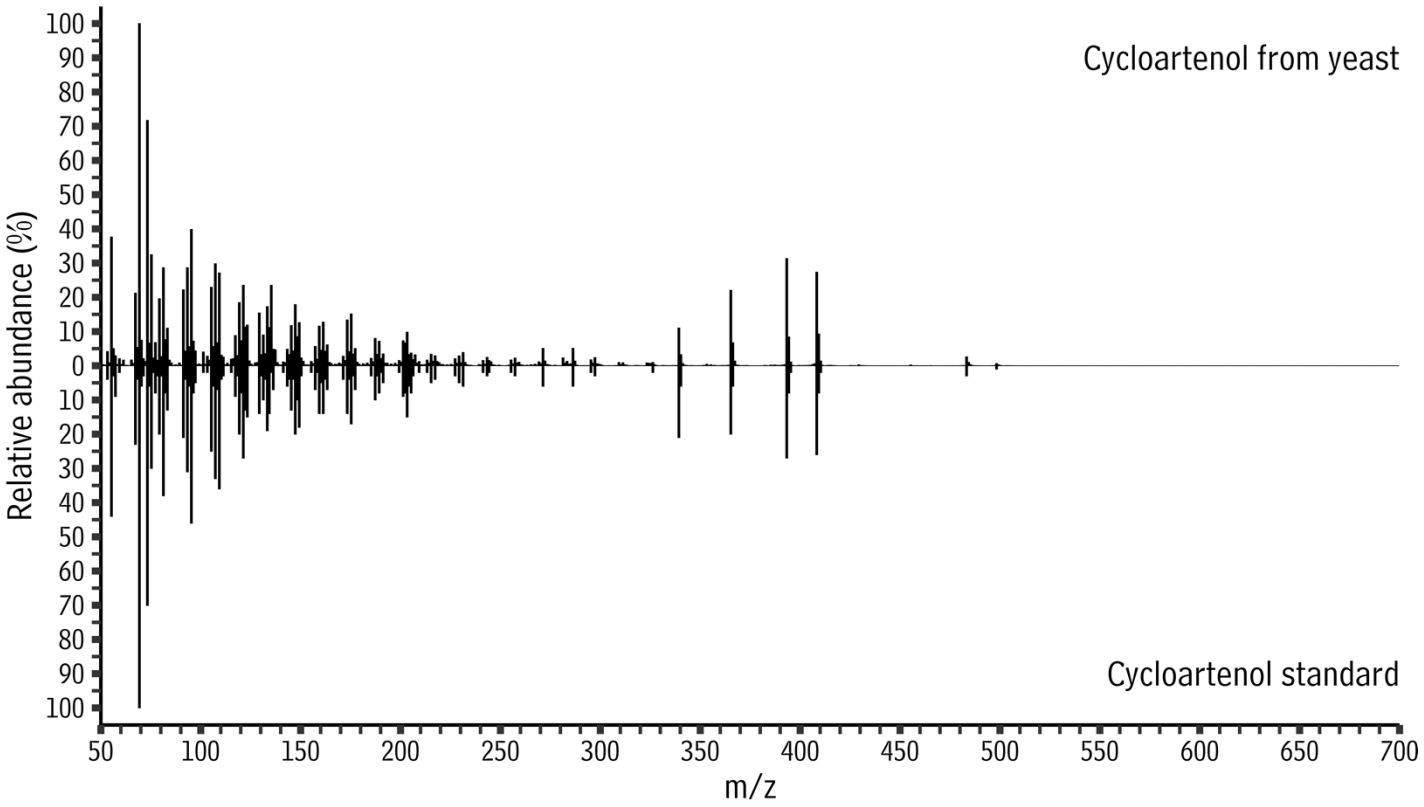
Supplementary Figure S7: Principal components analysis of wax compounds on *Sorghum bicolor* leaves. A. Bar chart showing the percent of variance explained (y-axis) by the first five dimensions of the principal component space (x-axis). **B.** Variable correlation plot showing the relationships between the variables (chemical compounds) and their relationships to the first two principal components. Each variable is represented with a line. Positively correlated variables are grouped together on the polar coordinate axis. The position of each variable line's tip between the origin and the outer circle indicates how much of the variance exhibited by that compound is encapsulated in the first two dimensions of the principal component space, with the variance of variables whose lines extend all the way to the outer circle being completely encapsulated by the two dimensional space. Each line is color coded according its compound class, yellow: alkanes, blue: aldehydes, orange: triterpenoids, red: primary alcohols, purple: methyl alkylresorcinols, pink: fatty acids, and grey: unidentified. **C.** Ordination plot showing the position of each sample (colored rings) within the space defined by the first principal component (x-axis) and second principal component (y-axis). The relative sizes of the two axes are scaled according to the amount of variance encapsulated by the dimension it represents. The rings representing each sample are colored according to the age of the plant from which the sample was taken, with samples from the youngest plants colored grey and samples from the most mature plants colored black, as indicated in the legend.



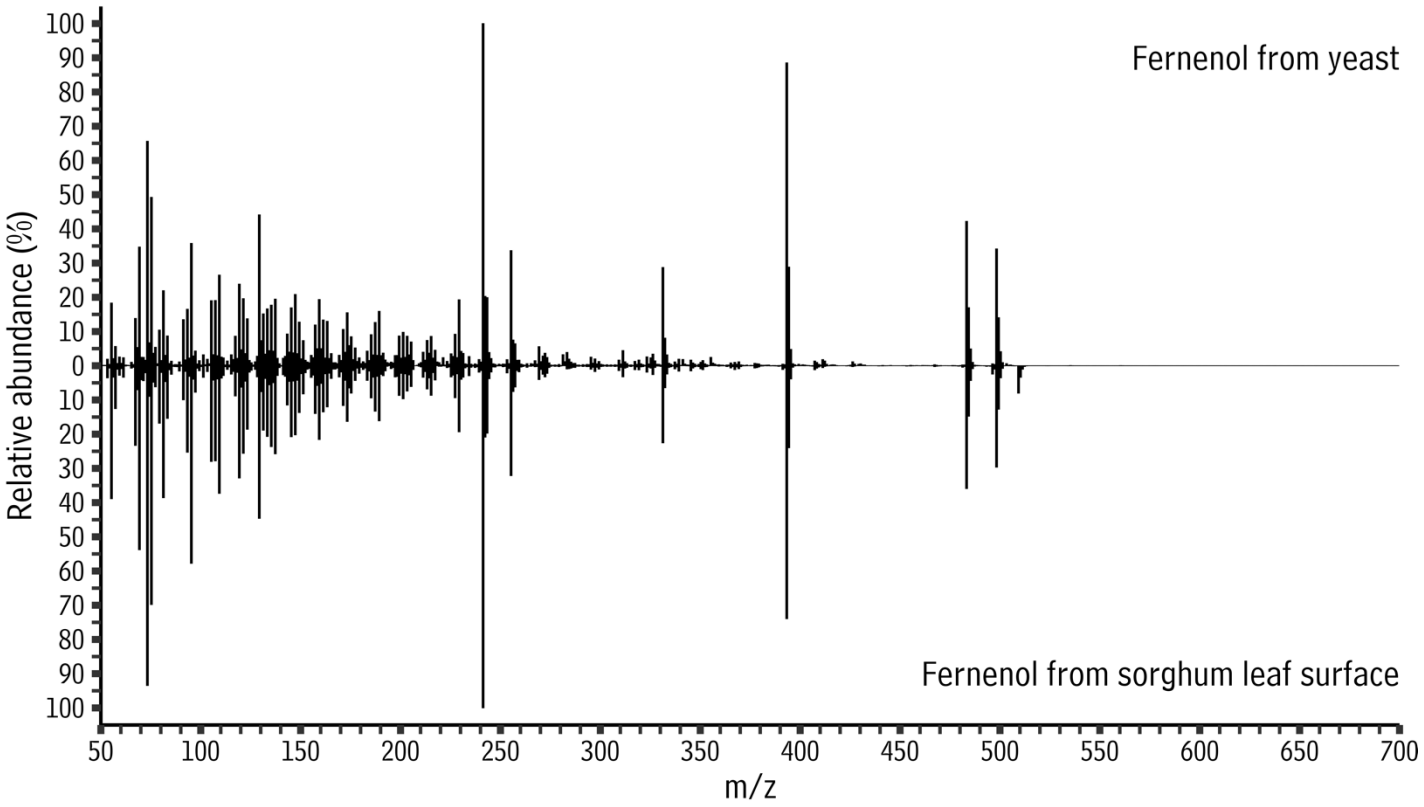
Supplementary Figure S8: Bar chart showing the mass spectrum of β-amyrin produced by transgenic yeast cultures (top) and the mass spectrum of β-amyrin found on sorghum leaf surfaces. The peak in the top spectrum at *m/z* 265 is from a partially co-eluting compound (bottom).



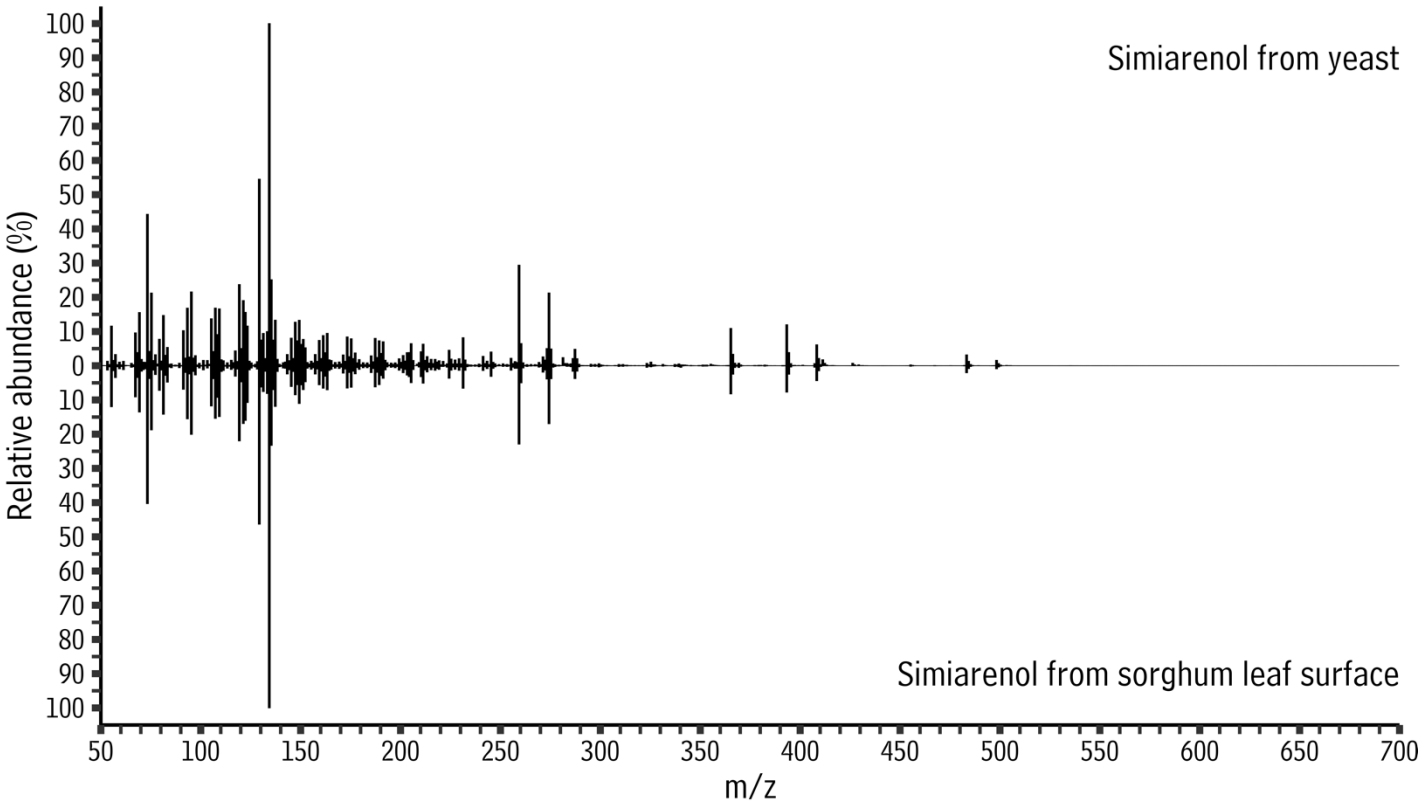
Supplementary Figure S9: Bar chart showing the mass spectrum of α-amyrin produced by transgenic yeast cultures (top) and the mass spectrum of α-amyrin found on sorghum leaf surfaces (bottom).



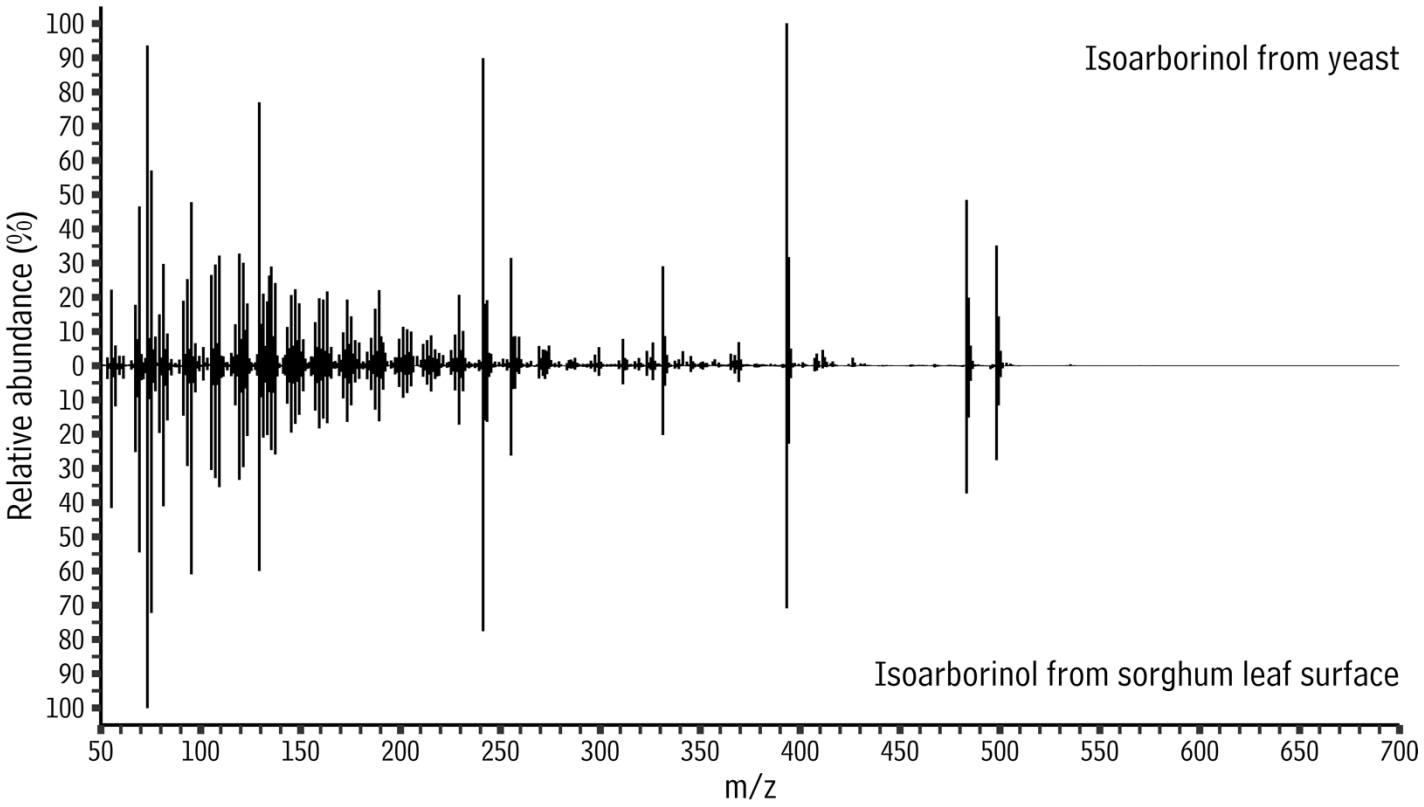
Supplementary Figure S10: Bar chart showing the mass spectrum of cycloartenol produced by transgenic yeast cultures (top) and the mass spectrum of a cycloartenol standard (bottom).



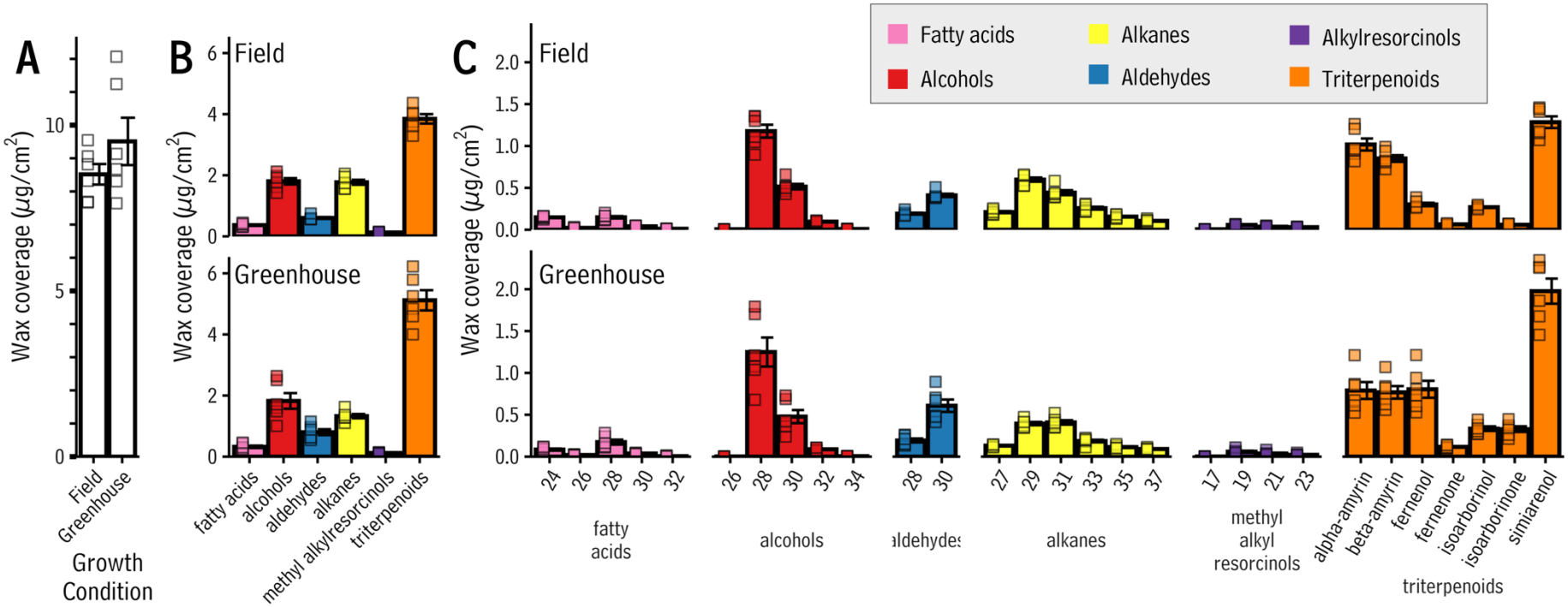
Supplementary Figure S11: Bar chart showing the mass spectrum of fernenol produced by transgenic yeast cultures (top) and the mass spectrum of fernenol from the sorghum leaf surface (bottom).



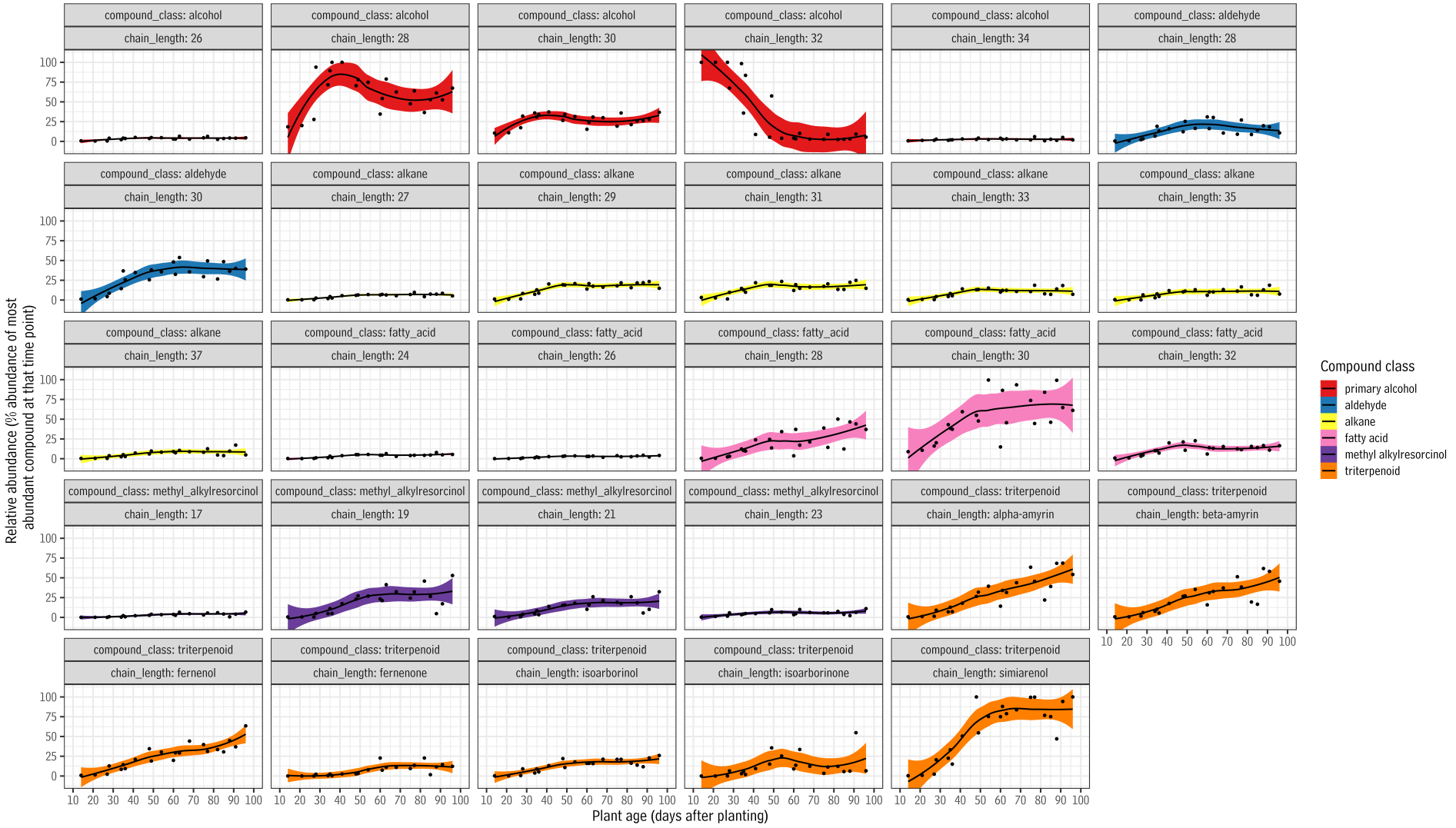
Supplementary Figure S12: Bar chart showing the mass spectrum of simiarenol produced by transgenic yeast cultures (top) and the mass spectrum of simiarenol from the sorghum leaf surface (bottom).



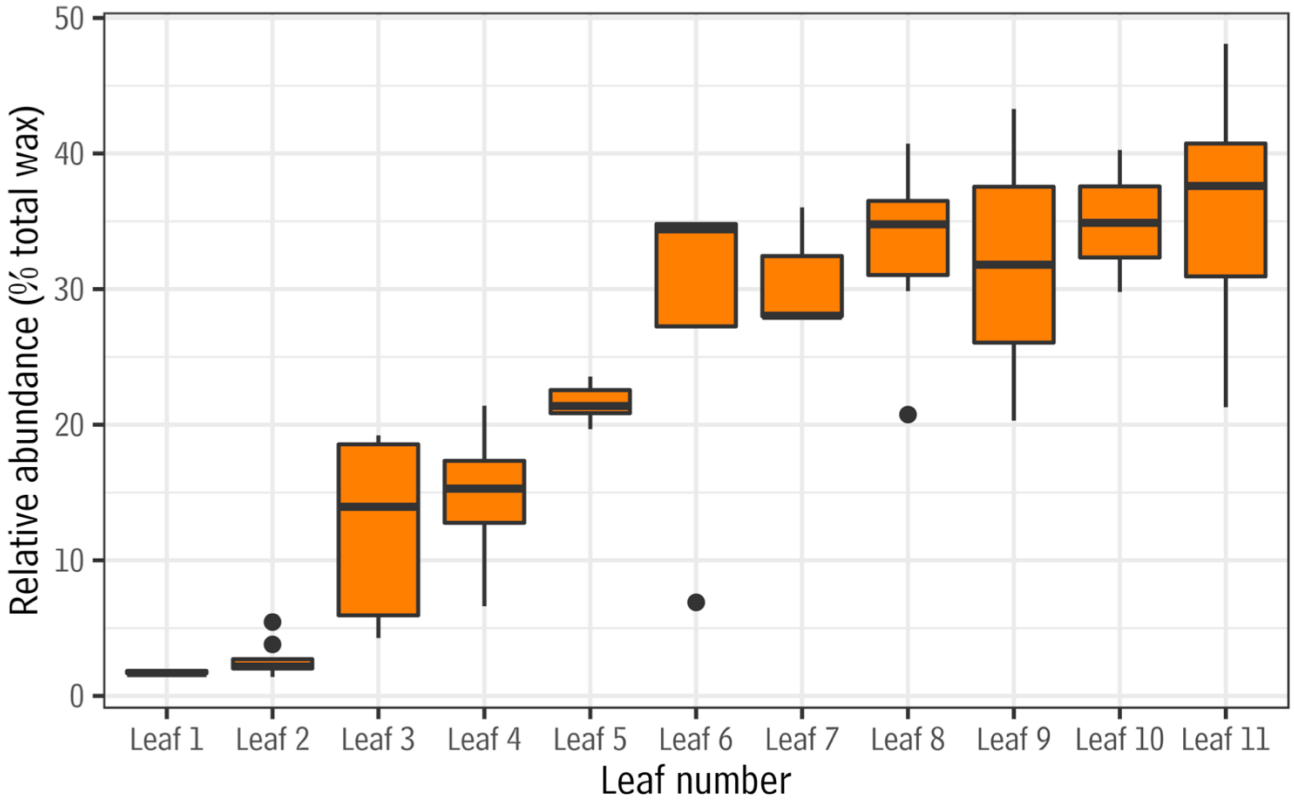
Supplementary Figure S13: Bar chart showing the mass spectrum of isoarborinol produced by transgenic yeast cultures (top) and the mass spectrum of isoarborinol from the sorghum leaf surface (bottom).



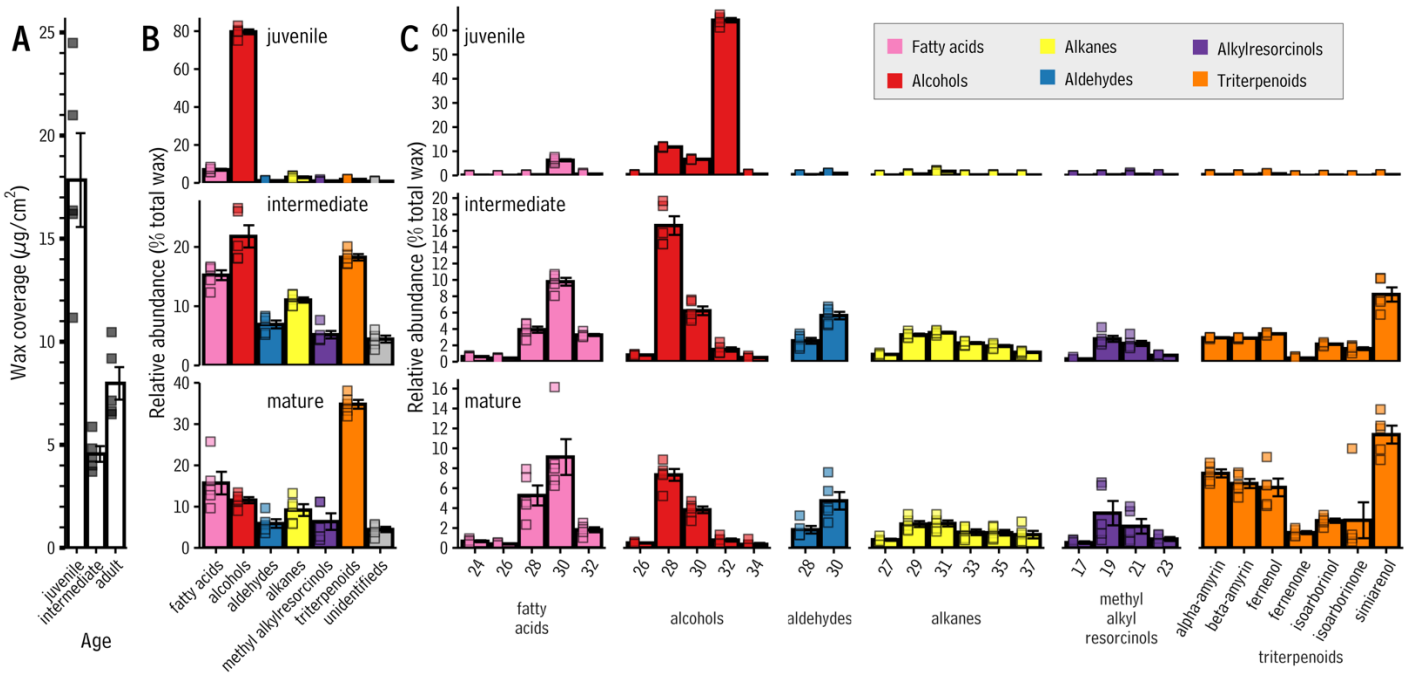
Supplemental Figure S14: Coverage and composition of wax on leaves from mature *Sorghum bicolor* plants grown in greenhouse and field conditions. A. Bar chart showing total wax coverage (in ug/cm2; y-axis) on leaves produced by mature plants in field (left bar) and greenhouse (right bar) conditions. **B.** Bar chart showing the abundance (in ug/cm^2; y-axis) of each compound class (x-axis) present on leaves produced by mature plants in field (top panel) and greenhouse (bottom panel) conditions. Bars are color-filled according to compound class, and as defined in Fig. 1A: yellow: alkanes, blue: aldehydes, orange: triterpenoids, red: primary alcohols, purple: methyl alkylresorcinols, pink: fatty acids, and grey: unidentified. **C.** Bar chart showing the abundance (in ug/cm^2; y-axis) of each compound identified on mature plants in field (top panel) and greenhouse (bottom panel) conditions. The chemical structures of these compounds are presented in Fig. 1B. Throughout, bar heights and error bars represent the mean and standard error of six independent samples, respectively.



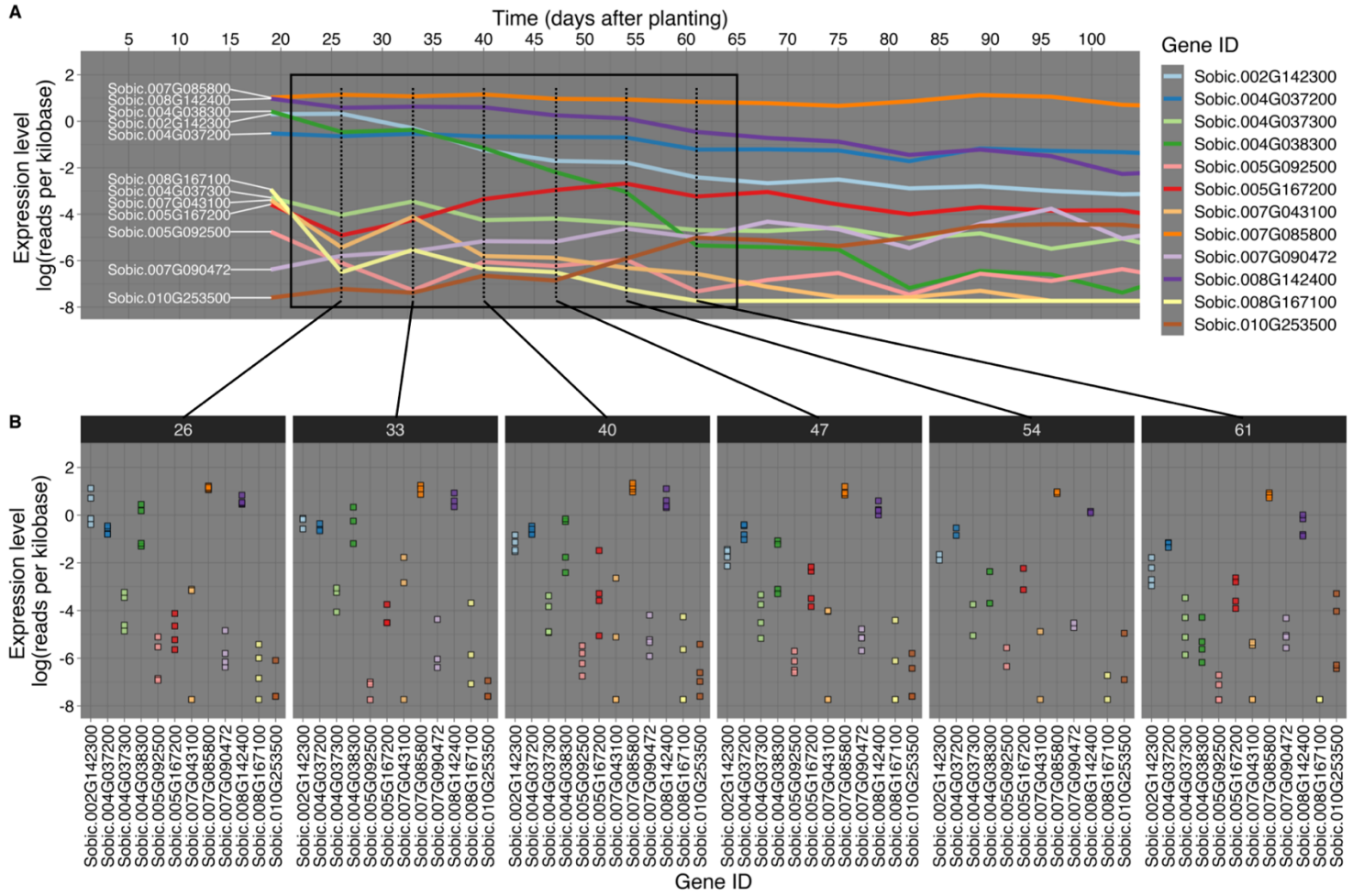
Supplemental Figure S15: Relative abundances of wax compounds on leaves of *Sorghum bicolor* over the course of plant development. Scatter plots showing the relative abundance (as percent of the most abundant compound at that time point, y-axis) of each wax compound on sorghum leaves from 14 to 98 days after planting (x-axis). Each point represents the abundance of a particular wax compound on the highest ligulated leaf of a plant of particular age. The trendlines and confidence intervals are meant merely to guide the eye in seeing trends. Confidence intervals are colored according to the compound class of the specific compound they represent, as indicated in the legend.



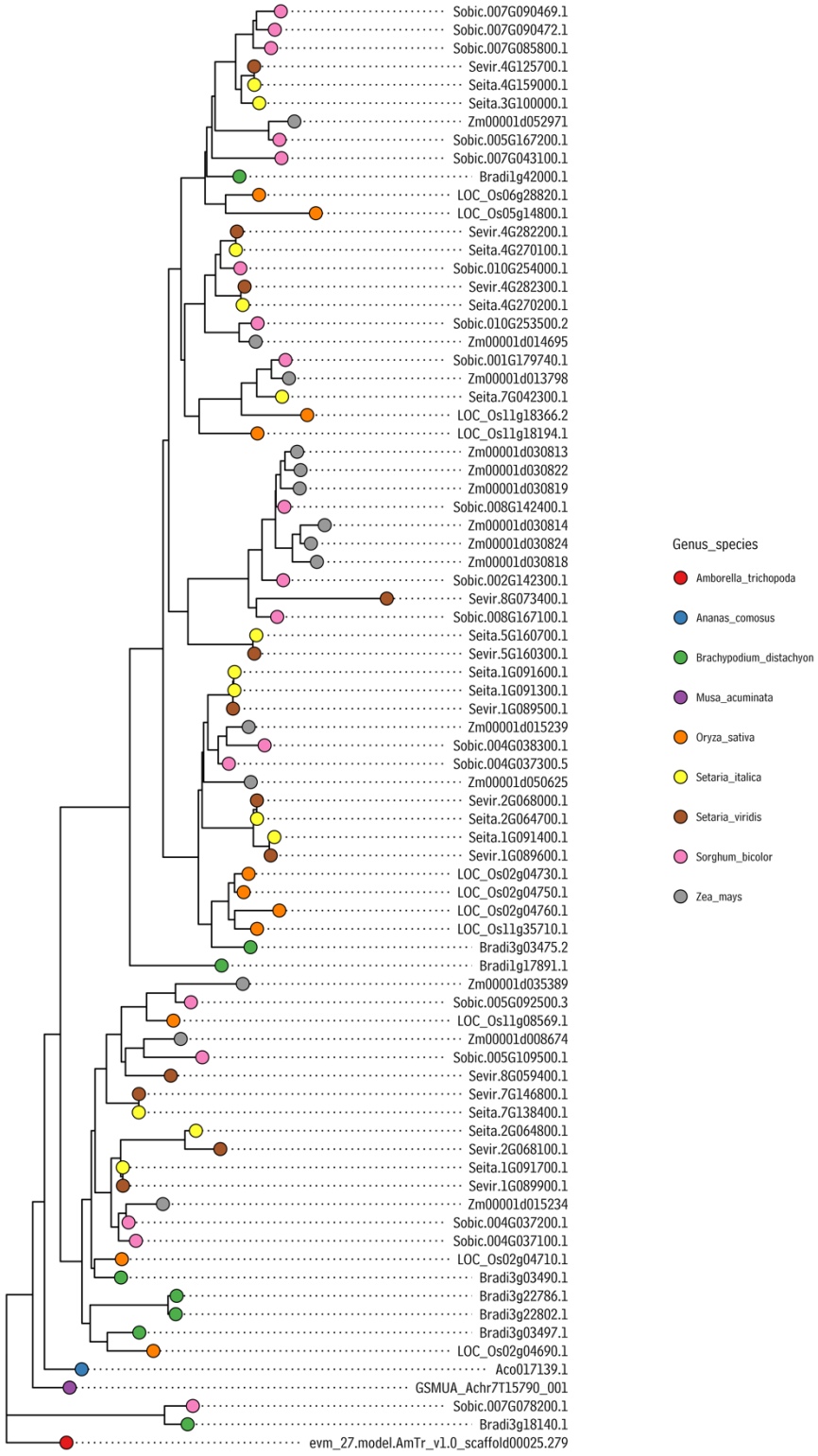
Supplemental Figure S16: Relative abundance of triterpenoids on leaves of *Sorghum bicolor*. Box plot showing the percent of the total wax mixture made up by triterpenoids (y-axis) on each leaf (x-axis) of sorghum.



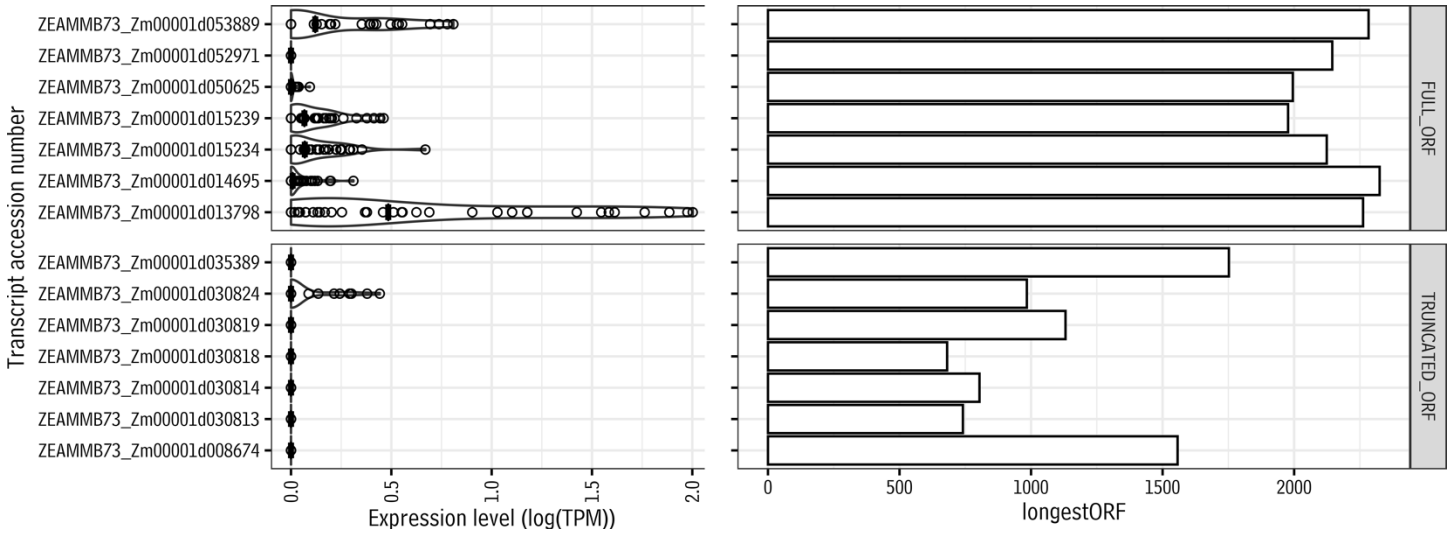
Supplemental Figure S17: Comparison of wax chemistry on juvenile, intermediate, and mature leaves of *Sorghum bicolor*. **A.** Bar chart showing total wax coverage (in ug/cm^2; y-axis) on leaves produced by juvenile (left bar), intermediate (middle bar), and adult (right bar) *S. bicolor* plants. **B.** Bar chart showing the relative abundance (percent total wax from that leaf; y-axis) of each compound class (x-axis) present on leaves produced by juvenile (top panel), intermediate (middle panel), and adult (bottom panel) *S. bicolor* plants. Bars are color-filled according to compound class, and as defined in Fig. 1A: yellow: alkanes, blue: aldehydes, orange: triterpenoids, red: primary alcohols, purple: methyl alkylresorcinols, pink: fatty acids, and grey: unidentified. **C.** Bar chart showing the relative abundance (percent total wax from that leaf; y-axis) of each compound identified on juvenile (top panel), intermediate (middle panel), and adult (bottom panel) leaves of *S. bicolor*. The chemical structures of these compounds are presented in Fig. 1B. Throughout, bar heights and error bars represent the mean and standard error of five independent samples, respectively.



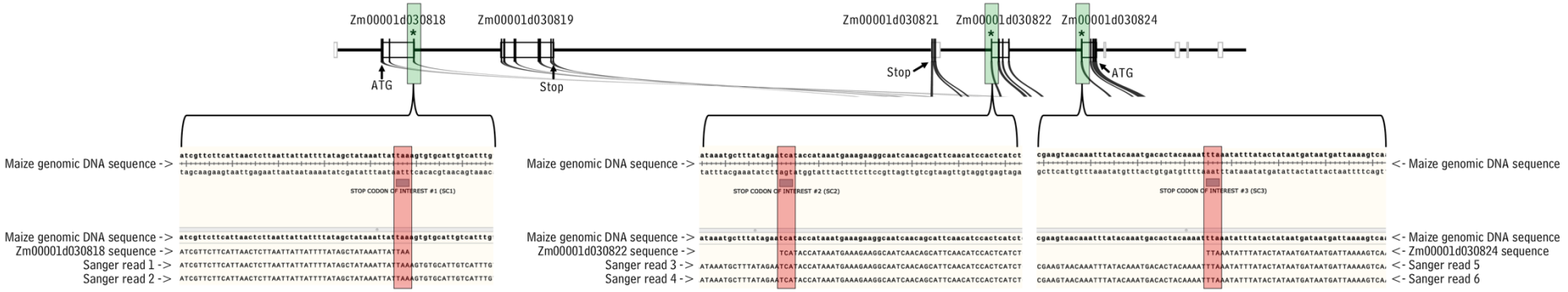
Supplemental Figure S18: Expression of OSC genes in sorghum leaves during leaf development. Data in this figure are from Varoquaux et al, 2019, downloaded from the Gene Expression Omnibus as normalized read counts and normalized to gene length in kilobases. **A**. Line plot showing mean expression level in leaves (as log of reads per kilobase, y-axis) of sorghum OSCs as a function of time (in days after planting, x-axis). Each line represents the mean expression level of a single OSC gene, as indicated with the white text labels and with line color as denoted in the shared legend. The black box shows the region of time leading up to flowering, a period of time during which leaves are undergoing considerable growth and cuticular waxes are being synthesized. **B**. Detailed depiction of expression levels of sorghum OSCs at six time points leading up to flowering. Each panel shows the expression level (as log of reads per kilobase, y-axis) of sorghum OSCs (x-axis) using squares colored by gene, as indicated in the shared legend. The time point (in days after planting) detailed by each panel is shown in white text at the top of each panel. Note: some genes detected in our analysis (Sobic.010G254000, Sobic.005G109500, Sobic.007G090469, Sobic.001G179740, Sobic.004G037100, and Sobic.007G078200), each of which had very low expression levels in our analysis (median expression levels below mean OSC expression level, Fig. 4), are not present in the Varoquaux dataset and are not shown in this supplemental figure.



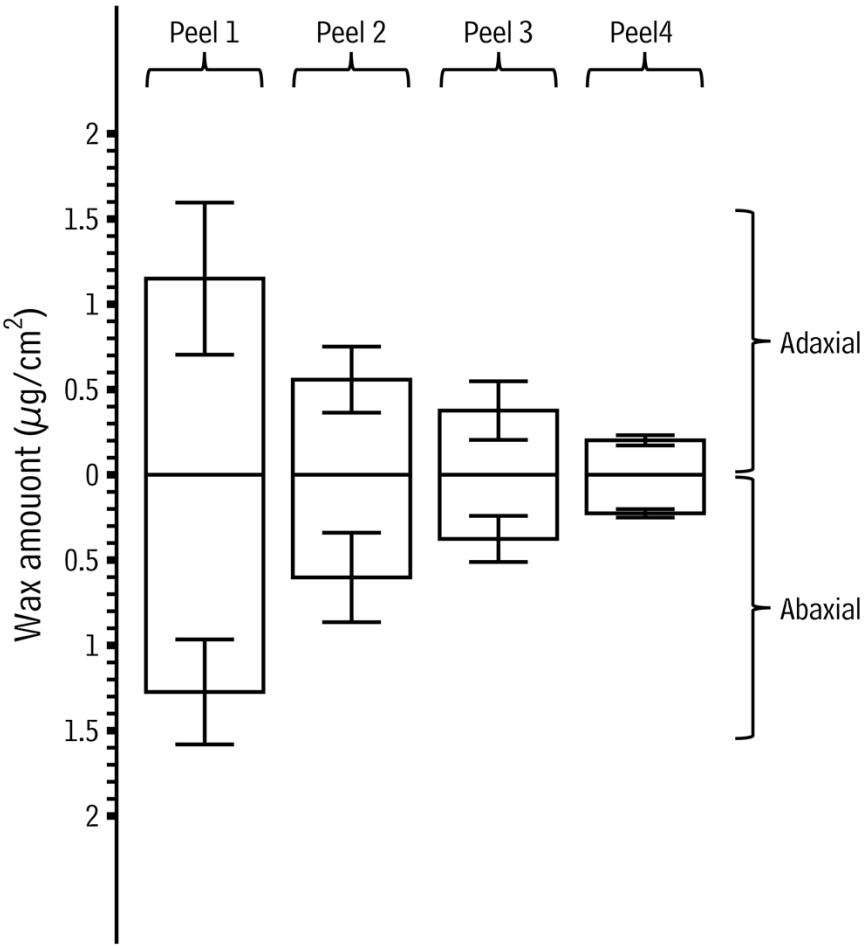
Supplemental Figure S19: Phylogeny of monocot oxidosqualene cyclases. Rooted phylogenetic tree showing relationships between annotated oxidosqualene cyclase-like gene models in six grass genomes (*Brachypodium distaychon*, accessions "Bradi[…]"; *Oryza sativa*, accessions "Loc_Os[…]"; *Zea mays*, accessions "Zm[…]"; *Sorghum bicolor*, accessions "Sobic.[…]"; *Setaria viridis*, accessions "Sevir.[…]"; *Setaria italica*, accessions "Seita.[…]"; *Musa acuminata*, accessions "GSMUA[…]"; and *Ananas comosus*, accessions "Aco[…]"). Circles at tree tips denote species as indicated in the legend.



Supplementary Figure S20: Violin plot (left) comparing distribution in expression levels of multiple oxidosqualene cyclase-like genes (annotated by accession number) as a measure of transcripts per million (TPM), in *Zea mays*. Median expression level of each investigated gene is shown with a vertical cross bar inside each violin. For each oxidosqualene-like gene, the length of the longest-open reading frame is shown in a horizontal bar chart (right). The entire figure is separated into two vertical facets according to whether the gene has a complete ORF ("FULL_ORF", top) or a truncated ORF ("TRUNCATED_ORF", bottom).



Supplemental Figure S21: Diagram showing PCR verification of in frame stop codons in *Zea mays* triterpenoid synthases. The top is a representation of maize genes. The long, horizontal black line indicates the chromosome, and black bars superimposed on top white boxes indicate coding sequence segments. Green boxes highlight regions with stop codon mutations, which are expanded to the sequence level in the diagrams below. In those, red regions indicate the stop codons themselves. Sanger reads from PCR reactions are aligned with the genomic sequence in each diagram, verifying the presence of the stop codons in the maize genome.



Supplementary Figure S22: Bar chart showing the amounts of wax recovered from successive gum arabic peels of *Sorghum bicolor* cuticular wax. Bars and error bars show the mean and standard deviation of four independent samples.

Supplementary Methods

Bulk Extraction of Sorghum Leaf Waxes and Thin Layer Chromatography Separation

The wax residue was resuspended in chloroform (4 mL), an aliquot was spotted onto TLC plate (EMD, 10 × 20 cm, 200 μm thickness), the plate was developed in chloroform, dried, then bands were visualized by applying primuline (Sigma; 99:1 acetone saturated with primuline:acetone) with a reagent sprayer. The silica that contained each band of compounds was scraped off using a razor blade, placed in test tubes, the silica in each tube was extracted using chloroform and the solvent from each extract was evaporated. The residue from each band was prepared for GC analysis: the residue was resuspended in $CHCl_3$ (200 μl) transferred to a GC vial insert, the solvent was evaporated at 70 °C under a steadily, directed stream of N_2. Then, the insert's contents were derivatized by dissolving the contents in pyridine (25 μL) and N,O-bis(trimethylsilyl)trifluoroacetamide (BSFTA, 25 μL, Sigma), capping each vial, and incubating (70 °C, 45 minutes). Following the incubation period, the solvent is evaporated under N_2, again. Finally, the residue was resuspended in $CHCl_3$ (50 mL) to create a gas chromatography-ready sample, and 2 μL of each sample is injected into the GC for chemical analysis.

Preparation of Wax Samples

Each square of tissue was placed into its own wide-mouthed, glass container holding 500 mL of $CHCl_3$ and tetracosane internal standard (25 μL; 0.1 mg/mL). Subsequently, the contents of the tissue were extracted into solution over the course of 30 seconds, aided by a rocker and swirling the solution around the tissue. Immediately following the extraction period, the extract was removed using a Pasteur pipette and transferred to a labelled GC vial. Then, the same leaf tissue is transferred to a new wide-mouth container of $CHCl_3$ (500 mL), where the extraction process was repeated: swirling and rocking the solution over the tissue for another 30 seconds. Extract from the secondary extraction course was removed and combined with the first 500 mL extract by transferring it to the same GC vial. The solvent was allowed to evaporate at room temperature overnight, then the residue was prepared for GC analysis as described for thin layer chromatography fractions.

Preparation of Epicuticular and Intracuticular Wax Samples

Gum arabic paste (1g gum arabic/1mL of ddH_2O) is prepared. Then, the paste is applied to defined areas of the leaf with a paint brush - roughly 50 cm^2 areas of undamaged leaf tissue- until a thin, even layer is covering the defined leaf area (approximately 2 mm thick). Then, the painted leaves are left to dry. Once hardened, the paste was peeled away from the leaf and collected. Gum arabic paste was then reapplied to the same area, left to dry, peeled, and collected. To prepare the peelings as GC samples, for each sample, the peelings were crushed into a fine powder, transferred to a scintillation vial, then dissolved in the addition of water (10 mL). To ensure all waxes were dissolved from the peel into the water, scintillation vials are sonicated for 30 minutes. After sonication, the epicuticular waxes are extracted from the water using 4 mL of $CHCl_3$, followed by vigorous vortex mixing. Then, the vial contents are separated carefully, by undisturbed rest

or low-speed centrifuging (≤ 2000 x g), forming two, distinct layers. Due to the density of the extracting solvent, the aqueous layer resides at the bottom of the vial, which can be removed with a pipette, and collected by transferring the extract to a new collection test tube. Once the removed, another aliquot of chloroform is added to the scintillation vial, which is vigorously mixed, carefully separated, then collected and pooled with the previous extracts. Once the extracts are collected, the collected volume is evaporated under N_2, until a volume appropriate for a GC vial insert is reached (≤ 200 μL). Then, the insert volume is completely dried under N_2 and subjected to 50 μL of derivatization reagents (25 μL of BSTFA and pyridine), sealed with GC vial caps, and incubated for 45 minutes at 70°C. Afterwards, the derivatization reagents are completely evaporated from the insert under N_2, the residues are resuspended in 50 μL of chloroform, and 2 μL of this volume is subsequently injected into the GC column for chemical analysis.

Once epicuticular waxes had been removed, it was possible to isolate the underlying intracuticular waxes. For this the previously peeled area of the leaf is the defined area of extraction and subjected to solvents in two stages. First, a clean glass tube of is pressed onto the surface of the leaf, creating a seal between it and the leaf. Then the glass tube is filled with the first aliquot of solvents: 1 mL of $CHCl_3$ and 25 μL of tetracosane internal standard (0.1 mg/mL solution). This solution is allowed to extract intracuticular wax constituents for 30 seconds, then it is removed from the leaf surface, and collected in a new test tube. Afterwards, a second aliquot of solvent (1 mL of $CHCl_3$ with no additional internal standard) is added to glass tube at the same location for 30 seconds, then the extracts are removed and pooled with extracts from the first aliquot. If another sample is desired, a new location, within the previously defined space, should be utilized. This process is repeated for each leaf and can be repeated on an individual leaf as many times as the defined space will allow. Following the two-stage extraction, the pooled extracts is dried under N_2 until the volume remaining is appropriate for a GC vial insert (≤ 200 μL). Once this volume is reached, the extracts are transferred to a GC vial, completely evaporated, then derivatized with 50 μL of reagents (25 μL of BSTFA and pyridine). Each sample is then sealed and incubate for 45 minutes at 70°C. After this period, each sample is opened, placed under a steady stream of N_2, and completely dried. Finally, each dried sample is resuspended in 50 μL of $CHCl_3$, and 2 μL of the sample is injected into the GC for chemical analysis.

Gas Chromatography-Mass Spectrometry Analyses

Samples were injected using an on-column injector in track oven mode into an HP-1 capillary column (Agilent, 30m, 0.25 mm inner diameter, 0.25 μm film thickness) with He carrier gas (1.4 mL/min) at 50°C. The oven was programmed as follows: 50°C for 2 minutes, 40°C/min ramp to 200°C, hold at 200°C for 2 min, increase by 3°C/min to 320°C, and held at 320°C for 30 min. Analytes were detected with a mass selective detector (Agilent 5975C EI/CI MS; m/z 40–600, 1 scan/s). Some chemical compound classes are more efficiently detected by mass spectral detectors than others. To determine if there was bias for or against any of the sorghum wax compound classes, representative samples were further analyzed using an identical GC system with an HP5 column that passed eluting compounds to a mass selective detector and a flame ionization detector in a 1:1 ratio. No statistically significant differences in detection efficiency were observed for any of the sorghum wax compound classes reported here.

Confirmation of Stop Codons In Maize Oxidosqualene Cyclase Genes

The following primer pairs were used for amplification of sequence encompassing each of the stop codons: Stop Codon 1, SC1_F 5'- CACTATCAACTACAAGTAACATCAGAAGGG-3', SC1_R 5'-GGGGTCCATGTTCCTTTCATCTATCAATAT-3'; Stop Codon 2, SC2_F 5'-CCAGAAACGGCCCTGATATTGATATTATAT-3', SC2_R 5'- TAAGTGACATTACTGACTTTAGAGCAATCC-3'; and Stop Codon 3, SC3_F 5'-AGGAAGATTTGCCAAAACAATTCAATGGAG-3', SC3_R 5'-GGGGTCAAAATTATTGCTAGACCTTAGTAT-3'.