Molecular Ecology (2006) **15**, 1275–1297 doi: 10.1111/j.1365-294X.2006.02824.x

Genomics of hybrid poplar (*Populus trichocarpa* × *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

STEVEN RALPH,* CLAIRE ODDY,* DAWN COOPER,* HESTHER YUEH,* SHARON JANCSIK,* NATALIA KOLOSOVA,* RYAN N. PHILIPPE,* DANA AESCHLIMAN,* RICK WHITE,* DEZENE HUBER,* CAROL E. RITLAND,‡ FRANÇOIS BENOIT,§ TRACEY RIGBY,§ ANDRÉ NANTEL,§ YARON S. N. BUTTERFIELD,¶ ROBERT KIRKPATRICK,¶ ELIZABETH CHUN,¶ JERRY LIU,¶ DIANA PALMQUIST,¶ BRIAN WYNHOVEN,¶ JEFFREY STOTT,¶ GEORGE YANG,¶ SARAH BARBER,¶ ROBERT A. HOLT,¶ ASIM SIDDIQUI,¶ STEVEN J. M. JONES,¶ MARCO A. MARRA,¶ BRIAN E. ELLIS,* CARL J. DOUGLAS,† KERMIT RITLAND‡ and JÖRG BOHLMANN*†‡

**Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia, Canada, V6T 1Z3, †Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada, V6T 1Z4, ‡Department of Forest Sciences, University of British Columbia, Vancouver, British Columbia, Canada, V6T 1Z4, §Biotechnology Research Institute, Montreal, Québec, Canada, H4P 2R2, ¶BC Cancer Agency Genome Sciences Centre, Vancouver, British Columbia, Canada, V5Z 4E6*

Abstract

As part of a genomics strategy to characterize inducible defences against insect herbivory in poplar, we developed a comprehensive suite of functional genomics resources including cDNA libraries, expressed sequence tags (ESTs) and a cDNA microarray platform. These resources are designed to complement the existing poplar genome sequence and poplar (*Populus* spp.) ESTs by focusing on herbivore- and elicitor-treated tissues and incorporating normalization methods to capture rare transcripts. From a set of 15 standard, normalized or full-length cDNA libraries, we generated 139 007 3′- or 5′-end sequenced ESTs, representing more than one-third of the *c.* 385 000 publicly available *Populus* ESTs. Clustering and assembly of 107 519 3′-end ESTs resulted in 14 451 contigs and 20 560 singletons, altogether representing 35 011 putative unique transcripts, or potentially more than three-quarters of the predicted *c.* 45 000 genes in the poplar genome. Using this EST resource, we developed a cDNA microarray containing 15 496 unique genes, which was utilized to monitor gene expression in poplar leaves in response to herbivory by forest tent caterpillars (*Malacosoma disstria*). After 24 h of feeding, 1191 genes were classified as up-regulated, compared to only 537 down-regulated. Functional classification of this induced gene set revealed genes with roles in plant defence (e.g. endochitinases, Kunitz protease inhibitors), octadecanoid and ethylene signalling (e.g. lipoxygenase, allene oxide synthase, 1-aminocyclopropane-1-carboxylate oxidase), transport (e.g. ABC proteins, calreticulin), secondary metabolism [e.g. polyphenol oxidase, isoflavone reductase, (–)-germacrene D synthase] and transcriptional regulation [e.g. leucine-rich repeat transmembrane kinase, several transcription factor classes (zinc finger C3H type, AP2/EREBP, WRKY, bHLH)]. This study provides the first genome-scale approach to characterize insect-induced defences in a woody perennial providing a solid platform for functional investigation of plant–insect interactions in poplar.

Keywords: forestry, herbivory, plant–insect interactions, transcriptome, tree genomics

Received 26 August 2005; revision accepted 26 October 2005

Correspondence: Dr Jörg Bohlmann, Fax: 604-822-2114; E-mail: bohlmann@interchange.ubc.ca

© 2006 Blackwell Publishing Ltd

Introduction

The genus *Populus*, consisting of *c.* 40 species of poplars and aspen distributed in diverse habitats throughout the Northern Hemisphere, has been firmly established as a system for genomic research of angiosperm tree biology (Taylor 2002; Bhalerao *et al*. 2003a; Brunner *et al*. 2004). With an estimated size of 485 ± 10 Mb, the genome of *Populus* is only 4.5× larger than the *Arabidopsis* genome, and is roughly 40× smaller than genomes of members of the pine family (Pinaceae), which includes many of the economically important gymnosperm tree species. The genome of a female *Populus trichocarpa* tree (Nisqually 1) has recently been shotgun sequenced to a depth of 7.5× coverage (http://genome.jgi-psf.org/Poptr1/Poptr1.home.html), with the assembly and annotation, and generation of supporting physical and genetic maps, being contributed by members of the International Poplar Genome Consortium (www.ornl.gov/ipgc/).

Complementary to complete genome sequencing, the large-scale sequencing of expressed genes permits analysis of the transcriptome of an organism. Sampling of the transcriptome can be performed using high-throughput single-pass sequencing of cDNA libraries constructed from different tissues and developmental stages, or from plants subjected to different environmental conditions or stress treatments to generate expressed sequence tags (ESTs; Adams *et al*. 1993). The application of normalization techniques to reduce the frequency of highly expressed genes can increase the rate of gene discovery, permitting the identification of rare transcripts (Soares *et al*. 1994; Bonaldo *et al*. 1996). When our poplar EST project was initiated in 2002, other large-scale poplar EST sequencing efforts were already established that focused primarily on wood formation, dormancy and floral development (Sterky *et al*. 1998; Bhalerao *et al*. 2003b; Schrader *et al*. 2004; Sterky *et al*. 2004). In addition, other small-scale gene discovery activities have developed poplar cDNA libraries and EST sequences focusing on wood formation (Dejardin *et al*. 2004), root development (Kohler *et al*. 2003), and stress response (Christopher *et al*. 2004; Nanjo *et al*. 2004; Rishi *et al*. 2004). In order to maximize gene discovery both within the large-scale EST programme described here, and relative to the *c.* 247 000 *Populus* ESTs in the public domain (27 May 2005 dbEST release of GenBank), we have focused our efforts on normalized cDNA libraries and included a variety of insect-induced and biotic elicitor-induced tissues with the goal to complement previous large-scale poplar EST activities.

ESTs are also the starting reagents for the construction of cDNA microarrays for transcriptome profiling studies (Schena *et al*. 1995). A major emphasis of our programme in forest health genomics is to generate and utilize genomics resources to investigate how tree genomes respond to attack by herbivorous insects, which is relatively poorly understood in contrast to plant responses to abiotic stress or pathogens. Insect-induced defence responses identified by microarray transcript profiling have recently been described for a few herbaceous species, such as the wild tobacco *Nicotiana attenuata* (Hui *et al*. 2003; Heidel & Baldwin 2004; Voelckel *et al*. 2004), *Arabidopsis thaliana* (Reymond *et al*. 2000, 2004), and *Sorghum bicolor* (Zhu-Salzman *et al*. 2004). In contrast, insect-induced responses in poplar, which provides a unique system to study genomics of plant–insect interactions in a long-lived woody perennial, have only been studied for a small number of genes. The newly developed poplar genomics resources now provide the first opportunity for genome-wide transcriptome analysis of insect-induced defence systems in an angiosperm tree.

Forest insect pests pose a challenge to the sustainability of both natural and planted forests. The risk of forest insect pest epidemics, which cannot be addressed with short-term crop rotation or pesticide application, as is possible in agriculture, is increasing with the introduction of exotic pest species and with global climate changes. The larvae of several insect herbivores [forest tent caterpillar (*Malacosoma disstria*), gypsy moth (*Lymantria dispar*), aspen blotch leafminer (*Phyllonorycter tremuloidiella*), large aspen tortrix (*Choristoneura conflictana*)] can cause extensive defoliation to stands of *Populus* species, particularly the trembling aspen (*Populus tremuloides*), during outbreak periods. Other insects, such as the larvae of the willow weevil (*Cryptorhynchus lapathi*) affect stem tissues of poplar trees. Forest tent caterpillars (FTCs) are distributed throughout North America and Eurasia. Larvae hatch in early spring and immediately begin to feed on the leaves of their hosts. By their final instar, larvae grow to over 1000 times their mass at hatching and consume more than 15 000 times their initial body weight in leaf tissue (Fitzgerald 1995). During population outbreaks, FTCs commonly defoliate trees occurring over millions of hectares, with a density as high as 20 000 caterpillars per tree (Stairs 1972; Fitzgerald 1995). Defoliated trees have reduced photosynthetic capacity and may produce less wood, but only in extreme cases are trees killed directly due to repeated episodes of defoliation by FTC larvae (Gregory & Wargo 1986). However, repeated and prolonged attack by FTCs may result in an increased incidence of fungal disease and infestation by other insects (Churchill *et al*. 1964; Hogg *et al*. 2002).

The first lines of defence against insect herbivores are constitutive chemical and physical barriers; however, if these barriers are breached, inducible defences are of central importance in reducing herbivory (Karban & Baldwin 1997; Agrawal 1998). In poplar trees, the new foliage that FTCs feed upon undergoes profound physical and chemical changes that render maturing leaves increasingly less acceptable to the caterpillars. Slow growth, and even population collapse, may result if caterpillars fail to

synchronize their development with that of the host tree (Fitzgerald 1995; Parry *et al*. 1998). Compared to young, emerging poplar leaves, mature leaves contain lower water and nitrogen content, a higher content of non-nutritive fibre, possess increased toughness and increased levels of phenolic compounds, which combined deters caterpillar feeding, reduces the digestibility of leaf protein, and leads to reduced caterpillar growth (Fitzgerald 1995). Constitutive levels of phenolic compounds in aspen leaves, including phenolic glucosides such as salicortin and tremulacin, and to a lesser extent condensed tannins, are strongly influenced by genotype and nutrient availability, and have been demonstrated to negatively impact growth and performance of FTCs and other herbivores (Hwang & Lindroth 1997; Osier & Lindroth 2001, 2004).

In addition to constitutive defences, herbivores trigger at least two types of inducible defence responses in poplars: direct defences that can result in the inhibition of insect growth or development and indirect defences consisting of volatiles emitted from plants that can serve as airborne signals that deter herbivores or attract predators and parasites of herbivores. Inducible direct defences in poplars involve a broad range of proteins (e.g. protease inhibitors, oxidative enzymes) and phytochemicals (e.g. phenolics) (Constabel 1999; Huber *et al*. 2004). In herbaceous plant-herbivore defence systems, constitutive and induced defence mechanisms appear to be tightly regulated, permitting economy when active defence is not required, and presenting a shifting defence profile when herbivores are present (Karban & Baldwin 1997; Kessler & Baldwin 2002). It is therefore a priority to identify the signalling systems and the transcriptional and other insect-induced changes that regulate defence responses. Relatively few studies of the induced defence response have been conducted in *Populus* species at the molecular level. To date, targeted studies have identified induced genes encoding trypsin protease inhibitors (Bradshaw *et al*. 1990; Hollick & Gordon 1993; Haruta *et al*. 2001a), endochitinases (Parsons *et al*. 1989; Davis *et al*. 1991), vegetative storage proteins (Davis *et al*. 1993), polyphenol oxidases (Constabel *et al*. 2000; Haruta *et al*. 2001b), dihydroflavonol reductase (Peters & Constabel 2002) and genes of terpenoid metabolism, including a sesquiterpene synthase involved in FTC-induced systemic volatile emissions (Arimura *et al*. 2004). In addition, a small-scale array consisting of 569 cDNA clones identified a set of 85 cDNAs that were differentially and systemically expressed in leaves 24 h after applying mechanical wounding (Christopher *et al*. 2004).

We have recently established a programme targeted at genome-wide discovery and expression profiling of insect-induced defence genes in poplar. We describe here results from the development of 15 standard, normalized or full-length cDNA (FLcDNA) libraries that were sequenced from the 5′- and 3′-ends of cDNA clones to generate 139 007 ESTs from poplar. Assembly of high-quality (hq) 3′-end sequences has identified 35 011 putative unique transcripts. We demonstrate greatly enhanced gene discovery by focusing on normalized, rather than standard cDNA libraries. Using this EST resource we have constructed a cDNA microarray consisting of 15 496 non-redundant ESTs, which has been applied to an initial study of the transcriptional response in poplar leaves to feeding by FTC larvae.

Materials and methods

Plant material and insects

Populus trichocarpa Torr. & Gray × *P. deltoides* Bartr. (Salicaceae), H11-11 genotype, was grown on the University of British Columbia South Campus farm. Cuttings of 30–100 cm were taken in February of 2003 from previous year shoots, placed in soil (35% peat, 15% perlite, 50% pasteurized mineral soil, 250 g^{-3} Osmocote™ 13-13-13 plus micronutrients) in 2-gallon pots (Stuewe & Sons Inc.), and watered daily. Trees were maintained in a greenhouse under constant summer conditions where a constant 16/8-h photoperiod was provided by high-pressure sodium lamps. Trees of 60–70 cm in height were used in experiments in August 2003. Average greenhouse temperature during the month was 23.8 °C (21.3 °C minimum and 28.9 °C maximum), with an average relative humidity of 62.7%. Forest tent caterpillars, *Malacosoma disstria* Hübner (Lepidoptera: Lasiocampidae), were from the Great Lakes Forestry Centre (NRCan, Sault Ste Marie, Canada). FTCs were reared and maintained on artificial diet (Addy 1969) at 27 °C, 50–60% relative humidity, 16/8-h photoperiod.

cDNA libraries

For a description of plant materials used in the construction of cDNA libraries please see Table 1. Total RNA was isolated according to the protocol of Kolosova *et al*. (2004), followed by poly(A)+ RNA purification with oligo d(T) cellulose using the Poly(A) Pure Kit (Ambion), following the manufacturer's instructions. Total RNA was quantified and quality checked by spectrophotometer and agarose gel. RNA was also evaluated for integrity and the presence of contaminants using reverse-transcription with Superscript II reverse transcriptase (Invitrogen) with an oligo d(T_{18}) primer and αP^{32} dGTP incorporation. After removal of unincorporated nucleotides using gel filtration columns (Microspin S-300 HR columns, Amersham Pharmacia Biotech), the resulting cDNA smear was resolved using a vertical 1% agarose alkaline gel and visualized using a Storm 860 phosphorimager (Amersham Pharmacia Biotech). Standard cDNA libraries were directionally constructed (5′ *Eco*RI and 3′ *Xho*I) using 5 μg of poly(A)+ RNA

Table 1 Libraries, tissue sources and species for EST sequences described in this study

cDNA library	Tissue/developmental stage	Species (genotype)
PT-X-FL-A-1‡	Outer xylem§	*Populus trichocarpa* (Nisqually 1)
PT-P-FL-A-2‡	Phloem and cambium§	*P. trichocarpa* (Nisqually 1)
PT-GT-FL-A-3‡	Young and mature leaves, along with green shoot tips§	*P. trichocarpa* (Nisqually 1)
PTxD-IL-FL-A-4‡	Local and systemic (above region of feeding) mature leaves harvested after continuous feeding by forest tent caterpillars, *Malacosoma disstria*. Local tissue was collected 4, 8 and 24 h post-treatment and systemic tissue 4, 12 and 48 h post-treatment.¶	*P. trichocarpa* × *deltoides* (H11-11)
PTxD-IL-A-5*	Local mature leaves harvested after continuous feeding by *M. disstria*. Tissue was collected 2, 12 and 24 h post-treatment.¶	*P. trichocarpa* × *deltoides* (H11-11)
PTxN-IB-A-6*	Bark (with phloem and cambium attached) harvested after continuous feeding by willow weevil, *Cryptorhynchus lapathi*. Tissue was collected 2, 6 and 48 h post-treatment.¶	*P. trichocarpa* × *nigra* (NxM6)
PT-DX-A-7*	Outer xylem harvested biweekly between April and October 2002††	*P. trichocarpa* (VT-125)
PTxD-NR-A-8*	Three-month old sapling trees grown in aerated hydroponic media in growth chambers. Roots were harvested from trees grown in media without a nitrogen source for 24 and 48 h, as well as trees grown in regular media.	*P. trichocarpa* × *deltoides* (H11-11)
PTxD-IL-N-A-9†	Local mature leaves harvested after continuous feeding by *M. disstria*. Tissue was collected 2, 12 and 24 h post-treatment.¶	*P. trichocarpa* × *deltoides* (H11-11)
PT-DX-N-A-10†	Outer xylem harvested biweekly between April and October 2002††	*P. trichocarpa* (VT-125)
PTxN-IB-N-A-11†	Bark (with phloem and cambium attached) harvested after continuous feeding by *C. lapathi*. Tissue was collected 2, 6 and 48 h post-treatment.¶	*P. trichocarpa* × *nigra* (NxM6)
PTxD-ICC-A-12*	Cultured cells grown in media supplemented with salicylic acid, benzothiadiazole, methyl jasmonate, chitosan or *Pollacia radiosa* extract. Cells harvested 3 h post-treatment, along with untreated control cells.	*P. trichocarpa* × *deltoides* (H11-11)
PT-MB-A-13*	Terminal vegetative buds‡‡	*P. trichocarpa* (wild genotype)
PTxD-ICC-N-A-14†	Cultured cells grown in media supplemented with salicylic acid, benzothiadiazole, methyl jasmonate, chitosan or *Pollacia radiosa* extract. Cells harvested 3 h post-treatment, along with untreated control cells.	*P. trichocarpa* × *deltoides* (H11-11)
PT-MB-N-A-15†	Terminal vegetative buds‡‡	*P. trichocarpa* (wild genotype)

*Standard cDNA library.
†Normalized cDNA library.
‡Full-length cDNA library.
§Harvested 15 May 2001 from 8-year-old trees within the Boise Cascade region of Washington State.
¶One- or 2-year old saplings grown in potted soil under greenhouse conditions at the University of British Columbia.
††Five-year old trees grown outdoors under natural conditions at the University of British Columbia South Campus farm.
‡‡Harvested 19 September 2001 from 20-year-old trees near Corvallis, Oregon.

and the pBluescript II XR cDNA library construction kit, following manufacturer's instructions (Stratagene) with modifications. Briefly, first strand synthesis was performed using Superscript II reverse transcriptase (Invitrogen) and an anchored oligo d(T) primer (5′-GAGAGAGAGAGAGA-GAGAGAACTAGTCTCGAGTTTTTTTTTTTTTTTTTVN-3′). Size fractionation was performed on *Eco*RI and *Xho*I-digested cDNA immediately prior to ligation into the vector using a 1% NuSieve GTG low melting point agarose gel (BioWhittaker Molecular Applications) and β-agarase (New England Biolabs) to isolate cDNAs ranging from 300 bp to 5 kb. Select cDNA libraries were normalized to Cot = 5 by using the Soares method (Soares *et al*. 1994; Bonaldo *et al*. 1996). The average insert size of cDNA libraries was routinely determined by performing colony polymerase chain reaction (PCR) on 48 randomly selected bacterial colonies from the amplified library using –21M13 forward (5′-TGTAAAACGACGGCCAGT-3′) and M13 reverse (5′-CAGGAAACAGCTATGAC-3′) primers. PCR amplicons were resolved on 1% agarose gels and visually compared to DNA size markers λ *Hin*dIII and 1 kb ladder (Invitrogen). FLcDNA libraries were constructed according to the methods of Carninci *et al*. (1996), with modifications, and will be described in detail elsewhere. Unless otherwise mentioned, all other reagents and solvents were from Fischer Scientific, Sigma-Aldrich, EM Science or Invitrogen.

DNA sequence analysis and contig assembly

For complete details of bacterial transformation with library plasmids, colony picking, DNA purification and evaluation, and DNA sequencing, please see Appendix S1

(Supplementary materials). DNA sequence chromatograms were processed using the PHRED software (versions 0.000925.c and 0.020425.c) (Ewing & Green 1998; Ewing *et al*. 1998). Sequences were quality-trimmed according to the high-quality contiguous region determined by PHRED and then vector-trimmed using CROSS_MATCH software (www.phrap.org). Sequences with less than 100 quality bases (PHRED 20 or better) after trimming were discarded. Sequences having polyA tails of 100 bases or more were eliminated from analysis. These sequences consistently resolve as poor quality sequences resulting from *Taq* polymerase slippage across a low-complexity region. EST sequences representing contamination from bacterial, yeast or fungal sources were identified using BLAST (Altschul *et al*. 1990, 1997) and removed from analysis. The *Escherichia coli* K12 DNA sequence (GI: 6626251) was used for bacterial contamination monitoring. The TIGR Gene Index database for *Aspergillus nidulans* (ANGI.060302), a GenBank *Saccharomyces cerevisiae* database (ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/yeast.nt.gz) (26 November 2003) and a custom database of *Agrobacterium tumefaciens* generated using SRS (Lion Biosciences) were used to monitor and filter out additional contamination. Sequences were also compared to the GenBank nonredundant (nr) database using BLASTX. BLAST hits with expect values (E) < 10^{-10} and a top rank among scores relative to a range of other plant sequence databases classified ESTs as contamination and these were removed prior to EST assembly. CAP3 (Huang & Madan 1999) was used to assemble ESTs into contigs using the parameters of 40 bp overlap and 95% identity.

Treatment of trees with FTC

For insect treatments, groups of third to fifth instar larvae were used in the experiments. Prior to being placed on plants, FTCs were starved for 40 h on moist filter paper. Five FTCs were added to each of five trees under mesh bags on individual trees. Insects were caged on the five lowest, fully expanded, healthy leaves of each tree. An additional set of five trees was covered with mesh bags, but were otherwise left untreated. Leaves with petioles removed were individually harvested from each FTC-treated and untreated control tree 24 h after the onset of FTC feeding, and separately flash frozen in liquid nitrogen and stored at –80 °C prior to RNA isolation following the protocol of Kolosova *et al*. (2004). RNA quality and quantity were evaluated as described above.

Microarray hybridization and analysis

For complete details of cDNA microarray fabrication and quality control, along with a detailed hybridization protocol, please see Appendix S1. 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 ImaGene-quantified data files are available at http://douglas.bcgsc.bc.ca. Ten hybridizations were performed comparing FTC-treated poplar leaves after 24 h of continuous feeding and untreated control leaves harvested at the same time. RNA isolated from each of the five individual FTC-treated trees was compared directly against the five individual untreated control trees using two hybridizations with a dye flip for each tree pair. Similarly, total RNA from a single pool of untreated control leaves from 5 trees was compared in self-self hybridizations with the same RNA in each channel, independently converted to cDNA.

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 nonlinearity of intensity distributions using the variance stabilizing normalization (vsn) method (Huber *et al*. 2002). In order to assess performance of the poplar 15.5K microarray, a model containing a dye effect and a treatment effect for untreated control (C) minus C′ was fit using data from four microarray slides. The C minus C′ effect was computed by assigning two of the self-hybridized arrays to be Cy5–Cy3 and two to be Cy3–Cy5. Three analyses were done such that each self-hybridized array was paired with each other self-hybridized array only once, either as a Cy5−Cy3 or a Cy3−Cy5 combination. Expression variance was derived from technical variance between slides. The ratio of the C minus C′ parameter estimate to the standard error was used to calculate a *t* statistic, from which a *P* value was obtained. In order to assess the biological response to FTC herbivory (H), a mixed-effects model containing a dye effect and a treatment effect for H minus C was fit using data from 10 microarray slides derived from five pairs of H vs. C trees with a dye flip for each pair. Expression variance was obtained from two sources, biological and technical. The variance between hybridizations from the same H vs. C tree pair is solely technical, whereas the variance between hybridizations in different tree pairs is a combination of biological and technical. Dye and H minus C effects, as well as biological and technical variation were estimated using a mixed-effects model where the error term for the H minus C and dye effects was computed by pooling the biological and technical variation. Since all 10 slides contained technical variation but only five pairs have biological variation, the biological term was given twice the weight of the technical term in the pooled estimate. Next, the ratio of the H minus C parameter estimate to the standard error was used to calculate a *t* statistic and *P* value. The *Q* value for each effect and gene was calculated for each of the two models to adjust for the false discovery rate (Storey & Tibshirani 2003). All statistical analyses were performed within the R statistical package (www.r-project.org/).

Real-time PCR

Prior to reverse transcription, 15 μg total RNA per tree was treated with DNaseI (Invitrogen) digestion 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(T_{18}) primer according to the 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 (Table S3, Supplementary materials) using a stringent set of criteria including predicted melting temperature of 64 ± 2 °C, primer lengths of 20–24 nucleotides, guanine-cytosine contents of 40–60% and PCR amplicon lengths of 100–350 bp. In addition, when possible, at least one primer of a pair was designed to cover an exon-exon junction according to the gene structure models at http://genome.jgi-psf.org/Poptr1/Poptr1.home.html. Primer specificity (single product of expected length) was confirmed by analysis on a 2% agarose gel, by melting curve analysis and for at least one PCR per gene, by sequence verification of PCR amplicons (data not shown). Primers for poplar translation initiation factor 5A (TIF5A) were designed (GenBank Accession no. CV251327; poplar EST WS0116_J23) and served as a quantification control. Real-time PCR 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 the 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 H_2O or 10 ng of nonreverse transcribed RNA were run with each primer pair as a control. To further evaluate the efficiency and amplification performance of each primer pair, a tenfold-dilution series of corresponding DNA plasmids (10^{-2} to 10^{-6} ng template) was analysed with a minimum of three independent technical replicates per dilution (data not shown). The programme for all PCRs 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 analysed using the Opticon Monitor 2 version 2.02 software (MJ Research). For each primer pair and tree a minimum of three independent technical replicates were performed. To generate a baseline-subtracted plot of the logarithmic increase in fluorescence signal (ΔR_n) vs. cycle number, baseline data were collected between cycles 3 and 10. All amplification plots were analysed with an R_n threshold of 0.003 to obtain C_t (threshold cycle) values. Transcript abundance for each FTC-responsive gene was normalized to TIF5A by subtracting the C_t value of TIF5A from the C_t value of each FTC-responsive transcript, where $\Delta C_t = C_{t\ \text{transcript}} - C_{t\ \text{TIF5A}}$. Transcript abundance of FTC-responsive genes in control and FTC-treated samples were obtained from the equation $(1 + E)^{-\Delta Ct}$, 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 response to FTC herbivory (H), a mixed-effects model for each gene containing a treatment effect for H minus untreated control (C) was fit using data from at least three independent technical replicates derived from each of the five FTC-treated and five untreated control trees. H minus C effects, as well as biological and technical variation, were estimated using a mixed-effects model where the error term for the H minus C effect was computed by pooling the biological and technical variation. The ratio of the H minus C parameter estimate to the standard error was used to calculate a t statistic and P value.

Results and discussion

Sequencing and assembly of poplar ESTs

Large-scale EST sequencing has proven to be an efficient approach to capture much of the expressed gene catalogue of an organism. The aims of this project were to advance gene discovery and transcript profiling for poplar tissues with an emphasis on biotic stress response. We obtained ESTs from a set of 15 unidirectional standard, normalized or FLcDNA libraries generated from tissues at various developmental stages and treatments sampled from five *Populus* genotypes (*P. trichocarpa*, Nisqually 1; *P. trichocarpa*, VT-125; *P. trichocarpa*, wild genotype; *P. trichocarpa* × *deltoides*, H11-11; and *P. trichocarpa* × *nigra*, NxM6; Table 1). Libraries were constructed using trees grown in the wild, in the greenhouse, under hydroponic conditions, or as cultured cells. Tissues included in libraries were subjected to a variety of stress treatments including nitrogen deprivation; elicitors such as salicylic acid, benzothiadiazole, methyl jasmonate, chitosan or *Pollacia radiosa* fungal extract; and herbivory by stem-boring willow weevils or defoliating FTC larvae.

A total of 139 007 sequences were generated consisting of 107 519 3′-end reads and 31 488 5′-end reads (Table 2). Trimming low-quality (see Table 2 for quality criteria) and vector sequences, and removing contaminant bacterial, yeast or fungal sequences provided a data set of 90 368 hq 3′ ESTs with a minimum length of 100 bp (Table 2). The 84.0% success rate of hq 3′ ESTs compares favourably with other recent large-scale tree EST projects [e.g. loblolly pine 79.6% (Kirst *et al*. 2003); poplar 72.8% (Sterky *et al*. 2004)]. We focused initially on 3′-end sequencing to minimize separating EST sequences representing the same transcript into different contigs, which often occurs with 5′ ESTs from standard cDNA libraries due to variable truncation positions

Table 2 Poplar EST summary*

Total sequences	139 007
Number of 5′ sequences	31 488
Number of 3′ sequences	107 519
Average assembled EST length (bp)†	643
Number of high-quality 3′ sequences‡	90 368
Number of contigs§	14 451
Number of singletons	20 560
Number of putative unique transcripts¶	35 011
Average number of contig members	4.83
Number of contigs containing	
2 ESTs	5 632
3–5 ESTs	5 693
6–10 ESTs	2 062
11–20 ESTs	775
21–50 ESTs	252
> 50 ESTs	37

*Assembled from the 18 March 2005 version of the poplar EST database using CAP3.
†High-quality (hq) sequences only.
‡A sequence is considered hq if its trimmed PHRED 20 length is > 100 bases after vector only, low-quality and contaminating yeast, bacterial or fungal sequences are removed.
§A contig (contiguous sequence) contains two or more ESTs; 3′ sequences only.
¶Number of putative unique transcripts among assembled 3′ ESTs equals the number of contigs plus the number of singletons.

at the 5′-end of cDNAs. The average read length of these hq 3′ ESTs was 643 bp (Table 2), which is substantially longer than other large-scale tree EST sequencing programs (364 bp, Kirst *et al*. 2003; 470 bp, Sterky *et al*. 2004). In addition to EST sequencing, we have also obtained high accuracy complete insert sequences for *c*. 4600 putative poplar FLcDNAs that will be reported in detail elsewhere (S. Ralph, R. Kirkpatrick & J. Bohlmann, in prep.).

The 90 368 hq 3′-end ESTs were assembled using the CAP3 program. Among these ESTs, 69 808 assembled into a total of 14 451 contigs, and the remaining 20 560 ESTs were classified as singletons, suggesting a combined total of 35 011 putative unique transcripts (Table 2). These sequences represent a substantial portion of the complete gene content in poplar, which is estimated at *c*. 45 000 proteins from a first draft of the poplar genome sequence. Contigs contained an average of 4.83 assembled EST sequences. Only 37 contigs consist of greater than 50 ESTs (Table 2) and the five largest contigs contain 591 (unknown function), 437 (metallothionein), 283 (ribulose biphosphate carboxylase small subunit), 198 (metallothionein) and 167 (metallothionein) ESTs, respectively. Mitochondrial and chloroplast RNA sequences were not filtered, but they contribute 126 (0.14%) and 367 (0.41%) ESTs to the entire data set, respectively. All hq sequences have been deposited in the dbEST division of GenBank (Accession nos DT469172–DT526799; CV225307–CV284047).

Quality and complexity of cDNA libraries and gene discovery

Sequences from each cDNA library were closely monitored to assess library complexity and sequence quality to gauge overall suitability for further sequencing. From each cDNA library, between 1536 and 18 432 clones were 3′-end sequenced, and from selected libraries, between 1536 and 8448 of these same clones were also sequenced from the 5′-end (Table 3). The rate of hq 3′ ESTs obtained from cDNA libraries ranged from 59.5% to 96.8%, with the lower pass rates resulting from FLcDNA libraries due to the frequent occurrence of long polyA tails. The average length of hq 3′-end ESTs among cDNA libraries ranged from 540 bp to 711 bp.

EST sequencing is a powerful method for gene discovery, but there are limitations to this approach. One limitation, in particular, is the redundant generation of ESTs derived from the most common transcripts that can reduce the overall efficiency of gene discovery that relies solely on the generation of ESTs from standard (i.e. non-normalized) cDNA libraries. Accordingly, the application of normalization strategies to equalize the abundance of all transcripts (Soares *et al*. 1994; Bonaldo *et al*. 1996) has proven to be advantageous (Marra *et al*. 1999; Scheetz *et al*. 2004). We assessed the rate of gene discovery for each poplar library by calculating the number of unique transcripts within the set of ESTs derived from each library, the average number of contig members, the percentage of ESTs with no BLASTN match to public *Populus* ESTs, the percentage of singleton ESTs, and the percentage of library-specific unique transcripts (Table 3). By using any of these five metrics, four of our five normalized libraries are of higher complexity (and higher rates of gene discovery) than the corresponding non-normalized libraries (Table 3). The one exception is library PTxD-IL-N-A-9 for which normalization was likely ineffective due to insufficient hybridization stringency (Cot = 5).

Among the four successfully normalized libraries, the percentage of unique transcripts identified within the first 1000 hq 3′ sequences averaged 94.5% and ranged from 93.0% to 96.4%, whereas among the corresponding standard libraries constructed from the same RNA sources the average was only 85.2% and ranged from 83.2% to 86.9% (Table 3). Similarly, the percentage of singletons within the first 1000 hq 3′ sequences averaged 89.9% from the four normalized libraries (range of 87.2% to 93.2%) compared to only 76.2% (range of 74.1% to 78.3%) among corresponding standard libraries. The average number of contig members was the same for standard libraries (2.89 average) and normalized libraries (2.88 average), even though a considerably larger number of EST sequences were generated from normalized libraries (Table 3). Likewise, within the set of ESTs derived from each library, the percentage of 3′ ESTs with

Table 3 Poplar cDNA library summary statistics[a]

cDNA library	Total of no. seq. (5′ ESTs)	No. of high-quality seq.[d]	Pass rate %[e]	Avg. EST length (bp)[f]	No. and % of unique transcripts[g]	Avg. no. of contig members[h]	% no BLASTN match vs. public poplar ESTs[i]	% singleton[j]	% library specific transcripts[k]	No. on 15.5K array
PT-X-FL-A-1[b]	3 072 (1536)	914	59.5	598	586 (n.a)	3.67	23.31/29.35	50.6 (n.a)	0.43 (1.01)	368
PT-P-FL-A-2[b]	15 744 (7296)	6 219	73.6	622	2 931 (75.9)	4.25	2.59/7.62	30.8 (62.8)	2.83 (6.88)	2439
PT-GT-FL-A-3[b]	19 968 (8448)	6 871	59.6	652	3 400 (77.2)	3.98	4.38/11.53	32.5 (64.7)	3.78 (7.60)	1765
PTxD-IL-FL-A-4[b]	9 216 (4608)	2 812	61.0	699	1 305 (54.0)	5.26	2.70/6.76	33.8 (43.2)	1.29 (3.11)	299
PTxD-IL-A-5[b]	3 072	2 777	90.3	655	1 518 (63.8)	5.53	1.83/6.29	44.6 (54.6)	1.39 (3.07)	975
PTxD-IL-N-A-9[c]	4 608	4 461	96.8	680	2 141 (62.1)	5.66	1.52/5.03	36.8 (52.7)	2.08 (4.93)	1013
PTxN-IB-A-6[b]	3 072	2 816	91.6	633	2 122 (86.9)	2.90	3.70/11.86	62.3 (78.3)	1.59 (3.11)	1496
PTxN-IB-N-A-11[c]	18 432	17 314	93.9	617	9 540 (93.1)	3.38	4.91/14.32	36.2 (87.2)	13.77 (19.15)	1691
PT-DX-A-7[b]	3 072	2 765	90.0	593	2 026 (86.1)	2.90	4.37/13.55	59.2 (77.0)	1.75 (3.05)	1387
PT-DX-N-A-10[c]	7 680	6 457	84.0	540	4 712 (93.0)	2.56	5.34/16.99	55.6 (87.2)	5.52 (7.14)	1351
PTxD-ICC-A-12[b]	1 536	1 215	79.1	636	993 (83.2)	2.83	6.38/15.30	71.7 (74.1)	0.65 (1.34)	414
PTxD-ICC-N-A-14[c]	20 352 (4992)	13 929	90.6	643	9 236 (95.8)	2.76	10.57/22.93	47.1 (92.1)	13.25 (15.41)	434
PT-MB-A-13[b]	1 536	1 302	84.7	674	1 059 (84.6)	2.96	6.93/15.35	71.8 (75.7)	0.78 (1.44)	494
PT-MB-N-A-15[c]	23 039 (4608)	16 901	91.6	711	10 819 (96.4)	2.85	11.91/23.69	44.5 (93.2)	16.29 (18.70)	0
PTxD-NR-A-8[b]	4 608	3 615	78.4	622	2 631 (85.2)	3.22	7.79/16.44	60.5 (76.3)	3.33 (4.00)	1360

[a]Assembled from the 18 March 2005 version of the poplar EST database using CAP3.
[b]Standard or full-length cDNA library (white background).
[c]Normalized cDNA library (grey background).
[d]A sequence is considered high-quality (hq) if its trimmed PHRED 20 length is > 100 bases after vector only, low-quality and contaminating yeast, bacterial or fungal sequences are removed; 3′ sequences only.
[e]Number of hq 3′ sequences as a percentage of total 3′ sequences.
[f]Hq 3′ sequences only.
[g]Number of putative unique transcripts among assembled hq 3′ ESTs equals the number of contigs plus the number of singletons. The number in parentheses is the percentage of unique transcripts among the first 1000 hq 3′ sequences assembled.
[h]Average number of contig members where a contig (contiguous sequence) contains two or more hq 3′ ESTs.
[i]Percentage of hq 3′ ESTs > 400 bp in length with no significant BLASTN similarity vs. a set of 247 353 *Populus* ESTs (excluding ESTs described in this study) from the 27 May 2005 version of the dbEST division of GenBank at low (left, expect value < $1e^{-20}$) and high (right, expect value < $1e^{-125}$) match stringency.
[j]Percentage singletons is the number of singletons in a library divided by the total number of hq 3′ sequences. The number in parentheses is the percentage of singletons among the first 1000 hq 3′ sequences assembled.
[k]Percentage of library specific unique transcripts calculated by adding the number of contigs and singletons that were present only in a single cDNA library divided by the total number of putative unique transcripts in the hq 3′ EST collection (35 011). The value in parentheses is the number of hq 3′ sequences derived from a given library expressed as a percentage of all hq 3′ sequences (90 368).

no BLASTN similarity to a collection of 247 353 *Populus* ESTs in the public domain (i.e. dbEST division of GenBank, excluding the ESTs described in this paper) at both low (E value < $1e^{-20}$) and high (E value < $1e^{-125}$) stringency was higher for normalized libraries (8.1% and 19.4% average, respectively) than standard libraries (5.3% and 14.0% average, respectively; Table 3). In addition, the percentage of library-specific transcripts was considerably higher for the four normalized libraries (80.5% average) compared to standard libraries (52.7% average; Table 3).

Collectively, these results indicate that the preparation of normalized cDNA libraries has greatly improved the complexity and rate of gene discovery within our poplar EST project. With 71% of hq 3′ ESTs derived from normalized cDNA libraries, it is not surprising that no unique transcripts were identified that contained ESTs obtained from all 15 cDNA libraries sequenced, and only three transcripts were present in 14 of 15 libraries, which likely represent 'housekeeping' genes (Table 4). Among this set, the most abundant housekeeping gene was identified as cyclophilin (152 ESTs), followed by elongation factor 1α (64 ESTs), and elongation factor 1β–α subunit (42 ESTs).

Table 4 Distribution of ESTs in multiple cDNA libraries

Libraries represented	Number of unique transcripts
15	0
14	3
13	5
12	7
11	30
10	46
9	78
8	122
7	235
6	326

Comparison against public Populus *ESTs, the poplar genome, and* Arabidopsis

In order to minimize redundant EST development relative to other poplar EST projects (Kohler *et al*. 2003; Sterky *et al*. 2004) and to support a genomics platform for the study

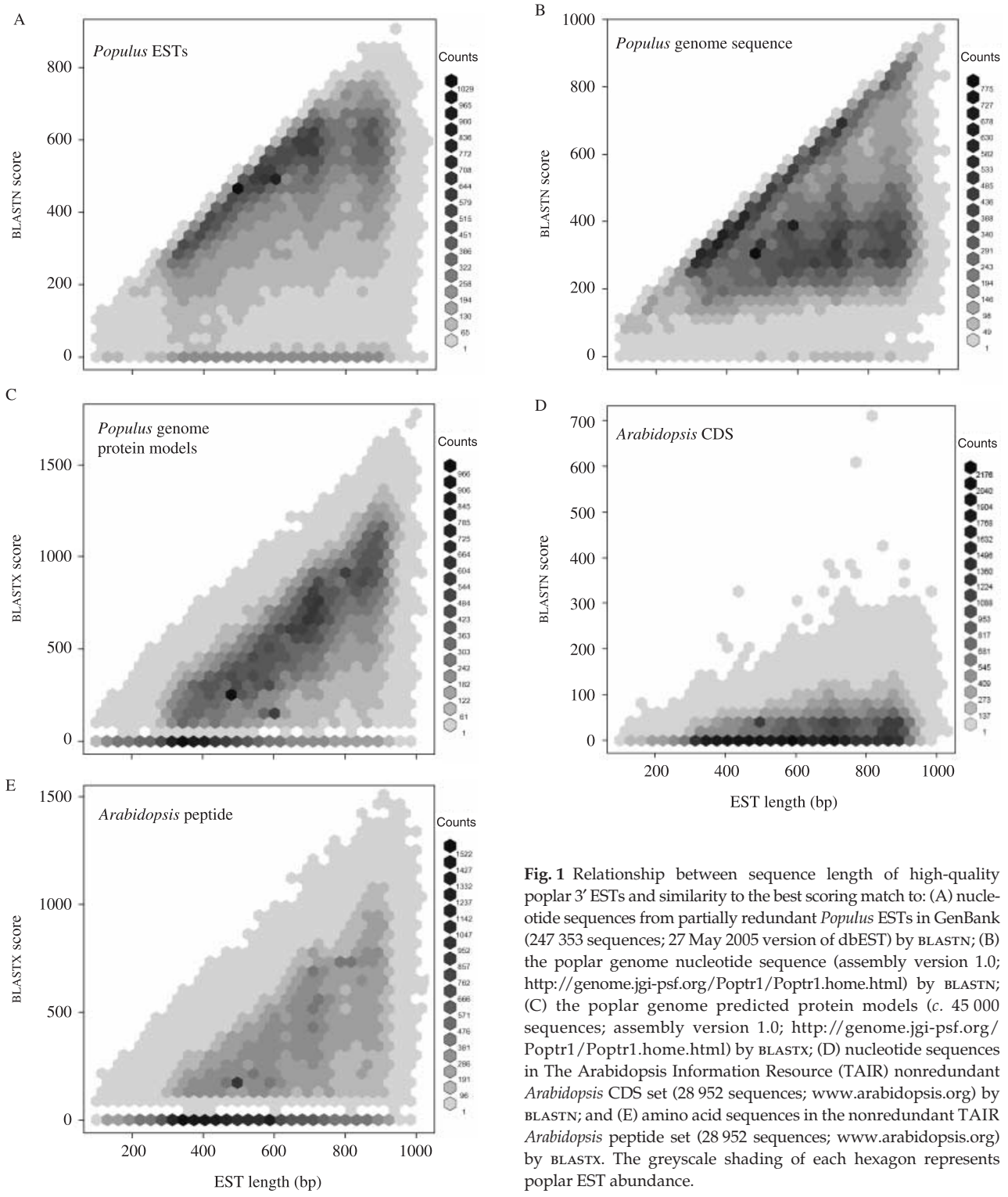

Fig. 1 Relationship between sequence length of high-quality poplar 3′ ESTs and similarity to the best scoring match to: (A) nucleotide sequences from partially redundant *Populus* ESTs in GenBank (247 353 sequences; 27 May 2005 version of dbEST) by BLASTN; (B) the poplar genome nucleotide sequence (assembly version 1.0; http://genome.jgi-psf.org/Poptr1/Poptr1.home.html) by BLASTN; (C) the poplar genome predicted protein models (*c.* 45 000 sequences; assembly version 1.0; http://genome.jgi-psf.org/Poptr1/Poptr1.home.html) by BLASTX; (D) nucleotide sequences in The Arabidopsis Information Resource (TAIR) nonredundant *Arabidopsis* CDS set (28 952 sequences; www.arabidopsis.org) by BLASTN; and (E) amino acid sequences in the nonredundant TAIR *Arabidopsis* peptide set (28 952 sequences; www.arabidopsis.org) by BLASTX. The greyscale shading of each hexagon represents poplar EST abundance.

of poplar–insect interactions, we developed normalized cDNA libraries to capture rare genes missed in other EST collections, and included tissues subjected to herbivory, elicitor, or pathogen treatment. Overall, among the 79 338 hq 3′ ESTs of more than 400 bp in length analysed in this project, there are 73 863 (93.0%) with similarity at low stringency (E value < $1e^{-20}$) to public *Populus* ESTs, and 66 589 (83.9%) with similarity at high stringency (E value < $1e^{-125}$) (Fig. 1A). These values increase to 93.7% (low stringency) and 87.4% (high stringency) when only hq 3′

ESTs > 800 bp in length (22 515) are considered. This suggests that *c.* 16% of the newly generated poplar ESTs are unique relative to any *Populus* EST previously in the public domain.

We then compared the new hq poplar 3′ EST sequences against the recently completed draft version of the poplar genome sequence and predicted genomic ORFs by BLASTN and BLASTX, respectively (Fig. 1). As expected, the similarity between ESTs and the poplar genome sequence is high, with 98.5% (89 081) of all hq 3′ ESTs having a BLASTN E value < $1e^{-5}$ (score > 25), and this increases only slightly to 98.7% (22 243) when only ESTs > 800 bp are considered (Fig. 1B). When analysing similarity between ESTs and predicted genome ORFs, we observed 87.0% (78 662) of hq 3′ ESTs with BLASTX E values < $1e^{-5}$ (score > 110), compared to 95.9% (21 604) for ESTs > 800 bp (Fig. 1C). The lack of sequence similarity between a small percentage of our ESTs and the poplar genome sequence and predicted genome ORFs could be due to contaminating species in cDNA libraries not removed through *in silico* screens (e.g. insects, yeast, bacteria, fungi, etc.), artefacts of annotation, and/or the different methodological limitations of either EST or genomic sequencing and assembly (e.g. long 3′ UTRs in EST sequences, ESTs representing organelle genes not represented in the nuclear genome sequence, gaps in the genomic DNA sequence assembly, incorrectly predicted genomic ORFs, alternate splice forms, etc.).

We also compared hq poplar 3′ ESTs generated in this project against The Arabidopsis Information Resource (TAIR) *Arabidopsis* CDS and peptide sets (28 952 sequences, www.arabidopsis.org; Fig. 1) by BLASTN and BLASTX, respectively. Overall, the similarity between poplar ESTs and *Arabidopsis* CDS sequences is relatively low with only 49.3% (44 600) of all hq 3′ ESTs having a BLASTN E value < $1e^{-5}$ (score > 25; Fig. 1D). This value increases to 64.5% (14 539) when only poplar ESTs of more than 800 bp are considered (Fig. 1D). The vast majority of ESTs less than 300 bp in length have low similarity to *Arabidopsis* peptides, whereas for ESTs more than 300 bp there is a positive correlation between EST length and BLASTX score (Fig. 1E). Overall, 76.6% (69 258) of hq 3′ ESTs have BLASTX E values < $1e^{-5}$ (score > 110) vs. *Arabidopsis* peptides, compared to 91.0% (20 508) for ESTs > 800 bp. This analysis identifies 2007 hq poplar ESTs in the sequences described here with BLASTX E values > $1e^{-5}$ and length > 800 bp (Table S1, Supplementary materials) that may represent genes lost during *Arabidopsis* evolution, gained during poplar evolution, or otherwise sufficiently diverged in sequence to no longer be recognizable as similar sequences since *Arabidopsis* and *Populus* diverged *c.* 100 million years ago. In a comparable analysis performed by Sterky *et al.* (2004) using *Populus* contig sequences derived from ESTs, they found 97.9% of contigs > 1000 bp (1089 of 1112) had a BLASTX score > 100 vs. *Arabidopsis* peptides.

Development of a poplar cDNA microarray

Based on the generation of poplar ESTs, we have developed a poplar cDNA microarray composed of 15 496 cDNA elements selected from 14 cDNA libraries representing leaves, buds, phloem, xylem, bark and root tissues, as well as cultured cells (Table 3). Clones on the array were selected from a CAP3 assembly of *c.* 37 000 3′-end ESTs and are enriched for EST sequences from elicitor- or herbivore-treated libraries (i.e. 6322 ESTs or 40.7%). Functional annotation of array elements has been assigned according to the TAIR *Arabidopsis* protein set using BLASTX, as well as using BLASTX vs. the set of *c.* 45 000 protein models predicted from the draft version of the poplar genome sequence. Overall, 11 418 (73.6%) of 15 496 spotted cDNAs have similarity to the TAIR *Arabidopsis* protein set by BLASTX (E < $1e^{-5}$), compared to 12 947 (83.6%) cDNAs with similarity to predicted poplar genome ORFs by BLASTX (E < $1e^{-5}$).

To perform an initial validation of the poplar 15.5K cDNA microarray performance, total RNA was isolated from five fully developed leaves from each of five 1-year-old poplar trees, pooled, and used to evaluate the false change rate. Four technical replicate hybridizations were performed using this same RNA source, independently converted to cDNA, and labelled with dendrimer 350 Cy3 or Cy5 fluorescent labels. Among the four hybridizations, the median foreground signal intensity for blank elements (Cy3-73.8; Cy5-112.7), DMSO buffer-only elements (Cy3-174.3; Cy5-279.4) and DNA-based negative control elements (Cy3-156.1; Cy5-241.7) was low compared to that of EST elements (Cy3-4371.4; Cy5-3759.3), indicating a low level of nonspecific hybridization. Among the three possible combinations for randomly assigning RNA in each channel to treatment groups when comparing the four slides in self-self hybridizations, the number of differentially expressed (fold-change > 1.5×; P value < 0.05) EST array elements was 94 (0.60%), 130 (0.83%) and 131 (0.84%) of 15 496 total. Differential expression in this case is due to technical variation between hybridizations. To estimate the false discovery rate (FDR), we calculated Q values (Storey & Tibshirani 2003) and found the FDR for the three possible self-self combinations was 100.0%, 86.2% and 56.2% at $P = 0.05$, and 100%, 86.2% and 48.9% at $P = 0.001$, respectively, suggesting there are no genes reliably differentially expressed when self-self hybridizations are performed (Table S2, Supplementary materials).

Microarray transcriptome profiling of FTC herbivory of poplar leaves

We utilized the poplar 15.5K cDNA microarray to examine global changes in gene expression in poplar in response to insect herbivory. Clonal hybrid poplar (*P. trichocarpa* × *deltoides*, H11-11 genotype) saplings were subjected to

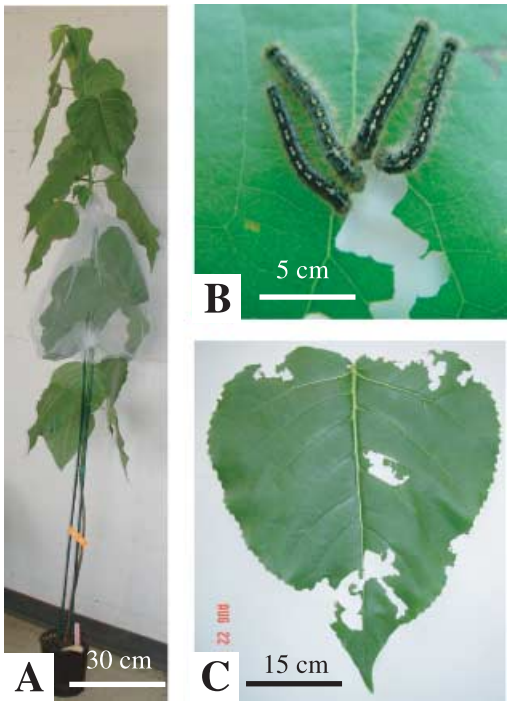

Fig. 2 Herbivory experiment set-up under greenhouse conditions. Insects were caged under mesh bags placed on the lowest five healthy mature leaves of 6-month-old poplar trees (A). Caterpillars inflict damage to leaf tissue by feeding from the outer edges (B). Representative image of leaf damage inflicted after 24 h of caterpillar feeding (C). Scale bars indicate approximate size.

feeding by FTC larvae caged on trees using mesh bags (Fig. 2). Differentially expressed genes were selected using two criteria: fold-change between FTC herbivory and untreated control > 1.5× and Student's *t*-test *P* value < 0.05. For a complete list of expression data for all genes represented on the microarray, see Table S2. Using these criteria, after 24 h of FTC feeding 1191 microarray elements were classified as up-regulated, compared to 557 down-regulated elements. We determined the FDR to be 7.5% at $P = 0.05$, diminishing to 0.3% at $P = 0.001$ (Table S2). As demonstrated by the boxplots in Fig. S1 (Supplementary materials), the majority of variation in our microarray experiments is derived from technical sources rather than biological, with the ANOVA estimate of technical variation being greater than biological variation for 91.1% of array elements (Table S2). This likely reflects the use of genetically identical plants that were treated uniformly under controlled greenhouse conditions. As illustrated in Fig. 3, representative scatter plots for self-self hybridizations (upper panel) and the response to FTC herbivory after 24 h of feeding (lower panel) demonstrate the broad dynamic range of the poplar 15.5K microarray in the detection of differentially expressed transcripts.

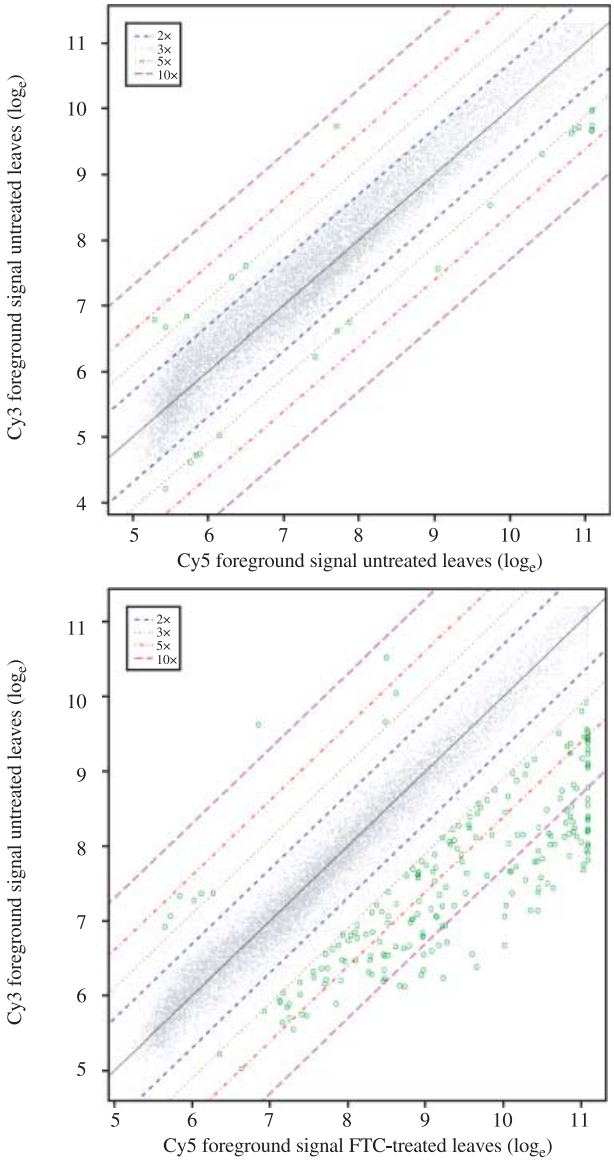

Fig. 3 Experiments conducted to assess the performance of the poplar 15.5K cDNA microarray. Upper panel: scatter plot illustrating the technical reproducibility and dynamic range of the poplar 15.5K microarray. Total RNA from untreated mature poplar leaves was labelled with either Cy3 or Cy5 fluors and directly compared on the same slide in four separate hybridizations (representative slide shown). Lower panel: scatter plot illustrating the detection of genes differentially expressed between total RNA from untreated mature leaves and leaves exposed to caterpillar feeding for 24 h (ten hybridizations performed, with dye-flips; representative slide shown). Two-, three-, five- and tenfold changes in gene expression are indicated by the parallel lines that flank the probe set data. Genes differing by greater than threefold between treatments are indicated by green open circles.

FTC-responsive genes identified in this analysis were classified into 10 functional groups based on annotation to *Arabidopsis*. The most prominent FTC-responsive genes of these 10 groups and all differentially expressed transcription

Table 5 Selected forest tent caterpillar-responsive array elements. A complete list of array elements is given in Table S2. Abbreviations: FC, fold-change; P, *P* value; Q, *Q* value. Colour scale from dark green to dark red correlates with fold-change expression. *Also analysed by real-time PCR (see Table 7 and Fig. 4).
Fold-change at least: −12, −6, −3, −1.5, ---, +1.5, +3, +6, +12

Clone ID	BLASTX vs. *Arabidopsis*	AGI code	E value	BLASTX vs. Poplar protein model	E value	FTC at 24 h FC	P	Q
No significant match to *Arabidopsis*								
WS0151_C13	No significant match	n.a.	n.a.	estExt_fgenesh4_pg.C_LG_X0093	2e-64	32.09	< 0.001	< 0.001
WS0192_L21	No significant match	n.a.	n.a.	estExt_Genewise1_v1.C_LG_VIII1172	1e-11	0.29	< 0.001	0.001
Biological process unknown								
WS0156_N20	Expressed protein	At5g58990	2e-41	estExt_Genewise1_v1.C_LG_I1066	2e-61	33.03	< 0.001	< 0.001
WS02010_G16	Senescence-associated protein	At1g53885	9e-25	gw1.158.158.1	1e-51	4.75	< 0.001	< 0.001
WS0153_M02	Induced upon wounding	At4g24220	8e-30	grail3.0040028802	5e-48	3.39	< 0.001	< 0.001
General metabolism								
WS0212_I21	Aminopeptidase M	At4g33090	3e-66	gw1.VI. 1753.1	8e-84	41.64	< 0.001	< 0.001
WS0124_K08	Apyrase	At5g18280	6e-20	eugene3.00190357	6e-38	40.92	< 0.001	< 0.001
WS0146_L23	Phosphorylase family protein	At4g24340	3e-29	eugene3.00870003	4e-99	33.35	< 0.001	< 0.001
WS0124_G12	Acid phosphatase	At4g25150	3e-53	estExt_Genewise1_v1.C_LG_I0436	2e-98	29.09	< 0.001	< 0.001
WS0156_L05	Thymidylate kinase	At5g59440	1e-58	gw1.I.2892.1	5e-80	25.28	< 0.001	< 0.001
WS01223_F23	Lycopene beta cyclase	At3g10230	1e-64	eugene3.00090104	1e-110	14.46	< 0.001	< 0.001
WS0132_A15	Lipase	At2g31100	3e-42	estExt_Genewise1_v1.C_LG_IV2794	4e-87	11.68	< 0.001	0.001
WS0145_F03*	Phytoene synthase	At5g17230	8e-77	fgenesh4_pg.C_LG_II000533	5e-88	5.97	< 0.001	0.002
WS01116_C06	Expansin	At4g38210	2e-56	estExt_fgenesh4_pg.C_LG_IX0014	8e-77	5.20	< 0.001	< 0.001
WS0154_D01	Carbonic anhydrase	At3g01500	7e-23	grail3.0005010201	3e-49	3.65	0.001	0.006
WS0231_K19	Galactinol synthase	At1g56600	7e-51	estExt_fgenesh4_pm.C_LG_XIII0025	4e-63	0.52	0.003	0.012
Photosynthesis								
WS02011_K12	Photosystem II protein	At1g03600	5e-44	estExt_Genewise1_v1.C_LG_V0741	4e-64	0.59	0.001	0.006
WS0142_N19	Chlorophyll A-B binding protein	At1g61520	7e-92	grail3.0012036701	1e-119	0.64	0.006	0.017
WS01224_M03	Ferredoxin reductase	At5g23440	7e-22	gw1.VIII.1629.1	1e-40	0.64	< 0.001	0.002
Transport								
WS0212_O05	ATPase	At2g24520	3e-97	estExt_fgenesh4_pg.C_1470038	1e-107	28.13	< 0.001	< 0.001
WS0156_A09	Metal transporter	At2g30080	9e-51	fgenesh4_pg.C_LG_IX000872	7e-65	14.92	< 0.001	< 0.001
WS0156_O13*	ABC transporter family protein	At1g65410	8e-32	gw1.179.28.1	3e-36	7.02	< 0.001	< 0.001
WS0114_H12	Major intrinsic protein	At4g01470	2e-85	estExt_fgenesh4_pg.C_LG_X1886	2e-97	6.35	< 0.001	< 0.001
WS0114_D04*	Calreticulin	At1g09210	6e-95	estExt_Genewise1_v1.C_LG_XIII0635	1e-112	5.09	< 0.001	< 0.001
WS0142_J05	Sec23A transport protein	At4g14160	2e-81	estExt_fgenesh4_pg.C_LG_XIV1163	1e-110	4.36	< 0.001	< 0.001
WS0144_M07	Lipid transfer protein	At2g44300	2e-16	grail3.0136006501	1e-77	0.63	0.001	0.006
Transcriptional regulation and signalling (also see Table 6)								
WS0205_I02	Leucine-rich repeat transmembrane protein kinase	At5g51560	2e-57	grail3.0033029601	3e-95	38.22	< 0.001	< 0.001
WS01223_D01	Choline kinase	At1g74320	2e-80	gw1.III.525.1	1e-104	28.47	< 0.001	< 0.001
Octadecanoid and ethylene signalling								
WS01212_L05	Lipoxygenase	At1g72520	1e-51	fgenesh4_pg.C_LG_I000375	1e-142	17.68	< 0.001	< 0.001
WS0155_D02*	Allene oxide cyclase	At1g13280	3e-69	gw1.IV.4073.1	4e-97	14.42	< 0.001	< 0.001
WS0145_B07*	Allene oxide synthase	At5g42650	1e-25	grail3.0040015701	2e-34	14.19	< 0.001	< 0.001
WS0123_G08	S-adenosylmethionine synthase	At1g02500	9e-30	estExt_Genewise1_v1.C_LG_XIII0489	2e-34	2.62	< 0.001	< 0.001
WS02011_I02*	1-amimocyclopropane-1-carboxylate oxidase	At1g05010	3e-34	eugene3.00002047	2e-45	2.41	< 0.001	0.001
Response to stress								
WS0134_G14*	Kunitz trypsin inhibitor	At1g17860	2e-05	estExt_fgenesh4_pg.C_LG_X0093	3e-40	32.31	< 0.001	< 0.001
WS0205_P08	Glycosyl hydrolase	At4g38650	2e-46	gw1.IX.1264.1	2e-39	23.80	< 0.001	< 0.001
WS0143_A03*	Basic endochitinase	At3g12500	4e-75	grail3.0001024001	1e-121	23.62	< 0.001	< 0.001
WS0144_M15	Stable protein 1	At3g17210	6e-36	estExt_fgenesh4_pm.C_LG_X0585	2e-58	14.91	< 0.001	< 0.001
WS0204_D16	Jacalin/lectin family protein	At1g19715	3e-21	estExt_Genewise1_v1.C_640646	7e-89	7.59	< 0.001	< 0.001
Secondary metabolism								
WS01212_M19	Cytochrome P450	At1g33730	1e-41	estExt_fgenesh4_pg.C_1580005	7e-64	12.39	< 0.001	0.002
WS01212_B20*	Isoflavone reductase	At4g39230	4e-37	eugene3.22230002	2e-99	10.49	< 0.001	< 0.001
TPS1*	(–)-germacrene D synthase (GenBank AAR99061)	At5g23960	1e-125	eugene3.00130815	0	7.07	< 0.001	< 0.001
PPO1*	Polyphenol oxidase PPO1 (GenBank AAG21983)	n.a.	n.a.	gw1.XI.3509.1	0	6.93	< 0.001	< 0.001
WS01214_G19	3-deoxy-D-arabino-heptulosonate 7-phosphate synthase	At4g39980	5e-16	gw1.V.2753.1	2e-24	5.33	< 0.001	0.001
WS0201_G13*	4-coumarate CoA ligase	At1g20510	3e-53	fgenesh4_pg.C_LG_V001627	1e-61	3.47	< 0.001	< 0.001
WS02011_F21	Laccase/diphenol oxidase	At5g09360	1e-56	estExt_Genewise1_v1.C_LG_XIX2228	1e-101	3.26	0.004	0.013
WS0168_K06*	Chorismate synthase	At1g48850	8e-15	grail3.0049006403	9e-23	3.25	< 0.001	< 0.001
WS01110_A05*	5-enolpyruvylshikimate3-phosphate synthase	At1g48860	3e-19	eugene3.00021350	2e-22	3.05	< 0.001	< 0.001
WS0154_H05	Phenylalanine ammonia- lyase	At2g37040	8e-57	estExt_Genewise1_v1.C_280658	3e-67	3.00	0.011	0.027
WS01224_J04	Chalcone synthase	At5g13930	2e-70	eugene3.00140920	3e-78	2.99	0.003	0.010
Detoxification, redox processes								
WS0145_I01	Superoxide dismutase copper chaperone	At1g12520	5e-36	gw1.III.445.1	6e-53	28.74	< 0.001	< 0.001
WS0178_N24	Glutathione S-transferase	At2g29420	4e-27	eugene3.00161127	2e-78	24.25	< 0.001	< 0.001
WS0162_C15	Thioredoxin	At2g01270	2e-41	estExt_fgenesh4_pm.C_LG_VIII0540	3e-69	16.57	< 0.001	0.001
WS0144_C22	Peroxidase	At5g05340	6e-82	estExt_Genewise1_v1.C_LG_XIII0228	1e-114	11.25	< 0.001	< 0.001

Table 6 Selected forest tent caterpillar-responsive transcription factors. A complete list of array elements is given in Table S2. Abbreviations: FC, fold-change; P, *P* value; Q, *Q* value. Colour scale from dark green to dark red correlates with fold-change expression. *Also analysed by real-time PCR (see Table 7 and Fig. 4).
Fold-change at least: −12, −6, −3, −1.5, ---, +1.5, +3, +6, +12

Clone ID	BLASTX vs. *Arabidopsis*	AGI code	E value	BLASTX vs. Poplar protein model	E value	FTC at 24 h		
						FC	P	Q
WS0171_K21	Zinc finger C3H type	At3g58030	6e-19	estExt_Genewise1_v1.C_LG_XVI1842	7e-64	11.96	< 0.001	< 0.001
WS0144_I19	Zinc finger C3H type	At1g19310	5e-19	eugene3.00021229	3e-40	4.40	< 0.001	< 0.001
WS0232_C14	Zinc finger C3H type	At5g48655	8e-16	fgenesh4_pg.C_LG_III001250	1e-94	2.28	< 0.001	0.001
WS01119_O05	Zinc finger C3H type	At5g59000	1e-41	gw1.I.1077.1	3e-88	1.82	< 0.001	0.001
WS0168_E11	Zinc finger C3H type	At3g55530	1e-48	estExt_fgenesh4_pm.C_LG_IX0591	2e-74	0.65	0.002	0.008
WS01218_M01	AP2-EREBP	At1g77200	3e-43	gw1.XVIII.2541.1	5e-52	3.83	< 0.001	0.003
WS0206_N03	AP2-EREBP	At3g23240	4e-35	gw1.V.1199.1	2e-50	2.11	< 0.001	< 0.001
WS0223_O23	AP2-EREBP	At1g78080	4e-33	fgenesh4_pm.C_LG_II000437	4e-49	2.07	< 0.001	< 0.001
WS0205_P06	AP2-EREBP	At2g28550	2e-07	gw1.X.2501.1	5e-07	1.51	0.041	0.066
WS0183_E12	Trihelix	At5g14540	2e-29	fgenesh4_pm.C_LG_I000853	3e-71	3.27	< 0.001	0.003
WS0193_E15	Basic helix-loop-helix	At4g17880	3e-29	estExt_fgenesh4_pm.C_1180004	1e-47	3.17	< 0.001	< 0.001
WS0205_O10	Basic helix-loop-helix	At1g05710	6e-39	eugene3.01420061	5e-61	2.12	0.002	0.010
WS0204_K10	Basic helix-loop-helix	At2g46510	8e-21	fgenesh4_pg.C_LG_II001585	1e-40	2.06	< 0.001	0.002
PX0011_G13	Basic helix-loop-helix	At1g69010	4e-15	eugene3.00150276	3e-97	1.66	< 0.001	0.001
WS0148_I15	Basic helix-loop-helix	At2g46510	3e-25	fgenesh4_pg.C_LG_II001585	9e-59	1.59	0.001	0.005
WS0161_N01	Basic helix-loop-helix	At3g47640	7e-42	grail3.0005058901	1e-105	0.62	< 0.001	0.001
WS0181_F07*	Myb	At4g37260	2e-40	gw1.II.3970.1	4e-49	2.89	0.010	0.025
WS0232_N20	Myb	At2g37630	2e-23	gw1.123.183.1	8e-63	0.66	0.005	0.016
WS0158_M20	Auxin response factor	At5g62000	9e-43	eugene3.00150845	3e-84	2.69	< 0.001	0.001
WS0168_N18	Zinc finger C2C2-GATA	At3g54810	2e-36	eugene3.00080330	1e-74	2.66	< 0.001	< 0.001
WS0233_M05	Zinc finger C2C2-GATA	At3g54810	4e-34	estExt_fgenesh4_pg.C_LG_X2015	2e-58	2.05	< 0.001	< 0.001
WS0178_A09	Zinc finger C2C2-GATA	At4g24470	2e-67	eugene3.00050602	1e-125	0.60	0.003	0.010
WS0131_P03	Zinc finger C2C2-GATA	At5g56860	3e-06	grail3.0013019801	2e-29	0.36	< 0.001	0.001
WS0201_I18	WRKY	At1g80840	4e-38	eugene3.00061944	1e-65	2.46	< 0.001	0.003
WS0231_E05*	WRKY	At5g13080	1e-38	grail3.0005030601	7e-56	2.44	0.008	0.021
WS0214_A24	WRKY	At3g56400	3e-30	grail3.0023037401	1e-137	2.19	< 0.001	0.001
WS0183_I09	WRKY	At4g22070	2e-44	estExt_Genewise1_v1.C_LG_XIV3374	4e-72	1.70	0.001	0.006
WS0125_A17	WRKY	At1g80840	3e-26	estExt_fgenesh4_pm.C_LG_III0624	1e-41	1.63	0.003	0.010
WS0181_A04	bZIP	At5g28770	6e-30	gw1.70.190.1	4e-80	2.34	0.002	0.009
WS0221_P11	bZIP	At2g40950	2e-51	grail3.0024020901	1e-95	2.02	0.001	0.004
WS0206_K17	bZIP	At3g62420	2e-18	eugene3.00080995	2e-69	1.67	0.002	0.008
WS0151_G22	bZIP	At1g75390	2e-39	estExt_Genewise1_v1.C_LG_V1456	4e-75	0.61	0.005	0.016
WS0151_P20	bZIP	At5g28770	4e-40	estExt_Genewise1_v1.C_LG_XIII1499	1e-121	0.57	< 0.001	0.001
WS0206_N15	bZIP	At1g75390	2e-37	estExt_fgenesh4_pg.C_LG_IV1320	7e-73	0.55	< 0.001	< 0.001
WS0202_B20	Myb-related	At2g46410	1e-12	eugene3.00070231	5e-25	2.24	< 0.001	0.002
WS0205_L05	No apical meristem (NAM/NAC)	At4g27410	4e-07	estExt_Genewise1_v1.C_LG_XI3994	1e-81	2.23	0.019	0.038
WS0224_B09	No apical meristem (NAM/NAC)	At1g01720	2e-96	grail3.0003068301	1e-136	1.96	< 0.001	0.002
WS0214_L13	No apical meristem (NAM/NAC)	At1g01720	1e-14	eugene3.00050086	5e-55	1.82	0.002	0.009
WS0122_J23	No apical meristem (NAM/NAC)	At2g02450	5e-15	estExt_Genewise1_v1.C_LG_IV1433	3e-33	1.66	0.007	0.018
WS0234_I22	No apical meristem (NAM/NAC)	At5g08790	1e-69	grail3.0011008901	1e-85	1.61	0.002	0.007
WS0212_H16	No apical meristem (NAM/NAC)	At1g01720	6e-14	grail3.0003068301	8e-33	1.61	0.006	0.018
WS0175_H24	No apical meristem (NAM/NAC)	At3g57150	2e-14	gw1.VI.2308.1	7e-21	1.51	< 0.001	0.002
WS0222_C23	No apical meristem (NAM/NAC)	At5g13180	6e-57	estExt_Genewise1_v1.C_LG_I7833	1e-110	0.61	< 0.001	< 0.001
WS0183_C14	C2C2-YABBY	At2g45190	4e-14	grail3.0033028501	3e-23	1.94	0.006	0.017
WS0167_E22	ARR-B	At2g25180	4e-06	fgenesh4_pg.C_LG_VI001883	7e-21	1.91	0.002	0.008
WS0174_B01	GRAS/scarecrow-like	At1g07530	6e-83	fgenesh4_pm.C_LG_IX000604	1e-130	1.69	0.003	0.012
WS0214_F06	Homeobox leucine zipper	At4g00730	8e-14	estExt_fgenesh4_pg.C_LG_III0408	8e-25	1.63	0.003	0.012
WS0162_N05	Homeobox leucine zipper	At4g37790	5e-17	estExt_fgenesh4_pg.C_LG_II1036	5e-34	1.54	0.001	0.004
WS01214_J05	Homeobox lecuine zipper	At4g16780	3e-22	fgenesh4_pm.C_LG_III000187	9e-65	0.50	0.001	0.004
WS0204_I07	Homeobox leucine zipper	At4g16780	1e-18	fgenesh4_pm.C_LG_III000187	2e-40	0.48	0.001	0.006
WS0166_E10	TUB	At2g47900	3e-53	estExt_fgenesh4_pg.C_LG_VIII1741	1e-64	1.51	0.001	0.005
WS0126_P16	Zing finger C2C2-Dof	At5g60850	2e-07	No match	n.a.	0.64	0.004	0.012
WS0232_G07	CCAAT-HAP2	At3g20910	4e-16	eugene3.00011755	7e-72	0.63	0.001	0.006
WS0234_L03	Zinc finger C2H2	At5g12330	6e-19	grail3.0074009801	6e-63	0.59	0.028	0.050
WS0124_I04	Zinc finger C2C2-CO-like	At3g02380	5e-49	gw1.123.49.1	1e-98	0.56	0.001	0.005
WS0141_N18	Zinc finger C2C2-CO-like	At3g02380	7e-44	estExt_fgenesh4_pm.C_LG_IV0339	1e-99	0.54	< 0.001	0.003

factors are shown in Tables 5 and 6, respectively. Viewed broadly, the FTC-responsive genes include genes involved in general metabolism, photosynthesis, transport, transcriptional regulation and signalling, octadecanoid and ethylene signalling, response to stress, secondary metabolism, detoxification and redox processes, as well as cDNAs with no significant match to *Arabidopsis* and genes of unknown biological processes. The data provided in Tables 5 and 6 and Table S2 provide a first insight into large-scale transcriptional changes in poplar leaves in response to insect feeding.

Genes of unknown functions affected by FTC

Examples of microarray cDNA elements with no similarity to *Arabidopsis* that represent FTC-induced or FTC-repressed poplar genes include ESTs WS0151_C13 and WS0192_L21 (Table 5). These ESTs have high similarity to other poplar ESTs in our collection, as well as to other angiosperm EST sequences in the public domain, confirming they are legitimate expressed sequences (data not shown). Even among ESTs with similarity to *Arabidopsis* genes, many of these genes are of unknown function. Examples of such poplar genes induced after FTC feeding include genes with similarities to *Arabidopsis*-expressed protein (poplar EST WS0156_N20), senescence-associated protein (WS02010_G16), and a protein induced upon wounding (WS0153_M02) (Table 5).

Genes of general metabolism affected by FTC

Several array elements for genes involved in general metabolism revealed induced transcript levels in poplar leaves upon FTC feeding (Table 5). These include an apyrase (WS0124_K08), thymidylate kinase (WS0156_L05), aminopeptidase M (WS0212_I21), an acid phosphatase (WS0124_G12), a phosphorylase family protein (WS0146_L23), a lipase (WS0132_A15), lycopene beta cyclase (WS01223_F23), phytoene synthase (WS0145_F03), carbonic anhydrase (WS0154_D01) and an expansin (WS01116_C06). Both apyrases and thymidylate kinase are involved in energy metabolism. The former hydrolyses nucleotide tri- and diphosphates to nucleomonophosphates and in plants are postulated to function in nodulation and energy metabolism (Cohn *et al*. 2001). The latter catalyses the phosphorylation of dTMP to form dTDP in both the *de novo* and salvage pathways of dTTP synthesis. Aminopeptidases are a diverse family of proteases that hydrolyse the amino-terminal residues of peptides or proteins and have been shown to perform important roles in protein synthesis and turnover. With regard to plant defence, two leucine aminopeptidases have been demonstrated to be induced locally or systemically in tomato in response to insect feeding, pathogen infection, wounding, methyl jasmonate, abscisic acid, ethylene and several abiotic treatments (Pautot *et al*. 1993; Chao *et al*. 1999). It has been proposed that aminopeptidases may modulate induced defences as activators of peptide hormones in plant defence signalling, or may facilitate turnover of proteins damaged during wounding, or may directly inactivate digestive enzymes and gut peptide hormones within insect guts, possibly in conjunction with plant protease inhibitors (Pautot *et al*. 1993). Both phytoene synthase and lycopene beta cyclase are part of the carotenoid biosynthesis pathway, the former being the first committed, and potentially rate-limiting step involving the condensation of two geranylgeranyl diphosphate molecules to produce 15-cis phytoene, and the latter catalyses a two-step reaction that creates β-carotene and its derivative xanthophylls (Hirschberg 2001). Although the direct connection to plant defence is not immediately evident for xanthophylls, whose primary role is photoprotection via energy dissipation from photosynthesis through nonphotochemical quenching; another possible product of this biosynthesis pathway is the hormone abscisic acid, which together with the jasmonate and ethylene signalling pathways, modulates defence gene expression (Anderson *et al*. 2004). During photosynthesis in C_4 plants, carbonic anhydrase is involved in converting CO_2 into bicarbonate for fixation by the primary carboxylating enzyme phosphoenolypyruvate carboxylase. Recently, a carbonic anhydrase in tobacco was shown to bind salicylic acid in chloroplasts, have antioxidant activity and function in the hypersensitive response in plant disease resistance (Slaymaker *et al*. 2002). Expansins are key regulators of cell wall extension during growth via disruption of hydrogen bonds between cellulose microfibrils and cross-linking glycans in the cell wall (Li *et al*. 2003), and likely contribute to the strengthening and/or repair of damaged cell walls during pathogen and insect attack. Among general metabolism genes down-regulated after FTC feeding we identified a galactinol synthase (WS0231_K19), which is responsible for raffinose family oligosaccharide production (Table 5).

Photosynthesis genes affected by FTC

In general, many genes associated with photosynthesis were down-regulated by FTC feeding including a photosystem II protein (WS02011_K12), a chlorophyll A–B binding protein (WS0142_N19) and a ferredoxin reductase (WS01224_M03) (Table 5). The inverse correlation between photosynthesis- and defence-related gene regulation has also been observed in other large-scale studies of the response to insect herbivory in angiosperms (Hermsmeier *et al*. 2001; Zhu-Salzman *et al*. 2004). This response presumably allows resource reallocation to defence responses, with reduced resource commitment to other primary functions.

Transport genes affected by FTC

Among genes that were induced by FTC feeding and associated with transport functions (Table 5), we identified several ATPases (e.g. WS0212_O05), a metal transporter (WS0156_A09); several ABC proteins (e.g. WS0156_O13), calreticulin (WS0114_D04), a sec23A transport protein (WS0142_J05), and a major intrinsic protein (WS0114_H12). Predicting a possible biological role for putative transport proteins in plant defence is particularly challenging due to the broad range of potential substrates. For example, ATPases actively transport a range of ions (e.g. H^+, Ca^{2+}, Na^+, K^+, Cl^-, Mg^{2+}) into or out of vacuoles and/or cells to support innumerable biological functions. In addition to their traditional role in detoxification processes, ABC proteins in plants have been demonstrated to participate in chlorophyll biosynthesis, formation of Fe/S clusters, stomatal movement and ion fluxes (Martinoia *et al*. 2002). ABC proteins may also be directly involved in plant defence via transport of signalling molecules such as jasmonate (Theodoulou *et al*. 2005), or transport of phytochemicals as has been shown for alkaloid (Shitan *et al*. 2003) and tepernoid (Jasiński *et al*. 2001) defence compounds, or reinforcement of cuticular waxes (Pighin *et al*. 2004).

Calreticulin is a highly conserved multifunctional protein, mainly localized to the endoplasmic reticulum, that has been suggested to be involved in many biological processes, chief among these being calcium binding, calcium signalling and as a chaperone. Plant calreticulins have been demonstrated to be up-regulated in response to pathogen-related signalling molecules including cell wall degrading enzymes of plant pathogenic bacteria and salicylic acid (Denecke *et al*. 1995), and after nematode infection of plant roots (Jaubert *et al*. 2002). Sec23A transport proteins are part of the coat protein complex II (COPII) that selectively incorporates cargo molecules and vesicle-targeting machinery into transport vesicles budding from the endoplasmic reticulum in the initial step of the secretory pathway (Movafeghi *et al*. 1999; Bickford *et al*. 2004); however, the direct connection to plant defence is unclear. Major intrinsic proteins facilitate the passive transport of small polar molecules such as water or glycerol across cell membranes (Johanson *et al*. 2001), and presumably are contributing to the redistribution of glycerol stores during the defence response.

One of the more abundant classes of genes represented on the poplar 15.5K microarray involved in transport or general metabolism is the lipid transfer proteins (LTP). We observed most LTP genes not to be responsive to FTC feeding; however, several genes within this large family were either induced or repressed after FTC attack (e.g. WS0144_M07). LTPs are small, basic proteins synthesized as precursors that transfer phospholipids between membranes and bind fatty acids *in vitro* and have been proposed to be involved in several processes potentially associated with plant–insect interactions including cutin biosynthesis (Kader 1996), pathogen-defence reactions (Garcia-Olmedo *et al*. 1995), and the recognition of intruders in plants and in systemic resistance signalling (Blein *et al*. 2002; Maldonado *et al*. 2002). The interpretation of function is complicated by the fact that LTPs are represented by a large number of genes with several subfamilies, and as of yet no systematic characterization of LTPs has been performed in any plant species.

Transcriptional regulation and signalling affected by FTC

Transcriptional regulation and intracellular signalling cascades for plant defence in general, and secondary metabolism in particular, are poorly understood. Among genes associated with signalling that were induced by FTC feeding we identified a leucine-rich repeat transmembrane protein kinase (WS0205_I02) and a choline kinase (WS01223_D01). Leucine-rich repeat containing transmembrane/receptor-like kinases (LRR-RLK) are represented by large gene families in plants and are able to perceive external signals at the plasma membrane and initiate signalling cascades via their cytoplasmic protein kinase domains (Diévart & Clark 2004). Thus far few LRR-RLKs have known functions, and there is little information concerning their ligands and which downstream signalling pathways are affected. In tomato, an LRR-RLK has been demonstrated to bind the systemic defence signalling peptide hormone systemin (Scheer & Ryan 2002); however, a systemin-like peptide hormone has yet to be identified in poplar. A novel role is emerging for phospholipids as second messengers in plant cells that are rapidly formed in response to a variety of stimuli via the activation of lipid kinases and phosphatases (Meijer & Munnik 2003). Phosphatidylcholine is a major phospholid of eurkaryotic membranes, which among other functions serves as a reservoir for lipid second messengers. Choline kinase is part of the biosynthesis pathway for phosphatidylcholine and has previously been demonstrated to be induced in response to salt stress in *Arabidopsis* (Tasseva *et al*. 2004).

Utilizing a set of 1618 transcription factors identified in *Arabidopsis* (Arabidopsis Gene Regulatory Information Server at the Ohio State University; http://arabidopsis.med.ohio-state.edu/AtTFDB/; 9 May 2005 download; Davuluri *et al*. 2003) we screened the BLASTX annotation to *Arabidopsis* to classify 458 ESTs on the poplar 15.5K array as putative transcription factors. Among this set, 56 transcription factors from 20 different families were differentially expressed in response to FTC feeding, with 40 transcription factors induced and only 16 repressed (Table 6). Among the more commonly induced transcription factor families were zinc finger C3H type, APETELA (AP2)/ethylene-responsive-element binding (EREBP), basic

helix-loop-helix (bHLH), WRKY, and no apical meristem (NAM/NAC) classes. Each of these transcription factor classes are represented by large gene families in *Arabidopsis* (i.e. zinc finger C3H, 33 genes; AP2/EREBP, 144 genes; bHLH, 139 genes; WRKY, 72 genes; NAM/NAC, 109 genes; Riechmann *et al*. 2000). To date most transcription factors linked to plant stress responses have been derived from the AP2/EREBP, WRKY, MYB and bZIP families (Stracke *et al*. 2001; Singh *et al*. 2002). The list of transcription factors in Table 6, especially those from transcription factor classes previously associated with plant defence, provide an interesting set of targets for further characterization in poplar insect defence.

Octadecanoid and ethylene pathway genes affected by FTC

Plant responses to biotic and abiotic stresses are regulated locally and systemically by a complex network of signalling cascades including peptide signals (e.g. systemin), salicylic acid, ethylene, H_2O_2, and fatty acid-derived oxylipins. Within this latter class, we observed several genes within the octadecanoid biosynthesis pathway to be strongly induced in poplar leaves in response to FTC feeding including a lipoxygenase (LOX; WS01212_L05), an allene oxide cyclase (AOC; WS0155_D02) and an allene oxide synthase (AOS; WS0145_B07), as well as two key genes involved in ethylene biosynthesis, S-adenosylmethionine synthase (SAM synthase; WS0123_G08) and 1-aminocyclopropane-1-carboxylate oxidase (ACC oxidase; WS02011_I02) (Table 5).

A role for jasmonates in plant defence was proposed by Farmer & Ryan (1992), who demonstrated that wounding led to the formation of jasmonates and the subsequent induction of genes for proteinase inhibitors that deter insect feeding. Since then a tremendous amount of work has substantiated the importance of octadecanoid-/oxylipin-signalling in plant-insect defence responses (Howe 2004; Halitschke & Baldwin 2005). The first biosynthesis step in oxylipin formation is catalysed by LOX enzymes, which introduces molecular oxygen at the C-13 position of linolenic acid, and which comprises a multigene family with specificity for production of either jasmonates or green leaf volatiles, the latter via the action of hydroperoxide lyase. In laboratory studies, plants deficient in the expression of jasmonates derived from LOX genes are impaired in their ability to produce chemical defences (e.g. protease inhibitors and nicotine) and are more susceptible to herbivore attack (Royo *et al*. 1999; Halitschke & Baldwin 2003). Moreover, in field studies comparing tobacco, *Nicotiana attenuata*, plants transformed with antisense LOX, hydroperoxide lyase or AOS genes, only the LOX-deficient plants were more vulnerable to *N. attentuata*'s adapted herbivores, as well as novel herbivore species (Kessler *et al*. 2004). Following the action of LOX enzymes, the next steps in jasmonate biosynthesis are the formation of an epoxide by AOS, ring formation by AOC, followed by a reduction step and three rounds of β-oxidation. An *Arabidopsis* knock-out mutant defective in AOS was unable to make endogenous jasmonates, even after wounding, and was defective in wound signal transduction for both vegetative storage protein and LOX genes, which are inducible by wounding and jasmonate treatment in wild-type plants (Park *et al*. 2002). In tomato, antisense AOC plants are also defective in wound signal transduction for protease inhibitors and are defective in jasmonic acid biosynthesis (Stenzel *et al*. 2003). Collectively, these studies imply that depending on the plant system, disruption of LOX, AOS and/or AOC activity can significantly reduce jasmonate production and impair wounding- and herbivory-responsive signal transduction in plant defence. The induction of LOX, AOS and AOC transcripts in poplar in response to FTC feeding suggests the importance of these enzymes, and jasmonates in general, in activating and/or modulating the poplar defence response.

Both SAM synthase and ACC oxidase are part of the ethylene biosynthesis pathway. Ethylene is an important modulator in defence signal transduction (Feys & Parker 2000) that has been demonstrated to be induced in response to insect herbivory in several plant systems (Arimura *et al*. 2002; Winz & Baldwin 2001), and has been demonstrated to regulate defence-orientated genes such as protease inhibitors (O'Donnell *et al*. 1996), defensin (Penninckx *et al*. 1998) and pathogenesis-related proteins (Díaz *et al*. 2002).

Stress response genes affected by FTC

Among genes involved directly in plant defence against insects, we identified several poplar transcripts induced by FTC feeding including basic endochitinases (e.g. WS0143_A03), Kunitz trypsin protease inhibitors (e.g. WS0134_G14), glycosyl hydrolases (e.g. WS0205_P08), a stable protein 1 (WS0144_M15), and lectin proteins (e.g. WS0204_D16) (Table 5). Endochitinases represent a large and diverse group of enzymes that catalyse the cleavage of internal β-1,4-glycoside bonds present in the biopolymers of *N*-acetylglucosamine found in chitin, a major component of fungal, bacterial and invertebrate cell walls. Plant chitinases are classified as pathogenesis-related (PR) proteins and have been demonstrated in many plant systems, including poplar (Clarke *et al*. 1994), to be transcriptionally induced in response to both biotic (e.g. viruses, bacteria, fungi, insect pests, etc.) and abiotic (e.g. drought, salinity, wounding, plant hormones, etc.) stress (Kasprzewska 2003). Kunitz protease inhibitors (PIs) are small proteins present at high concentrations in storage tissues that are also inducible in poplar leaves in response to attack by insects and pathogenic organisms (Bradshaw *et al*. 1990;

Hollick & Gordon 1993; Haruta *et al*. 2001a). PIs are postulated to contribute to plant defence by forming a stable complex with protease enzymes in the insect gut, inhibiting protease activity and thereby reducing absorption of amino acids from consumed leaves, causing a reduction in insect growth. In a separate study, we identified *c*. 30 Kunitz PIs in the poplar genome, organized in gene clusters, and demonstrated that the majority of these transcripts are inducible in leaves in response to wounding, caterpillar feeding, and methyl jasmonate; although a direct impact on insect performance remains to be determined (S. Ralph & J. Bohlmann, in prep.). Glycosyl hydrolases have previously been identified as PR proteins because they are rapidly induced during fungal attack and are proposed to contribute to plant defence by digesting wall components of the fungal pathogen. Furthermore, glycosyl hydrolases can be involved in the release of aglycons and possible activation of a wide array of small molecules. Lectins are carbohydrate-binding proteins, many of which have insecticidal activity (Peumans & Van Damme 1995) and are frequently induced after wounding.

Secondary metabolism genes affected by FTC

We observed the activation in FTC-treated poplar leaves of several genes involved in phenolic and terpenoid secondary metabolism including polyphenol oxidase (PPO1), several cytochrome P450 genes (e.g. WS01212_M19), isoflavone reductase (WS01212_B20), several terpene synthases [e.g. (–)-germacrene D synthase TPS1], 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase (WS01214_G19), chorismate synthase (WS0168_K06), 4-coumarate CoA ligase (4CL) (WS0201_G13), several laccases (e.g. WS02011_F21), 5-enolpyruvylshikimate 3-phosphate (EPSP) synthase (WS01110_A05), phenylalanine ammonia-lyase (PAL) (WS0154_H05), and chalcone synthase (WS01224_J04) (Table 5).

Phenolic secondary metabolites have been proposed to play a variety of roles in defence as phytoalexins, radical scavengers or structural barriers. Phenylpropanoid metabolism builds on the shikimate pathway which links the metabolism of carbohydrates to the biosynthesis of aromatic compounds. In a series of seven metabolic steps, phosphoenolpyruvate (PEP) and erythrose 4-phosphate (E4P) are converted to chorismate, which is the precursor of the aromatic amino acids tryptophan, phenylalanine and tyrosine (Herrmann & Weaver 1999). Genes representing several enzymatic steps within the pathway were induced by FTC feeding including: DAHP synthase, the first step involving condensation of PEP and E4P; EPSP synthase, which is the sixth step of the pathway that catalyses the condensation of PEP and shikimate 3-phosphate to produce 5-enolpyruvylshikimate 3-phosphate; and chorismate synthase, the final step eliminating phosphate from EPSP to produce chorismate. The most abundant classes of secondary phenolic compounds are subsequently derived from phenylalanine, which is converted to a variety of phenolic defence compounds including flavonoids, stilbenoids, condensed tannins and other polyphenolics, along with the structural polymer lignin, via the action of a complex metabolic grid of different enzyme activities (Dixon *et al*. 2001). Among this group of enzymes, we observed both PAL and 4CL to be induced after insect feeding, whereas other steps in the metabolic grid were generally unresponsive to insect attack (Table S2). This is in agreement with earlier studies using hybrid poplar suspension-cultures that showed induction of PAL and 4CL mRNAs in response to fungal elicitor treatment (Moniz de Sá *et al*. 1992).

Among genes encoding enzymes of branch pathways of phenylpropanoid metabolism, we observed several laccases/diphenol oxidases and PPOs to be induced after FTC feeding. Laccases are proposed to be involved in the polymerization of monolignols to produce lignin and lignans based on their ability to oxidize monolignols and their close spatial and temporal correlation with lignin deposition, and they are organized as a multigene family in poplar (Ranocha *et al*. 1999). Increased gene expression of laccases could lead to strengthening of cell walls during insect attack via increased lignin deposition and/or increased production of toxic lignans. PPOs catalyse the oxidation of *o*-diphenols to *o*-diquinones, which are highly reactive with phenolic substrates and are proposed to cross-link with dietary proteins of feeding insects, resulting in decreased amino acid assimilation (Felton *et al*. 1992). In poplar, it has been demonstrated that a PPO mRNA is inducible by herbivores, wounding and methyl jasmonate (Constabel *et al*. 2000), and that overexpression in transgenic poplar reduces FTC performance (Wang & Constabel 2004). Another important branch of phenylpropanoid metabolism for plant defence is the production of flavonoids, the first step of which is catalysed by chalcone synthase, which is frequently induced at the transcript level in plant species in response to a variety of stresses (Dixon & Paiva 1995). Similarly, isoflavonoid phytoalexins are low molecular weight antimicrobial compounds synthesized in response to abiotic and biotic stress, involving among other enzymes the activity of isoflavone reductase, which is induced after FTC feeding.

In addition to phenolic metabolites, the isoprenoid biosynthesis pathway leads to the production of hundreds or possibly thousands of tepenoid compounds that are important components in many plant defence systems (Huber *et al*. 2004). Specifically in poplar, recent work has demonstrated the systemic induction of terpenoid volatile emission from trees under FTC attack (Arimura *et al*. 2004). Such volatiles can serve as signals to attract predators and parasites of herbivores. The sequiterpene synthase germacrene D was induced both locally and systemically upon

Table 7 Fold-change differences measured using real-time PCR between five trees subjected to FTC herbivory for 24 h and five untreated control trees

Clone ID	BLASTX vs. *Arabidopsis*	FTC at 24 h			
		FC	LLCI	ULCI	*P* value
WS0143_A03	Basic endochitinase	239.98	103.58	555.98	< 0.001
PPO1	Polyphenol oxidase PPO1	129.83	68.74	245.19	< 0.001
WS0181_F07	Myb transcription factor	72.13	27.39	189.96	< 0.001
WS0134_G14	Kunitz trypsin inhibitor	54.66	17.44	171.32	< 0.001
WS01212_B20	Isoflavone reductase	30.90	17.13	55.75	< 0.001
WS0145_B07	Allene oxide synthase	16.98	8.42	33.92	< 0.001
WS0231_E05	WRKY transcription factor	15.91	7.21	35.08	< 0.001
TPS1	(–)-germacrene D synthase	11.32	5.09	25.13	< 0.001
WS0155_D02	Allene oxide cyclase	10.22	5.83	17.89	< 0.001
WS0114_D04	Calreticulin	5.86	3.48	9.84	< 0.001
WS0201_G13	4-coumarate CoA ligase	4.39	3.12	6.16	< 0.001
WS0156_O13	ABC transporter family protein	4.30	1.76	10.49	0.005
WS01110_A05	EPSP synthase	4.01	1.58	10.16	0.008
WS0168_K06	Chorismate synthase	2.13	1.18	3.84	0.017
WS0145_F03	Phytoene synthase	1.65	0.95	2.86	0.069
WS02011_I02	ACC oxidase	1.07	0.59	1.95	0.779

Abbreviations: FC, fold-change; LLCI, lower limit 95% confidence interval; ULCI, upper limit 95% confidence interval; EPSP synthase, 5-enolpyruvylshikimate 3-phosphate synthase. For further details see Materials and methods and Fig. 4.

FTC feeding on poplar leaves, resulting in systemic diurnal emission profiles of (–)-germacrene D (Arimura *et al*. 2004). FTC-induced terpene synthases identified in the new poplar EST resources could contribute to additional components of the FTC-induced blend of terpenoid emissions (Arimura *et al*. 2004).

Oxidative stress genes affected by FTC

Oxidative stress is known to be caused by the damage imposed by herbivore feeding and therefore, it is not surprising that the transcript abundance of several proteins that contribute to cellular survival after oxidative damage was increased. These included thioredoxin (WS0162_C15), several glutathione S-transferases (e.g. WS0178_N24), a superoxide dismutase copper chaperone (WS0145_I01) and several peroxidases (e.g. WS0144_C22).

Refined gene-specific expression using real-time PCR

In order to validate our microarray results and obtain more refined gene expression data, we designed gene-specific primers for 16 transcripts selected from Tables 5 and 6 and analysed their expression using real-time PCR (Table 7 and Fig. 4). These genes were chosen to represent a variety of functional classifications and range from 2.41-fold (i.e. ACC oxidase; WS02011_I02) to 32.31-fold (i.e. Kunitz protease inhibitor; WS0134_G14) induction according to microarray analysis (Table 5). Among these 16 transcripts, 14 were induced (fold-change > 1.5×, *P* value < 0.05) in response to FTC feeding, in agreement with results obtained using microarrays. In general, we observed larger changes in gene expression using real-time PCR, likely reflecting the greater dynamic range of detection and sensitivity of this method compared to cDNA microarrays. Significant FTC-induced gene expression ranged from 2.13-fold for chorismate synthase (WS0168_K06) to 239.98-fold for a basic endochitinase (WS0143_A03). Phytoene synthase was also weakly induced (1.65-fold), but lacked significant statistical support. Although ACC oxidase was weakly induced according to our microarray results, real-time PCR analysis indicates this transcript is not induced after FTC feeding. Since ACC oxidase is represented by a multigene family in many plant species the induced gene expression obtained using cDNA microarrays may reflect induction of a closely related gene family member. In agreement with the relatively low biological variation observed in our microarray analysis (Fig. S1), very consistent levels of expression were observed for each transcript among untreated control trees, and gene induction in response to FTC feeding was also very similar among trees (Fig. 4).

Conclusions

In summary, we have developed and applied a comprehensive set of functional genomics resources that form the foundation for functional characterization of defence mechanisms against insect herbivory in poplar. The large scale of our EST sequencing programme, combined with the application of normalization strategies during library

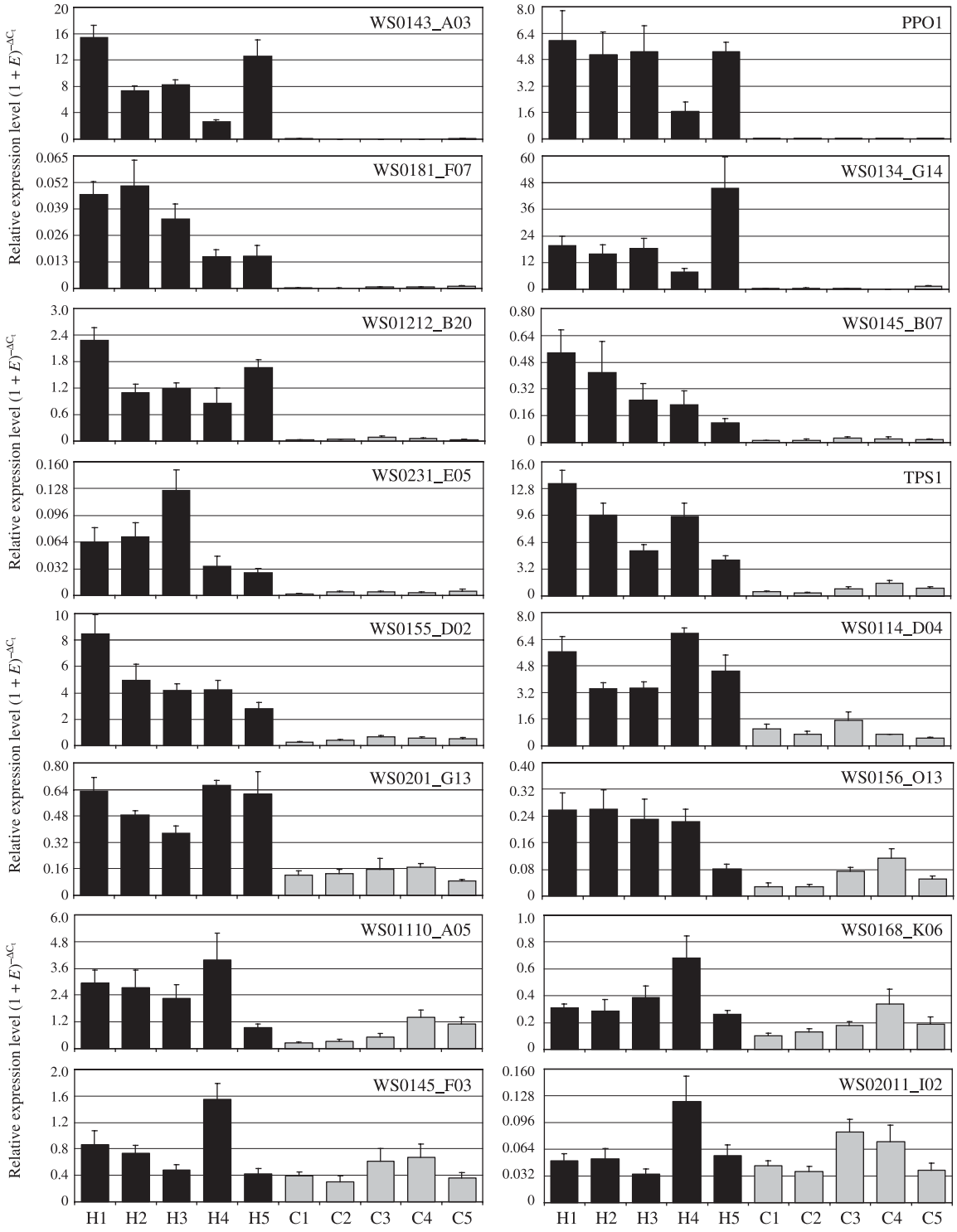

Fig. 4 Real-time PCR analysis of gene expression in poplar leaves in response to forest tent caterpillar herbivory (FTC). Values were determined using real-time PCR and represent fold-change differences between five trees subjected to FTC herbivory (H1 to H5) for 24 h and five trees left untreated as a control (C1 to C5). Gene expression was determined in each tree using at least three independent technical replicates. Transcript abundance for each gene was normalized to translation initiation factor 5 A (TIF5A; WS0116_J23) by subtracting the Ct value of each transcript, where $\Delta C_t = C_{t\ transcript} - C_{t\ TIF5A}$. Transcript abundance of genes in control and FTC 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. Statistical significance of expression differences was determined using a mixed-effects model (see Materials and methods).

construction, and a focus on herbivore- and elicitor-treated cDNA libraries has enabled us to capture potentially three-quarters of the *c.* 45 000 genes in the poplar genome and complement existing public *Populus* ESTs. Based on our EST resources we have developed the poplar 15.5K cDNA microarray, which when applied to the study of FTC-treated leaves revealed more than 1700 differentially expressed genes. This set of defence response genes contains several genes previously identified as components of the induced defence response to defoliating insects in poplar (e.g. endochitinases, Kunitz protease inhibitors, polyphenol oxidases). In addition, our transcriptome profiling revealed many genes not previously associated with induced poplar defence (e.g. ABC proteins, calreticulin, carotenoid biosynthesis, LRR-RLK, choline kinase) and emphasizes the potential importance of jasmonates in poplar defence signalling. Of special interest among this defence gene set are the 40 transcription factors induced after FTC feeding that potentially represent master switches for regulating the induced defence profile of poplar in response to defoliating insects.

Acknowledgements

We thank Michael Friedmann, Natalie Mattheus and Brian King for technical assistance with greenhouse experiments, Jürgen Ehlting for providing nitrogen-depleted hydroponic roots, David Kaplan for greenhouse support, Thomas Mitchell-Olds at Duke University for helpful comments on an earlier version of this paper, Bob McCron from the Canadian Forest Service for access to forest tent caterpillars, Peter Constabel and Manoela Miranda at the University of Victoria for providing elicitor- and pathogen-treated cultured poplar cells, Amy Brunner and Steven Strauss at Oregon State University for providing vegetative buds, Cindy Broberg and John Borden at Simon Fraser University for providing *Populus trichocarpa* × *nigra* trees and willow weevils, and the many staff at the Genome Sciences Centre for technical assistance with large-scale EST sequencing. The work was supported by Genome British Columbia, Genome Canada and the Province of British Columbia (grant to J.B., K.R., C.J.D., B.E.E.), and by the Natural Science and Engineering Research Council of Canada (NSERC, grant to J.B.).

Supplementary materials

The supplementary materials are available from http://www.blackwellpublishing.com/products/journals/suppmat/MEC/MEC2824/MEC2824sm.htm

Appendix S1 Supplemental materials and methods.

Figure S1 Boxplots showing the distribution of technical and biological variation among the 15 496 array elements using 10 microarray slides to assess the biological response to forest tent caterpillar herbivory, and a mixed-effects model to estimate variance (see Materials and methods).

Table S1 High-quality poplar ESTs (> 800 bp) with no similarity to *Arabidopsis* peptides.

Table S2 Genome-wide transcript profile of poplar leaves 24 h after the onset of forest tent caterpillar feeding using a 15.5K cDNA microarray.

Table S3 Primer sequences used for real-time PCR (5′ to 3′ orientation).

References

Adams MD, Soares MB, Kerlavage AR, Fields C, Venter JC (1993) Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. *Nature Genetics*, **4**, 373–380.

Addy ND (1969) Rearing the forest tent caterpillar on an artificial diet. *Journal of Economic Entomology*, **62**, 270–271.

Agrawal AA (1998) Induced responses to herbivory and increased plant performance. *Science*, **279**, 1201–1202.

Altschul SF, Gish W, Miller W, Meyers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Altschul SF, Madden TL, Schäffer AA *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389–3402.

Anderson JP, Badruzsaufari E, Schenk PM *et al.* (2004) Antagonistic interaction between abscisic acid and jasmonate-ethylene signaling pathways modulates defense gene expression and disease resistance in *Arabidopsis*. *Plant Cell*, **16**, 3460–3479.

Arimura G, Ozawa R, Nishioka T *et al.* (2002) Herbivore-induced volatiles induce the emission of ethylene in neighboring lima bean plants. *Plant Journal*, **29**, 87–98.

Arimura G, Huber DPW, Bohlmann J (2004) Forest tent caterpillars (*Malacosoma disstria*) induce local and systemic diurnal emissions of terpenoid volatiles in hybrid poplar (*Populus trichocarpa* × *deltoides*): cDNA cloning, functional characterization, and patterns of gene expression of (–)-germacrene D synthase *PtdTPS1*. *Plant Journal*, **37**, 603–616.

Bhalerao R, Nilsson O, Sandberg G (2003a) Out of the woods: forest biotechnology enters the genomic era. *Current Opinion in Biotechnology*, **14**, 206–213.

Bhalerao R, Keskitalo J, Sterky F *et al.* (2003b) Gene expression in autumn leaves. *Plant Physiology*, **131**, 430–442.

Bickford LC, Mossessova E, Goldberg J (2004) A structural view of the COPII vesicle coat. *Current Opinion in Structural Biology*, **14**, 147–153.

Blein JP, Coutos-Thévenot P, Marion D, Ponchet M (2002) From elicitins to lipid-transfer proteins: a new insight in cell signalling involved in plant defence mechanisms. *Trends in Plant Science*, **7**, 293–296.

Bonaldo MF, Lennon G, Soares MB (1996) Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Research*, **6**, 791–806.

Bradshaw HD, Hollick JB, Parsons TJ, Clarke HRG, Gordon MP (1990) Systemically wound-responsive genes in poplar trees encode proteins similar to sweet potato sporamins and legume Kunitz trypsin inhibitors. *Plant Molecular Biology*, **14**, 51–59.

Brazma A, Hingamp P, Quackenbush J *et al.* (2001) Minimum information about a microarray experiment (MIAME): toward standards for microarray data. *Nature Genetics*, **29**, 365–371.

Brunner AM, Busov VB, Strauss SH (2004) Poplar genome sequence: functional genomics in an ecologically dominant plant species. *Trends in Plant Science*, **9**, 49–56.

Carninci P, Kvam C, Kitamura A *et al.* (1996) High-efficiency full-length cDNA cloning by biotinylated CAP trapper. *Genomics*, **37**, 327–336.

Chao WS, Gu YQ, Pautot V, Bray EA, Walling LL (1999) Leucine aminopeptidase RNAs, proteins, and activities increase in response to water deficit, salinity, and the wound signals systemin, methyl jasmonate, and abscisic acid. *Plant Physiology*, **120**, 979–992.

Christopher ME, Miranda M, Major IT, Constabel CP (2004) Gene expression profiling of systemically wound-induced defenses in hybrid poplar. *Planta*, **219**, 936–947.

Churchill GB, John HH, Duncan DP, Hodson AC (1964) Long-term effects of defoliation of aspen by the forest tent caterpillar. *Ecology*, **45**, 630–636.

Clarke HR, Davis JM, Wilbert SM, Bradshaw HD, Gordon MP (1994) Wound-induced and developmental activation of a poplar tree chitinase gene promoter in transgenic tobacco. *Plant Moleclar Biology*, **25**, 799–815.

Cohn JR, Uhm T, Ramu S *et al.* (2001) Differential regulation of a family of apyrase genes from *Medicago truncatula*. *Plant Physiology*, **125**, 2104–2119.

Constabel CP (1999) A survey of herbivore-inducible defensive proteins and phytochemicals. In: *Induced Plant Defenses Against Herbivores and Pathogens* (eds Agrawal AA, Tuzun S, Bent E), pp. 137–166. The American Phytopathological Society, St Paul, Minnesota.

Constabel CP, Yip L, Patton JJ, Christopher ME (2000) Polyphenol oxidase from hybrid poplar. Cloning and expression in response to wounding and herbivory. *Plant Physiology*, **124**, 285–296.

Davis JM, Clarke HRG, Bradshaw HD, Gordon MP (1991) *Populus* chitinase genes: structure, organization, and similarity of translated sequences to herbaceous plant chitinases. *Plant Molecular Biology*, **17**, 631–639.

Davis JM, Egelkrout EE, Coleman GD *et al.* (1993) A family of wound-induced genes in *Populus* shares common features with genes encoding vegetative storage proteins. *Plant Molecular Biology*, **23**, 135–143.

Davuluri RV, Sun H, Palaniswamy SK *et al.* (2003) AGRIS: *Arabidopsis* Gene Regulatory Information Server, an information resource of *Arabidopsis cis*-regulatory elements and transcription factors. *BMC Bioinformatics*, **4**, 25–35.

Dejardin A, Leple JC, Lesage-Descauses MC, Costa G, Pilate G (2004) Expressed sequence tags from poplar wood tissues — a comparative analysis from multiple libraries. *Plant Biology (Stuttg)*, **6**, 55–64.

Denecke J, Carlsson LE, Vidal S *et al.* (1995) The tobacco homolog of mammalian calreticulin is present in protein complexes in vivo. *Plant Cell*, **7**, 391–406.

Díaz J, ten Have A, van Kan JAL (2002) The role of ethylene and wound signaling in resistance of tomato to *Botrytis cinerea*. *Plant Physiology*, **129**, 1341–1351.

Diévart A, Clark SE (2004) LRR-containing receptors regulating plant development and defense. *Development*, **131**, 251–261.

Dixon RA, Paiva NL (1995) Stress-induced phenylpropanoid metabolism. *Plant Cell*, **7**, 1085–1097.

Dixon RA, Chen F, Guo D, Parvathi K (2001) The biosynthesis of monolignols: a 'metabolic grid', or independent pathways to guaiacyl and syringyl units? *Phytochemistry*, **57**, 1069–1084.

Ewing B, Green P (1998) Base-calling of automated sequencer traces using PHRED II. Error probabilities. *Genome Research*, **8**, 186–194.

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using PHRED. I. Accuracy assessment. *Genome Research*, **8**, 175–185.

Farmer EE, Ryan CA (1992) Octadecanoid precursors of jasmonic acid activate the synthesis of wound-inducible proteinase inhibitors. *Plant Cell*, **4**, 129–134.

Felton GW, Donato KK, Broadway RM, Duffey SS (1992) Impact of oxidized plant phenolics on the nutritional quality of dietary protein to a noctuid herbivore, *Spodoptera exigua*. *Journal of Insect Physiology*, **38**, 277–285.

Feys BJ, Parker JE (2000) Interplay of signaling pathways in plant disease resistance. *Trends in Genetics*, **16**, 449–455.

Fitzgerald TD (1995) *The Tent Caterpillars*. Cornell University Press, Ithaca, New York.

Garcia-Olmedo F, Molina A, Segura A, Moreno M (1995) The defensive role of nonspecific lipid-transfer proteins in plants. *Trends in Microbiology*, **3**, 72–74.

Gregory RA, Wargo PM (1986) Timing of defoliation and its effect on bud development, starch reserves and sap sugar concentration in sugar maple. *Canadian Journal of Forest Research*, **16**, 10–17.

Halitschke R, Baldwin IT (2003) Antisense LOX expression increases herbivore performance by decreasing defense responses and inhibiting growth-related transcriptional reorganization in *Nicotiana attenuata*. *Plant Journal*, **36**, 794–807.

Halitschke R, Baldwin IT (2005) Jasmonates and related compounds in plant–insect interactions. *Journal of Plant Growth Regulation*, **23**, 238–245.

Haruta M, Major IT, Christopher ME, Patton JJ, Constabel CP (2001a) A Kunitz trypsin inhibitor gene family from trembling aspen (*Populus tremuloides* Michx.): cloning, functional expression, and induction by wounding and herbivory. *Plant Molecular Biology*, **46**, 347–359.

Haruta M, Pedersen JA, Constabel CP (2001b) Polyphenol oxidase and herbivore defense in trembling aspen (*Populus tremuloides*): cDNA cloning, expression, and potential substrates. *Physiologia Plantarum*, **112**, 552–558.

Heidel A, Baldwin IT (2004) Microarray analysis of salicylic acid- and jasmonic acid-signalling in responses of *Nicotiana attenuata* to attack by insects from multiple feeding guilds. *Plant, Cell and Environment*, **27**, 1362–1373.

Hermsmeier D, Schittko U, Baldwin IT (2001) Molecular interactions between the specialist herbivore *Manduca Sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*. I. Large-scale changes in the accumulation of growth- and defense-related plant mRNAs. *Plant Physiology*, **125**, 683–700.

Herrmann KM, Weaver LM (1999) The shikimate pathway. *Annual Review of Plant Physiology and Plant Molecular Biology*, **50**, 473–503.

Hirschberg J (2001) Carotenoid biosynthesis in flowering plants. *Current Opinion in Plant Biology*, **4**, 210–218.

Hogg EH, Brandt JP, Kochtubajda B (2002) Growth and dieback of aspen forests in northwestern Alberta, Canada, in relation to climate and insects. *Canadian Journal of Forest Research*, **32**, 823–832.

Hollick JB, Gordon MP (1993) A poplar tree proteinase inhibitor-like gene promoter is responsive to wounding in transgenic tobacco. *Plant Molecular Biology*, **22**, 561–572.

Howe GA (2004) Jasmonates as signals in the wound response. *Journal of Plant Growth Regulation*, **23**, 223–237.

Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. *Genome Research*, **9**, 868–877.

Huber DPW, Ralph S, Bohlmann J (2004) Genomic hardwiring and phenotypic plasticity of terpenoid-based defenses in conifers. *Journal of Chemical Ecology*, **30**, 2399–2418.

Huber W, von Heydebreck A, Sültmann H, Poustka A, Vingron M (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, **18**, S96–S104.

Hui D, Iqbal J, Lehmann K, Gase K, Saluz HP, Baldwin IT (2003) Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*: V. microarray analysis and further characterization of large-scale changes in herbivore-induced mRNAs. *Plant Physiology*, **131**, 1877–1893.

Hwang SY, Lindroth RL (1997) Clonal variation in foliar chemistry of aspen: effects on gypsy moths and forest tent caterpillars. *Oecologia*, **111**, 99–108.

Jasiński M, Stukkens Y, Degand H, Purnelle B, Marchand-Brynaert J, Boutry M (2001) A plant plasma membrane ATP binding cassette-type transporter is involved in antifungal terpenoid secretion. *Plant Cell*, **13**, 1095–1107.

Jaubert S, Ledger TN, Laffaire JB, Piotte C, Abad P, Rosso MN (2002) Direct identification of stylet secreted proteins from root-knot nematodes by a proteomic approach. *Molecular and Biochemical Parasitology*, **121**, 205–211.

Johanson U, Karlsson M, Johansson I *et al.* (2001) The complete set of genes encoding major intrinsic proteins in *Arabidopsis* provides a framework for a new nomenclature for major intrinsic proteins in plants. *Plant Physiology*, **126**, 1358–1369.

Kader JC (1996) Lipid-transfer proteins in plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, **47**, 627–654.

Karban R, Baldwin IT (1997) *Induced Responses to Herbivory*. University of Chicago Press, Chicago, Illinois.

Kasprzewska A (2003) Plant chitinases — regulation and function. *Cellular and Molecular Biology Letters*, **8**, 809–824.

Kessler A, Baldwin IT (2002) Plant responses to insect herbivory: the emerging molecular analysis. *Annual Review of Plant Biology*, **53**, 299–328.

Kessler A, Halitschke R, Baldwin IT (2004) Silencing the jasmonate cascade: induced plant defenses and insect populations. *Science*, **305**, 665–668.

Kirst M, Johnson AF, Baucom C *et al.* (2003) Apparent homology of expressed genes from wood-forming tissues of loblolly pine (*Pinus taeda* L.) with *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA*, **100**, 7383–7388.

Kohler A, Delaruelle C, Martin D, Encelot N, Martin F (2003) The poplar root transcriptome: analysis of 7000 expressed sequence tags. *FEBS Letters*, **542**, 37–41.

Kolosova N, Miller B, Ralph S *et al.* (2004) Isolation of high-quality RNA from gymnosperm and angiosperm trees. *BioTechniques*, **36**, 821–824.

Li Y, Jones L, McQueen-Mason S (2003) Expansins and cell growth. *Current Opinion in Plant Biology*, **6**, 603–610.

Maldonado AM, Doerner P, Dixon RA, Lamb CJ, Cameron RK (2002) A putative lipid transfer protein involved in systemic resistance signalling in *Arabidopsis*. *Nature*, **419**, 399–403.

Marra M, Hillier L, Kucaba T *et al.* (1999) An encyclopedia of mouse genes. *Nature Genetics*, **21**, 191–194.

Martinoia E, Klein M, Geisler M *et al.* (2002) Multifunctionality of plant ABC transporters — more than just detoxifiers. *Planta*, **214**, 345–355.

Meijer HJG, Munnik T (2003) Phospholipid-based signaling in plants. *Annual Review of Plant Biology*, **54**, 265–306.

Moniz de Sá M, Subramaniam R, Williams FE, Douglas CJ (1992) Rapid activation of phenylpropanoid metabolism in elicitor-treated hybrid poplar (*Populus trichocarpa* Torr. & Gray × *Populus deltoides* Marsh) suspension-cultured cells. *Plant Physiology*, **98**, 728–737.

Movafeghi A, Happel N, Pimpl P, Tai GH, Robinson DG (1999) *Arabidopsis* sec21p and sec23p homologs. Probable coat proteins of plant COP-coated vesicles. *Plant Physiology*, **119**, 1437–1446.

Nanjo T, Futamura N, Nishiguchi M, Igasaki T, Shinozaki K, Shinohara K (2004) Characterization of full-length enriched expressed sequence tags of stress-treated poplar leaves. *Plant and Cell Physiology*, **45**, 1738–1748.

O'Donnell PJ, Calvert C, Atzorn R, Wasternack C, Leyser HMO, Bowles DJ (1996) Ethylene as a signal mediating the wound response of tomato plants. *Science*, **274**, 1914–1917.

Osier TL, Lindroth RL (2001) Effects of genotype, nutrient availability, and defoliation on aspen phytochemistry and insect performance. *Journal of Chemical Ecology*, **27**, 1289–1313.

Osier TL, Lindroth RL (2004) Long-term effects of defoliation on quaking aspen in relation to genotype and nutrient availability: plant growth, phytochemistry and insect performance. *Oecologia*, **139**, 55–65.

Park JH, Halitschke R, Kim HB, Baldwin IT, Feldmann KA, Feyereisen R (2002) A knock-out mutation in allene oxide synthase results in male sterility and defective wound signal transduction in *Arabidopsis* due to a block in jasmonic acid biosynthesis. *Plant Journal*, **31**, 1–12.

Parry D, Spence JR, Volney WJA (1998) Budbreak phenology and natural enemies mediate survival of early-instar forest tent caterpillar (Lepidoptera: Lasiocampidae). *Environmental Entomology*, **27**, 1368–1374.

Parsons TJ, Bradshaw HD, Gordon MP (1989) Systemic accumulation of specific mRNAs in response to wounding in poplar trees. *Proceedings of the National Academy of Sciences, USA*, **86**, 7895–7899.

Pautot V, Holzer FM, Reisch B, Walling LL (1993) Leucine aminopeptidase: an inducible component of the defense response in *Lycopersicon esculentum* (tomato). *Proceedings of the National Academy of Sciences, USA*, **90**, 9906–9910.

Penninckx IAMA, Thomma BPHJ, Buchala A, Métraux JC, Broekaert WF (1998) Concomitant activation of jasmonate and ethylene response pathways is required for induction of a plant defensin gene in *Arabidopsis*. *Plant Cell*, **10**, 2103–2114.

Peters DJ, Constabel CP (2002) Molecular analysis of herbivore-induced condensed tannin synthesis: cloning and expression of dihydroflavonol reductase from trembling aspen (*Populus tremuloides*). *Plant Journal*, **32**, 701–712.

Peumans WJ, Van Damme EJM (1995) Lectins as plant defense proteins. *Plant Physiology*, **109**, 347–352.

Pighin JA, Zheng H, Balakshin LJ *et al.* (2004) Plant cuticular lipid export requires an ABC transporter. *Science*, **306**, 702–704.

Ranocha P, McDougall G, Hawkins S *et al.* (1999) Biochemical characterization, molecular cloning and expression of laccases — a divergent gene family — in poplar. *European Journal of Biochemistry*, **259**, 485–495.

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

Reymond P, Weber H, Damond M, Farmer EE (2000) Differential gene expression in response to mechanical wounding and insect feeding in *Arabidopsis*. *Plant Cell*, **12**, 707–720.

Reymond P, Bodenhausen N, Van Poecke RMP, Krishnamurthy V, Dicke M, Farmer EE (2004) A conserved transcript pattern in response to a specialist and a generalist herbivore. *Plant Cell*, **16**, 3132–3147.

Riechmann JL, Heard J, Martin G *et al.* (2000) *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science*, **290**, 2105–2110.

Rishi AS, Munir S, Kapur V, Nelson ND, Goyal A (2004) Identification and analysis of safener-inducible expressed sequence tags in *Populus* using a cDNA microarray. *Planta*, **220**, 296–306.

Royo J, León J, Vancanneyt G *et al.* (1999) Antisense-mediated depletion of a potato lipoxygenase reduces wound induction of proteinase inhibitors and increases weight gain of insect pests. *Proceedings of the National Academy of Sciences, USA*, **96**, 1146–1151.

Scheer JM, Ryan CA (2002) The systemin receptor SR160 from *Lycopersicon peruvianum* is a member of the LRR receptor kinase family. *Proceedings of the National Academy of Sciences, USA*, **99**, 9585–9590.

Scheetz TE, Laffin JJ, Berger B *et al.* (2004) High-throughput gene discovery in the rat. *Genome Research*, **14**, 733–741.

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, **270**, 467–470.

Schrader J, Moyle R, Bhalerao R *et al.* (2004) Cambial meristem dormancy in trees involves extensive remodelling of the transcriptome. *Plant Journal*, **40**, 173–187.

Shitan N, Bazin I, Dan K *et al.* (2003) Involvement of CjMDR1, a plant multidrug-resistance-type ATP-binding cassette protein, in alkaloid transport in *Coptis japonica*. *Proceedings of the National Academy of Sciences, USA*, **100**, 751–756.

Singh KB, Foley RC, Oñate-Sánchez L (2002) Transcription factors in plant defense and stress responses. *Current Opinion in Plant Biology*, **5**, 430–436.

Slaymaker DH, Navarre DA, Clark D, del Pozo O, Martin GB, Klessig DF (2002) The tobacco salicylic acid-binding protein 3 (SABP3) is the chloroplast carbonic anhydrase, which exhibits antioxidant activity and plays a role in the hypersensitive defense response. *Proceedings of the National Academy of Sciences, USA*, **99**, 11640–11645.

Soares MB, Bonaldo MF, Jelene P, Su L, Lawton L, Efstratiadis A (1994) Construction and characterization of a normalized cDNA library. *Proceedings of the National Academy of Sciences, USA*, **91**, 9228–9232.

Stairs GR (1972) Pathogenic microorganisms in the regulation of forest insect populations. *Annual Review of Entomology*, **17**, 355–372.

Stenzel I, Hause B, Maucher H *et al.* (2003) Allene oxide cyclase dependence of the wound response and vascular bundle-specific generation of jasmonates in tomato — amplification in wound signalling. *Plant Journal*, **33**, 577–589.

Sterky F, Regan S, Karlsson J *et al.* (1998) Gene discovery in the wood-forming tissues of poplar: Analysis of 5692 expressed sequence tags. *Proceedings of the National Academy of Sciences, USA*, **95**, 13330–13335.

Sterky F, Bhalerao RR, Unneberg P *et al.* (2004) A *Populus* EST resource for plant functional genomics. *Proceedings of the National Academy of Sciences, USA*, **101**, 13951–13956.

Storey JD, Tibshirani R (2003) Statistical significance for genome-wide studies. *Proceedings of the National Academy of Sciences, USA*, **100**, 9440–9445.

Stracke R, Werber M, Weisshaar B (2001) The *R2R3-MYB* gene family in *Arabidopsis thaliana*. *Current Opinion in Plant Biology*, **4**, 447–456.

Tasseva G, Richard L, Zachowski A (2004) Regulation of phosphatidylcholine biosynthesis under salt stress involves choline kinases in *Arabidopsis thaliana*. *FEBS Letters*, **566**, 115–120.

Taylor G (2002) *Populus*: *Arabidopsis* for forestry. Do we need a model tree? *Annals of Botany*, **90**, 681–689.

Theodoulou FL, Job K, Slocombe SP *et al.* (2005) Jasmonic acid levels are reduced in COMATOSE ATP-binding cassette transporter mutants. Implications for transport of jasmonate precursors into peroxisomes. *Plant Physiology*, **137**, 835–840.

Voelckel C, Weisser WW, Baldwin IT (2004) An analysis of plant–aphid interactions by different microarray hybridization strategies. *Molecular Ecology*, **13**, 3187–3195.

Wang J, Constabel CP (2004) Polyphenol oxidase overexpression in transgenic *Populus* enhances resistance to herbivory by forest tent caterpillar (*Malacosoma disstria*). *Planta*, **220**, 87–96.

Winz RA, Baldwin IT (2001) Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*. IV. Insect-induced ethylene reduces jasmonate-induced nicotine accumulation by regulating putrescine *N*-Methyltransferase transcripts. *Plant Physiology*, **125**, 2189–2202.

Zhu-Salzman K, Salzman RA, Ahn JE, Koiwa H (2004) Transcriptional regulation of sorghum defense determinants against a phloem-feeding aphid. *Plant Physiology*, **134**, 420–431.

The lead investigators of this study, Steven Ralph and Jörg Bohlmann, have developed a research program on functional genomics of forest health with a major emphasis on the genomic interactions of forest trees and insect pests. In addition, another research interest is the role of plant secondary metabolism in plant-insect defence systems.
