

RESEARCH ARTICLE

Open Access

Transcriptome analysis of *Pinus monticola* primary needles by RNA-seq provides novel insight into host resistance to *Cronartium ribicola*

Jun-Jun Liu^{*}, Rona N Sturrock and Ross Benton

Abstract

Background: Five-needle pines are important forest species that have been devastated by white pine blister rust (WPBR, caused by *Cronartium ribicola*) across North America. Currently little transcriptomic and genomic data are available to understand molecular interactions in the WPBR pathosystem.

Results: We report here RNA-seq analysis results using Illumina deep sequencing of primary needles of western white pine (*Pinus monticola*) infected with WPBR. *De novo* gene assembly was used to generate the first *P. monticola* consensus transcriptome, which contained 39,439 unique transcripts with an average length of 1,303 bp and a total length of 51.4 Mb. About 23,000 *P. monticola* unigenes produced orthologous hits in the *Pinus* gene index (PGI) database (BLASTn with E values < e-100) and 6,300 genes were expressed actively (at RPKM ≥ 10) in the healthy tissues. Comparison of transcriptomes from WPBR-susceptible and -resistant genotypes revealed a total of 979 differentially expressed genes (DEGs) with a significant fold change > 1.5 during *P. monticola*-*C. ribicola* interactions. Three hundred and ten DEGs were regulated similarly in both susceptible and resistant seedlings and 275 DEGs showed regulatory differences between susceptible and resistant seedlings post infection by *C. ribicola*. The DEGs up-regulated in resistant seedlings included a set of putative signal receptor genes encoding disease resistance protein homologs, calcineurin B-like (CBL)-interacting protein kinases (CIPK), F-box family proteins (FBP), and abscisic acid (ABA) receptor; transcriptional factor (TF) genes of multiple families; genes homologous to apoptosis-inducing factor (AIF), flowering locus T-like protein (FT), and subtilisin-like protease. DEGs up-regulated in resistant seedlings also included a wide diversity of down-stream genes (encoding enzymes involved in different metabolic pathways, pathogenesis-related -PR proteins of multiple families, and anti-microbial proteins). A large proportion of the down-regulated DEGs were related to photosystems, the metabolic pathways of carbon fixation and flavonoid biosynthesis.

Conclusions: The novel *P. monticola* transcriptome data provide a basis for future studies of genetic resistance in a non-model, coniferous species. Our global gene expression profiling presents a comprehensive view of transcriptomic regulation in the WPBR pathosystem and yields novel insights on molecular and biochemical mechanisms of disease resistance in conifers.

Keywords: Defence response, Disease resistance, RNA-seq analysis, Transcriptome profiling, White pine-blister rust (WP-BR) interaction

* Correspondence: Jun-Jun.Liu@NRCan-RNCan.gc.ca
Pacific Forestry Centre, Canadian Forest Service, Natural Resources Canada,
506 West Burnside Road, Victoria, BC V8Z 1 M5, Canada

Background

Genetic resistance to the white pine blister rust (WPBR) fungus (*Cronartium ribicola*) in western white pine (WWP, *Pinus monticola*) and other five-needle pines is an important and highly desired trait. Introduced to North America in the early 1900s, *C. ribicola* has decimated native white pines and significantly altered both forest ecosystems and the ability to manage the species for profitable timber production. White pine breeding and subsequent use of resistant germplasm for forest restoration is a long-term process; since the 1940s, it has required the attention of a few generations of forest geneticists [1].

Several types of DNA markers such as amplified fragment-length polymorphism (AFLP) markers [2], single nucleotide polymorphism (SNP) markers [3,4] and microsatellite (SSR) markers [4] have been developed and applied to WWP research, and there is some molecular information available for molecular breeding of white pine resistance against *C. ribicola* [5]. For example, the plant disease resistance (R) family of NBS-LRR proteins and several families of pathogenesis-related (PR) proteins, including chitinases (PR3), thaumatin like proteins -TLPs (PR5), intracellular ribonuclease-like proteins (PR10), and antimicrobial peptides/proteins (AMPs), have been shown to contribute to host resistance in WP-BR interactions [5,6]. A recent proteomic profiling uncovered over one hundred *P. monticola* proteins modulated by *C. ribicola* inoculation, which included heat shock proteins (HSPs), reactive oxygen species (ROS) scavenging enzymes, and intermediate factors functioning in the signal transduction pathways triggered by well-known plant R genes, as well as other defence-related proteins [7]. Histochemical analysis revealed that the resistance response to systemic *C. ribicola* spread is localized internally in needle and stem tissues and that the build-up of physical barriers and deposition of cell wall-bound phenolic compounds play a crucial role in the defense reaction [8,9]. Despite these important results, there is still much to learn about the genetic basis of host resistance to *C. ribicola* in WWP and other five-needle pines (such as whitebark pine and limber pine).

Even though there have been significant improvements in genomic sequencing techniques over the past decade, the full genome of a conifer species is still unavailable. As a group, white pines have one of the largest plant genomes (27.36–37.68 pg/C) [10]; the genome size of *P. monticola* is estimated at 28.25 pg/C with a calculated length of about 2.7×10^4 Mb per 1C genome. Full genome sequencing of any single white pine species would thus be very expensive. RNA sequencing (RNA-seq) is a recently developed, high-throughput method for profiling transcriptomes. RNA-seq is cost-economic and time-saving, especially compared to traditional expressed sequence tag (EST) sequencing, and it can generate

transcriptome data for non-model species using incomplete genome information [11]. In addition to profiling gene expression, RNA-seq has shown powerful applications in areas, such as cataloguing of non-coding RNAs, investigation of the transcriptional structure of genes and splicing patterns, and the study of posttranscriptional modification and mutations [12]. RNA-seq has also provided information on complex regulation networks for gene expression patterns and on gene variations (such as SNPs and SSRs) in an increasing number of non-model plants [13], but, to date has not been used in study of the WPBR pathosystem.

In this study, we used RNA-seq analysis to profile the transcriptome of *P. monticola* primary needles during early stages of infection by *C. ribicola*; seedlings with major gene resistant (*Cr2/-*) and susceptible (*cr2/cr2*) genotypes were used. With *de novo* assembly followed by gene annotation and functional classification, our RNA-seq analysis generated the first *P. monticola* consensus transcriptome. Comparison of RNA-seq data sets from resistant (*Cr2/-*) and susceptible (*cr2/cr2*) genotypes revealed significant expressional differences among genes involved in defense signalling pathways and metabolic pathways. The first-ever set of transcriptome and global gene expression data reported here on *C. ribicola*-infected white pine needles significantly expands our knowledge of the molecular framework of the WPBR pathosystem.

Results

Transcriptome sequencing by RNA-seq and *de novo* assembly

Three RNA-seq 76-bp paired end read libraries were prepared from total RNA extracted from primary needles of uninfected seedlings (control at time 0-day post *C. ribicola*-infection, dpi) and infected seedlings with resistant (*Cr2/-*) and susceptible (*cr2/cr2*) genotypes (at 4-dpi). For each of three cDNA libraries, 116.4, 123.5 and 141.9 million 76-bp paired end reads were collected respectively. An overview of the raw reads data is given in Additional file 1: Table S1. Only 0.2% of raw reads were discarded due to low quality bases and reads (Additional file 1: Table S1). A separate assembly was performed for each of the three libraries and produced 39,135 to 45,236 contigs with an average N50 of 1,488-bp and an average length of 875-bp (Table 1). We also performed a single assembly of the total reads from all three cDNA libraries, which generated a common transcriptome data set of 36,923 contigs having a minimum length of 300-bp and a minimum mapped read of 50.

Since the separate assembly approach aligned more raw reads (75.5% of total) into contigs and mapped them back to the assembled transcripts than did the single assembly approach (Additional file 1: Table S1), a

Table 1 Measurement of contigs from *de novo* assembly

Sample ID	N75	N50	N25	Mini	Maxi	Ave	Count	Total (bp)
Assembly of three libraries separately								
Un-infected	707	1,580	2,466	160	13,127	899	42,499	38,211,860
Res-Cr2/_ (4-dpi)	661	1,495	2,408	116	20,237	872	45,236	39,447,742
Sus-cr2/cr2 (4-dpi)	663	1,390	2,188	131	13,104	855	39,135	33,478,214
Average	677	1,488	2,354	136	15,489	875	42,290	37,045,939
Consensus transcriptome	1,094	1,814	2,696	197	20,237	1,303	39,439	51,405,388
Assembly of three libraries together								
	425	1113	2121	165	16,402	670	72,095	48,292,721
Common transcriptome^(a)	1712	2409	3376	300	16,402	1041	36,923	38,447,991

^acontigs were selected with minimum length of 300-bp and minimum mapped reads of 50.

reciprocal BLASTn approach was used for search of one cDNA library assembly with another. This process generated a consensus transcriptome of 39,439 unique genes with each contig present in at least two cDNA libraries. This assembly of 39,439 contigs had an average length of 1,303-bp and a total length of ~51-Mb. The remaining contigs specific for each cDNA library were searched again by BLASTn against the PGI database, and those sample-specific contigs with strong Blastn hits (E-value < e-20) were added to the consensus transcriptome, producing a transcriptome of 43,890 contigs as reference for further analysis.

Using BLAST programs to assess the *de novo* assembly quality, the consensus transcriptome of 39,439 unique transcripts was compared with the PGI and Spruce Gene Index (SGI) databases, the protein database of the poplar leaf rust fungus *Melampsora laricis-populina*, as well as a set of *P. monticola* EST data (Additional file 1: Table S2). BLAST analysis revealed that 85% (~33,000 contigs) of infected *P. monticola* consensus assembly showed significant homology to the PGI and SGI databases (tBLASTx with E value < e-10), about 23,000 contigs (59% of the total) having orthologous hits in the PGI database alone. Only 2.8% of the infected WWP consensus transcriptome had orthologous hits in the poplar leaf rust fungus genome (BLASTx with E value < e-100), suggesting that ~5% of the assembled transcripts may come from *C. ribicola*. BLASTn analysis also revealed that ~97% of the *P. monticola* ESTs had orthologous hit in the *P. monticola* consensus assembly (E value < e-100) (Additional file 1: Table S2). These results indicate that we obtained a high percentage of *P. monticola* expressed genes and thus further support application of the *de novo* assembled transcriptome for global gene expression profiling.

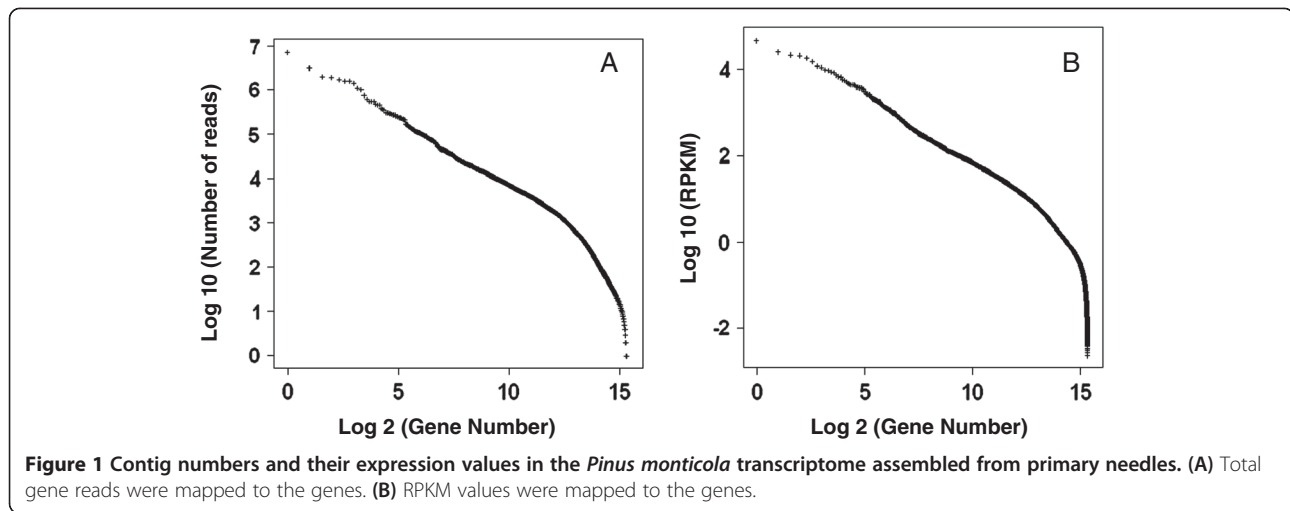
Transcriptional profile of western white pine primary needles

The top ten contigs with the highest expression values (Additional file 1: Table S3) of total gene reads accounted

for 30% of total mapped reads and they were followed by about 4,600 contigs that were mapped by 90% of total mapped reads (Figure 1A). When gene expression levels were normalised by the reads per kilobase of transcript per million mapped reads (RPKM) values to get a more reasonable index for relative levels of transcript expression, the top 10 contigs with the highest RPKM values (Additional file 1: Table S3) were still mapped by 26% of total mapped reads. There were ~1,500 contigs with highly abundant expression levels at RPKM ≥ 50 and ~6,300 contigs with transcript levels at RPKM ≥ 10 in the control primary needles (un-infected at 0-dpi) (Figure 1B). The transcriptome set with 43,890 contigs appeared to be saturated as estimated by either total gene reads or RPKM. The top 10 unigenes with most abundant transcript expression were *P. monticola* homologs encoding for ribulose bisphosphate carboxylase/oxygenase (RuBisCO), rRNA intron encoded homing endonuclease, a chloroplastic oxygen-evolving enhancer protein, a chlorophyll a/b binding protein, S-adenosyl-methionine-dependent methyltransferase, RuBisCO activase, fructose-bisphosphate aldolase, cell wall-associated hydrolase, as well as conserved hypothetical proteins with unknown function (Additional file 1: Table S3).

Annotation of the primary needle transcriptome

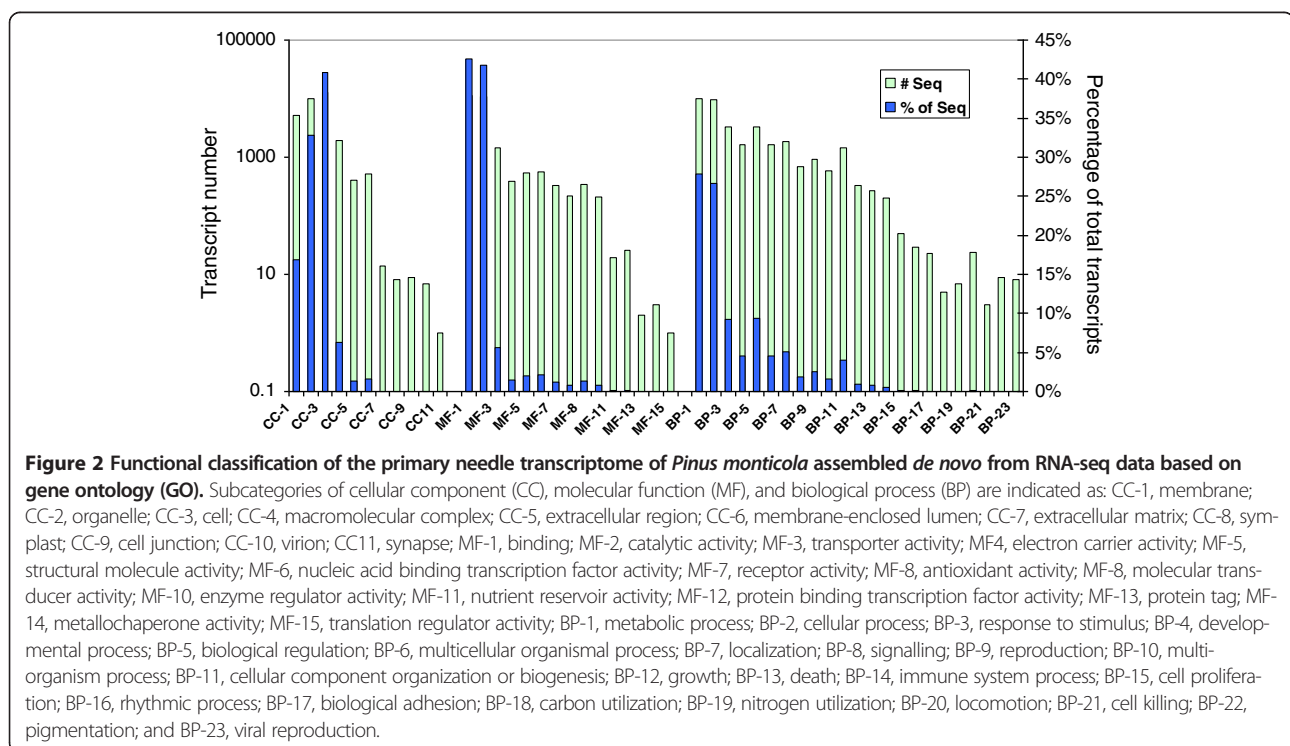
Gene annotation was performed for the *P. monticola* primary needle transcriptome assembly based on sequence homologies to the databases of the National Center for Biotechnology Information (NCBI) non-redundant sequences (nr), the Protein Information Resource (PIR), the Universal Protein Resource (UniProts), the Gene Ontology (GO), and the Kyoto Encyclopedia of Genes and Genomes (KEGG) using BLAST2GO. Gene names and GO terms were assigned to the contigs based on their homologies to these available databases. About 72% of total contigs (31,577 out of 43,890) were assigned gene names and 50% of them (21,577) were assigned at least one GO term. GO annotation assignments classified 21,577 unique contigs into 23 subcategories of the biological



process category, 15 subcategories of the molecular function category, and 11 subcategories of the cellular component category at level 2 (Figure 2). The four subcategories with the most highly abundant transcripts under the biological process category were metabolic process (27.8%), cellular process (26.6%), biological regulation (9.34%), and response to stimulus (9.2%). The subcategories with the most highly abundant transcripts in the molecular function category included binding (42.6%), catalytic activity (41.7%), and transporter activity (5.5%). Within the cellular components category, the four most common groups of proteins were assigned to

the subcategories of cell (40.8%), organelle (32.8%), membrane (16.9%), and macromolecular complex (6.2%) (Figure 2).

To further estimate the completeness of the transcriptome data and the effectiveness of annotations, we searched annotated contig sequences using GOslim_plant. More than one quarter (26.9%) of the sequences were localized to the plastid, 17.1% to the mitochondrion, 15.9% to the nucleus, and 13.8% to the plasma membrane. The extracellular space and cell wall were localized by about 4% of total sequences, contributing to the first layer of plant defence to outside stimuli (Figure 3).



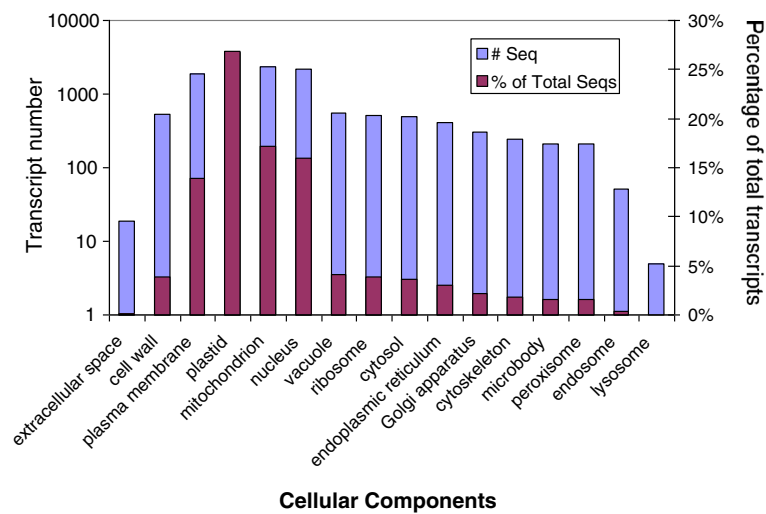


Figure 3 Distribution of unique transcripts in the primary needle transcriptome of *Pinus monticola* assigned to GO cellular component category using “goslim_plant” analysis in the BLAST2GO program. Sequences were assigned a putative function based on sequence identity to annotated proteins from other species.

Gene annotation conducted using enzyme code and KEGG databases revealed activities of many biological pathways in *P. monticola* primary needles. A total of 1,315 enzymes encoded by 7,561 transcripts were mapped to 136 metabolic pathways (Additional file 1: Table S4). Six pathways with the most abundant unique sequences included starch and sucrose metabolism (5.9%), purine metabolism (5.8%), phenylalanine metabolism (3.4%), methane metabolism (3.4%), phenylpropanoid biosynthesis (3.2%), and amino sugar and nucleotide sugar metabolism (2.6%). Each of these metabolic pathways was mapped with at least 200 unique transcripts.

Detection of differentially expressed genes (DEGs) in response to rust infection

A quality control test on the data assembled from each cDNA library confirmed that they were suitable for statistical analysis for DEG identification (Additional file 2: Figure S1). We compared three WWP primary needle transcriptome profiles (*Cr2*-resistant seedlings at 4-dpi vs. control at 0-dpi, *cr2*-susceptible seedlings at 4-dpi vs. control at 0-dpi, and *Cr2*-resistant seedlings at 4-dpi vs. *cr2*-susceptible seedlings at 4-dpi) to better understand the WPBR pathosystem at the transcriptome level. The reference transcriptome with 43,890 contigs was used to map raw reads for DEG detection between any two treatments. A total of 979 DEGs were revealed with a RPKM fold change > 1.5 and a cut-off of $p < 0.05$ with Z-test by Bonferroni-correction (Figure 4). We detected 562 DEGs in compatible WP-BR interaction (*cr2* vs. *avcr2*) and 789 DEGs in incompatible WP-BR interaction (*Cr2* vs. *avcr2*) (Figure 5). There were 310

DEGs regulated similarly after *C. ribicola* infection in both susceptible and resistant seedlings while there were 275 DEGs regulated differently (Figures 4 and 5).

The expression patterns were clustered into eight different types based on the K-means method: five types for up-regulation patterns (Figure 6A-E) and three types for down-regulation patterns (Figure 6F-H). The type I cluster included DEGs regulated positively in both resistant and susceptible seedlings at similar magnitudes (Figure 6A). While types II and III DEGs also showed up-regulation in both resistant and susceptible seedlings, they differed in degree of up-regulation (Figure 6B and C). The type IV cluster included DEGs with rust-enhanced transcript levels only in susceptible seedlings (Figure 6D) and type V included DEGs enhanced only in resistant seedlings (Figure 6E). Down-regulated patterns after *C. ribicola* infection are represented by types VI - VIII. DEGs down-regulated at similar levels in both resistant and susceptible seedlings were grouped into the type VI (Figure 6F). The type VII included DEGs with greater down-regulation levels in susceptible than in resistant seedlings (Figure 6G). The type VIII included DEGs regulated negatively by rust infection only in resistant seedlings (Figure 6H).

To confirm gene expression level measured by RPKM fold change, a subset of 26 contigs were subjected to analysis of quantitative reverse-transcriptase polymerase chain reaction (qRT-PCR). As shown in Figure 7, the relative transcript levels measured by qRT-PCR and RNA-seq were highly correlated ($R = 0.8$) with statistical significance ($p = 0$).

To explore potential functions of DEGs in response to *C. ribicola* invasion, GO-based classification was conducted.

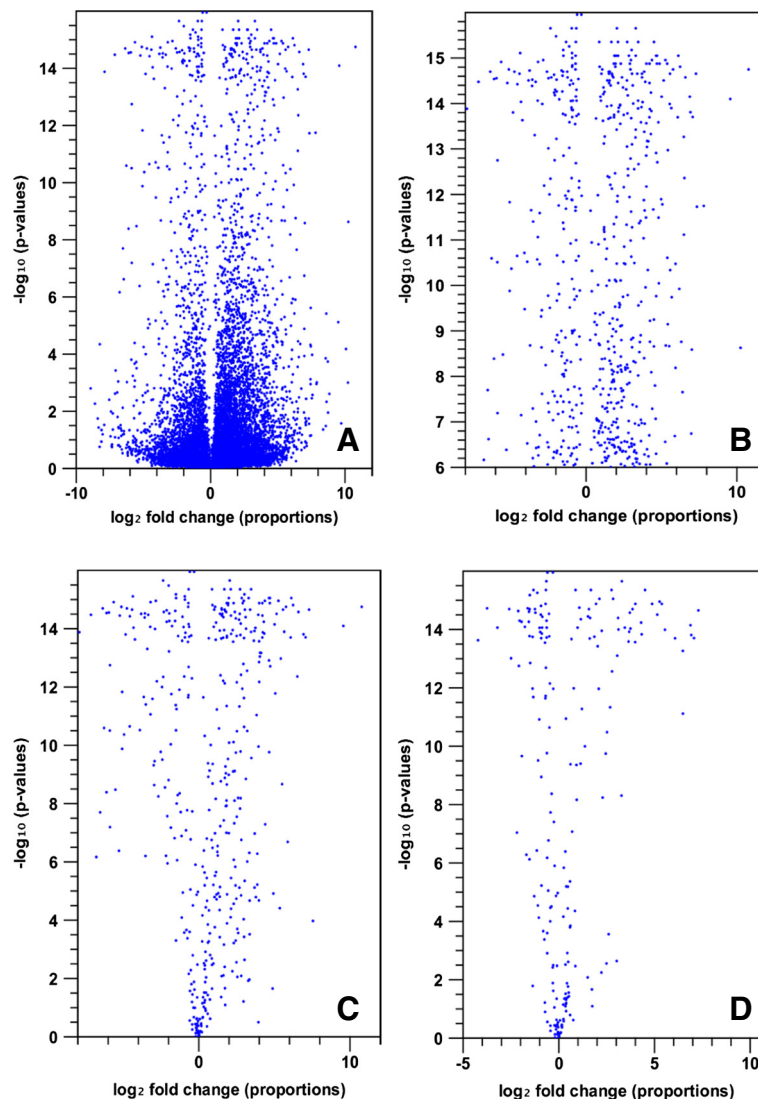
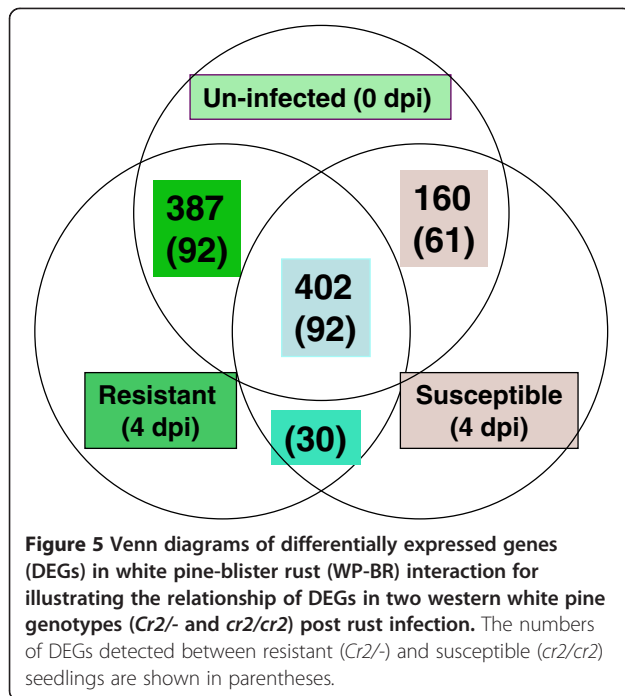


Figure 4 Global view of transcriptional changes by volcano plot using Kal's statistical test (Z-test). Log₂ fold change in RPKM expression values (x-axis) versus $-\log_{10}$ (Bonferroni corrected P-values) (y-axis) were computed in scatter plots: **(A)** All 43,980 contigs in comparisons of three cDNA libraries; **(B)** Differentially expressed genes (DEGs) detected between un-infected (0-dpi) and resistant seedlings post rust infection (4-dpi); **(C)** DEGs detected between un-infected (0-dpi) and susceptible seedlings post rust infection (4-dpi); and **(D)** DEGs detected between resistant and susceptible seedlings post rust infection.

Eighty-five percent (830/979) of all DEGs showed significant homologies with expressed genes from other organisms. GO analysis classified 508 of the annotated DEGs into 17, 13, and 8 subgroups (at level-2) of the biological process, molecular function, and cellular component categories, respectively (Additional file 3: Figure S2). The three biological processes most affected by *C. ribicola* infection were in the subcategories of metabolic process (345 DEGs), cellular process (303 DEGs) and response to stimulus (174 DEGs). Among the 136 metabolic pathways identified as active in infected white pine needles, 90

of them were affected significantly by rust infection. Only ~5.4% of total annotated contigs (410/7,561) were regulated by rust infection, but they encoded ~20.5% of total annotated enzymes (270/1,315) in the primary needle transcriptome. Based on their high sequence and enzyme numbers, the top 10 rust-modulated pathways were flavonoid biosynthesis, methane metabolism, carbon fixation in photosynthesis, starch and sucrose metabolism, phenylalanine metabolism, phenylpropanoid biosynthesis, amino sugar and nucleotide sugar metabolism, glycolysis/



gluconeogenesis, cysteine and methionine metabolism, and pyruvate metabolism (Additional file 1: Table S4).

DEGs specifically regulated in resistant seedlings

A total of 789 DEGs were detected in the comparison of the resistant genotype (*Cr2*⁻) at 4-dpi with control at 0-dpi. Of these 789 DEGs, 387 were regulated only in resistant seedlings while another 402 genes were also regulated by rust infection in susceptible (*cr2/cr2*) seedlings (Figure 5). Of the 387 resistance-specific DEGs, 245 were up-regulated with fold changes up to 1,217 and one transcript by *de novo* synthesis (Additional file 4: Table S5) and 142 were down-regulated with fold changes down to -88 by rust infection (Additional file 4: Table S6). The up-regulated DEGs included a set of putative signal receptor genes encoding putative R proteins with domains of nucleotide-binding site (NBS) and leucine-rich repeat (LRR), receptor-like protein kinases (RLK), calcineurin B-like (CBL)-interacting protein kinases (CIPK), F-box family proteins (FBP), and abscisic acid (ABA) receptor; transcriptional factor (TF) genes of multiple families (such as NAC, DOF, PLATZ, ARF2, zinc-finger HD, nuclear transcription factor Y, and other DNA-binding proteins); genes homologous to apoptosis-inducing factor (AIF), flowering locus T-like protein (FT), and subtilisin-like protease; as well as a wide diversity of down-stream genes (encoding enzymes involving different metabolic pathways, PR proteins of multiple families, and AMPs) (Additional file 4: Table S5). A large proportion of down-regulated DEGs were related to photosystem I and II, the metabolic pathway of carbon

fixation, and flavonoid biosynthesis (Additional file 1: Table S4 and Additional file 4: Table S6). Expansins are an example of cell-wall proteins down-regulated by rust infection in resistant seedlings.

DEGs specifically regulated in susceptible seedlings

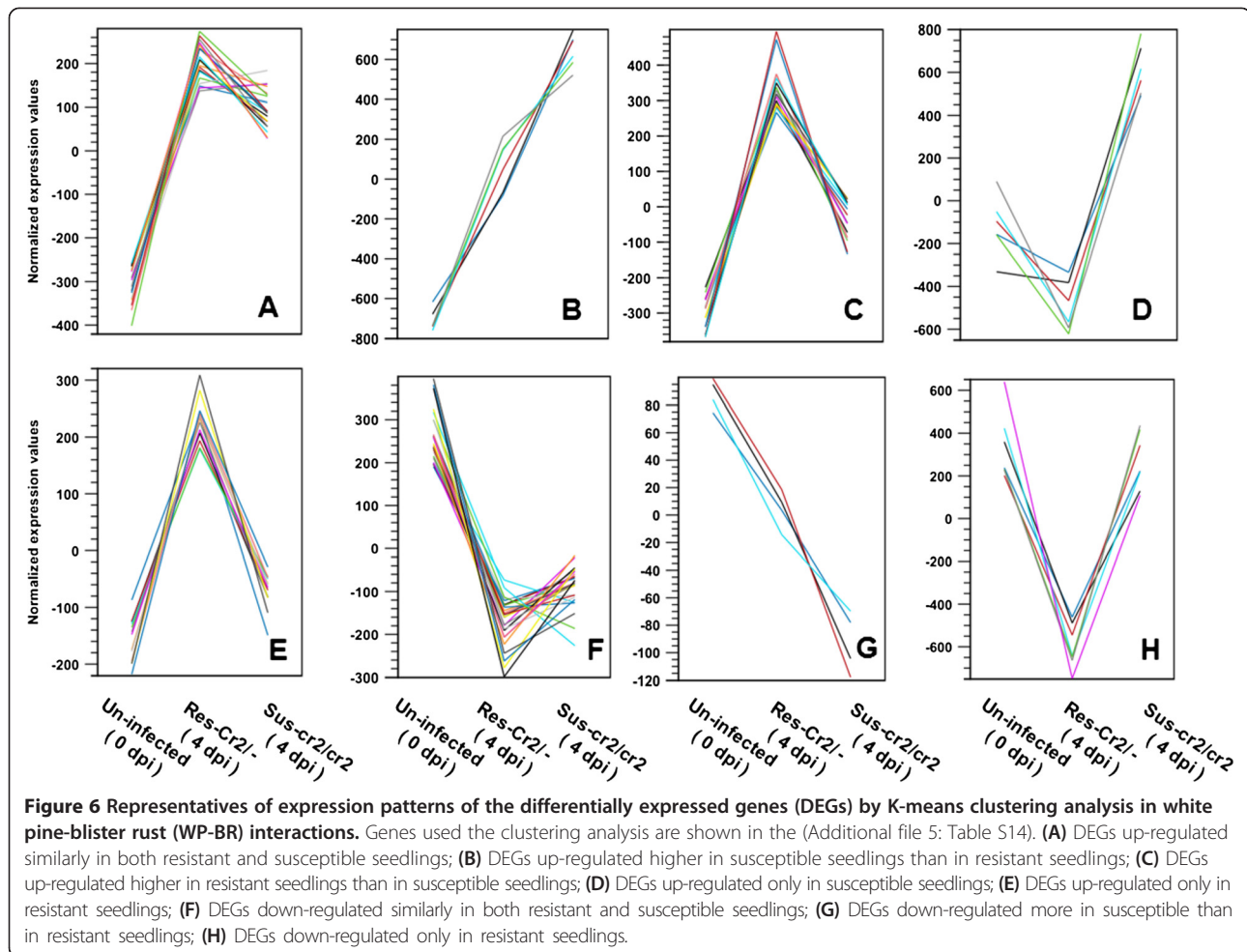
A total of 562 DEGs modulated by *C. ribicola* infection (at 4-dpi) were detected in the comparison of the susceptible (*cr2/cr2*) genotype with un-infected control seedlings. One hundred sixty of these DEGs were specifically regulated in susceptible seedlings, 114 genes of them up-regulated with fold changes up to 336 (Additional file 4: Table S7) and 46 genes down-regulated with fold changes down to -15 (Additional file 4: Table S8). In addition to a TF gene with a WRKY domain, we detected up-regulated genes for subsets of calcium-binding proteins, chitinases (PR3), TLPs (PR5), other PR proteins (PR1, PR4, and PR6), and myrosinase-binding proteins (MyroBP) specifically in susceptible seedlings (Additional file 4: Table S7). Apart from genes involved in photosystems, transcripts from multiple types of retrotransposons were noted among the down-regulated genes in the susceptible seedlings (Additional file 4: Table S8).

DEGs regulated similarly in both resistant and susceptible seedlings

Of the 402 DEGs co-modulated in both resistant and susceptible genotypes, 310 of them showed no significant difference between both genotypes; 204 of these DEGs were up-regulated and 106 DEGs down-regulated following rust infection (Additional file 4: Table S9 and S10). A number of zinc-finger, RING-finger, FBP genes, and TFs (C3HL, AP2-REF-B3, R2R3-MYB, CCCH type, and HD-leucine zipper) were up-regulated in both genotypes, as well as transcripts encoding for subsets of proteins regulated in auxin-mediated signalling, dnaj chaperone family protein, glutathione S-transferases (GST), peroxidases, thioredoxins, chitinases, ubiquitin-conjugating enzyme e2, β -glucanases, glucanase-inhibitors, and enzymes related to biosyntheses of ethylene, phenylpropanoid, and stilbenoid (Additional file 4: Table S9), suggesting their involvement in basal host response to *C. ribicola* invasion. In addition to transcripts for a group of small heat-shock proteins (HSP), genes for photoassimilate-responsive proteins and enzymes related to carbon fixation in photosynthesis and flavonoid biosynthesis were among the 106 DEGs negatively regulated in both resistant and susceptible seedlings (Additional file 1: Table S4 and Additional file 4: Table S10).

DEGs regulated differentially in both resistant and susceptible seedlings

A total of 275 DEGs were detected in the comparison between resistant and susceptible genotypes; 141 DEGs

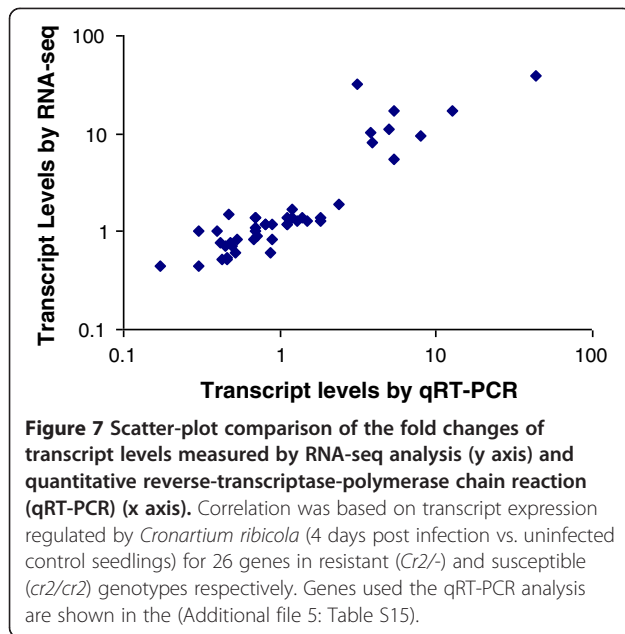


showed relatively higher transcript levels in resistant seedlings than in susceptible seedlings (Additional file 4: Table S11) while 134 DEGs showed the opposite expression pattern (Additional file 4: Table S12). Although carbon fixation in photosynthesis is down-regulated in both resistant and susceptible seedlings, the relative expression levels of related transcripts were higher in resistant seedlings, suggesting that rust infection caused less damage to photosynthesis in resistant seedlings than in susceptible seedlings. Relatively higher levels of transcripts related to biosynthesis of cellulose, flavonoid, flavone and flavonol in resistant seedlings suggested that more active production of these compounds may be beneficial to trees infected by WPBR. In contrast, relatively lower levels of transcripts in resistant than in susceptible seedlings indicated that compatible WP-BR interactions may lead to more active metabolisms for phenylalanine, tyrosine, starch and sucrose (Additional file 1: Table S4).

Discussion

We used RNA-seq technology to generate transcriptome data and examined global gene expression profiles to

identify defense-responsive genes in WP-BR interactions. This work demonstrates that RNA-seq is a useful and effective tool for *de novo* transcriptome assembly and discovery of candidate genes underlying host genetic resistance to pathogens, even in a non-model species without genome and complete EST databases. Enzyme annotation and pathway assignment of the *P. monticola* transcriptome provides a genomics resource for further investigating candidate genes involved in various metabolic pathways in a conifer species, such as those involved in physiological responses to environmental stresses. Comparison of sequence data from infected vs. non-infected and resistant vs. susceptible transcriptomes revealed almost one thousand DEGs using Z-test with a Bonferroni correction (2.5% of the total transcriptome assembly), and 85% of them were functionally annotated. It is noteworthy that a number of the DEGs revealed by RNA-seq include *P. monticola* defence-related genes/proteins investigated in previous studies, such as NBS-LRR, PR3, TLP (PR5), PR10, AMP, HSP, and other defence-related genes [5,7,14]. qRT-PCR analysis of a subset of DEGs further verified transcript expression



levels as revealed by transcriptome comparison from raw RNA-seq data. These consistent results from various research strategies indicate that RNA-seq provides a powerful tool for comprehensive transcriptome profiling that reveals important molecular interactions in the WPBR pathosystem.

Cr2 candidates for incompatible WP-BR interaction

Plants have evolved and adapted various defense mechanisms to protect themselves from invasions by microbial pathogens. Plant immunity is controlled by two layers of inducible responses: basal response triggered by conserved microbial features, and specific response triggered by gene-for-gene recognition. Specific responses to pathogen attack activate effector-triggered immunity (ETI) through direct or indirect interaction of host R proteins with pathogen effector/ avirulence proteins, which leads to the hypersensitive response (HR), a defense mechanism featured by programmed cell death (PCD) around the local pathogen infection sites [15]. White pine *Cr2*-seedlings display a typical HR in response to infection by *C. ribicola avcr2* strain, presumably by recognizing the *avcr2* product using an uncharacterized receptor encoded by the *P. monticola* R gene *Cr2* [16]. In other well-characterized pathosystems, the incompatible interaction usually leads to a series of dramatic molecular changes such as ion flux, a ROS burst, callose deposition, and eventually HR-like cell death, resulting in complete resistance to avirulent pathogenic strains [17]. In contrast, susceptible plants show only basal responses through interaction of host pattern recognition receptors (PRRs) with pathogen- or microbe-associated molecular patterns (PAMPs or

MAMPs). In PAMP- or MAMP-triggered immunity (PTI), pathogen avirulence factors (Avr) promote virulence by targeting other host proteins (in the absence of R proteins) to manipulate plant physiological processes more suitable to pathogen growth and multiplication in susceptible plants [15].

Deciphering the transcript fingerprint of *Cr2* is one of the key questions in studying the molecular WP-BR interactions. In this study, we identified two NBS-LRR genes (with the highest identities to *Picea sitchensis* proteins ABR16233 and ABR16103, respectively) and two RLK genes (one RLK with extracellular LRR domain and another one with cysteine-rich domain) with significant up-regulation only in resistant seedlings post *C. ribicola* infection (Additional file 4: Table S5), which provides novel insight into the expression profiles of these two plant superfamilies in the WPBR pathosystem [6,7,18]. Most characterized plant R genes belong to the NBS-LRR or RLK superfamily for plant R-Avr interactions [19,20]. Up-regulation of RLK and NBS-LRR homologous transcripts in resistant seedlings by infection of the *avcr2* rust strain suggests their involvement in white pine major gene (*Cr2*) resistance. SNP genotyping of both *P. monticola* NBS-LRR and RLK superfamilies for positional R candidates would help resolve the molecular characterization of *Cr2*.

Downstream signalling networks triggered by incompatible R-Avr interactions are mediated by calcium-activated protein kinases (such as CIPKs) and mitogen-activated PK (MAPK) cascades [21]. These signalling networks then control the activities and synthesis of a series of TFs, enzymes, phytohormones (such as jasmonic acid -JA and salicylic acid -SA), PR proteins, AMPs, phytoalexins and other secondary metabolites, which coordinately contribute to host resistance against pathogens [22]. Consistent with this widely accepted R-Avr interaction model, we detected significant up-regulation of transcripts for MAPK, CIPK, SNF1-related PK regulatory subunit, and other PKs in *P. monticola* resistant seedlings following rust infection. CBLs act as Ca²⁺ sensors to activate specific PKs. Rice CIPKs participate in various layers of MAMP-induced defense responses, leading to final PCD in rice [23]. Arabidopsis MAPKs (MPK3 and MPK6) are positive mediators of defense responses induced by MAMP and pathogen, controlling both basal resistance and elicitor-induced resistance to fungal pathogen *Botrytis cinerea* [24]. As a regulatory subunit of the SNF1-related protein kinase (SnRK) complex, SNF1-related protein kinase regulatory subunit beta-2 likely plays a role in a signal transduction cascade regulating gene expression and carbohydrate metabolism in plant response to environmental stresses. SnRK2s phosphorylate Ser/Thr residues in the R-X-X-S/T motif of the ABF type TFs for activation of a large number of

ABA/stress-responsive genes [25]. Rust-upregulated transcript expression suggests that several *P. monticola* PK families, including MAPKs, CIPKs and SnRK2s (Additional file 5: Table S13), may be intermediate factors involved in *Cr2*-mediated resistance to *C. ribicola*.

Novel *P. monticola* genes involved in the *Cr2*-mediated resistance

Among DEGs positively regulated in *Cr2*-mediated resistance, *P. monticola* genes homologous to genes encoding AIF, FT-like protein, subtilisin-like protease, RING/FYVE/PHD zinc finger-containing proteins, and membrane attack complex component/perforin (MACPF) domain proteins, attracted our attention (Additional file 5: Table S13). The MACPF domain proteins are well-known for their critical functions in innate and adaptive immunity, and they are capable of forming transmembrane lytic pores known as the membrane-attack complex (MAC) by interacting with other cell components for pathogen invasion or host protection. Arabidopsis MACPF proteins probably act as mediators that recognize plant signals for negative regulation of cell death programs and defense responses [26,27]. Both SA and a chitin elicitor promote expression of an Arabidopsis MACPF gene (*CAD1*) and the *cad1* mutant shows spontaneously activated expression of PR genes and greatly increased SA levels [26]. RING finger proteins constitute a large family and play key roles in regulating plant growth/developmental processes, hormone signalling, and responses to (a)biotic stresses. A rice RING-H2 finger gene (*OsBIRF1*) was induced differentially in an incompatible interaction with *Magnaporthe grisea*, and its constitutive expression led to enhanced disease resistance and elevated expression levels of defense-related genes encoding PR1, PR2, PR3 and PR5 proteins [28]. Up-regulation of a *P. monticola* MACPF homolog and differential expression of the *P. monticola* RING gene family in the incompatible WP-BR interaction suggest their potential roles in conifer defense against *C. ribicola*.

As a positive regulator of apoptosis in mammals, AIFs function in apoptotic and redox signalling: which enhances mitochondrial bioenergetics and complex I activity/assembly to help maintain proper cellular redox homeostasis in mitochondria and forms a chromatin degrading complex with other proteins in the nucleus [29]. In addition to up-regulation of an AIF homolog in the incompatible WP-BR interaction, we also observed differential expression of families of thioredoxins, GSTs and peroxidases (Additional file 5: Table S13), suggesting that redox signalling by oxidative burst is part of the defence mechanism of *P. monticola* against *C. ribicola* [8]. Recently a unique peptide signal (GmSubPep), embedded in a soybean subtilisin-like protein, was discovered to activate the transcription of defense genes against pathogens [30]. Although transcript levels of two subtilisin-like protein homologs were significantly up-

regulated in *Cr2*-mediated resistance (14- and 80-fold), whether these proteins are processed to generate signal peptides involved in host defence is still an interesting question to address in future studies. Plant FT RNA is able to travel by its *cis*-element from leaf to shoot apical meristem [31], suggesting its role in systemic signalling by intercellular RNA trafficking through phloem transport. RNA trafficking contributes to local and long-distance coordination of plant development and response to the environment [32]. A further functional analysis of these *P. monticola* novel defence-related genes would provide novel insight into resistance mechanisms of this conifer.

ABA signalling involved in the *Cr2*-mediated resistance

Among other intermediate factors potentially functioning in *Cr2*-triggered signalling, we detected significant up-regulation of transcripts for ABA receptor, ABA 8-hydroxylase, GRAM-containing ABA-responsive protein, and annexin proteins in resistant seedlings post *C. ribicola* infection (Additional file 5: Table S13). Apart from its central role in plant development, ABA plays a modulating role in diverse plant-pathogen interactions mediated at least in part by crosstalk with JA and SA [33]. ABA receptor proteins bind and respond to the hormone by activating the transcription of ABA-responsive genes involved in plant stress responses [34], including PR10 proteins [35].

The GRAM (for glucosyltransferases, Rab-like GTPase activators, and myotubularins) domain is ubiquitous in glucosyltransferases, myotubularins, and other membrane-associated proteins in eukaryotes [36]. The pepper GRAM domain-containing ABA-responsive protein, ABR1, negatively regulates ABA signaling by suppressing ABA biosynthesis, but promotes SA and ROS production, ultimately leading to cell death and disease resistance [37]. The mutant plants of an ABA 8'-hydroxylase gene (*cyp707a3-1*) accumulated a higher level of stress-induced ABA with exaggerated ABA-inducible gene expression. ABA treatment suppresses induction of systemic acquired resistance (SAR) by inhibiting the SA pathway [38]. As signaling components with phospholipid binding ability, some annexin proteins are implicated in responses to ABA, oxidative, saline, cold, and pathogenic stress [39,40]. Consistent with these studies, we observed a dramatic up-regulation of *P. monticola* transcripts for five proteins in resistant seedlings: ABA receptor *pyl8*, GRAM-containing ABA-responsive protein, ABA 8-hydroxylase, ABA-responsive protein, and annexin homolog (Additional file 5: Table S13), suggesting that ABA-signalling may play a part in the *Cr2*-mediated resistance.

Auxin signalling involved in the *Cr2*-mediated resistance

It is notable that 25 auxin-related transcripts were positively regulated by *C. ribicola* infection (Additional file 5: Table S13), and 15 of them were up-regulated

specifically in resistant seedlings or with significantly higher expression levels in resistant seedlings than in susceptible seedlings, including two TFs (ARF2 and IAA13). ARF2 belongs to a TF family that binds to auxin-responsive elements (AuxREs) in the promoters of auxin-regulated genes, and acts as the intermediate factor for crosstalk between the primary signaling pathways of auxin and ABA [41]. Auxin itself is known as a virulence factor to suppress host defense in some plant pathosystems [42,43]. Pathogen-secreted indole-3-acetic acid (IAA) increases rice disease symptoms by inducing expansins that cause loosening of the cell wall [44]. White pine bark tissues increase a few-fold in thickness as fusiform swelling in the cankered regions (especially at aeciopore stage) in susceptible seedlings, suggesting that *C. ribicola* may release plant hormone(s) to induce proliferation of host bark cells in responsive (i.e. susceptible) hosts [8]. In *Cr2*-resistant seedlings we found dramatic up-regulation (the highest with >1,000 folds increase) of a number of genes that down-regulate auxin and cellulose synthase-like protein (Additional file 4: Table S5), but expansin genes were highly suppressed (>10-fold decrease, Additional file 4: Table S6), indicating that auxin may have complex roles in the WP-BR interaction, and suppression of auxin action may be another important mechanism underlying the *Cr2*-mediated resistance by suppressing the loosening of the pine cell wall.

The ubiquitin–proteasome system (UPS) was regulated by *C. ribicola* infection

UPS is a protein degradation system worthy of special attention in the WPBR pathosystem because all 22 related DEGs were positively regulated by *C. ribicola* infection and 14 of them were up-regulated only in resistant seedlings (Additional file 5: Table S13). The UPS-related genes specifically regulated in the resistance response include various types of ubiquitin-conjugating enzymes and ubiquitin-associated (UBA) zinc-finger (ZZ) and phox/Bem1p (PB1) domain proteins. The UPS regulates proteins of the ABA receptor-signal complex and its downstream targets. Most F-box proteins (FBPs) are characterized as components of the SCF (SKp1, Cullin, F-box protein) E3 ubiquitin–ligase complex, which participates in the recognition and recruitment of target proteins for ubiquitination and degradation by the 26S UPS. Expression of many FBPs is transcriptionally regulated in a temporal and tissue specific manner, or in response to (a)biotic stressors [45]. Auxin binding to the SCF complex results in enhanced removal of TF repressors belonging to the AUX/IAA family through the SCF E3-ubiquitin ligase proteasome (26S) pathway [46]. The degradation of the AUX/IAA transcriptional repressors leads to the activation of ARFs and the expression of

auxin-responsive genes, which, in turn, positively regulate plant resistance to necrotrophic fungi in Arabidopsis [47]. The proteins of the plant p62/NBR1/Joka2 family contain PB1, ZZ and C-terminal UBA domains and presumably function as cargo receptors in the selectivity of autophagy, which may constitute an important part of plant response to environmental stresses [48]. Accumulated evidence has identified subunits and proposed regulators of SCF ubiquitin ligases as essential components of R gene-mediated resistance [49]. Up-regulation of the UPS- and FBP-related transcripts may help white pine to reprogram cellular homeostasis by recovering normal conformation of proteins and enzymes for resistance to *C. ribicola* invasion.

Differential regulation of multiple TF families in the WPBR pathosystem

Down-stream defense-responsive genes are normally regulated positively or negatively by different TFs that are direct or indirect targets of various signal transduction pathways. The *P. monticola* TFs differentially expressed between compatible and incompatible interactions belong to a few superfamilies with well-characterized structural domains such as ARF2, IAA13, AP2, zinc-finger CCCH type, zinc-finger-HD, PLATZ, NAC, DOF zinc-finger, AP2/ERF-B3, R2R3-MYB, WRKY, C3HL, HD-leucine-zipper, and tubby-like F-box. Other genes potentially involved in regulating transcription processes include nuclear TF Y subunit b-3, cold-shock DNA-binding protein, ethylene-responsive transcriptional co-activator, and other DNA-binding proteins (Additional file 5: Table S13).

In addition to ARF2 and IAA13 functioning in auxin-signalling, we found five other families of TFs (zinc finger HD, NAC, PLATZ-domain, Dof-zinc finger, and tubby-like F-box) responsive specifically in *Cr2*-resistance. A soybean zinc finger homeodomain TF, GmZF-HD1, activates the expression of the calmodulin GmCaM4 in response to pathogens by specifically interacting with A/T-rich repeats in the promoter [50]. PLATZ is a class of plant-specific zinc-dependent DNA-binding protein responsible for A/T-rich sequence-mediated transcriptional repression [51]. The Dof TF family is involved in the control of a variety of plant-specific processes, including defense response, phytochrome signaling, and phenylpropanoid metabolism in an environmental and tissue-specific manner [52]. Six rice NAC genes showed preferential expression under biotic stress [53]. Arabidopsis NAC TFs may play a dual role in regulating both JA- and ABA-dependent responses [54]. An Arabidopsis NAC TF (ATAF1) functions as an attenuator of ABA signalling for the mediation of efficient penetration resistance upon *Blumeria graminis* attack [55]. NAC TFs manipulate plant stress responses by activating other genes encoding R2R3-MYB TF,

amylase, cold responsive protein, dehydration responsive proteins, GST, and late embryogenesis abundant (LEA) proteins [56]. We observed significant regulation of two NAC homologs as well as GST, R2R3-MYB, and LEA genes following *C. ribicola* infection. Whether the latter are the targets of the *P. monticola* NAC TFs is an interesting question to address in future work.

P. monticola TF genes of six families (TF 3, C3HL domain class, AP2/ERF-B3 domain-containing, R2R3-MYB, CCCH type zinc finger, and HD-leucine zipper) were up-regulated in both resistant and susceptible seedlings. Two DEGs were detected in each family of the TFs with AP2/ERF-B3, R2R3-MYB, or CCCH type zinc finger domain, respectively. One family member was responsive in both resistant and susceptible seedlings while another was responsive only in resistant seedlings. Plant zinc-finger CCCH type TFs have been implicated to function in a series of plant developmental and adaptive processes, including plant defense responses to (a) biotic stresses, by regulating gene expression from the transcriptional to posttranscriptional levels [57,58]. A loss-of-function mutant of an Arabidopsis zinc finger CCCH domain-containing protein showed an increased local susceptibility to a fungal pathogen and sensitivity to seed germination in the presence of ABA [57]. Arabidopsis R2R3-MYB TF (MYB12) directly acts on the promoters of the flavonoid biosynthesis genes and it is placed at the downstream end of the signalling chain that causes flavonol-specific gene activation in phenylpropanoid biosynthesis [59]. Wheat R2R3 MYB TF (TaPIMP1) mediates host resistance to fungal pathogen and drought stresses by regulating defense-related genes (PR1 and PR5 genes) through ABA- and SA-signalling pathways [60]. Accompanying differential expression of *P. monticola* R2R3-MYB genes, we also observed differential expression of genes encoding enzymes related to biosynthesis of flavonol and phenylpropanoid, as well as genes encoding for multiple members of cytochrome p450, the TLP (PR5), PR10, and other PR families between resistant and susceptible genotypes.

Plant TFs with one AP2/ERF domain together with one B3 domain are assigned to the RAV (related to ABI3/VP1) family belonging to the plant AP2/ERF superfamily, and RAV family TFs have been reported to respond to hormones (ethylene and brassinosteroid) and (a)biotic stresses [61,62]. Over-expression of a RAV gene induced expression of ERF (ethylene-responsive factor) and PR5 genes and increased tolerance to bacterial pathogen in transgenic tomato [62]. *P. monticola* ACC oxidase genes were up-regulated and one ethylene-responsive transcriptional coactivator was down-regulated in similar patterns in both resistant and susceptible seedlings, suggesting ethylene-signalling may be

involved only in the basal response in the WPBR pathosystem. However, two *P. monticola* RAV (AP2/ERF-B3 domain) homologs showed differential expression between *Cr2*⁻ and *cr2/cr2* genotypes, suggesting their upstream intermediate factors (such as MAPK and CIPK as candidates) may be different between two genotypes.

Conclusions

In the present study WWP primary needles pooled from multiple seedlings at early stages (0- and 4-dpi) post *C. ribicola* infection were used for cDNA library construction. Secondary needle, stem (bark), and other tissues from more mature WWP trees or from tissues at relatively late pathogen infection stages could be investigated to detect tissue-specific defence responses and to identify host genes regulated specifically during other infection phases of the *C. ribicola* life cycle. These host genes may be found to contribute more to quantitative disease resistance. In this study, about 2.5% of the whole transcriptome assembly was identified as rust-responsive genes and 85% of them were functionally annotated in *P. monticola* defense, but their putative contribution to host resistance to *C. ribicola* awaits verification by functional genomics studies or association studies focused on exploring gene variation in *P. monticola* populations.

RNA-seq analysis of the WP-BR interactions revealed that (1) two types of plant R candidates (NBS-LRR and RLKs) were up-regulated specifically in resistant genotype following *C. ribicola* infection, suggesting a distinct role of these R candidates in *Cr2*-mediated resistance; (2) the biosynthesis and signalling pathways of multiple plant hormones (auxin, ABA, and ET) were coordinately regulated following rust infection, indicating that the auxin and ABA-mediated signaling pathways are involved in white pine resistance to rust; 3) a set of novel TFs were identified in response to *C. ribicola* infection, some of them (NAC, PLATZ, Dof, and ZF-HD TFs) specifically responsive in the incompatible WP-BR interaction; and 4) several families of PR proteins (PR1, PR2, PR3, PR5, PR6), ROS-related proteins (GSTs, thioredoxin-like and peroxidases), UPS proteins, and retrotransposons were differentially expressed at the transcriptional level between resistant and susceptible genotypes following *C. ribicola* infection.

Methods

Plant materials and rust inoculation

Inoculation of *P. monticola* seedlings with *C. ribicola* was the same as described previously [7]. In brief, 200 six-month-old seedlings of the open-pollinated seed lot #3926 (from a heterozygous resistant tree) were inoculated with *C. ribicola* in an inoculation chamber designed to facilitate basidiospore shed with day temperatures of 16°C and night temperatures of 12°C,

respectively. Basidiospores were shed from *C. ribicola*-infected *Ribes nigrum* leaves that were laid on a metal mesh above the white pine seedlings. Infected leaves were collected from a *Ribes* garden on Vancouver Island where only pathogenic *avcr2* isolates were available. Basidiospore density was monitored by placing glass slides at random underneath the mesh during rust inoculation. WWP needles were inoculated at a spore density higher than 3,000 spores per square centimeter. Primary needles were collected from at least 10 seedlings individually for each treatment at 0- and 4-dpi and stored at -80°C. Each seedling was pre-identified as resistant (*Cr2*⁻) or susceptible (*cr2/cr2*) genotype using *Cr2*-linked DNA markers as reported previously [63].

cDNA library construction

Total RNA was extracted from approximately five grams of needles pooled from at least 10 seedlings per treatment using a protocol described previously [64]. After eliminating traces of genomic DNA by treatment with DNase I (Invitrogen, Burlington, ON, Canada), an automated capillary gel electrophoresis was used to assess RNA quality and quantity using a Bioanalyzer 2100 with RNA 6000 Nano Labchips (Agilent Technologies Ireland, Dublin, Ireland). Total RNA samples with 28S/18S ratios in a range from 1.8 to 2.0 and RNA integrity index from 8.0 to 10.0 were selected for further processing.

PolyA + RNA fraction was purified from total RNAs (~5 µg) using a MACS mRNA isolation kit (Miltenyi Biotec, Bergisch Gladbach, Germany). Double stranded cDNA was synthesized using a Superscript cDNA synthesis kit (Invitrogen) with random hexamer primers at a concentration of 5 µM. The cDNA fragments were sheared using a Covaris E110 (Covaris, Woburn, MA, USA) for 75 sec under conditions "Duty cycle" of 20% and "Intensity" of 5, then the cDNA fraction with lengths of 200-250 bp was excised from 8% polyacrylamide gel electrophoresis (PAGE) for cDNA library construction using a paired-end sample prep kit (Illumina, San Diego, CA, USA). Briefly, the cDNAs were subject to end-repair and phosphorylation by T4 DNA polymerase, Klenow DNA polymerase, and T4 polynucleotide kinase respectively in a single reaction. The cDNAs with 3'-A overhangs, generated by Klenow fragment (3' to 5' exo minus), were ligated to paired-end adapters, which contain 5'-T overhangs. The adapter-ligated products were purified and enriched by PCR for 10 to 15 cycles using Phusion DNA polymerase and paired-end primer set (Illumina). PCR product of the desired size range was purified using 8% PAGE. The purified cDNA quality was assessed and quantified using an Agilent DNA 1000 series II assay kit on the Agilent 2100 Bioanalyzer (Agilent) and a Quant-IT ds-DNA HS assay kit on the Qubit

fluorometer (Invitrogen). The cDNA library was diluted to 8 nM and this final concentration was checked and determined again by Quant-IT dsDNA HS Assay before Illumina sequencing.

Illumina RNA-seq analysis

The Illumina GA IIX platform was used for deep sequencing of the cDNA libraries from both 5'- and 3'- end for 76-bp reads following the manufacturer's manual at the British Columbia Cancer Agency (Vancouver, BC, Canada). The deconvolution of fluorescent images to DNA sequences, base-calling and quality value calculation were performed using the Illumina data processing pipeline (version 1.4). The raw Illumina 76-bp pair-end sequences were deposited in the NCBI Sequence Read Archive (SRA) under accession numbers SRR1013833, SRR1013836, and SRR1013837.

Bioinformatics analyses

The CLC genomics workbench (version 5.5; CLC bio, Aarhus, Denmark) was selected for *de novo* transcriptome assembly in the present study because the CLC software has a faster computing pace with comparable or better assembly results than other bioinformatics programs [65]. Reads and read stretches of poor quality bases were removed and trimmed with filter threshold at quality score ($Q \geq 0.05$), ambiguous nucleotide ($N \leq 2$), and length trimming ($N \geq 30$) before *de novo* transcriptome assembly.

Using the BLASTn and tBLASTx algorithms, all non-redundant contigs were used for a BLAST search against the NCBI nr database (<http://www.ncbi.nlm.nih.gov/>), the PGI or SGI database (<http://compbio.dfci.harvard.edu/tgi/tgipage.html>) [66], a *Melampsora laricis populina* protein database (<http://genome.jgi-psf.org/Mellp1/Mellp1.download.ftp.html>), and a set of *P. monticola* ESTs (Girard-Martel M, personal communication). The PGI (Release 9.0, March 26, 2011) and SGI databases (Release 5.0, March 30, 2011) contained 77,326 and 79,409 unique ESTs respectively. GO annotation assignment [67] was used to perform functional gene annotation by mapping GO terms using databases of the NCBI nr, PIR (<http://pir.georgetown.edu/pirwww/>), GO (<http://www.geneontology.org/>), UniProts (<http://www.ebi.ac.uk/UniProt/>), and KEGG (<http://www.genome.jp/kegg/>) in the BLAST2GO program (Biobam Bioinformatics S. L., Valencia, Spain, <http://www.blast2go.com/b2ghome/about-blast2go>) with an E-value cutoff of 10^{-6} [68]. Biological pathways were identified by gene annotation using enzyme code and KEGG databases [69]. GOslims_Plant was used to generate a focused view of the plant GO categories (http://www.geneontology.org/GO_slims/). A consensus transcriptome was created by assembly comparison of three cDNA libraries

constructed from uninfected WWP needles (as control at 0-dpi), *C. ribicola*-infected needles from resistant (*Cr2/-*) and susceptible (*cr2/cr2*) genotypes at 4-dpi following a procedure of reciprocal BLASTn as described by Ness et al. [70].

Raw reads data were mapped back to the reference transcriptome for evaluation of gene expression levels and global transcript expression profiling among plant samples with different treatments. For statistical analysis to identify DEGs, RPKM was calculated as the normalized transcript expression value [71]. To determine whether cDNA libraries were compatible with each other, a box plot and a density plot were performed to evaluate the RPKM overall distribution, variability, and similarity in each sample using Bioconductor (version 2.12) software in conjunction with R software (version 3.0.0). A Z-test [72] was used to identify DEGs between any two experimental conditions (at a Bonferroni corrected *p*-value cut-off of 0.05) using the CLC genomics workbench. The expression patterns of DEGs were analyzed by a K-means clustering method [73] using euclidean distance based on gene expression values over all input samples (Additional file 5: Table S14).

Transcript expression analysis via qRT-PCR

A subset of contigs assembled from RNA-seq was used for transcript expression analysis via qRT-PCR. qRT-PCR analysis was performed as described previously [7,14,18]. Gene-specific primers of 26 genes were designed (Additional file 5: Table S15), including actin and tubulin genes as internal controls. Student *t* tests were used to analyze the significance of transcript differences between control and infected samples. Correlation/regression analyses were performed to compare fold changes of transcripts measured by qRT-PCR and RNA-seq analysis. ANOVA tests were used to estimate statistical significance of correlation between two sets of expression data generated by RNA-seq analysis and qRT-PCR.

Additional files

Additional file 1: Table S1. RNA-seq reads generated and analyzed in three pooled western white pine primary needles. **Table S2.** BLAST search to evaluate transcriptome data sets of western white pine primary needles. **Table S3.** Top ten sequences with most abundant expression levels based on total gene reads or their RPKM values. **Table S4.** Gene and enzyme numbers of metabolic pathways detected in the whole primary needle transcriptome and differentially expressed genes (DEGs) regulated in the white pine-blister rust (WP-BR) interactions.

Additional file 2: Figure S1. Plot analysis of transcript expression values (RPKM) in three western white pine cDNA libraries for quality control. The RPKM overall distribution and variability of three cDNA libraries/samples were similar, indicating that they were comparable for identification of differentially expressed genes (DEGs) at the transcriptome level. (A) A box plot analysis using CLC genomics workbench; (B) A density plot using Bioconductor (version 2.12) software in conjunction with R software (version 3.0.0).

Additional file 3: Figure S2. Functional classification of the differentially expressed genes (DEGs) in white pine-blister rust (WP-BR) interactions at an early stage (4-dpi) post *Cronartium ribicola* inoculation. Subcategories (A) for biological process (BP), (B) for molecular function (MF), and (C) for cellular component (CC).

Additional file 4: Table S5. A list of 245 contigs with significant up-regulation specifically in resistant (*Cr2/-*) seedling post *C. ribicola* infection (4-dpi). **Table S6.** A list of 142 contigs with significant down-regulation specifically in resistant (*Cr2/-*) seedling post *C. ribicola* infection (4-dpi).

Table S7. A list of 114 contigs with significant up-regulation specifically in susceptible (*cr2/cr2*) seedling post *C. ribicola* infection (4-dpi).

Table S8. A list of 46 contigs with significant down-regulation specifically in susceptible (*cr2/cr2*) seedling post *C. ribicola* infection (4-dpi).

Table S9. A list of 204 contigs with significant up-regulation in both resistant (*Cr2/-*) and susceptible (*cr2/cr2*) seedling post *C. ribicola* infection (4-dpi). **Table S10.** A list of 106 contigs with significant down-regulation in both resistant (*Cr2/-*) and susceptible (*cr2/cr2*) seedling post *C. ribicola* infection (4-dpi). **Table S11.** A list of 141 contigs with significantly higher expression levels in resistant (*Cr2/-*) seedlings than in susceptible (*cr2/cr2*) seedling post *C. ribicola* infection (4-dpi). **Table S12.** A list of 134 contigs with significantly higher expression levels in susceptible (*cr2/cr2*) seedling than in resistant (*Cr2/-*) seedlings post *C. ribicola* infection (4-dpi).

Additional file 5: Table S13. Functional grouping of differentially expressed genes (DEGs) during early stages of compatible and incompatible white pine-blister rust (WP-BR) interactions. **Table S14.** A list of contigs used in the K-means clustering analysis. **Table S15.** A list of primers used for quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JJL conceived the study, performed analysis of RNA-seq data, and drafted the manuscript. RB performed statistical analysis. RS conceived and helped to facilitate construction of the bioinformatics platform. All authors read and provided comments and approved the final manuscript.

Acknowledgements

This research was supported in part by the Canadian Forest Service and the CFS-Genomics R&D Initiative Fund awarded to J.-J. Liu. We thank PFC staff Gary Zhang and Aimin Guan for bioinformatic programming, Kristian Arndt for database management, Hao Chen for construction of the bioinformatic computing server, Marie Girard-Martel for EST data, and Holly William and Arezoo Zamany for critical review on the manuscript.

Received: 25 May 2013 Accepted: 2 December 2013

Published: 16 December 2013

References

1. Bingham RT, Squillace AE, Duffield JW: **Breeding blister-rust-resistant western white pine.** *J Forestry* 1953, **51**:163–168.
2. Kim MS, Richardson BA, McDonald GI, Klopfenstein NB: **Genetic diversity and structure of western white pine (*Pinus monticola*) in North America: a baseline study for conservation, restoration, and addressing impacts of climate change.** *Tree Genet Genom* 2011, **7**:11–21.
3. Jermstad KD, Sheppard LA, Kinloch BB, Delfino-Mix A, Ersoz ES, Krutovsky KV, Neale DB: **Isolation of a full-length CC-NBS-LRR resistance gene analog candidate from sugar pine showing low nucleotide diversity.** *Tree Genet Genom* 2006, **2**:1614–2942.
4. Liu J-J, Hammett C, Sniezko RA: ***Pinus monticola* pathogenesis-related gene *PmPR10-2* alleles as defense candidates for stem quantitative disease resistance against white pine blister rust (*Cronartium ribicola*).** *Tree Genet Genom* 2013, **9**:397–408.
5. Richardson BA, Ekramoddouh AK, Liu J-J, Kim M-S, Klopfenstein NB: **Current and future molecular approaches to investigate the white pine blister rust pathosystem.** *Forest Pathol* 2010, **40**:314–331.
6. Smith JA, Blanchette RA, Burnes TA, Jacobs JJ, Higgins L, Witthuhn BA, David AJ, Gillman JH: **Proteomic comparison of needles from blister rust-resistant and susceptible *Pinus strobus* seedlings reveals up regulation of**

- putative disease resistance proteins. *Mol Plant Microbe Interact* 2006, **19**:150–160.
7. Zamany A, Liu JJ, Ekramoddoullah AK: Comparative proteomic profiles of *Pinus monticola* needles during early compatible and incompatible interactions with *Cronartium ribicola*. *Planta* 2012, **236**:1725–1746.
 8. Hudgins JW, McDonald GI, Zambino PJ, Klopfenstein NB, Franceschi VR: Anatomical and cellular responses of *Pinus monticola* stem tissues to invasion by *Cronartium ribicola*. *Forest Pathol* 2005, **35**:423–443.
 9. Jacobs JJ, Burnes TA, David AJ, Blanchette RA: Histopathology of primary needles and mortality associated with white pine blister rust in resistant and susceptible *Pinus strobus*. *For Pathol* 2009, **39**:361–376.
 10. Grotkopp E, Rejmanek M, Sanderson MJ, Rost TL: Evolution of genome size in pines (*Pinus*) and its life-history correlates: supertree analyses. *Evolution* 2004, **58**:1705–1729.
 11. Wang Z, Gerstein M, Snyder M: RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009, **10**:57–63.
 12. Westermann AJ, Gorski SA, Vogel J: Dual RNA-seq of pathogen and host. *Nat Rev Microbiol* 2012, **10**:618–630.
 13. Strickler SR, Bombarely A, Mueller LA: Designing a transcriptome next-generation sequencing project for a nonmodel plant species. *Am J Bot* 2012, **99**:257–266.
 14. Liu J-J, Zamani A, Ekramoddoullah AK: Expression profiling of a complex thaumatin-like protein family in western white pine. *Planta* 2010, **231**:637–651.
 15. Jones JD, Dangl JL: The plant immune system. *Nature* 2006, **444**:323–329.
 16. Kinloch BB Jr, Sniezko RA, Dupper GE: Virulence gene distribution and dynamics of the white pine blister rust pathogen in western North America. *Phytopathology* 2004, **94**:751–758.
 17. Zipfel C, Robatzek S: Pathogen-associated molecular pattern-triggered immunity: veni, vidi...? *Plant Physiol* 2010, **154**:551–554.
 18. Liu J-J, Ekramoddoullah AK: Genomic organization, induced expression and promoter activity of a resistance gene analog (*PmTNL1*) in western white pine (*Pinus monticola*). *Planta* 2011, **233**:1041–1053.
 19. Elmore JM, Lin ZJ, Coaker G: Plant NB-LRR signaling: upstreams and downstreams. *Curr Opin Plant Biol* 2011, **14**:365–371.
 20. Fritz-Laylin LK, Krishnamurthy N, Tör M, Sjölander KV, Jones JDG: Phylogenomic analysis of the receptor-like proteins of rice and Arabidopsis. *Plant Physiol* 2005, **138**:611–623.
 21. Rodriguez MC, Petersen M, Mundy J: Mitogen-activated protein kinase signaling in plants. *Annu Rev Plant Biol* 2010, **61**:621–649.
 22. Tena G, Asai T, Chiu WL, Shen J: Plant mitogen-activated protein kinase signaling cascades. *Curr Opin Plant Biol* 2001, **4**:392–400.
 23. Kurusu T, Yagala T, Miyao A, Hirochika H, Kuchitsu K: Identification of a putative voltage-gated Ca²⁺ channel as a key regulator of elicitor-induced hypersensitive cell death and mitogen-activated protein kinase activation in rice. *Plant J* 2005, **42**:798–809.
 24. Galletti R, Ferrari S, De Lorenzo G: Arabidopsis MPK3 and MPK6 play different roles in basal and oligogalacturonide- or flagellin-induced resistance against *Botrytis cinerea*. *Plant Physiol* 2011, **157**:804–814.
 25. Kulik A, Wawer I, Krzywińska E, Bucholc M, Dobrowolska G: SnRK2 protein kinases - key regulators of plant response to abiotic stresses. *OMICS* 2011, **15**:859–872.
 26. Morita-Yamamuro C, Tsutsui T, Sato M, Yoshioka H, Tamaaki M, Ogawa D, Matsuura H, Yoshihara T, Ikeda A, Uyeda I, Yamaguchi J: The Arabidopsis gene *CAD1* controls programmed cell death in the plant immune system and encodes a protein containing a MACPF domain. *Plant Cell Physiol* 2005, **46**:902–912.
 27. Noutoshi Y, Kuromori T, Wada T, Hirayama T, Kamiya A, Imura Y, Yasuda M, Nakashita H, Shirasu K, Shinozaki K: Loss of necrotic spotted lesions 1 associates with cell death and defense responses in *Arabidopsis thaliana*. *Plant Mol Biol* 2006, **62**:29–42.
 28. Liu H, Zhang H, Yang Y, Li G, Yang Y, Wang X, Basnayake BMVS, Li D, Song F: Functional analysis reveals pleiotropic effects of rice RING-H2 finger protein gene *OsBIRF1* on regulation of growth and defense responses against abiotic and biotic stresses. *Plant Mol Biol* 2008, **68**:17–30.
 29. Hangen E, Blomgren K, Bénit P, Kroemer G, Modjtahedi N: Life with or without AIF. *Trends Biochem Sci* 2010, **35**:278–287.
 30. Pearce G, Yamaguchi Y, Barona G, Ryan CA: A subtilisin-like protein from soybean contains an embedded, cryptic signal that activates defense-related genes. *Proc Natl Acad Sci USA* 2010, **107**:14921–14925.
 31. Li C, Zhang K, Zeng X, Jackson S, Zhou Y, Hong Y: A cis element within Flowering Locus T mRNA determines its mobility and facilitates trafficking of heterologous viral RNA. *J Virol* 2009, **83**:3540–3548.
 32. LeBlanc M, Kim G, Westwood JH: RNA trafficking in parasitic plant systems. *Front Plant Sci* 2012, **3**:203.
 33. Fan J, Hill L, Crooks C, Doerner P, Lamb C: Abscisic acid has a key role in modulating diverse plant-pathogen interactions. *Plant Physiol* 2009, **150**:1750–1761.
 34. Klingler JP, Batelli G, Zhu J-K: ABA receptors: the START of a new paradigm in phytohormone signalling. *J Exp Bot* 2010, **61**:3199–3210.
 35. Iturriaga EA, Leech MJ, Barratt DHP, Wang TL: Two ABA responsive proteins from pea (*Pisum sativum* L.) are closely related to intracellular pathogenesis-related proteins. *Plant Mol Biol* 1994, **24**(Suppl 1):235–240.
 36. Jiang SY, Ramamoorthy R, Ramachandran S: Comparative transcriptional profiling and evolutionary analysis of the GRAM domain family in eukaryotes. *Dev Biol* 2008, **314**:418–432.
 37. Choi DS, Hwang BK: Proteomics and functional analyses of pepper abscisic acid-responsive 1 (ABR1), which is involved in cell death and defense signaling. *Plant Cell* 2011, **23**:823–842.
 38. Yasuda M, Ishikawa A, Jikumaru Y, Seki M, Umezawa T, Asami T, Maruyama-Nakashita A, Kudo T, Shinozaki K, Yoshida S, Nakashita H: Antagonistic interaction between systemic acquired resistance and the abscisic acid-mediated abiotic stress response in Arabidopsis. *Plant Cell* 2008, **20**:1678–1692.
 39. Lee S, Lee EJ, Yang EJ, Lee JE, Park AR, Song WH, Park OK: Proteomic identification of annexins, calcium-dependent membrane binding proteins that mediate osmotic stress and abscisic acid signal transduction in Arabidopsis. *Plant Cell* 2004, **16**:1378–1391.
 40. Mortimer JC, Laohavisit A, Macpherson N, Webb A, Brownlee C, Battey NH, Davies JM: Annexins: multifunctional components of growth and adaptation. *J Exp Bot* 2008, **59**:533–544.
 41. Wang L, Hua D, He J, Duan Y, Chen Z, Hong X, Gong Z: Auxin response factor2 (ARF2) and its regulated homeodomain gene *HB33* mediate abscisic acid response in Arabidopsis. *PLoS Genet* 2011, **7**:e1002172.
 42. Domingo C, Andrés F, Tharreau D, Iglesias DJ, Talón M: Constitutive expression of *OsGH3.1* reduces auxin content and enhances defense response and resistance to a fungal pathogen in rice. *Mol Plant Microbe Interact* 2009, **22**:201–210.
 43. Fu J, Liu H, Li Y, Yu H, Li X, Xiao J, Wang S: Manipulating broad-spectrum disease resistance by suppressing pathogen-induced auxin accumulation in rice. *Plant Physiol* 2011, **155**(1):589–602.
 44. Ding X, Cao Y, Huang L, Zhao J, Xu C, Li X, Wang S: Activation of the indole-3-acetic acid-amido synthetase *GH3-8* suppresses expansin expression and promotes salicylate- and jasmonate-independent basal immunity in rice. *Plant Cell* 2008, **20**:228–240.
 45. Jain M, Nijhawan A, Arora R, Agarwal P, Ray S, Sharma P, Kapoor S, Tyagi AK, Khurana JP: F-box proteins in rice. Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. *Plant Physiol* 2007, **14**:1467–1483.
 46. Kempinski S, Leyser O: The Arabidopsis F-box protein TIR1 is an auxin receptor. *Nature* 2005, **435**:446–451.
 47. Llorente F, Muskett P, Sánchez-Vallet A, López G, Ramos B, Sánchez-Rodríguez C, Jordá L, Parker J, Molina A: Repression of the auxin response pathway increases Arabidopsis susceptibility to necrotrophic fungi. *Mol Plant* 2008, **1**:496–509.
 48. Zientara-Ryttter K, Łukomska J, Moniuszko G, Gwozdecki R, Surowiecki P, Lewandowska M, Liszewska M, Wawrzyńska A, Sirko A: Identification and functional analysis of *Joka2*, a tobacco member of the family of selective autophagy cargo receptors. *Autophagy* 2011, **7**:1145–1158.
 49. Devoto A, Muskett PR, Shirasu K: Role of ubiquitination in the regulation of plant defence against pathogens. *Curr Opin Plant Biol* 2003, **6**:307–311.
 50. Park HC, Kim ML, Lee SM, Bahk JD, Yun D-J, Lim CO, Hong JC, Lee SY, Cho MJ, Chung WS: Pathogen-induced binding of the soybean zinc finger homeodomain proteins GmZF-HD1 and GmZF-HD2 to two repeats of ATTA homeodomain binding site in the calmodulin isoform 4 (*GmCaM4*) promoter. *Nucleic Acids Res* 2007, **35**:3612–3623.
 51. Nagano Y, Furuhashi H, Inaba T, Sasaki Y: A novel class of plant-specific zinc-dependent DNA-binding protein that binds to A/T-rich DNA sequences. *Nucleic Acids Res* 2001, **29**:4097–4105.
 52. Skiryicz A, Jozefczuk S, Stobiecki M, Muth D, Zanor MI, Witt I, Mueller-Roeber B: Transcription factor *AtDOF42* affects phenylpropanoid metabolism in *Arabidopsis thaliana*. *New Phytol* 2007, **175**:425–438.

53. Nuruzzaman M, Manimekalai R, Sharoni AM, Satoh K, Kondoh H, Ooka H, Kikuchi S: **Genome-wide analysis of NAC transcription factor family in rice.** *Gene* 2010, **465**:30–44.
54. Bu Q, Jiang H, Li CB, Zhai Q, Zhang J, Wu X, Sun J, Xie Q, Li C: **Role of the Arabidopsis thaliana NAC transcription factors ANAC019 and ANAC055 in regulating jasmonic acid-signaled defense responses.** *Cell Res* 2008, **18**:756–767.
55. Jensen MK, Hagedorn PH, de Torres-Zabala M, Grant MR, Rung JH, Collinge DB, Lyngkjær MF: **Transcriptional regulation by an NAC (NAM-ATAF1, 2-CUC2) transcription factor attenuates ABA signalling for efficient basal defence towards *Blumeria graminis* f. sp. *hordei* in Arabidopsis.** *Plant J* 2008, **56**:867–880.
56. Puranik S, Sahu PP, Srivastava PS, Prasad M: **NAC proteins: regulation and role in stress tolerance.** *Trends Plant Sci* 2012, **17**:369–381.
57. AbuQamar S, Chen X, Dhawan R, Bluhm B, Salmeron J, Lam S, Dietrich RA, Mengiste T: **Expression profiling and mutant analysis reveals complex regulatory networks involved in Arabidopsis response to *Botrytis* infection.** *Plant J* 2006, **48**:28–44.
58. Guo YH, Yu YP, Wang D, Wu CA, Yang GD, Huang JG, Zheng CC: **GhZFP1, a novel CCCH-type zinc finger protein from cotton, enhances salt stress tolerance and fungal disease resistance in transgenic tobacco by interacting with GZIRD21A and GZIPR5.** *New Phytol* 2009, **183**:62–75.
59. Mehrtens F, Kranz H, Bednarek P, Weisshaar B: **The Arabidopsis transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis.** *Plant Physiol* 2005, **138**:1083–1096.
60. Zhang Z, Liu X, Wang X, Zhou M, Zhou X, Ye X, Wei X: **An R2R3 MYB transcription factor in wheat, TaPIMP1, mediates host resistance to *Bipolaris sorokiniana* and drought stresses through regulation of defense- and stress-related genes.** *New Phytol* 2012, **196**:1155–1170.
61. Sohn KH, Lee SC, Jung HW, Hong JK, Hwang BK: **Expression and functional roles of the pepper pathogen-induced transcription factor RAV1 in bacterial disease resistance, and drought and salt stress tolerance.** *Plant Mol Biol* 2006, **61**:897–915.
62. Li CWS RC, Cheng CP, You SJ, Hsieh TH, Chao TC, Chan MT: **Tomato RAV transcription factor is a pivotal modulator involved in the AP2/EREBP-mediated defense pathway.** *Plant Physiol* 2011, **156**:213–227.
63. Liu J-J, Ekramoddoullah AK: **Development of leucine-rich repeat polymorphism, amplified fragment length polymorphism, and sequence characterized amplified region markers to the *Cronartium ribicola* resistance gene *Cr2* in western white pine (*Pinus monticola*).** *Tree Genet Genom* 2008, **4**:601–610.
64. Liu J-J, Xiang Y: **In silico mining and PCR-based approaches to transcription factor discovery in non-model plants: gene discovery of the WRKY transcription factors in conifers.** *Methods Mol Biol* 2011, **754**:21–43.
65. Brautigam A, Mullick T, Schliesky S, Weber APM: **Critical assessment of assembly strategies for non-model species mRNA-Seq data and application of next-generation sequencing to the comparison of C3 and C4 species.** *J Exp Bot* 2011, **62**:3093–31.
66. Quackenbush J, Cho J, Lee D, Liang F, Holt I, Karamycheva S, Parvizi B, Pertea G, Sultana R, White J: **The TIGR Gene Indices: analysis of gene transcript sequences in highly sampled eukaryotic species.** *Nucleic Acids Res* 2001, **29**:159–164.
67. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene Ontology: tool for the unification of biology.** *The Gene Ontology Consortium.* *Nat Genet* 2000, **25**:25–29.
68. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics* 2005, **21**:3674–3676.
69. Kanehisa M, Goto S: **KEGG: Kyoto encyclopedia of genes and genomes.** *Nucleic Acids Res* 2000, **28**:27–30.
70. Ness RW, Sjol M, Barrett SC: **De novo sequence assembly and characterization of the floral transcriptome in cross- and self-fertilizing plants.** *BMC Genomics* 2011, **12**:298.
71. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y: **RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays.** *Genome Res* 2008, **18**:1509–1517.
72. Kal AJ, van Zonneveld AJ, Benes V, van den Berg M, Koerkamp MG, Albermann K, Strack N, Ruijter JM, Richter A, Dujon B, Ansoorge W, Tabak HF: **Dynamics of gene expression revealed by comparison of serial analysis of gene expression transcript profiles from yeast grown on two different carbon sources.** *Mol Biol Cell* 1999, **10**:1859–1872.
73. Lloyd SP: **Least squares quantization in PCM.** *IEEE Trans Inf Theory* 1982, **28**:128–137.

doi:10.1186/1471-2164-14-884

Cite this article as: Liu et al.: Transcriptome analysis of *Pinus monticola* primary needles by RNA-seq provides novel insight into host resistance to *Cronartium ribicola*. *BMC Genomics* 2013 **14**:884.

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

