

Supporting Information

Figure legends S1–S4; Table legends S1–S6; Methods S1

Figure Legends

Figure S1: Schematic of microarray hybridization design. Experiments were designed to study temporal and spatial patterns of poplar herbivore defense responses. Each circle indicates a specific leaf type and time point of harvest. Blue circles indicate leaves from untreated control trees; orange circles indicate leaves from trees with local source (LSo) leaves at the base of the tree treated by mechanical wounding and FTC OS application. Each arrow represents a hybridization on a microarray slide; solid arrows indicate dye-flipped direct comparisons between treated and respective control leaves, dashed arrows indicate balanced loops between leaf types or time points for either treated or control samples. LSo, local source leaves; SSo, systemic source leaves; SSi, systemic sink leaves; 2 h, 6 h, or 24 h timepoints. A total of 54 hybridizations were performed.

Figure S2: Comparison of variance and response obtained by microarray hybridizations using samples from individual trees separately or as pooled samples (see Ralph *et al.*, 2006). (a) Similar level of variation of microarray data obtained with separate samples from individual trees (biological and technical variation) or with pooled samples (technical variation). (b) Plot of differential expression of genes measured in 2 h OS-treated samples with pooled tree samples compared to individual tree samples. Number/Count scale from light grey to black correlates with an increasing number of genes in that category.

Figure S3: Overall spatial and temporal patterns of differentially expressed (DE) genes in local and systemic leaves in response to FTC OS detected by transcriptome analysis on a microarray platform of 15,496 cDNA elements. Genes were classified as DE between treated (T) and control (C) leaves using a threshold of fold-change > 1.5 , $P < 0.050$, $Q < 0.050$. The change of transcript abundance for DE genes is shown on the x-axis as \ln of fold change between T and C leaves. The number of array elements for each level of change of expression is shown on the y-axis. The 5% most up- or down-regulated genes are outside of the dotted vertical lines. Solid vertical lines mark the minimum and maximum DE. The number of genes significantly up- and down-regulated is given for each leaf type and time point.

Figure S4: Amino acid sequence alignment of galactinol synthases (GOLS). Alignments of protein sequences predicted from cDNA and gDNA sequences were generated using ClustalW (BLOSUM matrix, gap open penalty = 50, gap extension penalty = 0.5) and Boxshade. Conserved similarity shading was based on 50% identity (black) and 50% similarity (grey). A black bar above the alignment indicates a conserved carboxyl terminal pentapeptide signature. The hypothetical manganese-binding motif DXD is indicated with '****'. The serine phosphorylation site is indicated with 'X'. Nomenclature is as follows: Pt, *Populus trichocarpa*; Ptd, *Populus trichocarpa* x *deltoides*; At, *Arabidopsis thaliana*; Ar, *Ajuga reptans*; Le, *Lycopersicon esculentum*; Os, *Oryza sativa*; Zm, *Zea mays*. The sequences with identifiers ending with 'g' are predicted gene models from the *P. trichocarpa* v2.0 genome (www.phytozome.net/poplar); corresponding full-length cDNA versions end with a 'period-number' designation. Accession numbers are listed in Table S5.

Table legends

Table S1. Microarray analysis of transcriptome response in LSo, SSo, and SSi leaves in response to simulated herbivory by addition of FTC OS to mechanical wounds, at 2 h, 6 h, and 24 h post-treatment. Abbreviations: FC, fold-change; Est, estimated effect; SE, standard error; *P*, *P* value; *Q*, *Q* value.

Table S2: Gene-specific oligonucleotide primers used in qPCR analyses.

Table S3. qPCR analysis of expression in LSo, SSo, and SSi leaves in response to simulated herbivory at 2 h, 6 h, and 24 h post-treatment. Values represent fold-change (FC) relative to untreated control trees. Values obtained by qPCR represent the mean of three or more independent technical replicates, each consisting of pooled RNA from five biological replicates. A Student's *t*-test (two-sample, unpaired, one-sided) was performed to test significance (*P*) of up- or down-regulation of each transcript between treated and control plants. For each transcript, qPCR data is shaded light-grey, with the corresponding microarray data provided below (see Table S1 for the complete microarray dataset). Abbreviations: LRR, leucine-rich repeat.

Table S4. Cluster analysis of transcriptome response in LSo, SSo, and SSi leaves in response to simulated herbivory at 2 h, 6 h, and 24 h post-treatment. Abbreviations: FC, fold-change; Est, estimated effect; SE, standard error; *P*, *P* value; *Q*, *Q* value.

Table S5. Sequence relatedness of plant GOL S proteins from pairwise amino acid sequence comparisons using full-length protein sequences is shown as percent identity. GOL S

nomenclature and GenBank protein accession numbers are as follows - *Poplar*: Full-length cDNAs: PtdGOL S1.2, WS0188_I19 (EU305720); PtdGOL S6.1, WS02432_H16 (EU305724); PtdGOL S3.1, WS0231_K19 (EU305723); PtdGOL S2.1, WS0141_H15 (EU305722); PtdGOL S1.1, WS0131_D05 (EU305719); PtdGOL S2.1, WS01126_I24 (EU305721); PtdGOL S1.1, WS0125_K23 (EU305718); *Genome Models*: PtGOL S1g (POPTR_0008s19370); PtGOL S8g (POPTR_0008s10040); PtGOL S2g (POPTR_0010s05170); PtGOL S5g (POPTR_0005s00850); PtGOL S7g (POPTR_0014s11210); PtGOL S4g (POPTR_0013s00720); PtGOL S3g (POPTR_0013s00730); PtGOL S6g (POPTR_0002s19230); PtGOL S9g (POPTR_0010s16020); from *P. trichocarpa* v2.0 genome (www.phytozome.net/poplar); *Arabidopsis thaliana*: AtGOL S1 (AAB63818), AtGOL S2 (AAG09103), AtGOL S3 (AAC33195), AtGOL S4 (NP_176248), AtGOL S5 (AAB71970), AtGOL S6 (AAN13051), AtGOL S7 (BAB10052), AtGOL S8 (NP_850902); *Ajuga reptans*: ArGOL S1 (CAB51533), ArGOL S2 (CAB51534); *Ammopiptanthus mongolicus*: AmGOL S1 (ABF66656); *Brassica napus*: BnGOL S1 (AAD26116); *Cucumis melo*: CmGOL S1 (AAL78687), CmGOL S2 (AAL78686); *Glycine max*: GmGOL S1 (AAM96867); *Lycopersicon esculentum*: LeGOL S1 (AF311943); *Medicago sativa*: MsGOL S1 (AAM97493); *Oryza sativa*: OsGOL S1 (BAA05538); *Pisum sativum*: PsGOL S1 (CAB51130); *Triticum aestivum*: TaGOL S1 (BAF51565), TaGOL S2 (BAF51566); *Thellungiella halophila*: ThGOL S1 (AAM19710); *Xerophyta viscosa*: XvGOL S1 (ABK27907); *Zea mays*: ZmGOL S1 (AF497507), ZmGOL S2 (AF497508), ZmGOL S3 (AF497509).

Table S6. qPCR analysis of GOLS systemic gene expression in LSo, SSo, and SSi leaves in response to simulated herbivory at 2 h, 6 h, and 24 h post-treatment. Values represent fold-change (FC) relative to control trees. Values obtained by qPCR represent the mean of three or more independent technical replicates, each consisting of pooled RNA from five biological replicates. A Student's *t*-test (two-sample, unpaired, one-sided) was performed to test significance (*P*) of up- or down-regulation of each transcript between treated and control plants. For each transcript, qPCR data are shaded light-grey, with the corresponding microarray data provided below (see Table S1 for the complete microarray dataset). '--' indicates that this gene is not represented on the 15.5K microarray.

Methods S1

Microarray hybridization and gene expression data analysis

Details of the hybridization protocol for the 15.5K poplar microarray, along with fabrication and quality control details, have been reported in Ralph *et al.* (2006). For each treatment, leaf group, and time point, equal amounts of total RNA were combined from each of the five biological replicate trees prior to cDNA microarray analysis. Samples were indirectly labelled with Cy5 and Cy3 fluorescent dyes using the 3DNA Array 350 kit (Genisphere, Hatfield, PA, USA). All microarray experiments were designed to comply with MIAME guidelines (Brazma *et al.*, 2001). All scanned microarray TIF images, an ImaGene grid, the gene identification file and Imagenet quantified data files have been submitted to GEO (www.ncbi.nlm.nih.gov/geo/; series GSE16383). A total of fifty-four hybridizations were performed using the hybridization scheme

shown in Figure S1. Total RNA samples from OS-treated leaves 2hrs, 6hrs, and 24hrs after treatment from all three leaf types were each compared with total RNA from the corresponding leaf types in control trees harvested at the same time point, using two slides each with a dye-flip [$2 \times (3 \times 3) = 18$ slides]. Balanced loops were also incorporated to compare both OS-treated or control trees, each either to the other two time points for a single leaf type, or to the other two leaf types at a single time point [$2 \times (3+3) = 12$ loops @ 3 slides/loop = 36 slides].

To validate our findings with pooled biological replicates, we also performed additional hybridizations using independent biological replicates to examine differential gene expression in SSi leaves at 2hrs in response to OS-treatment. Total RNA from four individual OS-treated trees was compared directly against four individual untreated control trees using four hybridizations, balanced to have each sample type labelled twice with each dye. Similarly, total RNA from a single pool of OS-treated leaves from each of five independent biological replicate trees (described above) was compared to untreated control leaves from five independent biological replicate trees using four hybridizations, again balanced to have each sample type labelled twice with each dye.

In order to assess the biological response to mechanical damage plus FTC OS (OS), in local source (LSO), systemic source (SSO), and systemic sink leaves (SSi), and through time (2hrs, 6hrs, 24hrs), a mixed-effects model containing a dye effect and a treatment effect for OS minus control (C) was fit using data from 54 microarray slides. In order to assess the differences in variance between microarray experiments involving pooled or individual tree samples, models containing a dye effect and a treatment effect for OS minus C were fit using data from four microarray slides for the individual replicate analysis, and four microarray slides for the pooled replicate analysis. Before data normalization, the lowest 10% of median foreground intensities

was subtracted from the median foreground intensities to correct for background intensity. After quantification of the signal intensities, data were normalized to compensate for non-linearity of intensity distributions using the variance stabilizing normalization (VSN) method (Huber *et al.*, 2002). To assess the transcriptional response to stress treatments, a linear mixed-effects model was fitted to the normalized intensities in the Cy3 and Cy5 channels. The model contained an adjustment for dye effect, an array effect indicating which Cy5/Cy3 pair was on each array, and a treatment effect indicating treatment and time point. For both microarray experiments, expression variance was derived from technical variance among slides. The ratio of each treatment parameter estimate to the standard error was used to calculate a t statistic, from which a P -value was obtained. The Q value for each effect and gene was calculated for each of the models to adjust for the false discovery rate (Storey & Tibshirani, 2003). Differentially expressed transcripts were defined as having fold-change ratios >1.5 , $P < 0.05$, and $Q < 0.05$. Cluster analysis was performed on a set of 6,472 array elements showing significant differential expression in at least one sample (i.e., fold-change between treated and control leaves > 1.5 x, $P < 0.05$, and $Q < 0.05$) using the divisive DIANA algorithm (Bryan, 2004), and the first eight clusters formed were examined in detail. All statistical analyses were performed within the R statistical package (www.r-project.org/), save for two-way ANOVAs and subsequent Tukey tests, which were performed with SYSTAT 11 (Systat Software Incorporated). Table S1 provides a complete list of microarray expression data for all genes represented on the microarray.

Quantitative real-time PCR (qPCR) and gene expression data analysis

qPCR was done as previously described (Philippe *et al.*, 2009; Ralph *et al.*, 2006). Prior to reverse transcription, 15 μ g total RNA per tree was treated with DNaseI (Invitrogen)

according to the manufacturer's instructions to remove genomic DNA. The resulting RNA was divided into three aliquots of 5 µg and independent cDNA synthesis reactions were performed using Superscript II reverse transcriptase (Invitrogen) with an oligo d(T₁₈) primer according to manufacturer's instructions. The efficiency of cDNA synthesis was assessed by gel electrophoresis prior to pooling of the three reactions per tree. Gene-specific primers were designed for microarray validation, for isoprene synthase expression verification, and for galactinol synthase expression analysis (Table S2) using a stringent set of criteria including predicted melting temperature of $64 \pm 2^{\circ}\text{C}$, primer lengths of 20-24 nucleotides, guanine-cytosine contents of 40-60% and PCR amplicon lengths of 100-350 bp. Primer specificity (single product of expected length) was confirmed by analysis on a 2% (w/v) agarose gel, by melting curve analysis and, for at least one PCR reaction per gene, by sequence verification of PCR amplicons (data not shown). Primers for poplar translation initiation factor 5A (TIF5A) were designed (GenBank accession number CV251327; poplar EST WS0116_J23 [POPTR_0006s19870]) and served as a quantification control (forward primer 5'-GACGGTATTTTAGCTATGGAATTG-3', reverse primer 5'-CTGATAACACAAGTTCCTGC-3), as in Ralph *et al.* (2006).

qPCR was conducted on a DNA Engine Opticon 2 in an optical 96-well plate (MJ Research) using the DyNAmo HS SYBR green kit (Finnzymes) according to manufacturer's instructions. Reaction mixtures contained 10 ng cDNA as template, 0.3 µM of each primer and 10 µL of DyNAmo master mix in a final volume of 20 µL. Reactions with the cDNA template replaced by nuclease-free H₂O or 10 ng of non-reverse transcribed RNA were run with each primer pair as a control. The program for all PCR reactions was: 95°C for 15 min; 40 cycles of 10 s at 95°C, 30 s at 60°C, and 30 s at 72°C. Data were analyzed using the Opticon Monitor 2

version 2.02 software (MJ Research). For each primer pair a minimum of three independent technical replicates was performed. To generate a baseline-subtracted plot of the logarithmic increase in fluorescence signal (ΔR_n) versus cycle number, baseline data were collected between cycles 3 and 10. All amplification plots were analyzed with an R_n threshold of 0.005 to obtain C_t (threshold cycle) values. Transcript abundance for each OS-induced gene was normalized to TIF5A by subtracting the C_t value of TIF5A from the C_t value of each OS-induced transcript, where $\Delta C_t = \Delta C_{t_{\text{transcript}}} - \Delta C_{t_{\text{TIF5A}}}$. Transcript abundance of OS-induced genes in control and OS-treated samples were obtained from the equation $(1 + E)^{-\Delta C_t}$, where E is the PCR efficiency, as described by Ramakers *et al.* (2003). A transcript with a relative abundance of one is equivalent to the abundance of TIF5A in the same tissue. In order to assess the biological fold-change response to OS treatment, a mixed effects model for each gene containing a treatment effect for OS minus untreated control (C) was fit using data from at least three independent technical replicates for each of the three leaf groups in OS-treated and untreated control cDNA pools. The ratio of the OS minus C parameter estimate to the standard error was used to calculate a t statistic and P value.

References

- Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FCP, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M. 2001.** Minimum information about a microarray experiment (MIAME) - toward standards for microarray data. *Nature Genetics* **29**(4): 365-371.
- Bryan J. 2004.** Problems in gene clustering based on gene expression data. *Journal of Multivariate Analysis* **90**(1): 44-66.
- Huber W, Von Heydebreck A, Sultmann H, Poustka A, Vingron M. 2002.** Variance stabilization applied to microarray data calibration and the quantification of differential expression. *Bioinformatics* **18**(Supplement 1): S96-S104.
- Philippe RN, Ralph SG, Kulheim C, Jancsik SI, Bohlmann J. 2009.** Poplar defense against insects: Genome analysis, full-length cDNA cloning, and transcriptome and protein analysis of the poplar Kunitz-type protease inhibitor family. *New Phytologist* **184**(4): 865-884.
- Ralph S, Oddy C, Cooper D, Yueh H, Jancsik S, Kolosova N, Philippe RN, Aeschliman D, White R, Huber D, Ritland CE, Benoit F, Rigby T, Nantel A, Butterfield YSN, Kirkpatrick R, Chun E, Liu J, Palmquist D, Wynhoven B, Stott J, Yang G, Barber S, Holt RA, Siddiqui A, Jones SJM, Marra MA, Ellis BE, Douglas CJ, Ritland K, Bohlmann J. 2006.** Genomics of hybrid poplar (*Populus trichocarpa* x *deltoides*) interacting with forest tent caterpillars (*Malacosoma disstria*): Normalized and full-

length cDNA libraries, expressed sequence tags, and a cDNA microarray for the study of insect-induced defences in poplar. *Molecular Ecology* **15**(5): 1275-1297.

Ramakers C, Ruijter JM, Deprez RHL, Moorman AFM. 2003. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neuroscience Letters* **339**(1): 62-66.

Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America* **100**(16): 9440-9445.