

RESEARCH ARTICLE

Open Access



CrossMark

Meta-transcriptomics indicates biotic cross-tolerance in willow trees cultivated on petroleum hydrocarbon contaminated soil

Emmanuel Gonzalez^{1†}, Nicholas J. B. Brereton^{1*†}, Julie Marleau¹, Werther Guidi Nissim², Michel Labrecque^{1,2}, Frederic E. Pitre^{1,2} and Simon Joly^{1,2}

Abstract

Background: High concentrations of petroleum hydrocarbon (PHC) pollution can be hazardous to human health and leave soils incapable of supporting agricultural crops. A cheap solution, which can help restore biodiversity and bring land back to productivity, is cultivation of high biomass yielding willow trees. However, the genetic mechanisms which allow these fast-growing trees to tolerate PHCs are as yet unclear.

Methods: *Salix purpurea* 'Fish Creek' trees were pot-grown in soil from a former petroleum refinery, either lacking or enriched with C10-C50 PHCs. De novo assembled transcriptomes were compared between tree organs and impartially annotated without *a priori* constraint to any organism.

Results: Over 45 % of differentially expressed genes originated from foreign organisms, the majority from the two-spotted spidermite, *Tetranychus urticae*. Over 99 % of *T. urticae* transcripts were differentially expressed with greater abundance in non-contaminated trees. Plant transcripts involved in the polypropanoid pathway, including phenylalanine ammonia-lyase (PAL), had greater expression in contaminated trees whereas most resistance genes showed higher expression in non-contaminated trees.

Conclusions: The impartial approach to annotation of the de novo transcriptomes, allowing for the possibility for multiple species identification, was essential for interpretation of the crop's response treatment. The meta-transcriptomic pattern of expression suggests a cross-tolerance mechanism whereby abiotic stress resistance systems provide improved biotic resistance. These findings highlight a valuable but complex biotic and abiotic stress response to real-world, multidimensional contamination which could, in part, help explain why crops such as willow can produce uniquely high biomass yields on challenging marginal land.

Keywords: *Salix*, Biomass, Transcriptomics, Meta-transcriptomics, Plant abiotic stress, Plant biotic stress, *Tetranychus*, Crop physiology, Phytoremediation, RNA-seq

Background

The ubiquitous use of oil and petroleum products in society has led to localised pollution of land with by-products from petroleum refining, such as aliphatic and polycyclic aromatic hydrocarbons that can be carcinogenic to humans and toxic to most agricultural crops [1]. The number of contaminated sites is thought to be as high as 30,000 across Canada [2], 384,400 across the US [3] and 342,000 across

the EU (although *potentially* contaminated sites have been estimated at 2.5 million) [4].

Such large resources of "marginal" land have been recognised as both socially and economically important to rejuvenate and could become more important if predicted increases in the vulnerability of food security, driven by climate change associated weather extremity, are realised [5, 6]. The ability to cultivate efficient and resilient perennial crops on marginal land, thus bypassing the food vs fuel debate and maximising land-use, is essential for the future of agriculture [7]. An emerging approach for cheap and environmentally sustainable land decontamination

* Correspondence: nicholas.brereton@umontreal.ca

†Equal contributors

¹Institut de recherche en biologie végétale, University of Montreal, 4101 Sherbrooke E, Montreal, QC H1X 2B2, Canada

Full list of author information is available at the end of the article

is phytoremediation, which exploits natural physiological mechanisms of plants to degrade, immobilise and/or selectively uptake contaminants from soil and water [8]. Fast growing short-rotation-coppice willows have emerged as a promising phytoremediation crop, capable of producing high biomass yields ($\sim 15 \text{ t ha}^{-1} \text{ year}^{-1}$) on polluted or degraded land [9, 10] as well as an attractive potential lignocellulosic crop [11], which can benefit local biodiversity [12] and does not require high fertiliser application [13]. Some varieties of willow are highly tolerant to organic contaminants such as: polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs) and petroleum hydrocarbons, and inorganic contaminants such as: As, Cd, Co, Cr, Cu, Hg, Mn, Ni, Pb, Sn and Ti [14, 15]. High biomass yields have also been maintained on land contaminated with high amounts of C10-C50 hydrocarbons, such as that of a petrochemical plant in Varennes, Quebec, (957 mg kg^{-1}) [16]. Whilst the physiology of willow has been studied in the presence of common soil contaminants, the molecular mechanisms underpinning this tolerance are still largely unknown.

Several trials have investigated the gene expression or transcriptomic response of willow, and similar crops, to a number of inorganic contaminants applied in isolation, such as: chromium [17], copper [18], zinc [19] and cadmium [20]. Little research, however, has investigated the transcriptomic response of woody crops to organic contamination. Recent research has explored rhizospheric bacteria and their role, hypothesised from expression profiles, in organic contamination tolerance by trees; Yergeau *et al.* [21] found overrepresentation of transcripts implying widespread changes in localised microbial competition. The expression profiles of endophytic bacteria in the above-ground tissues of these trees have been potentially hampered by traditional culturing of bacterial communities, which can limit the potential to identify some of the organisms of interest. However, research has identified some above-ground phytoremediation endophytes; Kang *et al.* [22] isolated a poplar endophyte that can degrade trichloroethylene, an organic contaminant.

To better understand how fast-growing trees can tolerate real-world polluted land, we assessed the changes in expression profiles of willow trees grown using soil from the site of a former petroleum refinery; either contaminated or non-contaminated with high levels of C10-C15 petroleum hydrocarbons. Recent evidence suggests endophytic bacteria are intimately involved in tree stress responses [23–25], and could potentially play an important role in the phytoremediation process. It also seems, as a more general factor to consider, that almost all organisms are involved in complex interdependent relationships as part of a “metaorganism” [26, 27]; we therefore assessed differentially expressed sequences without the limitation of direct mapping or annotation to *Salix*

purpurea genome alone, providing the opportunity for an unconstrained view of the origin of sequenced RNA.

Results

Tree growth

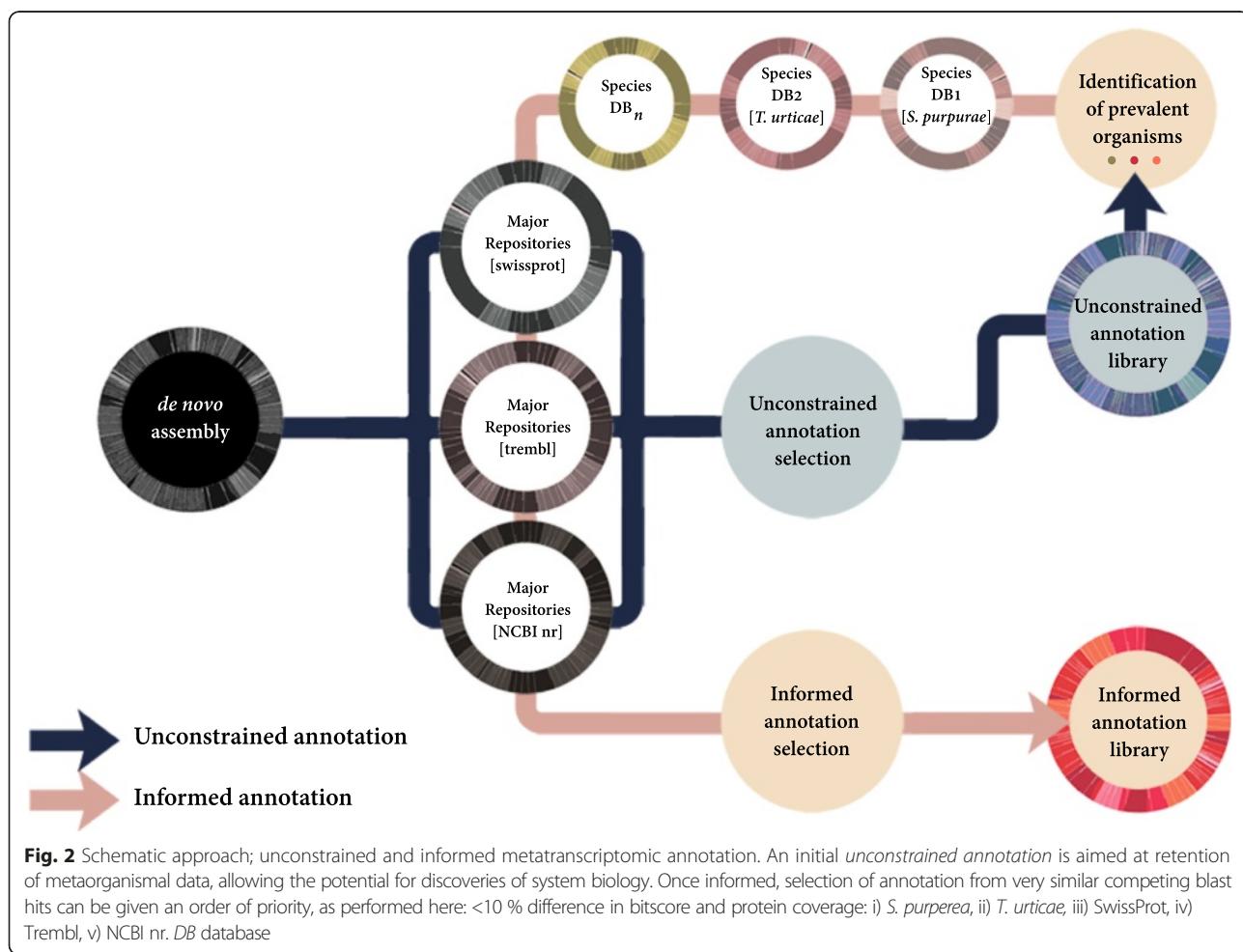
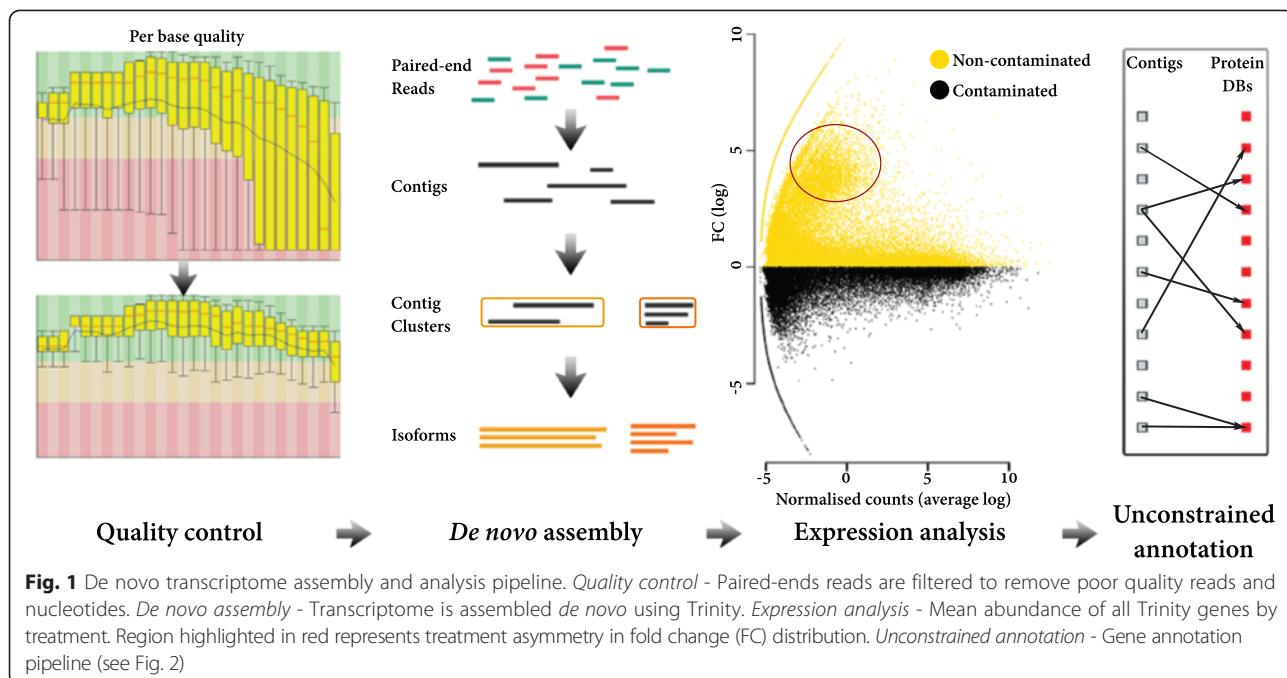
Composition of contaminated soil from Varennes was highly enriched with C10-C50's, PCB's and PAH's. The specific concentration level of each contaminant was assessed for each pot in the trial. In contaminated pots, C10-C50's concentration was on average 837.5 mg kg^{-1} , PAH's (collectively) averaged 62.5 mg kg^{-1} and PCB's averaged 0.2 mg kg^{-1} (Additional file 1: Table S1). The composition of non-contaminated soil, taken from the same site, did not have abundant amounts of these contaminants: C10-C50's $<100 \text{ mg kg}^{-1}$, PAH's $<0.1 \text{ mg kg}^{-1}$, and PCB's $<0.017 \text{ mg kg}^{-1}$. Trees did not show any observable difference in phenotype between treatments over the 6 months of the experiment, this included height which was an average of 223 cm at the time of harvest. Oven dried biomass yields were not significantly different (students *t*-test, $p = 0.67$) at an average of 47.31 g ($sd = 8.66$) and 45.09 g ($sd = 4.48$) for non-contaminated and contaminated trees, respectively.

De novo transcriptome assembly and annotation

RNA sequencing was used to estimate transcript abundance for three organs (buds, leaves or stems) sampled from eight trees, four grown on contaminated and four grown on non-contaminated soil, resulting in a total of 24 samples. Depth was a total of 2.4 billion reads, assembled into 424,752 isoforms (1,064,713,007 bp in total) that belonged to 68,191 Trinity genes. N50 of the isoforms was 3,392 bp and the average GC content 39.3 %. At this stage of *de novo* transcriptome analysis the transcripts of non-plant origin are often discarded prior to differential gene expression analysis. Here all assembled transcripts were retained, whatever their origin. A degree of asymmetry was observed within the entire gene expression dataset, with a large group of transcripts substantially upregulated in non-contaminated trees (Fig. 1). Two major species were identified by an initial impartial annotation of DE genes, here termed unconstrained annotation, *S. purpurea* and *T. urticae*. Species-specific sequence databases were added (to Swiss-Prot, TrEMBL and NCBI nr), allowing for a second, final annotation, here termed “informed annotation” (Fig. 2).

Differentially expressed gene origin

Out of the total 7275 unique genes that were identified as significantly differentially expressed, 6536 were assigned protein identity whilst 739 were classified as unknown (Additional file 2: file S1). Using this annotation approach (Fig. 2), transcripts were best annotated from a total of 138 different organisms.



The most prominent plant species was *S. purpurea*, as expected, with 41 % of all unique DE genes (Fig. 3a). We also found 0.3 % potential bacterial genes and 0.1 % potential fungal genes. Unexpectedly, a very high number

of DE transcripts were best annotated from animal species, with 44 % of all unique DE genes identified as from the polyphagous herbivore *T. urticae*, or two-spotted spidermite.

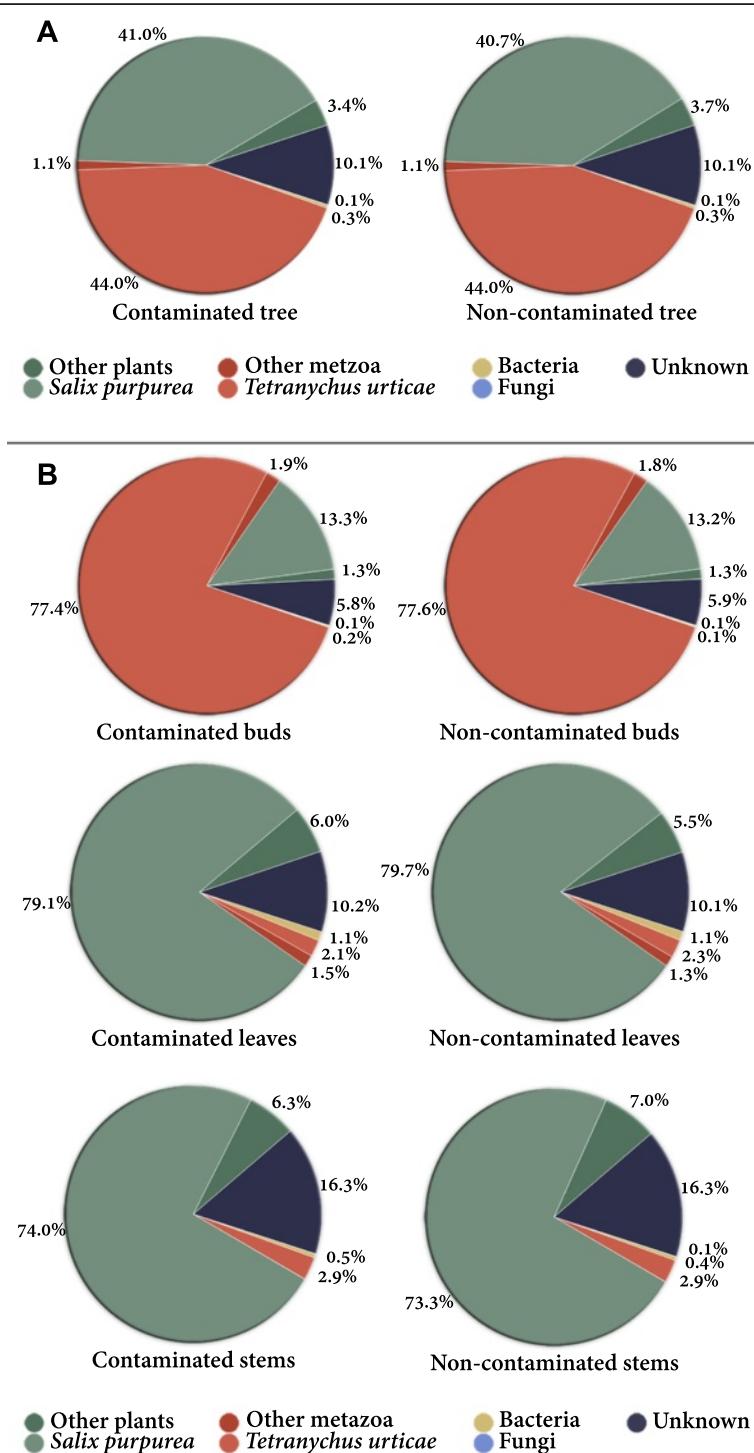


Fig. 3 Origin of unique genes differentially expressed due to treatment. **a** The origin of DE transcripts across the entire transcriptome and **b** separated by organ tissue after the best hit was selected from blast querying of *S. purpurea*, *T. urticae*, Swiss-Prot, TrEMBL and NCBI nr protein databases. PPDE >=0.95

Large variation in the species origin of DE genes was observed between the different tree organs, however, species composition varied very little between treatments (Fig. 3b). Although most plant DE genes were best annotated from the *S. purpurea* genome (Clone 94006), the likely source of all plant genes being the *S. purpurea* cultivar 'Fish Creek' grown here, around 8 % were preferentially annotated (>10 % better identity) from non-*Salix* plant species. The proportion of *S. purpurea* genes within the respective organ transcriptomes remained similar in leaves and stems, with leaves having 79–80 %, and stems having 73–74 %, in contaminated trees and non-contaminated trees respectively (Fig. 3b). Bud DE genes, however, differed markedly from stems and leaves in terms of species of origin. Only 13 % of DE genes in buds were from *S. purpurea*, over 77 % of the transcripts being annotated from *T. urticae*. The number of unique DE genes also varied substantially between organs but remained very similar between treatments; with stems, for example, having roughly four times more diverse *S. purpurea* transcripts than leaves (Additional file 2: file S1). One of the important aspects here was that, even though almost all the unique spidermite contigs were present in buds of both contaminated and non-contaminated trees, the abundance was so different (99 % of transcripts were in higher abundance in non-contaminated trees) (fragments per kilobase of transcript per million - fpkm) willow genes would have been highly biased if spidermite RNA was ignored. The importance of identifying as much RNA as possible is clearly of real consequence, not only in terms of the systems biological interactions, but also for the accurate technical quantification willow gene expression.

Gene ontology is often used as a means to reduce complexity of large'omic data to give clues as to trends of physiological response. A very large proportion of differentially expressed plant genes did not have GO (47.2 %), KEGG (96.5 %), KOG (57.3 %) or Panther (41.4 %) classification terms, so caution is taken in relying on partial data. However, a substantial number of genes involved in plant stress responses were differentially expressed, including large overrepresentation of oxidoreduction and defence proteins (Additional file 2: file S1).

Plant abiotic stress gene expression responses

When plant DE genes were investigated directly, those involved in responses to general and abiotic stress were identified in both contaminated and non-contaminated trees; the most prominent of which were involved in oxidoreduction mechanisms, drought stress and salinity stress responses.

A number of genes involved in reactive oxygen species (ROS) production and ROS scavenging mechanisms were differentially expressed (Additional file 1: Table S2). A cytosolic ascorbate peroxidase 1 (APX1) was present in very

high abundance in trees grown under both types of treatment (340 fpkm). A number of peroxidases were expressed in very high abundance relative to the contaminated transcriptome as a whole, *SapurV1A.0209s0110.1* and *SapurV1A.1899s0050.1*. A larger number of peroxidases were upregulated, to a lesser degree, in non-contaminated trees. Variation in redox mechanics, identified with genes potentially involved in ROS scavenging, such as glutamate synthase (*SapurV1A.0807s0060.1* and *SapurV1A.0174s0260.1*), were exclusively present in greater abundance in contaminated trees; whereas three OG FeII oxidoreductase gene transcripts, *SapurV1A.0587s0100.1*, *SapurV1A.0006s0890.1* and *SapurV1A.0020s0550.1*, were in greater abundance exclusively in non-contaminated trees. Nine DE gene transcripts with putative roles in drought resistance or response were identified as in greater abundance in contaminated trees. Highly abundant with respect to the transcriptome as a whole was a gene transcript coding for an early-responsive to dehydration protein (*SapurV1A.0030s0160.1*) that presents at an average of 360 fpkm in contaminated trees as well as dehydrin (*SapurV1A.0016s0910.1*), an abiotic responsive group of late embryogenesis abundant proteins, at 120 fpkm.

Genes involved in cell wall construction were differentially expressed between the two treatments; of 78 annotated cell wall polysaccharide genes, 38 had transcripts in higher abundance in contaminated trees and 40 in non-contaminated trees (Additional file 1: Table S3). Notably, cellulose synthase subunit A (*SapurV1A.0027s0400.1*) gene transcripts were in high abundance in contaminated trees. All of the seven DE xyloglucan endotransglycosylase/hydrolase (XET) gene transcripts were in higher abundance in non-contaminated trees. A substantial array of ethylene, auxin, abscisic acid, gibberellin and brassinosteroid, biosynthesis and receptor DE genes were also identified (Additional file 1: Table S4). Ethylene overproducer-like gene transcripts were uniformly in greater abundance in contaminated trees (*SapurV1A.0376s0050.2*, *SapurV1A.3328s0010.1*, *SapurV1A.0376s0050.2* and *SapurV1A.0006s1540.1*) as well as transcripts from a 1-aminocyclopropane-1-carboxylate oxidase (ACO - *SapurV1A.0829s0140.1*) gene that was in very high abundance.

Both transcript fold change and relative abundance between treatments are potentially, but not necessarily, a valuable means by which to interrogate large transcriptomic datasets for DE genes of interest. A useful way to integrate these two approaches is by using abundance weighted fold change (Fig. 4, Additional file 2: file S1). The transcript with the strongest homology to phenylalanine ammonia lyase (PAL), the third most abundant DE plant transcript in the contaminated trees across all tissues, was the only well annotated transcript to prominently stand out using weighted fold change. Two PAL

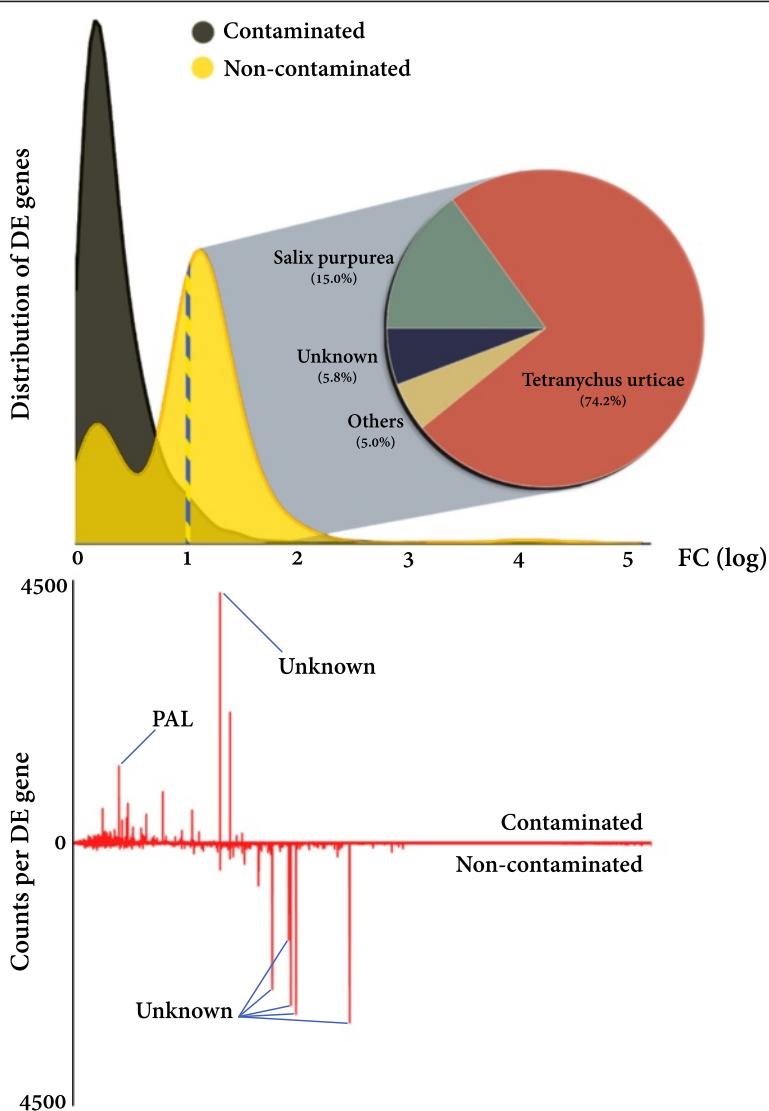


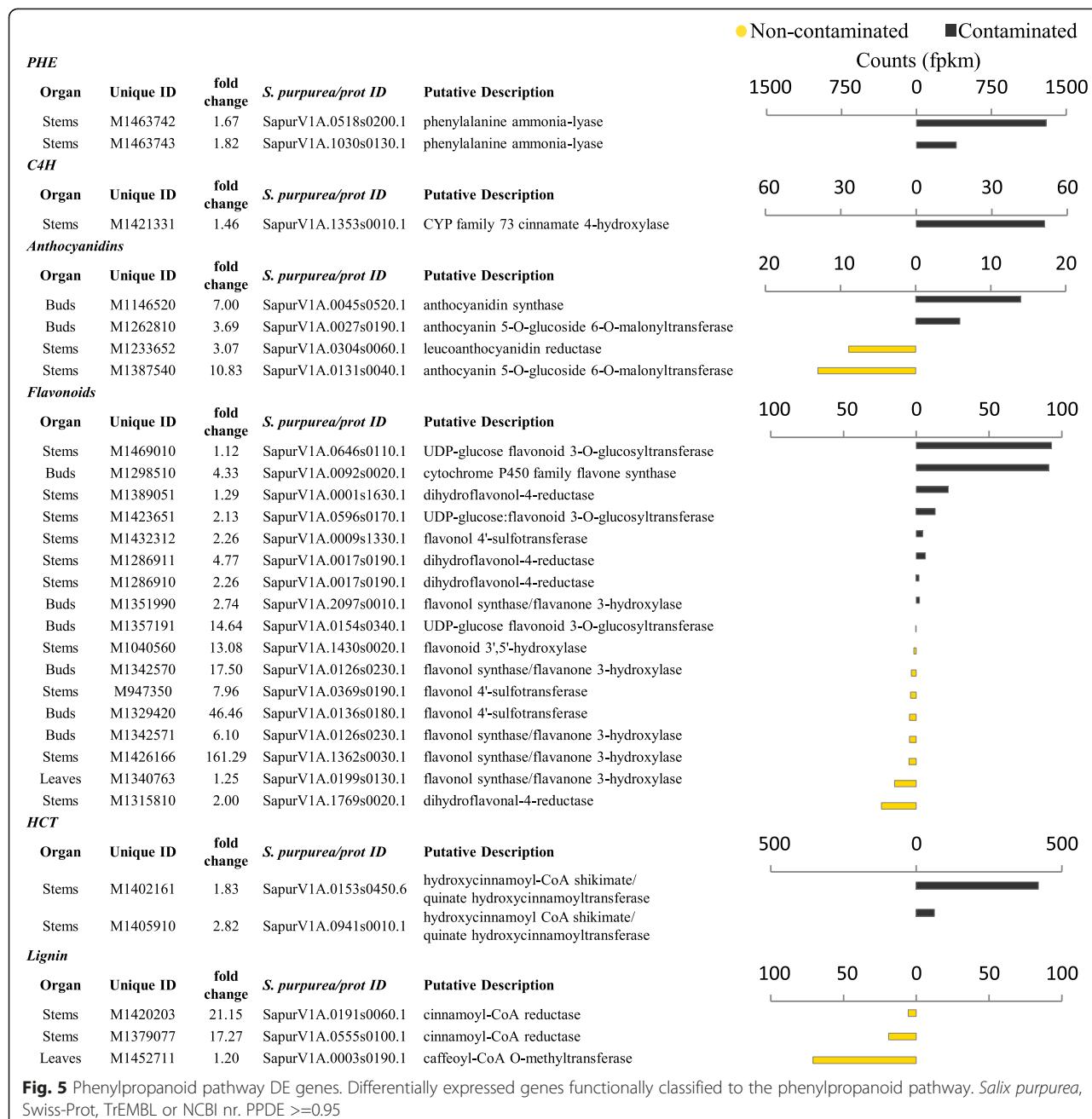
Fig. 4 Differentially expressed gene distribution and abundance weighted fold change. Fold change (FC) distribution of DE genes (top) per treatment. The origin (organism) of genes with FC increases of >1 (\log_{10}) in non-contaminated trees is highlighted (dashed line and pie chart). Individual (normalised mean) transcript counts (fpkm) per differentially expressed gene (bottom) are segregated by fold change for a weighted view of differential expression. The higher of the two treatments is shown for each DE gene. Major peaks in abundance weighted FC are highlighted and labelled. PAL phenylalanine ammonia lyase. PPDE ≥ 0.95

gene transcripts, with strong homology to PAL in the *S. purpurea* genome, were present in high abundance in the contaminated trees (Fig. 5). The principal route for plant defence metabolite production, the phenylpropanoid pathway, begins with deamination of phenylalanine by PAL. A number of genes downstream of PAL in the polypropanoid pathway were also differentially expressed. Genes involved in production of defence secondary metabolites, such as phenolic glycosides and condensed tannins, were upregulated in both treatments. Notably transcripts from a cytochrome P450 flavone synthesis gene were in high abundance in contaminated trees (*SapurV1A.0092s0020.1*). Genes directly involved in

lignin subunit biosynthesis, cinnamoyl-CoA reductase (*SapurV1A.0191s0060.1* and *SapurV1A.0555s0100.1*) and caffeoyl-CoA O-methyltransferase (*SapurV1A.0003s0190.1*) had transcripts consistently in higher abundance in non-contaminated.

Plant biotic stress gene expression responses

The clearest treatment-specific effect was the very high abundance of non-plant DE gene transcripts in non-contaminated trees. We observed that 44 % of the unique DE transcripts across all tissues were best identified as *T. urticae* genes (when blasted without substantial bias towards any given organism); with the buds having

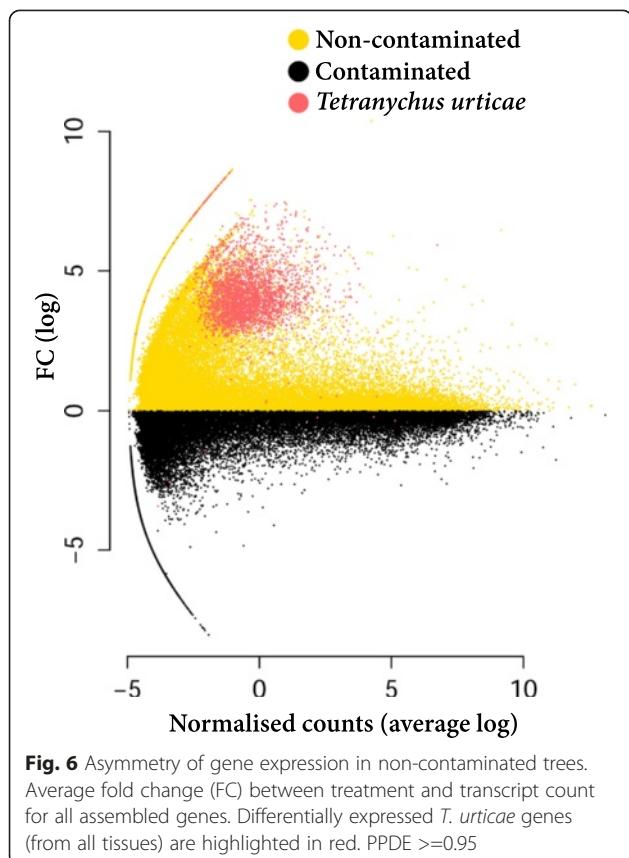


the majority of metazoan sequence, 97 % of which was *T. urticae* (Fig. 3b). *Tetranychus urticae* genes had very stark treatment-specific expression, with 99 % in higher abundance in non-contaminated trees (Figs. 4 and 6). Under the assumption that *T. urticae* preferentially infested non-contaminated trees, we investigated whether *Salix* resistance gene (R-gene) expression reflected an increase in biotic challenge in a treatment-specific manner. Of the R-genes, 12 coiled-coil nucleotide-binding site leucine-rich repeat genes (CC-NBS-LRR) were identified as differentially expressed and were in greater abundance

in non-contaminated trees without exception (Fig. 7). Twelve toll-interleukin (TIR)-NB-LRR DE genes were identified with transcripts in greater abundance in non-contaminated trees with one exception (*tr|Q1KT00_poptr*). There were also 16 BED finger NBS-LRR genes identified, 14 with transcripts in greater abundance in non-contaminated trees.

Tetranychus urticae gene expression

A number of *T. urticae* transcripts whose abundance is thought specific to a given developmental stage were



found to be differentially expressed and were in higher abundance in the buds of non-contaminated trees (Fig. 8) but also at high levels compared to the rest of all the DE genes in the non-contaminated tree transcriptome. Three larva specific markers were identified: *tetur20g00200*, *tetur02g10770* and *tetur20g00200* as well as one adult specific marker *tetur01g02670*. Four Embryo specific markers were upregulated: *tetur04g01610*, *tetur04g01580*, *tetur11g00600* and *tetur34g00420*. Only the Embryonic marker *tetur34g00420* was present at the extraordinary high abundance expected as characteristic of a dominant developmental stage.

Transcripts encoding *T. urticae* detoxification proteins, characteristic of arthropod responses to toxic plant secondary metabolites, were identified as present in high abundance in the non-contaminated trees: these included fifteen cytochrome P450 genes (including clan's 2, 3 and 4 as well as mitochondrial), nine glutathione S-transferases genes (of classes omega, mu and delta), 11 Carboxyl/cholinesterase genes and 10 ABC C transporters (and 1 ABC B transporter) (Fig. 9). A suite of cysteine peptidases also had transcripts in greater abundance in the buds of non-contaminated trees. These cover the four major classes identified as the spidermite's proteolytic digesting equipment: papains (C1A), legumains (C13), caspases (C14) and calpains (C2). Transcripts from four C1A

papain (*tetur08g05010*, *tetur12g01860*, *tetur12g04631* and *tetur25g00650*) and two C13 legumain genes (*tetur05g04550* and *tetur28g01760*) were present in the highest abundance.

Microorganism differentially expressed genes

Thirty-two transcripts derived from microorganisms were differentially expressed in response to treatment. Although only 25 % of the transcripts annotated from microorganisms had informative functional description, the taxonomic origin of transcripts can be informative. Six unique transcripts, derived from *Propionibacterium acnes*, were all in greater abundance in buds of non-contaminated trees expressed alongside the very high *T. urticae* gene expression (Additional file 1: Table S5). These included a glycosyl hydrolase (GH65) and a protein with an alpha/beta hydrolase domain. A number of bacteria, such as *Bacillus stratosphericus* and *Klebsiella* sp., were putatively identified as the origin of transcripts in greater abundance in trees cultivated on contaminated soil.

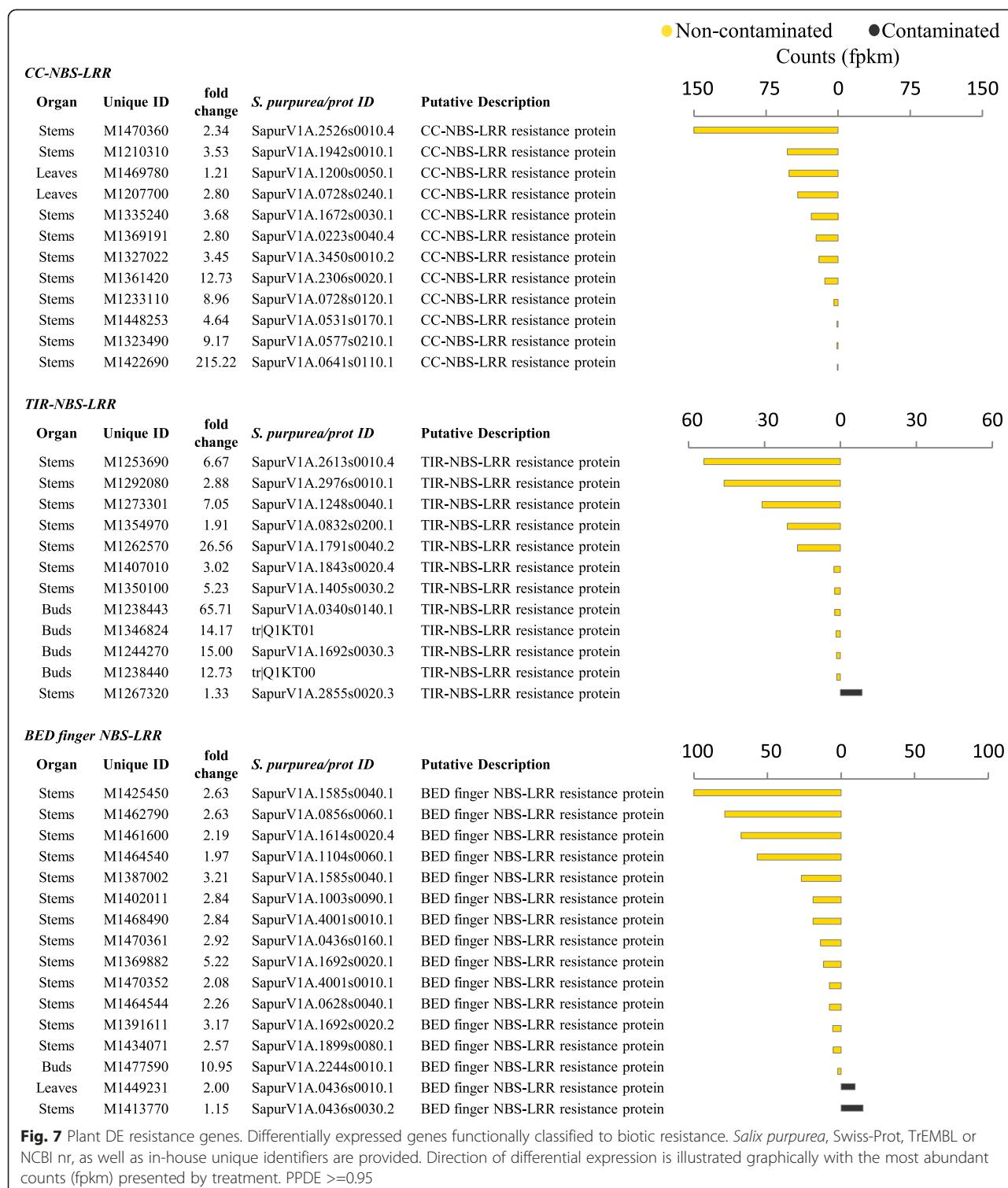
Discussion

Willow tolerance of petroleum hydrocarbons

The cultivar 'Fish Creek' (*S. purpurea*) has been shown to tolerate high levels of C10-C50 petroleum hydrocarbon soil contamination without substantial losses to biomass yields [16]. The capacity of this crop to tolerate real-world levels of petroleum by-product contamination establishes the potential to utilise and derive value from this type of unused marginal land via biomass and bioproduct production. This tolerance was further demonstrated here by the lack of significant reduction in biomass yield or even any clear, observable phenotypic in pot-grown greenhouse trees established in contaminated soil.

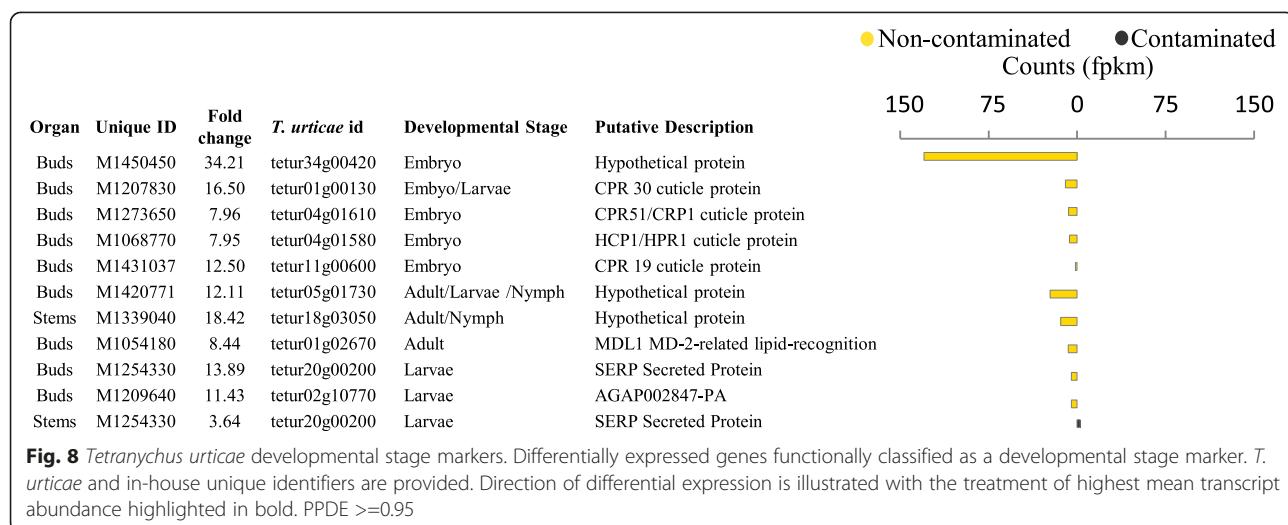
De novo transcriptome assembly and unconstrained annotation

Although effective *de novo* approaches exist for transcriptome assembly, the annotation step is often a bottleneck for non-model organism studies, especially if multiple accessions (or species) are being compared. The complexity of extra-laboratory biological systems, and concepts such as the meta-organism [26, 27], further exacerbates this problem as foreign organisms (such as endophytes) are excluded *ipso facto* from analysis during mapping of RNAseq data to a reference genome or are deliberately discarded as foreign upon annotation. That is, even when a reference genome is available, using it for quantifying gene expression alone by direct mapping can be a risky, or even flawed, approach. Software such as Trinity allows comparison of *de novo* assemblies from different cultivars (or accessions) without bias due to genetic distance from a reference genome. Transcripts can then be annotated by blasting against a reference genome but those assembled transcripts which



are not present in the reference genome, such as those unique to an accession, are also preserved and can be annotated independently. Here *de novo* assembly was preferentially utilised even though a single willow cultivar was being assessed and the annotated genome sequence

for that species (*S. purpurea* clone 94006) was available. This was due partly to the broad diversity of willow in mind but also because of compelling research performed by Doty *et al.* [22–24, 28] identifying endophytes involved in phytoremediation in *Populus trichocarpa*, indicating



treatment-specific variation in expression from foreign organisms is important to maintain and reveal. Any DE transcripts unique to *S. purpurea* cv. 'Fish Creek' but not present in *S. purpurea* cv. Clone 94006 (the reference genome) are also retained, an important choice given that 8 % of plant transcripts we identified as better annotated by plants other than this reference *Salix* genome (Fig. 3a). More importantly, over 50 % of the unique transcripts differentially expressed between treatments would have been lost if RNA reads would have been mapped directly to the reference *Salix* genome.

Plant organ species profiles

The substantial variation in types of non-*salix* species present between the organs suggests that these different tissues represent highly distinctive environments. From a metaorganismal perspective, it would be surprising if the multispecies makeup was constant throughout a plant, as different tissues are hospitable to different organisms. This is supported by the similarity of species (transcript) distribution between treatments within an organ and that very few DE transcripts were shared between organs (412 shared between two organs). Only a very small number of DE genes were unique to treatment within an organ, variation coming from abundance of DE transcripts. Bao *et al.* [29] recently demonstrated that at least 36 % of genes in poplar (which shares close macrosynteny with willow [30]) undergo alternate splicing, analysis of isoform variation was beyond the scope of the work here but such variation would seem likely.

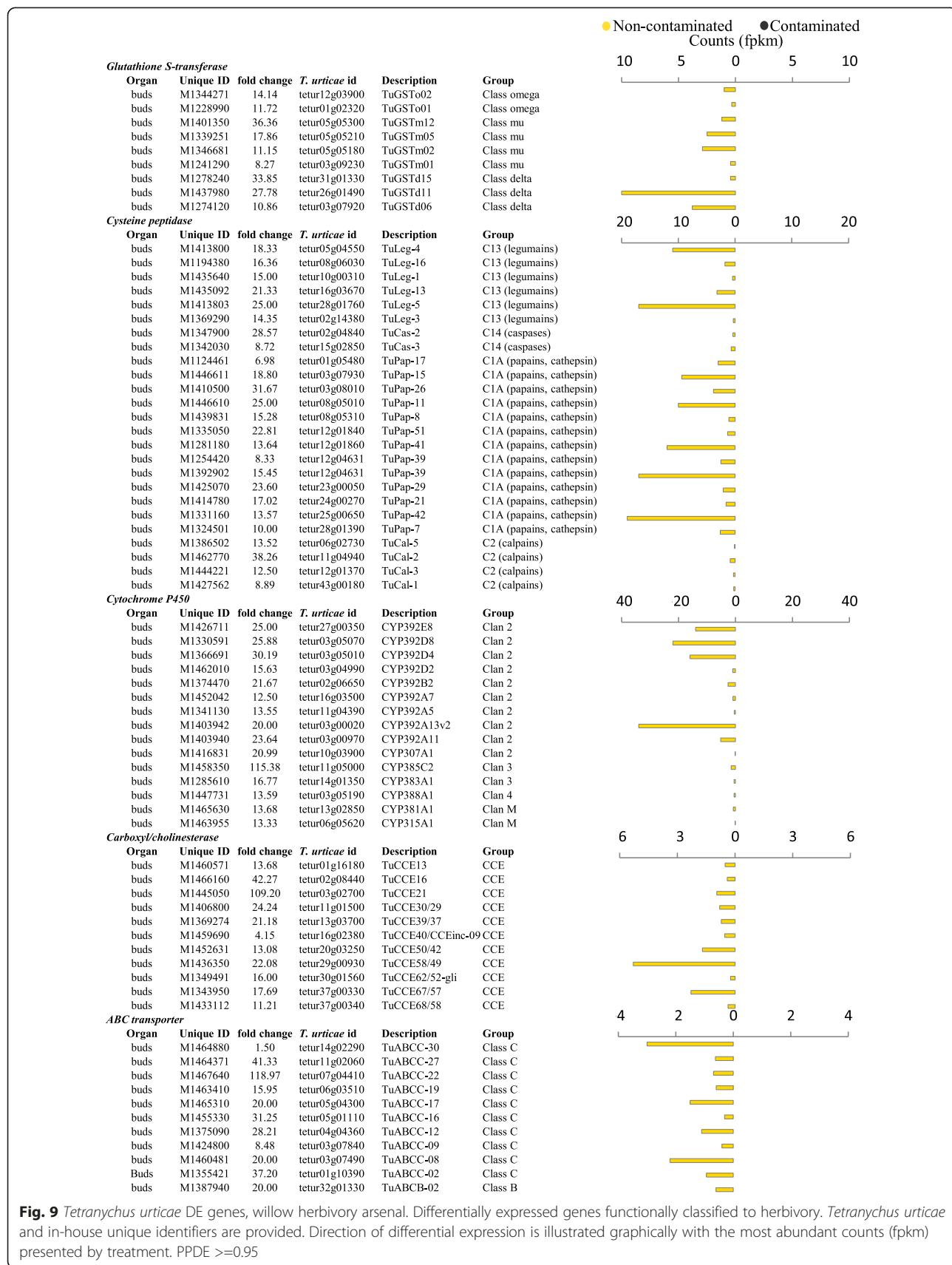
The willow transcriptome exhibited differential expression of genes implying substantial abiotic or biotic stress in both treatments. Plant genes involved in important pathways, such as ROS synthesis and scavenging mechanisms, were present in both treatments. However, differential expression of genes involved in discrete physiological

responses could be identified. Using gene ontology analysis of DE genes to reduce complexity, there were broad changes in nucleotide binding, oxidoreduction and defence protein biosynthesis (Additional file 2: file S1); however, whilst used extensively in model species, gene ontology is less powerful in non-model crops as GO, KEGG or Panther terms were not available for many genes (<60 % annotation of unique transcripts). Similar to the problems of mapping reads to the reference *Salix* genome alone, utilising only 60 % of the data runs the risk of being arbitrarily reductionist. Instead, specific gene expression was investigated for these broad categories; in doing so the authors recognise the risks of interpreting phenotype through transcript abundance alone, beyond successful contamination tolerance, and do so with care.

Plant abiotic stress gene expression

In high biomass yielding crops, such as willow, the vascular cambium represents the majority of cellular division, and therefore, mass of the tree; so strictly regulated stress mechanisms and responses are not unexpected in stem tissue. The involvement of ROS in stress resistance (and many plant physiological processes) is well established [31, 32] as well as, more specifically, in resistance of willow and closely related organisms [33, 34]. In high abundance in both contaminated and non-contaminated trees (significant DE but very low fold change) was APX1 (Additional file 1: Table S2), a central component of the reactive oxygen gene network in the cytosol of *Arabidopsis* leaves (Davletova *et al.* [35]). Intriguingly, APX1 [36, 37], alongside AP2/ERF transcription factors [38], have also been implicated in biotic and abiotic cross-tolerance.

A number of *Salix* genes likely to be involved in detoxification of ROS as well as general response to the stress induced by C10-C50 contamination were differentially expressed and in greater abundance in contaminated



tress. These included a number of glutamate synthase genes, glutamate being a constituent of glutathione, an important molecule for oxidative stress tolerance via ROS detoxification [39], as well as a wide number of drought responsive genes implicit in broad abiotic resistance and also associated, in poplar [40], with production of antioxidant phenolic compounds such as kaempferol and catechin.

Major alterations in cell wall construction can be expected between the two treatments as high numbers of cell wall genes were differentially expressed (Additional file 1: Table S3). XET expression showed a clear treatment specific pattern of great abundance in non-contaminated trees. Research performed by Albert *et al.* [41] revealed XET mRNA accumulation in response to biotic stress from the parasite *Cuscuta reflexa*. The same uniform pattern of higher expression in non-contaminated trees was observed in three transcripts encoding two enzymes directly involved in cell wall lignin monomer biosynthesis (Fig. 5), cinnamoyl-CoA reductase and caffeoyl-CoA O-methyltransferase.

A wide spectrum of differentially expressed brassinosteroid, ethylene, abscisic acid and auxin biosynthesis and receptor genes are to be expected in response to extensive (treatment-specific) biotic and abiotic stress. Genes involved in ethylene signalling were strongly differentially expressed with ethylene overproducer-like transcripts present in high abundance in contaminated trees, along with others involved in the ethylene biosynthesis pathway, including S-adenosyl-L-methionine (SAM), 1-aminocyclopropane-1-carboxylate (ACC) and ACC oxidase (ACO). Interestingly, when viewed alongside the very high overexpression of FLA12 in the stems of contaminated trees (Additional file 1: Table S3), a characteristic marker of tension wood tissue in reaction wood producing poplar [42–44] and willow (Brereton, unpublished) trees, an expression profile very similar to tension wood production emerges. Adding to this is a concurrent increase in the abundance of a CesA in stems of contaminated trees as well as the very high abundance of ACO (Additional file 1: Table S4), thought to result in the secondary xylem tissue localisation (or asymmetry) of ethylene biosynthesis from ubiquitous (non-localised) ACC [45]. A link between salinity and drought stress, and the formation of a tension wood has recently been suggested [46, 47] although as an antagonistic response, where tension wood markers such as FLA12 were down-regulated.

One of the most abundant DE transcripts in contaminated trees (the third for plant transcripts) was PAL: representing one of a number of genes involved in the phenylpropanoid-flavonoids pathways known to drive secondary metabolite production in response to abiotic stress [48, 49]. Transcripts encoding proteins at the beginning of the polypropanoid pathway, including PAL and a

cinnamate 4-hydroxylase (C4H) (Fig. 5), were strongly up-regulated in contaminated trees, perhaps expected due to high contamination and potential drought/osmotic stress (Additional file 1: Table S2) [33, 34, 40, 50–52]. The relationship between abiotic stress induced increases in PAL and C4H with subsequent increases in phenolics, such as catechin, have previously been observed in other crop species such as tea [53], tomato [54], potato [55] and lettuce [56] as well as being well reported in poplar [57]. Although it is common knowledge that large amounts of salicylic acid can be present in willow, a wide spectrum of other phenolic compounds can be present in very high abundance in *Salix* [58–60]. Many of these are induced defence compounds in response to arthropod herbivory in *Salix* [61, 62] and, interestingly, a number of compounds downstream of PAL catalysis have been shown to be highly induced in response to *T. urticae* feeding [63]. This agrees with the potential for cross-tolerance as common phenolic production is elicited by either abiotic or biotic stress. Recent evidence implies this is likely as PAL can be directly induced as a defence response to arthropod herbivory [64] in addition to its role in abiotic stress response.

***Tetranychus urticae* developmental stage expression profile**

Tetranychus urticae life cycle progresses over four stages: embryo, larval, nymph and then adult [65, 66]. *T. urticae* DE transcripts, 99 % of which were in higher abundance in buds of non-contaminated trees, were screened for the genes identified as markers for specific developmental stages when highly abundant [65]. Clear, very high abundance of *tetur34g00420* gene, one of the 10 embryo markers, suggests an early developmental stage of *T. urticae* at this time (Fig. 8). Larval markers were also present in relatively small abundance and adult markers in medium abundance.

This important treatment-specific interaction between willow and *T. urticae* would have been lost if the de novo assembled transcriptome were annotated using the *S. purpurea* genome alone (or if reads had been mapped to the reference genome). A drawback of such discoveries, based solely on next-generation sequencing data, is that phenotypic data on *T. urticae* infection is challenging to gather post-hoc. It is interesting to note that, if the harvest was indeed performed at an initial phase of embryonic infection, phenotype might have been difficult to recognise without prior knowledge of the interaction. This is because the microscopic eggs need to be deliberately sought and the characteristic webbing, clearly visible in abundance at later infestation stages, may not yet be established.

Plant biotic stress gene expression

The majority of biotic or disease resistance genes (R-genes) encode nucleotide binding sequence leucine

rich repeat proteins in plants [67]. These ancient proteins represent the incredibly diverse and adaptable frontline defence machinery in plants capable of recognising foreign molecules and then eliciting appropriate responses. Recently Yang *et al.* [68] identified 330 NBS-LRR encoding genes in poplar. In similar research Kohler *et al.* [69] identified 119 CC-NBS-LRR, 64 TIR-NBS-LRR and 34 BED-finger-NBS-LRR genes, suggesting a diverse and extensive resistance arsenal is present. As 99 % of *T. urticae* transcripts were in greater abundance in non-contaminated trees, we considered what the plant response to highly increased infestation would be and investigated the differential expression of R-genes between the two treatments. The hypothesis being that a substantial increase of *T. urticae* challenge in non-contaminated trees would induce distinct immune responses.

A very stark pattern of expression was observed; a wide spectrum of intracellular NBS-LRR encoding R-genes were uniformly in greater abundance in non-contaminated, potentially *T. urticae* infected trees (Fig. 7). These included increased abundance of the largest classes of R-gene intracellular receptors; 100 % of the CC-NB-LRR resistance genes were in greater abundance in non-contaminated trees, 92 % of TIR-NB-LRR resistance genes and 88 % of BED finger NB-LRR resistance genes, providing evidence of a more robust biotic resistance response in non-contaminated trees.

***Tetranychus urticae* herbivory gene expression**

Upregulation of classical *T. urticae* detoxification genes was observed from non-contaminated trees (Fig. 9). These genes included glutathione-S-transferases (GST), carboxyl/cholinesterase [70], ABC transporters and a large number of cytochrome P450's [65]. GST's can conjugate toxic compounds, such as exuded plant secondary metabolites (ie. anthrocyclins and alkaloids), into glutathione (GSH). Large numbers of ABC transporters have been shown as present in *T. urticae* (at least 103 genes) with the ABC C transporters identified here (which forms a sister clade to multidrug resistance-associated proteins), known to transport neutral conjugates such as GSH [71]. The spidermite's digestive arsenal used against willow consists predominantly of cysteine proteases [72, 73], this is reflected in the abundant transcripts present in buds of non-contaminated trees of two classes of cysteine proteases: papain (C1A), likely to have a role in the digestive process, and legumain (C13), potentially involved in feeding and digestion of plant material, were numerous and had transcripts in high abundance [73, 74].

There is an established history of increased herbivory resistance in poor nutrient quality and abiotically challenging environments [75, 76], as well as preferential herbivory, from *T. urticae*, of crops with high nutrient content [77, 78]. Without direct assessment of *T. urticae*

infestation we cannot conclusively say whether this increase in plant biotic resistance machinery (R-genes) was because of preferential *T. urticae* herbivory of non-contaminated trees or a compromised biotic response in contaminated trees (such as *T. urticae* effector driven suppression of R-genes). We favour the likelihood of willow cross-tolerance, driven by the soil contamination, as *T. urticae* gene expression was so comprehensively upregulated in non-contaminated trees (99 % of genes were in greater abundance). To test this, targeted research in both greenhouse and field environments is required and could have important implications, both for fundamental biology of abiotic and biotic resistance mechanisms as well as for the phytoremediation industry. Additional evidence supporting this hypothesis comes from Grenier *et al.* [16] where, as an independent observation of mature *S. purpurea* 'Fish Creek' phenotype grown on the site where contaminated soil came from (Varennes, QC), it was noted that the insect *Tuberolachnus salignus* (large willow aphid) preferentially infested the willows established on non-contaminated over contaminated land.

Microorganism differentially expressed genes

Whilst little information can be derived regarding microorganism function due to poor annotation description, the type of bacteria present is interesting as the community make-up is thought to drastically change in response to contamination [79]. The most prominent organism in DE transcripts upregulated in the buds of non-contaminated trees, concurrent with 99 % of *T. urticae* transcripts, was *Propionibacterium acnes* (Additional file 1: Table S5). *Propionibacterium acnes* is well known as an endophyte in grapevine [80], *Thlaspi goesingense* [81] and poplar [82]; a less well recognised role is as part of mite bacterial flora. Known to associate with the *Psoroptes ovis* mite [83], *P. acnes* grows in the midgut where it's thought to be involved in the digestive process. As the bacteria produces propionic acid from carbohydrate, it is of potential interest that one of the DE transcripts was a glycosyl hydrolase (GH65) which could relate to plant cell wall depolymerisation. It would be of interest to establish if any of the microorganisms represented in non-contaminated trees here were directly part of a *T. urticae* salivary bacterial population as such strategies are used, in some cases, to undermine plant defences [84].

Out of the 14 DE transcripts in greater abundance in contaminated trees, a number were associated with microorganisms commonly found in wastewater. *Bacillus stratosphericus* is characterised as a bacteria which inhabits the stratosphere however, many *B. stratosphericus* isolates come from unclean water sources, perhaps reflecting the contaminated soil here. Zhang *et al.* [85] isolated *B. stratosphericus* from river water (Tyne, UK) and groups such as Gupta *et al.* [86], who demonstrated heavy metal

remediation properties in *B. stratosphericus*, isolated from soil rich in municipal waste. Klebsiella sp., also in greater abundance in contaminated trees, have been isolated specifically from sludge in wastewater plants [87]. Along these same lines; one of these 14 transcripts characteristic of contaminated trees was best annotated with an unknown sequence isolated from an (arsenic rich) mine drainage bacterial metagenome [88].

Cross-tolerance potential of *S. purpurea*

If the increased *T. urticae* infestation of non-contaminated trees is accepted from gene expression analysis alone then a number of potential causes related to treatment can be considered. As discussed above, a possible reason may relate to variation in abiotic stress responses as induced by soil contamination. Differential expression of genes involved in specific biotic and abiotic responses were identified between treatments. Although such relationships are often considered as antagonistic there are many examples of *increases* in biotic resistance due to abiotic stress response, frequently achieved through expression of phenolic defence compounds, in *Arabidopsis* [89], tobacco [38], tomato [90], potato [50]. Genes involved in abscisic acid and salicylic acid, hormones involved in complex stress response regulatory networks, were differentially expressed and are presented as potential candidates for such interactions. This “real world” soil presents a disadvantage for unpicking such interactions but grants advantage in retaining some complexity of these crops systems, applicable to the field, and thus, a wealth of gene candidates with hypotheses of potential value to pursue.

A means by which cross-tolerance could be achieved by *S. purpurea* cultivated on contaminated land could be the increase production of compounds downstream in the polypropanoid pathway, such as condensed-tannins, often produced in high abundance in *Salix* [59]. Kao *et al.* [91] demonstrated strong association between PAL expression and condensed tannin accumulation in poplar. If the same mechanisms exist in willow, including the production of extractable compounds such as condensed tannins [92], then further value to the willow phytoremediation system can be established.

The differential expression of genes due to treatment does suggest a complex interplay between abiotic and biotic stress; this needs to be robustly tested before certainty can be established and the authors speculate about such interactions with care. If true, presence of *T. urticae* would compel an extension of the view of tolerant phytoremediating species such as willow; not only do species such as willow find reduced competition from other plant species in these difficult environments but may also benefit, in principle, from the advantage of reduced biotic challenge.

Conclusions

Over half of the differentially expressed genes would have been discarded if RNA reads were mapped directly to a reference *Salix* genome and, importantly, a crucial factor would have been overlooked with the potential to interact with the entirety of the meta-transcriptomic system. Roughly equal amounts of unique genes from spidermite were identified as from willow. Close to all genes deriving from spidermites had transcripts in greater abundance in non-contaminated trees, which in turn had a suite of plant resistance genes (also) uniformly in greater abundance than in contaminated trees. We propose that abiotic stress responses in contaminated trees, such as increased PAL expression, could interact with spidermite infestation.

It is an ambitious target to mitigate environmental impact of the petrochemical industry by rejuvenating polluted land as well as adding value to that land, or marginal land capitalisation, through the production of biomass. In terms of that added value, these results suggest that the very act of tolerating pollution may provide willows with a secondary competitive advantage, facilitating high biomass yields challenging environments, through cross-tolerance against biotic attack.

Methods

Experimental design, plant material and soil composition

Soil was collected from the site of a former petrochemical refinery at Varennes, Canada, either enriched in (contaminated) or lacking (non-contaminated) C10-C50 petroleum hydrocarbons (Additional file 1: Table S1). Cuttings of *S. purpurea* cv. ‘Fish Creek’ were sampled as part of a larger split-plot sampling designed consisting of six blocks [21] of which we only investigated the (randomised) contamination effect. The 20 cm cuttings were first established for 8 weeks in conventional potting media before being transfer to the treatment soil in 20 l pots. Six replicates per treatment were sampled for RNAseq (but only four were sequenced downstream). Growth conditions were 16 h 20 °C day and 8 h 18 °C night with excess watering and individual plant pot saucers to reduce leaching. Trees were grown for 6 months before harvesting: leaves, stem and buds. The leaves sampled were from the 5 to 15th fully unfurled leaf from the top of the stem. The term buds here refers to the developing tip of the stems, not axillary buds. Leaves and buds were flash frozen in liquid nitrogen, stem vascular cambium samples were taken, from stem sections towards the base of the stem, by peeling off bark and scraping the inner developing xylem before flash freezing [93, 94].

RNA extraction, sequencing and quality control

Total RNA was extracted using the modified CTAB protocol [95, 96] with RNA quantity and quality assessed with a BioAnalyser (Agilent). The four best replicates, in

terms of potential degradation, were used for downstream RNAseq (selected from the six samples based on highest RNA integrity number - RIN). TrueSeq 100 bp paired-ends libraries were constructed from extracted mRNA (Illumina® TruSeq® RNA Sample Preparation Kit v1, purifying PolyA containing mRNA using poly-Toligo attached magnetic beads before random hexamer pairing for the cDNA synthesis). Samples were sequenced (four per lane) using an Illumina HiSeq 2000 sequencing system. Sequence data was quality controlled using Trimmomatic [97].

De novo transcriptome assembly and differential expression analysis

Trinity software [98] was used to reconstruct the transcriptome *de novo* using default settings and discarding transcripts <201 bp. Sequences qualified as a “gene” were the union of transcripts similar enough to be considered by Trinity as putative isoforms of the same gene. Bowtie2 [99] software was used to map RNA-seq reads to the *de novo* assembled transcriptome with strict parameters (–no-discordant –no-mixed). A median of 84 % of RNA-Seq reads mapped back to our transcriptome, suggesting the transcriptome correctly represents our data (see Additional file 2: file S1). Raw and normalised transcript abundance was calculated using eXpress [100] with default parameters.

Given the open nature of the annotation we proceeded with these Trinity genes for differential expression analysis instead putative isoforms. It should be noted that these Trinity genes can also summarise allelic variants and duplicate genes and that, ideally, all isoforms should be analysed individually so biological information of greater resolution can be preserved. To assess differential expression, EBSeq [101] was used to identify genes for which transcript abundance had changed significantly between the two experimental conditions for each tissue. EBSeq is less prone to adjust or to discard data due to assumed distribution or FDR (false discovery rate) control [102], exhibiting more consistent behaviour than other DE software when using these real datasets (data not shown). Default parameters were used with 15 iterations for convergence at an FDR of 5 % with significance identified and expressed as posterior probability differential expression (PPDE) greater or equal to 0.95 [101].

We first performed an *unconstrained* annotation of the transcriptome (Fig. 2) using three major protein databases: nr, SwissProt and TrEMBL. This was deliberately performed without assumption as to the origin of sequence. All DE genes sequences passed through each of the three databases using blastx [103] with the following parameters: e-value 1e⁻⁴, gapopen 11, gapextend 1, word_size 3, matrix BLOSUM62. For each query, up to 20 results were retained to ensure presence of the highest protein similarities without a sole reliance on e-value, unsuitable for comparison of

proteins from different databases [104]. Bitscore and protein coverage were used as selection parameters for the best hit from each database and again for selecting the best hit across databases to build the unconstrained annotation library (Fig. 2).

The most prevalent species revealed using unconstrained annotation of DE genes were then identified: in this case being *S. purpurea* and *T. urticae* (the two-spotted spidermite). Recently, the complete genome and annotation of *S. purpurea* genome has been made available on phytozome (*Salix purpurea* v1.0, DOE-JGI, http://phytozome.jgi.doe.gov/pz/portal.html#?info?alias=Org_Spurpurea), drafted by Smart *et al.* (unpublished) in collaboration with the DOE-JGI. The *T. urticae* genome is available at EnsemblMetazoa (http://metazoa.ensembl.org/Tetranychus_urticae/Info/Index) [65]. Once the most prominent organisms had been identified, a second *informed* annotation (Fig. 2) was performed which included these additional species-specific protein databases. Just less than 43.8 % of the final annotation derived from the *T. urticae* database (3193 unique genes), 40.6 % from *S. purpurea* (2960 unique genes), 0.5 % from Swiss-Prot (33 unique genes), 4.5 % from TrEMBL (324 unique genes), 0.4 % from nr (26 unique genes) and 10.2 % unknown (no hit; 739 unique genes).

Additional files

Additional file 1: Table S1. Soil contaminants. Table S2 Plant stress-related DE genes. Table S3 Plant carbohydrate/cell wall DE genes. Table S4 Plant hormone/signalling DE genes. Table S5 Differentially expressed transcripts from microorganisms. Table S6 Resistance related DE genes (DOCX 111 kb)

Additional file 2: File S1. Containing abundance and annotation of 7486 differentially expressed genes. RNAseq data has been deposited in the European Nucleotide Archive (ENA) and can be found via the unique study accession number: PRJEB10287. (XLSX 1320 kb)

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

SJ, FP and ML designed the study. SJ, FP, ML, JM and WGN performed the plant growth trials and sample preparation. EG, NJBB, ML, FEP and SJ interpreted the data and drafted the manuscript. All authors read and approved the final manuscript.

Authors' information

Not applicable.

Availability of data and materials

Not applicable.

Acknowledgements

We are grateful to Péromont Inc. for allowing access to the Varennes site. We thank the Genome Quebec Innovation Centre for support and Calcul Québec for computing resources.

Funding

The project was funded by the GenoRem Project (Genome Canada and Genome Québec).

Author details

¹Institut de recherche en biologie végétale, University of Montreal, 4101 Sherbrooke E, Montreal, QC H1X 2B2, Canada. ²Montreal Botanical Gardens, 4101 Sherbrooke E, Montreal, QC H1X 2B2, Canada.

Received: 7 May 2015 Accepted: 30 September 2015

Published online: 12 October 2015

References

1. Henner P, Schiavon M, Morel JL, Lichfouse E. Polycyclic aromatic hydrocarbon (PAH) occurrence and remediation methods. *Analisis*. 1997;25(9–10):M56–9.
2. De Sousa C. Contaminated sites: the Canadian situation in an international context. *J Environ Manage*. 2001;62(2):131–54.
3. Hamlin EM. Turning brownfields into greenbacks. *J Am Plann Assoc*. 1999;65(2):236–7.
4. Panagos P, Van Liedekerke M, Yigini Y, Montanarella L. Contaminated sites in Europe: review of the current situation based on data collected through a European network. *J Environ Public Health*. 2013;2013:158764.
5. Gelfand I, Sahajpal R, Zhang XS, Izaurralde RC, Gross KL, Robertson GP. Sustainable bioenergy production from marginal lands in the US Midwest. *Nature*. 2013;493(7433):514–7.
6. Forster P, Ramaswamy V, Artaxo P, Berntsen T, Betts R, Fahey DW, et al. Changes in atmospheric constituents and in radiative forcing. In: Climate change 2007: the physical science basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. 2007. p. 212.
7. Graham-Rowe D. Beyond food versus fuel. *Nature*. 2011;474(7352):S6–8.
8. Fulford ID, Watson C. Phytoremediation of heavy metal-contaminated land by trees - a review. *Environ Int*. 2003;29(4):529–40.
9. Kuzovkina YA, Volk TA. The characterization of willow (*Salix* L.) varieties for use in ecological engineering applications: co-ordination of structure, function and autecology. *Ecol Eng*. 2009;35(8):1178–89.
10. Labrecque M, Teodorescu T, Daigle S. Effect of wastewater sludge on growth and heavy metal bioaccumulation of two *Salix* species. *Plant Soil*. 1995;171(2):303–16.
11. Stephenson AL, Dupree P, Scott SA, Dennis JS. The environmental and economic sustainability of potential bioethanol from willow in the UK. *Bioresour Technol*. 2010;101(24):9612–23.
12. Haughton AJ, Bond AJ, Lovett AA, Dockerty T, Sunnenberg G, Clark SJ, et al. A novel, integrated approach to assessing social, economic and environmental implications of changing rural land-use: a case study of perennial biomass crops. *J Appl Ecol*. 2009;46(2):315–22.
13. Karp A, Shield I. Bioenergy from plants and the sustainable yield challenge. *New Phytol*. 2008;179(1):15–32.
14. Mleczek M, Riessmann I, Rutkowski P, Kaczmarek Z, Golinski P. Accumulation of selected heavy metals by different genotypes of *Salix*. *Environ Exp Bot*. 2009;66(2):289–96.
15. Pitre FE, Teodorescu TI, Labrecque M. Brownfield phytoremediation of heavy Metals using *Brassica* and *Salix* supplemented with EDTA: results of the first growing season. *J Environ Sci Eng*. 2010;4:51–9.
16. Grenier V, Pitre FE, Guidi Nissim W, Labrecque M. Genotypic differences explain most of the response of willow cultivars to petroeleum contaminated soil. *Trees-Struct Funct*. 2015.
17. Quaggiotti S, Barcaccia G, Schiavon M, Nicole S, Galla G, Rossignolo V, et al. Phytoremediation of chromium using *Salix* species: cloning ESTs and candidate genes involved in the Cr response. *Gene*. 2007;402(1–2):68–80.
18. Turchi A, Tamantini I, Camussi AM, Racchi ML. Expression of a metallothionein A1 gene of *Pisum sativum* in white poplar enhances tolerance and accumulation of zinc and copper. *Plant Sci*. 2012;183:50–6.
19. Di Bacco D, Galla G, Bracci T, Andreucci A, Barcaccia G, Tognetti R, et al. Transcriptome analyses of *Populus x euramericana* clone I-214 leaves exposed to excess zinc. *Tree Physiol*. 2011;31(12):1293–308.
20. Konlechner C, Turktas M, Langer I, Vaculik M, Wenzel WW, Puschenreiter M, et al. Expression of zinc and cadmium responsive genes in leaves of willow (*Salix caprea* L.) genotypes with different accumulation characteristics. *Environ Pollut*. 2013;178:121–7.
21. Yergeau E, Sanschagrin S, Maynard C, St-Arnaud M, Greer CW. Microbial expression profiles in the rhizosphere of willows depend on soil contamination. *Isme J*. 2014;8(2):344–58.
22. Kang JW, Khan Z, Doty SL. Biodegradation of trichloroethylene by an endophyte of hybrid poplar. *Appl Environ Microbiol*. 2012;78(9):3504–7.
23. Doty SL. Enhancing phytoremediation through the use of transgenics and endophytes. *New Phytol*. 2008;179(2):318–33.
24. Doty SL, Oakley B, Xin G, Kang JW, Singleton G, Khan Z, et al. Diazotrophic endophytes of native black cottonwood and willow. *Symbiosis*. 2009;47(1):23–33.
25. Khan Z, Roman D, Kintz T, delas Alas M, Yap R, Doty S. Degradation, phytoprotection and phytoremediation of phenanthrene by endophyte *pseudomonas putida*, PD1. *Environ Sci Technol*. 2014;48(20):12221–8.
26. Bosch TCG, McFall-Ngai MJ. Metaorganisms as the new frontier. *Zoology*. 2011;114(4):185–90.
27. Bell TH, Joly S, Pitre FE, Yergeau E. Increasing phytoremediation efficiency and reliability using novel omics approaches. *Trends Biotechnol*. 2014;32(5):271–80.
28. Khan Z. Endophyte assisted phytoremediation of trichloroethylene (TCE) - an environmental contaminant. *Comp Biochem Phys A*. 2007;146(4):S273–4.
29. Bao H, Li EY, Mansfield SD, Cronk QCB, El-Kassaby YA, Douglas CJ. The developing xylem transcriptome and genome-wide analysis of alternative splicing in *Populus trichocarpa* (black cottonwood) populations. *BMC Genomics*. 2013;14.
30. Hanley SJ, Mallott MD, Karp A. Alignment of a *Salix* linkage map to the *Populus* genomic sequence reveals macrosynteny between willow and poplar genomes. *Tree Gen Genomes*. 2006;3(1):35–48.
31. Bolwell GP, Wojtaszek P. Mechanisms for the generation of reactive oxygen species in plant defence - a broad perspective. *Physiol Mol Plant P*. 1997;51(6):347–66.
32. Gill SS, Tuteja N. Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. *Plant Physiol Biochem*. 2010;48(12):909–30.
33. Tegelberg R, Julkunen-Tiitto R. Quantitative changes in secondary metabolites of dark-leaved willow (*Salix myrsinifolia*) exposed to enhanced ultraviolet-B radiation (vol 113, pg 541, 2001). *Physiol Plant*. 2002;115(1):174–4.
34. Xiao XW, Yang F, Zhang S, Korpelainen H, Li CY. Physiological and proteomic responses of two contrasting *Populus cathayana* populations to drought stress. *Physiol Plant*. 2009;136(2):150–68.
35. Davletova S, Rizhsky L, Liang HJ, Zhong SQ, Oliver DJ, Coutu J, et al. Cytosolic ascorbate peroxidase 1 is a central component of the reactive oxygen gene network of *Arabidopsis*. *Plant Cell*. 2005;17(1):268–81.
36. Pastorius GM, Foyer CH. Common components, networks, and pathways of cross-tolerance to stress. The central role of "redox" and abscisic acid-mediated controls. *Plant Physiol*. 2002;129(2):460–8.
37. Storozhenko S, De Pauw P, Van Montagu M, Inze D, Kushnir S. The heat-shock element is a functional component of the *Arabidopsis APX1* gene promoter. *Plant Physiol*. 1998;118(3):1005–14.
38. Park JM, Park CJ, Lee SB, Ham BK, Shin R, Paek KH. Overexpression of the tobacco *Tsi1* gene encoding an EREBP/AP2-Type transcription factor enhances resistance against pathogen attack and osmotic stress in tobacco. *Plant Cell*. 2001;13(5):1035–46.
39. May MJ, Vernoux T, Leaver C, Van Montagu M, Inze D. Glutathione homeostasis in plants: implications for environmental sensing and plant development. *J Exp Bot*. 1998;49(321):649–67.
40. Barchet GLH, Dauwe R, Guy RD, Schroeder WR, Soolanayakanahally RY, Campbell MM, et al. Investigating the drought-stress response of hybrid poplar genotypes by metabolite profiling. *Tree Physiol*. 2014;34(11):1203–19.
41. Albert M, Werner M, Proksch P, Fry SC, Kaldenhoff R. The cell wall-modifying xyloglucan endotransglycosylase/hydrolase LeXTH1 is expressed during the defence reaction of tomato against the plant parasite *Cuscuta reflexa*. *Plant Biol*. 2004;6(4):402–7.
42. Andersson-Gunneras S, Mellerowicz EJ, Love J, Segerman B, Ohmiya Y, Coutinho PM, et al. Biosynthesis of cellulose-enriched tension wood in *Populus tremula*: global analysis of transcripts and metabolites identifies biochemical and developmental regulators in secondary wall biosynthesis.(vol 45, pg 144, 2005). *Plant J*. 2006;46(2):349–9.
43. Lafarguette F, Leple JC, Dejardin A, Laurans F, Costa G, Lesage-Descaves MC, et al. Poplar genes encoding fasciclin-like arabinogalactan proteins are highly expressed in tension wood. *New Phytol*. 2004;164(1):107–21.
44. Ramos P, Valenzuela C, le Provost G, Plomion C, Gantz C, Moya-Leon MA, et al. ACC oxidase and ACC synthase expression profiles after Leaning of Young Radiata (*P. radiata* D. Don) and Maritime Pine (*P. pinaster* Ait.) seedlings. *J Plant Growth Regul*. 2012;31(3):382–91.
45. Andersson-Gunneras S, Hellgren JM, Bjorklund S, Regan S, Moritz T, Sundberg B. Asymmetric expression of a poplar ACC oxidase controls ethylene production during gravitational induction of tension wood. *Plant J*. 2003;34(3):339–49.

46. Janz D, Lautner S, Wildhagen H, Behnke K, Schnitzler JP, Rennenberg H, et al. Salt stress induces the formation of a novel type of 'pressure wood' in two *Populus* species. *New Phytol.* 2012;194(1):129–41.
47. Pitre FE, Lafarguette F, Boyle B, Pavé N, Caron S, Dallaire N, et al. High nitrogen fertilization and stem leaning have overlapping effects on wood formation in poplar but invoke largely distinct molecular pathways. *Tree Physiol.* 2010;30(10):1273–89.
48. Dixon RA, Paiva NL. Stress-induced phenylpropanoid metabolism. *Plant Cell.* 1995;7(7):1085–97.
49. Jeong MJ, Choi BS, Bae DW, Shin SC, Park SU, Lim HS, et al. Differential expression of kenaf phenylalanine ammonia-lyase (PAL) ortholog during developmental stages and in response to abiotic stresses. *Plant Omics.* 2012;5(4):392–9.
50. Arasimowicz-Jelonek M, Floryszak-Wieczorek J, Drzewiecka K, Chmielowska-Bak J, Abramowski D, Izbianska K. Aluminum induces cross-resistance of potato to Phytophthora infestans. *Planta.* 2014;239(3):679–94.
51. Paakkonen E, Seppanen S, Holopainen T, Kokko H, Karenlampi S, Karenlampi L, et al. Induction of genes for the stress proteins PR-10 and PAL in relation to growth, visible injuries and stomatal conductance in birch (*Betula pendula*) clones exposed to ozone and/or drought. *New Phytol.* 1998;138(2):295–305.
52. MacDonald MJ, D'Cunha GB. A modern view of phenylalanine ammonia lyase. *Biochem Cell Biol.* 2007;85(3):273–82.
53. Singh K, Kumar S, Rani A, Gulati A, Ahuja PS. Phenylalanine ammonia-lyase (PAL) and cinnamate 4-hydroxylase (C4H) and catechins (flavan-3-ols) accumulation in tea. *Funct Integr Genomics.* 2009;9(1):125–34.
54. Guo J, Wang MH. Characterization of the phenylalanine ammonia-lyase gene (SIPAL5) from tomato (*Solanum lycopersicum* L.). *Mol Biol Rep.* 2009;36(6):1579–85.
55. Navarre DA, Payyavula RS, Shakya R, Knowles NR, Pillai SS. Changes in potato phenylpropanoid metabolism during tuber development. *Plant Physiol Biochem.* 2013;65:89–101.
56. Tomás-Barberán FA, Loaiza-Velarde J, Bonfanti A, Saltveit ME. Early wound- and ethylene-induced changes in phenylpropanoid metabolism in harvested lettuce. *J Am Soc Hortic Sci.* 1997;122(3):399–404.
57. Subramanian R, Reinold S, Molitor EK, Douglas CJ. Structure, inheritance, and expression of hybrid poplar (*Populus trichocarpa* × *Populus deltoides*) phenylalanine ammonia-lyase genes. *Plant Physiol.* 1993;102(1):71–83.
58. Julkunen-Inttilä R. Phenolic Constituents of Salix - a chemotaxonomic survey of further Finnish species. *Phytochemistry.* 1989;28(8):2115–25.
59. Orians CM, Fritz RS. Secondary chemistry of hybrid and parental willows - phenolic glycosides and condensed tannins in *Salix Sericea*, *S-Eriophyllum*, and Their Hybrids. *J Chem Ecol.* 1995;21(9):1245–53.
60. Elmquist T, Cates RG, Harper JK, Gardjell H. Flowering in males and females of a Utah willow, *Salix-Rigida* and effects on growth, tannins, Phenolic Glycosides and sugars. *Oikos.* 1991;61(1):65–72.
61. Denno RF, Larson S, Olmstead KL. Role of enemy-free space and plant-quality in host-plant selection by willow beetles. *Ecology.* 1990;71(1):124–37.
62. Hallgren P, Ikonen A, Hjalten J, Roineinen H. Inheritance patterns of phenolics in F1, F2, and back-cross hybrids of willows: implications for herbivore responses to hybrid plants. *J Chem Ecol.* 2003;29(5):1143–58.
63. Van Den Boom CEM, Van Beek TA, Posthumus MA, De Groot A, Dicke M. Qualitative and quantitative variation among volatile profiles induced by *Tetranychus urticae* feeding on plants from various families. *J Chem Ecol.* 2004;30(1):69–89.
64. Castee CL, Hansen AK, Walling LL, Paine TD. Manipulation of plant defense responses by the tomato psyllid (*Bactericerca cockerelli*) and its associated endosymbiont *candidatus liberibacter psyllaurois*. *Plos One.* 2012;7(4):e35191.
65. Grbic M, Van Leeuwen T, Clark RM, Rombauts S, Rouze P, Grbic V, et al. The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature.* 2011;479(7374):487–92.
66. Shih CIT, Poe SL, Cromroy HL. Biology, life table, and intrinsic rate of increase of *Tetranychus-Urticae* Acarina-Tetranychidae. *Ann Entomol Soc Am.* 1976;69(2):362–4.
67. McHale L, Tan XP, Koehl P, Michelmore RW. Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* 2006;7(4):21.
68. Yang S, Zhang X, Yue JX, Tian D, Chen JQ. Recent duplications dominate NBS-encoding gene expansion in two woody species. *Mol Gen Genomics.* 2008;280(3):187–98.
69. Kohler A, Rinaldi C, Duplessis S, Baucher M, Geelen D, Duchaussoy F, et al. Genome-wide identification of NBS resistance genes in *Populus trichocarpa*. *Plant Mol Biol.* 2008;66(6):619–36.
70. Tsubota T, Shiotsuki T. Genomic analysis of carboxyl/cholinesterase genes in the silkworm *Bombyx mori*. *BMC Genomics.* 2010;11:377.
71. Dermauw W, Osborne EJ, Clark RM, Grbic M, Tirry L, Van Leeuwen T. A burst of ABC genes in the genome of the polyphagous spider mite *Tetranychus urticae*. *BMC Genomics.* 2013;14.
72. Michaud D, Cantin L, Raworth DA, Vrain TC. Assessing the stability of cystatin cysteine proteinase complexes using mildly-denaturing gelatin-polyacrylamide gel electrophoresis. *Electrophoresis.* 1996;17(1):74–9.
73. Santamaría ME, Hernandez-Crespo P, Ortego F, Grbic V, Grbic M, Diaz I, et al. Cysteine peptidases and their inhibitors in *Tetranychus urticae*: a comparative genomic approach. *BMC Genomics.* 2012;13.
74. Nisbet AJ, Billingsley PF. A comparative survey of the hydrolytic enzymes of ectoparasitic and free-living mites. *Int J Parasitol.* 2000;30(1):19–27.
75. Coley PD, Bryant JP, Chapin FS. Resource availability and plant antiherbivore defense. *Science.* 1985;230(4728):895–9.
76. Yamamura N, Tsuji N. Optimal strategy of plant antiherbivore defense - implications for apparency and resource-availability theories. *Ecol Res.* 1995;10(1):19–30.
77. Culliney TW, Pimentel D. Ecological effects of organic agricultural practices on insect populations. *Agric Ecosyst Environ.* 1986;15(4):253–66.
78. Wilson LT, Smilanick JM, Hoffmann MP, Flaherty DL, Ruiz SM. Leaf nitrogen and position in relation to population parameters of pacific spider mite, *Tetranychus pacificus* (Acar: Tetranychidae) on grapes. *Environ Entomol.* 1988;17(6):964–8.
79. Bell TH, Hassan SE, Lauron-Moreau A, Al-Otaibi F, Hijri M, Yergeau E, et al. Linkage between bacterial and fungal rhizosphere communities in hydrocarbon-contaminated soils is related to plant phylogeny. *Isme J.* 2014;8(2):331–43.
80. Campisano A, Ometto L, Compant S, Panche M, Antonielli L, Yousaf S, et al. Interkingdom transfer of the acne-causing agent, *Propionibacterium acnes*, from human to grapevine. *Mol Biol Evol.* 2014;31(5):1059–65.
81. Idris R, Trifonova R, Puschenerreiter M, Wenzel WW, Sessitsch A. Bacterial communities associated with flowering plants of the Ni hyperaccumulator *Thlaspi goesingense*. *Appl Environ Microbiol.* 2004;70(5):2667–77.
82. Ervin JS. Changes in hybrid poplar endophytic microbial diversity in response to trichloroethylene exposure. *Logan: Utah State University;* 2010.
83. Hogg JC, Lehane MJ. Microfloral diversity of cultured and wild strains of *Psoroptes ovis* infesting sheep. *Parasitology.* 2001;123:441–6.
84. Chung SH, Rosa C, Scully ED, Peiffer M, Tooker JF, Hoover K, et al. Herbivore exploits orally secreted bacteria to suppress plant defenses. *Proc Natl Acad Sci U S A.* 2013;110(39):15728–33.
85. Zhang JW, Zhang ER, Scott K, Burgess JG. Enhanced electricity production by use of reconstituted artificial consortia of estuarine bacteria grown as biofilms. *Environ Sci Technol.* 2012;46(5):2984–92.
86. Gupta K, Chatterjee C, Gupta B. Isolation and characterization of heavy metal tolerant Gram-positive bacteria with bioremedial properties from municipal waste rich soil of Kestopur canal (Kolkata), West Bengal, India. *Biologia.* 2012;67(5):827–36.
87. Ghanem I, Orfi M, Shamma M. Biodegradation of Chlorpyrifos by *Klebsiella* sp isolated from an activated sludge sample of waste water treatment plant in Damascus. *Folia Microbiol.* 2007;52(4):423–7.
88. Bertin PN, Heinrich-Salmeron A, Pelletier E, Gouhen-Chollet F, Arsene-Ploetz F, Gallien S, et al. Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics. *Isme J.* 2011;5(11):1735–47.
89. Sharma YK, Leon J, Raskin I, Davis KR. Ozone-induced responses in *Arabidopsis thaliana*: the role of salicylic acid in the accumulation of defense-related transcripts and induced resistance. *Proc Natl Acad Sci U S A.* 1996;93(10):5099–104.
90. Orsini F, Cascone P, De Pascale S, Barbieri G, Corrado G, Rao R, et al. Systemin-dependent salinity tolerance in tomato: evidence of specific convergence of abiotic and biotic stress responses. *Physiol Plant.* 2010;138(1):10–21.
91. Kao YY, Harding SA, Tsai CJ. Differential expression of two distinct phenylalanine ammonia-lyase genes in condensed tannin-accumulating and lignifying cells of quaking aspen. *Plant Physiol.* 2002;130(2):796–807.
92. Graham N, Gang F, Fowler G, Watts M. Characterisation and coagulation performance of a tannin-based cationic polymer: a preliminary assessment. *Colloid Surface A.* 2008;327(1–3):9–16.
93. Ko JH, Kim HT, Hwang I, Han KH. Tissue-type-specific transcriptome analysis identifies developing xylem-specific promoters in poplar. *Plant Biotechnol J.* 2012;10(5):587–96.

94. Plavcova L, Hacke UG, Almeida-Rodriguez AM, Li EY, Douglas CJ. Gene expression patterns underlying changes in xylem structure and function in response to increased nitrogen availability in hybrid poplar. *Plant Cell Environ.* 2013;36(1):186–99.
95. Chang S, Puryear J, Cairney J. A simple and efficient method for isolating RNA from pine trees. *Plant Mol Biol Rep.* 1993;11(2):113–6.
96. Gambino G, Perrone I, Gribaudo I. A rapid and effective method for RNA extraction from different tissues of grapevine and other woody plants. *Phytochem Anal.* 2008;19(6):520–5.
97. Lohse M, Bolger AM, Nagel A, Fennie AR, Lunn JE, Stitt M, et al. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res.* 2012;40(W1):W622–7.
98. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011;29(7):644–U130.
99. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9(4):357–U354.
100. Roberts A, Pachter L. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nat Methods.* 2013;10(1):71–U99.
101. Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BMG, et al. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics.* 2013;29(8):1035–43.
102. Soneson C, Delorenzi M. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics.* 2013;14:91.
103. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–10.
104. Altschul SF. Amino-acid substitution matrices from an information theoretic perspective. *J Mol Biol.* 1991;219(3):555–65.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

