New Phytologist

Conservation and divergence of gene expression plasticity following *c.* 140 million years of evolution in lodgepole pine (*Pinus contorta*) and interior spruce (*Picea glauca* × *Picea engelmannii*)

Sam Yeaman[1,2]*, Kathryn A. Hodgins[1,2,3]*, Haktan Suren[4,5], Kristin A. Nurkowski[2], Loren H. Rieseberg[1], Jason A. Holliday[4] and Sally N. Aitken[2]

[1]Department of Botany, 6270 University Blvd, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada; [2]Department of Forest and Conservation Sciences, University of British Columbia, 2424 Main Mall, Vancouver, BC, V6T 1Z4, Canada; [3]School of Biological Sciences, Monash University, Building 18, Melbourne, Vic. 3800, Australia; [4]Department of Forest Resources and Environmental Conservation, Virginia Polytechnic Institute and State University, 304 Cheatham Hall, Blacksburg, VA 24061, USA; [5]Genetics, Bioinformatics and Computational Biology Program, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

Author for correspondence:
Sally N. Aitken
Tel: +1 604 822 6020
Email: sally.aitken@ubc.ca

Received: *27 January 2014*
Accepted: *20 March 2014*

New Phytologist (2014) **203:** 578–591
doi: 10.1111/nph.12819

Key words: climate, divergence, Engelmann spruce, gene expression, lodgepole pine, plasticity, RNAseq, white spruce.

Summary

- Species respond to environmental stress through a combination of genetic adaptation and phenotypic plasticity, both of which may be important for survival in the face of climatic change.
- By characterizing the molecular basis of plastic responses and comparing patterns among species, it is possible to identify how such traits evolve. Here, we used *de novo* transcriptome assembly and RNAseq to explore how patterns of gene expression differ in response to temperature, moisture, and light regime treatments in lodgepole pine (*Pinus contorta*) and interior spruce (a natural hybrid population of *Picea glauca* and *Picea engelmannii*).
- We found wide evidence for an effect of treatment on expression within each species, with 6413 and 11 658 differentially expressed genes identified in spruce and pine, respectively. Comparing patterns of expression among these species, we found that 74% of all orthologs with differential expression had a pattern that was conserved in both species, despite 140 million yr of evolution. We also found that the specific treatments driving expression patterns differed between genes with conserved versus diverged patterns of expression.
- We conclude that natural selection has probably played a role in shaping plastic responses to environment in these species.

Introduction

Anthropogenic climate change is predicted to disproportionately impact high latitudes in the Northern Hemisphere over the next century (IPCC, 2007), with significant consequences for boreal and temperate ecosystems (Bonan, 2008; Allen *et al.*, 2010). These regions have already experienced significant climate change (Mbogga *et al.*, 2009), which has been linked to the devastating effects of pests and pathogens (Woods *et al.*, 2005), as well as drought-related dieback in boreal forests (Hogg *et al.*, 2008; Allen *et al.*, 2010). Continuing changes in climate will almost certainly affect the yield of commercially managed boreal tree species (e.g. Wang *et al.*, 2006, 2012), and may threaten the survival of vulnerable species and populations (Alberto *et al.*, 2013; Hamann & Aitken, 2013). Plasticity may be especially important for response to climate change in conifers and other forest trees because of their long generation times and the rapid pace of contemporary climatic change (Franks *et al.*, 2013). Understanding the genetic basis of local adaptation and the capacity for plastic and adaptive responses to climate will be critical to predicting and managing the results of these changes in both natural and managed populations (Aitken *et al.*, 2008; Alberto *et al.*, 2013; Sork *et al.*, 2013).

While there is a large body of theoretical predictions and empirical results concerning the interplay between plasticity and local adaptation (Franks *et al.*, 2013), more empirical data are needed to comprehensively evaluate how plants respond to changes in environment (Snell-Rood *et al.*, 2010) and many questions remain unanswered. How often are plastic responses to environment conserved across species? Are plastic responses more conserved for some types of environmental stresses or gene functional categories? Do adaptive genetic responses involve the same genes that are involved in plasticity? Studies of gene expression using microarrays and RNAseq can enable quantitative assessment of plastic molecular responses to abiotic stress across a large number of genes, to help answer these questions.

By examining variation in gene expression within and among species, it is possible to investigate whether any observed plastic

*These authors contributed equally to this work.

responses are conserved or evolving divergently. In conjunction with studies of local adaptation within species and sequence divergence among species, we can assemble a more complete picture of the genomic basis of adaptation. Patterns of gene expression have now been characterized in a large number of species (reviewed in Snell-Rood *et al.*, 2010), but relatively few have compared species to identify which genes have conserved or divergent patterns of expression. While genes with conserved patterns can be identified by comparing co-expression networks characterized in different experiments (reviewed in Movahedi *et al.*, 2012), such meta-analyses have less power to assess divergent expression in response to environment, because experimental conditions differ among studies.

Here, we describe the results of a study on gene expression in lodgepole pine (*Pinus contorta*) and interior spruce (natural hybrid *Picea engelmannii* × *Picea glauca*) seedlings in response to a range of climatic and photoperiodic treatments. Lodgepole pine and interior spruce are two of the most widespread conifers in Western North America, and both are of considerable ecological and economic importance, with over 200 million trees planted annually in Western Canada (https://www.for.gov.bc.ca/hre/forgen/interior/interior.htm). The interior spruce complex is managed as a single taxon over large areas, and the seeds used in this study were sampled from a population within the hybrid zone (see the Materials and Methods section); hereafter we refer to it as a single 'species' for simplicity. Pine and spruce lineages both belong to the family Pinaceae, with an estimated divergence time of *c.* 140 million yr (Savard *et al.*, 1994; Wang *et al.*, 2000). Previous research has found extensive evidence for local adaptation to climate at the phenotypic level in both lodgepole pine (Rehfeldt *et al.*, 1999, 2001; Wu & Ying, 2004; Wang *et al.*, 2006; Eckert *et al.*, 2012) and white and Englemann spruce (Roche, 1969; Rweyongeza *et al.*, 2007; Ukrainetz *et al.*, 2011). Recently, studies have begun to characterize the genomic basis of adaptation in these species, in terms of both sequence variation (Namroud *et al.*, 2008; Eckert *et al.*, 2012; Parchman *et al.*, 2012; De La Torre *et al.*, 2014) and functional genomics (Joosen *et al.*, 2006; Holliday *et al.*, 2008; El Kayal *et al.*, 2011). In a study of white spruce (*Picea glauca*), Sitka spruce (*Picea sitchensis*), and black spruce (*Picea mariana*), 5407 genes were differentially expressed by tissue type, but only 60 of these genes had different patterns of differential expression when compared among these closely related species (Raherison *et al.*, 2013). Other studies in these species have found evidence for differential expression attributable to season (Holliday *et al.*, 2008; Reid *et al.*, 2013), pest resistance (Porth *et al.*, 2012), and even gravity in space (Beaulieu *et al.*, 2013), but as yet there has been little investigation of changes in gene expression in response to climatic conditions in these species and no comparative studies across conifer genera.

Our aim in this experiment was to characterize and compare patterns of expression in lodgepole pine and interior spruce to identify genes that are involved in plastic responses to climate, and explore how such responses have diverged among species. To this end, seedlings from a single population of each species were grown under one of seven different environmental treatments, following which their mRNA was extracted and sequenced. We began our analyses by assembling a *de novo* reference transcriptome for each species and estimating expression levels for all genes within these references. We identified thousands of genes within each species that showed significant patterns of differential expression by treatment, often with associated patterns of over-representation of gene ontology (GO) terms related to environmental conditions. From a set of *c.* 14 000 orthologs common to both species, we then identified genes that had highly conserved or highly divergent patterns of expression between these taxa. In combination with work currently underway on the genomic basis of local adaptation to climate, the results of this experiment will help us to understand how these species respond to their complex and heterogeneous environments.

Materials and Methods

Plant material and experimental treatments

We obtained 1-yr-old container-grown seedlings of lodgepole pine (*Pinus contorta* Douglas var. latifolia (Engelm.) Critchfield) and interior spruce (hybrid *Picea glauca* (Moench) Voss × *Picea engelmannii* Parry ex Engelm.) from the BC Ministry of Forests, Mines and Lands (Canada), on 3 May 2011. The pine originated from BC orchard seed lot number 63 019; the spruce originated from BC orchard seed lot number 63 060 (second generation seed orchards from the Nelson seed planning zones for each species containing genotypes from southeastern British Columbia at low to medium elevations). These seedlings had been grown under standard commercial production conditions for reforestation and cold-stored over winter. We potted the samples in small D16 cells (262 ml), using standard potting soil. We applied 1.5 g l^{-1} of liquid fertilizer (N-P-K; 20-8-20) to all samples, and moved all the plants to a single growth chamber with mild spring temperatures (5–15°C) and a 14 h day length. All plants were watered daily for 6 d until treatments were initiated (see Supporting Information Methods S1 'Experimental design and sampling' for further details).

The environmental treatments used in this study were designed to represent a range of stressful and nonstressful conditions as a way to stimulate expression responses, rather than attempt to accurately simulate specific climatic scenarios. We used five chambers to simulate different environmental conditions: cold and wet (CW) with temperatures of 5–15°C; mild and wet (MW), mild and dry (MD), mild and wet with 18 h daylight (MW18), and mild and wet plus heat (MWh), each with temperatures of 15–25°C; and hot and dry (HD), with temperatures of 25–35°C. In one additional treatment (MWbs), plants were grown under the same regime as MW but were sampled after budset, as described below in this section. All 'wet' treatments were watered every 24 h; 'dry' treatments were watered every 36–72 h, after visible signs of water stress. We maintained MW18 at 18 h day length throughout the experiment; in all other treatments, day length was varied gradually to mimic normal seasonal patterns for 52°N. We maintained the light

intensity at 400 lux throughout the experiment in all chambers. The MWh treatment included a biweekly 3-h heat treatment at midday; the peak temperature of this treatment was progressively increased from 35°C up to 45°C. In total, we had seven treatments that varied in temperature, moisture, photoperiod and developmental stage sampled (Table 1). We randomly assigned plants from each species to each of the seven treatments on 9 May, moving the plants to their respective growth chambers, with the exception of the HD plants, which first acclimatized at 15–25°C. We transferred the HD plants on 16 May to the 25–35°C chamber, and began restricting water (HD and MD) and applying heat stress (MWh) at this time. For all treatments except MWbs, we harvested tissues for RNA extraction at the end of the treatment cycle and before watering, with two to three sampling dates for each treatment, beginning on 16 June and ending on 30 June 2011. For MWbs, we sampled tissues 2 and 4 wk after budset (16 and 30 August 2011).

RNA extraction, library preparation and sequencing

We extracted RNA from needle samples in 44 lodgepole pine and 39 interior spruce plants, with three to eight individuals of each species represented in each treatment (all of which were nonnormalized biological replicates; Table S1). While we attempted to extract RNA from stem and roots, we obtained suitable yields from just four individuals per species (see Methods S1 'Experimental design and sampling'). These root plus stem libraries were only used for reference assembly and a limited comparison of differential expression in needle versus root plus stem, but were excluded from analysis of treatment effects. Samples were extracted using the protocol of Kolosova *et al.* (2004), in a random order with respect to treatment to avoid batch effects. We used a Nanodrop (Nanodrop Products, Wilmington, DE, USA) to ensure that each sample had a concentration of > 5 ng μl^{-1} (for a sample submission of 1 μg of total RNA), an $A_{260}/A_{280} > 1.8$, and an $A_{260}/A_{230} > 2.0$, as well as an Agilent bioanalyzer (Agilent, Santa Clara, CA, USA) to check that our samples were not degraded (RNA integrity number > 7.0). Standard paired end 100-bp Illumina RNAseq libraries (Illumina, San Diego, CA, USA) were prepared at Genome Quebec (Montreal, QC, Canada) and sequenced on 11 lanes of an Illumina HiSeq.

Table 1 A description of the treatments used for the RNAseq study of gene expression in lodgepole pine and interior spruce (see Supporting Information Methods S1 'Experimental design and sampling' for further details)

Treatment name	Treatment description
Cold wet (CW)	5–15°C; watered every *c.* 24 h, never water-stressed; photoperiod simulating 52°N
Mild wet (MW)	15–25°C; watered every *c.* 24 h, never water-stressed; photoperiod simulating 52°N
Mild dry (MD)	15–25°C; watered every 36–72 h, with visible signs of water stress; photoperiod simulating 52°N
Mild wet heat (MWh)	15–25°C; biweekly 3-h heat treatment during midday (35–45°C); watered every *c.* 24 h, never water-stressed; photoperiod simulating 52°N
Mild wet budset (MWbs)	15–25°C; watered every *c.* 24 h, never water-stressed; photoperiod simulating 52°N; sampled after budset
Mild wet 18 h (MW18)	15–25°C; watered every *c.* 24 h, never water-stressed; constant 18-h day lengths
Hot dry (HD)	25–35°C; watered every *c.* 36 h, with visible signs of water stress; photoperiod simulating 52°N

Reference assembly

Reference transcriptomes for each species were assembled using Trinity (Grabherr *et al.,* 2011), based on sequence data pooled from seven (pine) or eight (spruce) individuals from each species, including at least one sample from each treatment and multiple tissue types wherever possible (see Methods S1 'Reference assembly' for detailed methods). Contamination was removed by performing BLAST searches against several databases and a range of quality metrics was used to assess build quality (see Methods S1 'Reference assembly' for details). We also filtered out contigs with little to no detectable expression (described in the next section, 'Expression analysis'), and retained only those contigs that mapped successfully (using GMAP; Wu & Watanabe, 2005) to either the draft white spruce genome (Birol *et al.*, 2013) or the loblolly pine genome (Wegrzyn *et al.*, 2014). These filtering steps yielded a final 'reduced reference transcriptome' for each species that we used for all subsequent analysis.

Expression analysis

We used RSEM to estimate gene expression levels in each individual (Li & Dewey, 2011), aligning each library back to our initial Trinity assemblies. We then analyzed patterns of gene expression using the EDGER software package, which applies an over-dispersed Poisson model to account for both biological and technical variability in the count data (Robinson *et al.*, 2010). We used expression data to filter our initial Trinity assemblies by retaining only those contigs that had at least one count per million reads in at least three (spruce) or four (pine) libraries. These cutoffs are lenient enough to retain genes that are expressed in only one treatment (as the smallest treatments have three and four libraries in spruce and pine, respectively), but strict enough to eliminate genes with little to no detectable expression in most individuals. To examine patterns of differential expression among treatments, we excluded the four root plus stem tissue libraries from each species, and only analyzed data from the 39 spruce and 44 pine needle tissue libraries. Within EDGER, we used glmFit and glmLRT to estimate differential expression using tagwise estimates of dispersion. Qualitative patterns of differential expression were only slightly affected by changing the prior.df parameter used to estimate tagwise dispersion (not shown), so we used the default setting of prior.df = 20. Patterns of differential expression were assessed both across all treatments

(one factor with seven levels; model fit with intercept term) and as pairwise comparisons between each pair of treatments (using 'contrasts'; model fit without intercept term). Unless otherwise indicated, we used a cutoff of a false discovery rate (FDR) = 0.01 to identify genes with significant differential expression.

Comparison of interior spruce and lodgepole pine

Orthologs were identified between pine and spruce by using an all-versus-all TBLASTX and then clustering genes into orthogroups using OrthoMCL version 2.0.8 (Li *et al.*, 2003) with default parameters. To minimize the inclusion of alternatively spliced transcripts as different genes, we used the longest isoform of each gene identified by Trinity (i.e. the longest contig from each component group). We also only included genes that had detectable expression in both species, retaining those genes that had at least one count per million in one or more library from each species. For all identified orthologs, we used EdgeR to fit a model with species, treatment, and species × treatment interaction terms to the pooled gene expression data by using glmFit and glmLRT with tagwise dispersion to test for significant effects of each term. To examine whether having a diverged or conserved pattern of expression was associated with differential expression in certain treatments, we fit models to subsets of the pooled gene expression data and used 'contrasts' to identify genes with significant differential expression in pairwise contrasts. Conserved genes with significant pairwise differential expression between treatments were identified by fitting a model with only a treatment effect to the subset of genes that had a significant treatment effect but a nonsignificant treatment × species interaction term (the conserved expression genes (CEGs)). Genes that had diverged patterns of expression between species were identified by fitting a model with only a treatment × species interaction effect to the subset of genes that had a significant interaction term (the diverged expression genes (DEGs)). Because of nonindependence between pairwise contrasts, we calculated χ^2 contingency test statistics for the numbers of DEG versus CEG in each pairwise contrast (excluding cells with expected counts of < 5), and then used nonparametric tests to assess significance, by permuting the membership of each gene (whether CEG or DEG), calculating χ^2, and repeating 10 000 times. By this approach, *P*-values represent the proportion of times that $\chi^2_{permute} > \chi^2_{test}$.

Co-expression analysis

Weighted gene co-expression network analysis (WGCNA) is used to identify clusters of co-expressed genes (Langfelder & Horvath, 2008). We created co-expressed gene networks for both pine and spruce transcriptomes (analyzed separately) following the standard procedure of WGCNA. We calculated pair-wise Pearson correlations between members of each gene pair and transformed this into an adjacency matrix. The adjacency matrix was used to calculate gene similarity based on the co-expression relationship of the genes to one another. The clusters were then selected using the dynamic treecut algorithm.

Annotation

We identified GO terms using a combination of BLASTX searches against the TAIR 10 protein database (The Arabidopsis Information Resource) and the NR database (NCBI) screened for green plants. Annotations were assigned using Blast2Go (see Methods S1 'Annotation' for details). We performed an enrichment analysis to test for an excess or paucity of gene classes (based on GO terms) in our significant sets of genes relative to all the other expressed genes using Fisher's exact test.

Results

Reference assembly

Of the 373.3 Gb of sequence received from Genome Quebec, our filtering pipeline reduced the amount of sequence data by 6.7% to a total of 348.0 Gb (Table 2). In our *de novo* assemblies, we had a total of 340 157 contigs for spruce and 338 996 for pine, with many of these contigs probably representing alternatively spliced transcripts. Based on our BLAST results and gene expression analysis (see 'Expression analysis within pine and spruce'), we found that many of these were lowly expressed noncoding sequences of questionable biological relevance. A small subset were also found to be contamination from other species based on BLAST and alignments to the draft genomes. All of the 357 ultra conserved orthologs (UCOs; see Methods S1 'Reference assembly' for details) from *Arabidopsis thaliana* were present in both assemblies. Most of these UCOs were at least 80% of the length of the *A. thaliana* transcript (spruce mean length 84.5%; pine 83.5%). This suggests that our reference assemblies have captured most genes, and that the majority are close to the full length of the actual transcript. After filtering our references according to the inclusion criteria described in the 'Expression analysis' section in the Methods, our final pine and spruce references contained 23 519 and 23 889 contigs, respectively; we refer to these as 'reduced references', which we used for all subsequent expression analysis.

Assessment of assembly strategy

Many factors have been shown to impact the number of genes identified, such as sequencing depth, normalization, and the life stages, organs or tissues sampled (Wall *et al.*, 2009; Ekblom & Galindo, 2010; Lai *et al.*, 2012; Hodgins *et al.*, 2014). While the number of contigs assembled by Trinity increased with the number of sequence reads used as input, we found no substantial effect of using samples grown under different environmental conditions compared to samples from the same treatment (Fig. 1a; see Methods S1 'Assessment of assembly strategy' for details on methods). Assemblies built using reads from individuals grown under one environmental treatment (MD) or seven different treatments (treat7) contained similar numbers of contigs (Fig. 1b) and similar distributions of contig size (Fig. S2). By contrast, assemblies built from multiple sample types (root plus stem and foliar) contained 20% more contigs by number, or 14% more assembled sequence by total length (Fig. 1a).

We found that a little over one lane of an Illumina HiSeq provided a sufficient amount of sequence (*c.* 40 Gb) to assemble a reasonably complete transcriptome using libraries from both different treatments and different plant organs.

Expression analysis within pine and spruce

Across all seven treatments, we found that 11 658 genes in pine and 6413 genes in spruce were differentially expressed at FDR < 0.01. We also detected a large number of genes that were differentially expressed between organ types (root plus stem versus foliar; across all treatments), with 8131 genes in pine and 6695 genes in spruce. Based on heatmap plots, the patterns of differential gene expression by experimental treatment for the two species appear broadly similar in many ways (Fig. 2). To quantify the effect of individual treatments, we also examined the number of genes with differential expression in pairwise contrasts of all pairs of treatments (Fig. 3). We found patterns that were similar in pine and spruce for certain contrasts: the CW treatment resulted in considerable differential expression, while there was almost no detectable differential expression among comparisons of the MW, MD, and MW18 treatments (Fig. 3). In contrast, the HD and MWbs treatments resulted in much more differential expression in pine than in spruce, but this may partly reflect the lower power in spruce because of smaller sample sizes. Finally, in most contrasts the numbers of up- and downregulated genes were roughly similar, with the exception of more gene upregulation in HD in spruce and more gene upregulation in MWh in pine.

Table 2 The assembly metrics for lodgepole pine and interior spruce

Assembly metric	Pine	Spruce
Filtered sequence (Gb)	39.45	45.70
Total number of contigs	338 996	340 157
Component number[1]	227 588	226 113
Assembly length (Mb)	178.0	182.2
N50 (bp)	657	690
Average contig length (bp)	525.1	535.5
Per cent UCOs[2] present	100	100
Per cent UCOs > 80% *A. thaliana* transcript	83.5	84.5
Number of contaminating contigs[3]	7936	5215
Number of reduced reference contigs	23 889	23 519

The *de novo* assemblies were produced using the Trinity assembler.
[1]Trinity groups similar contigs together into components, representing alternatively spliced transcripts, close paralogs, or alleles (Grabherr *et al.*, 2011).
[2]Ultra conserved orthologs (UCO) are 357 single-copy genes that are shared by *Arabidopsis thaliana*, humans, mice, yeast, fruit flies, and *Caenorhabditis elegans* (A. Kozik, unpublished; http://compgenomics.ucdavis.edu/compositae_reference.php).
[3]Contigs from species other than green plants based on BLAST. Most were of fungal origin (Fig. S1).

Conservation and divergence of plasticity in gene expression

Using TBLASTX and OrthoMCL on the pine and spruce reduced references, we identified 14 691 one-to-one orthologs. By analyzing gene expression across both species and treatments for these orthologs, it was possible to identify numerous genes with: (1) differences in expression among treatments that are conserved in their pattern and average expression level between species (only treatment term significant); (2) conserved expression across treatments but different average levels of expression between species (only species term significant); (3) altered patterns of expression among treatments and species as well as differences in average amounts of expression (all terms significant); or (4) altered patterns but conserved average amounts of expression (only treatment × species interaction term significant; Fig. 4). Out of 5794 orthologs that had a significant effect of treatment on expression, 4298 (*c.* 74%, or *c.* 29% of all orthologs) did not have a significant treatment × species interaction term, indicating that for these genes the pattern of differential expression was conserved in both species (hereafter CEGs). Of these CEGs, 2207 (*c.* 38% or *c.* 15% of all orthologs) also did not have a significant species term, indicating that the baseline amount of expression was also conserved. In contrast, 2428 orthologs (17% of all orthologs) had a significant treatment × species interaction term, indicating that gene expression had diverged between species in a detectable way (hereafter DEGs). We found a significant

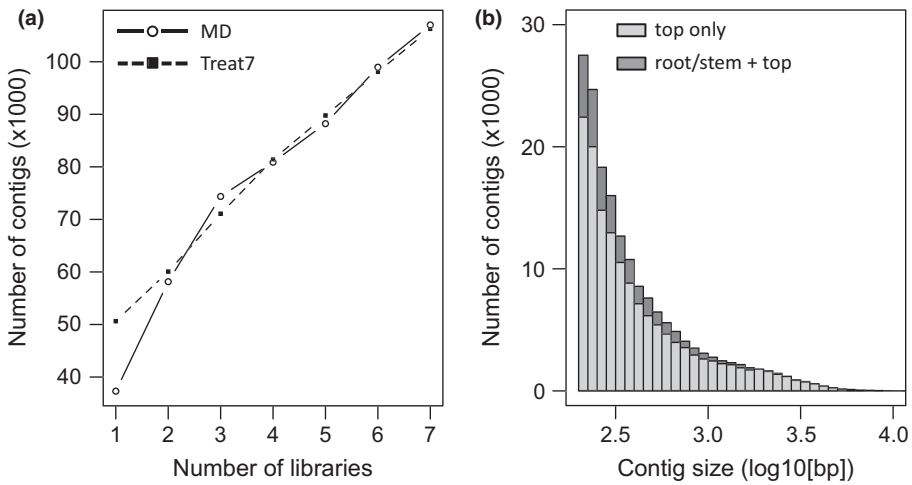

Fig. 1 (a) Number of lodgepole pine contigs yielded with increasing numbers of individuals combined together before assembly, with individuals all taken from the same treatment (MD) or from different treatments (treat7). (b) Number of lodgepole pine contigs yielded by assemblies built using equal numbers of reads but with only foliar samples or with a combination of foliar and root plus stem tissues.

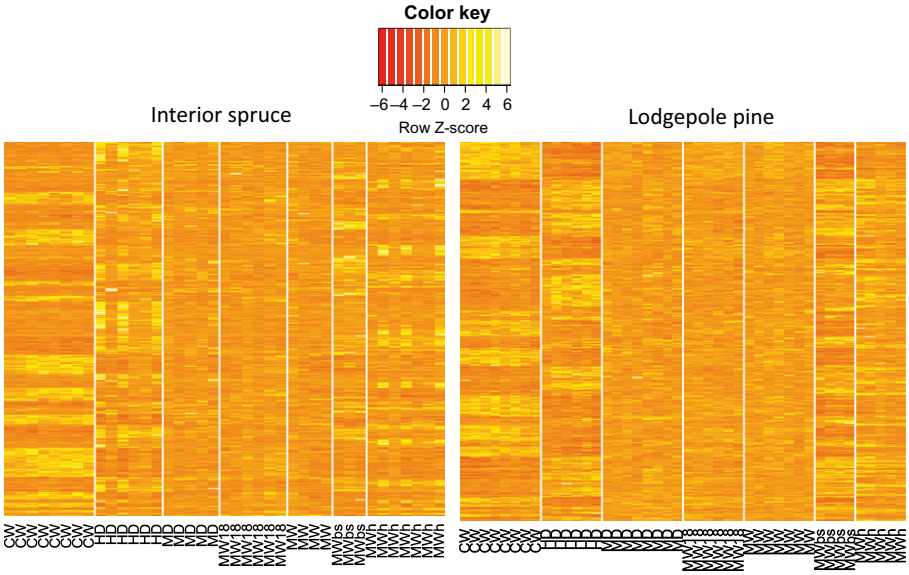

Fig. 2 Heatmaps showing patterns of expression for the 2567 most differentially expressed genes in interior spruce (left) and the 6711 most differentially expressed genes in lodgepole pine (right), at a false discovery rate (FDR) < 0.0001. Experimental treatments for each individual are shown along the x-axis: CW, cold wet; HD, hot dry; MD, mild dry; MW18, mild wet 18 h; MW, mild wet; MWbs, mild wet budset; MWh, mild wet heat.

		CW	HD	MD	MW18	MW	MWbs	MWh
CW	P		4745	1623	379	853	3743	2991
	S		4336	1912	338	1058	3497	2788
HD		2051		1201	1906	1766	1407	961
		2739		1330	2051	1781	1678	915
MD		527	103		0	1	1012	699
		754	45		0	1	1514	222
MW18		119	663	0		3	1282	988
		300	234	0		0	1770	559
MW		376	146	2	1		1311	794
		491	77	1	1		1836	269
MWbs		951	121	59	143	55		1755
		1150	124	77	176	141		1404
MWh		1590	564	190	236	130	371	
		2245	375	234	447	192	252	

Fig. 3 Number of genes with differential expression in pairwise contrasts among treatments in lodgepole pine (above diagonal) and interior spruce (below diagonal), at a false discovery rate (FDR) < 0.01. Upper numbers in each cell indicate the number of genes that are upregulated in the treatment listed in the column; lower numbers indicate the number of genes that are downregulated for the treatment listed in the column. Darker shades of orange indicate greater numbers of differentially expressed genes. Experimental treatments are indicated by: CW, cold wet; HD, hot dry; MD, mild dry; MW18, mild wet 18 h; MW, mild wet; MWbs, mild wet budset; MWh, mild wet heat.

difference between species in average expression level in 7418 orthologs, which represents approximately half of all identified orthologs.

To further explore the effects of specific treatments on these patterns, we examined the numbers of DEGs and CEGs that had significant pairwise contrasts among treatments (Fig. 5). We found evidence that DEGs and CEGs had significantly different distributions across the pairwise contrasts for both pine ($P < 10^{-4}$) and spruce ($P < 10^{-4}$) using nonparametric randomization tests (see the Materials and Methods section for details). In spruce, DEGs were more commonly upregulated in MWbs and MWh, while CEGs were more commonly downregulated in these treatments (Fig. 5). In pine, the data suggest the opposite pattern for MWbs and MWh, although much less strongly and consistently than in spruce. In both species, we found that more of the genes that were differentially expressed in MW18 had diverged expression patterns, regardless of the direction of regulation. However, more of the genes involved in differential expression in CW versus MWh treatments were conserved in their patterns of expression.

We also found significant differences in the proportion of CEGs versus DEGs in the co-expression clusters identified by WGCNA in both spruce ($\chi^2 = 100.9$; df = 11; $P < 10^{-15}$) and pine ($\chi^2 = 52.8$; df = 6; $P < 10^{-8}$; see the 'Annotation and co-expression' section for further description of the co-expression clusters and Table S2 for raw data). In spruce, DEGs were strongly overrepresented in the relatively small S6, S7, S8, S11, and S13 clusters (with 126%, 104%, 97%, 76% and 55% more DEGs than expected, respectively), while CEGs were weakly overrepresented in the relatively large S2 and S14 clusters (7% and 6% more CEGs than expected, respectively) and the smaller S10 cluster (21% more CEGs than expected). In pine, only the relatively small P1 cluster was heavily biased in its membership, with 133% more DEGs than expected (see Fig. 6 for plots showing expression patterns in each cluster and the 'Annotation and co-expression' section for GO term analysis).

Annotation and co-expression

Among the 23 889 expressed genes in spruce, 17 131 (71.7%) had a hit in the NR green plant database and 14 832 (62%) had a hit in the TAIR database. Of the 23 519 expressed genes in pine, 17 491 (67.9%) had a hit matching our criteria in the NR database, and 15 453 (65%) were found in the TAIR database. The highest percentage of top BLAST hits for both species was to *Picea sitchensis* (47.7% pine; 48.6% spruce). For spruce, 13 459

New Phytologist

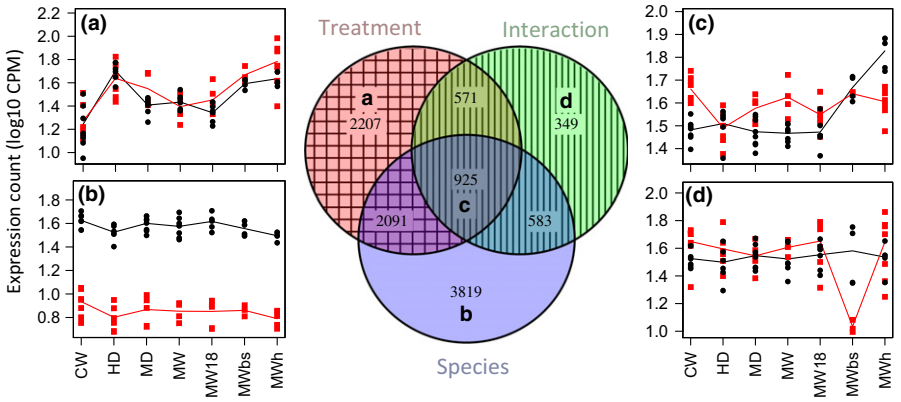

Fig. 4 The number of orthologs (out of 14 691 total) with significant patterns of differential expression at a false discovery rate (FDR) < 0.01, based on a model with species, treatment, and species × treatment interaction terms. Plots (a–d) each show an example of the pattern of expression for a pair of orthologs with characteristic patterns of expression typical to orthologs in the corresponding region of the Venn diagram, in counts per million reads (CPM) for pine (black lines) and spruce (red lines). Experimental treatments are indicated by: CW, cold wet; HD, hot dry; MD, mild dry; MW18, mild wet 18 h; MW, mild wet; MWbs, mild wet budset; MWh, mild wet heat.

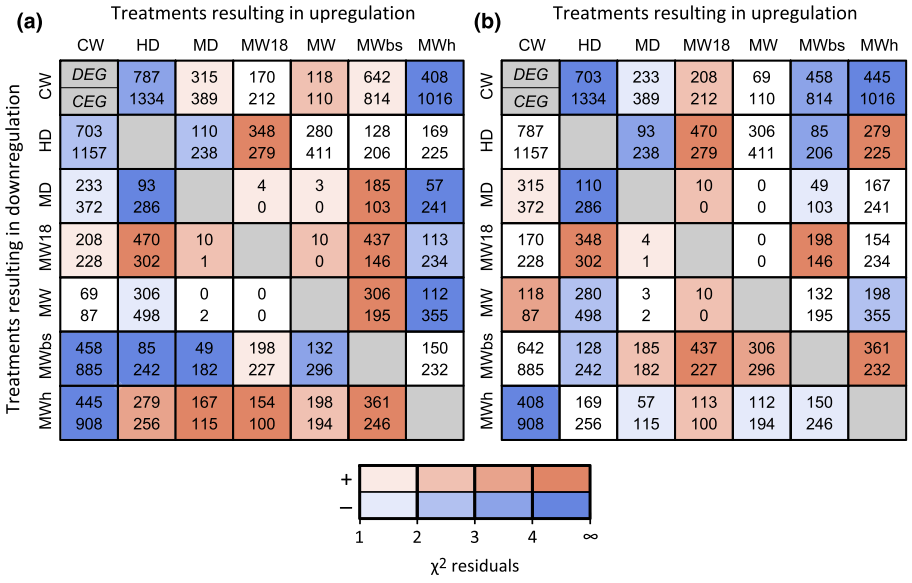

Fig. 5 Number of orthologs that have a conserved (conserved expression genes (CEGs); lower number in each cell) or diverged (diverged expression genes (DEGs); upper number in each cell) pattern of differential expression when compared between pine and spruce for each pairwise contrast among treatments. (a) Results for spruce orthologs; (b) results for pine orthologs. Colors indicate whether there are more (orange) or fewer (blue) DEGs than expected, based on a χ^2 contingency test of the number of DEGs versus CEGs in each pairwise contrast (colors reflect the magnitude of the DEG residuals). For a given cell, orthologs are upregulated in the treatment marked in the column, relative to the treatment marked in the row. Experimental treatments are indicated by: CW, cold wet; HD, hot dry; MD, mild dry; MW18, mild wet 18 h; MW, mild wet; MWbs, mild wet budset; MWh, mild wet heat.

genes were assigned GO annotations and 16 173 had IPS results (InterProScan from Blast2go; see Methods S1). For pine, 13 749 were assigned GO annotations and 16 240 had IPS results. Of the 6413 genes that differed between the treatments for spruce, 3588 were assigned GO annotations and 4351 had IPS results. Of the 11 658 genes that differed significantly between the treatments for pine, 7237 were assigned GO annotations and 8378 had IPS results.

For spruce, we found 15 GO slim terms over-represented in the set of differentially expressed genes among the seven treatments, including several biological processes such as response to abiotic stimulus (Tables 3a, S3). Of the top 100 differentially expressed genes, seven were annotated as heat shock proteins in *A. thaliana* and 13 were assigned GO terms relating to stress response or response to stimulus (Table S3). For pine, we found 14 GO slim terms over-represented, including response to endogenous stimulus (Tables 3b, S4). In addition, of the top 100 differentially expressed genes, 14 were identified as heat shock proteins in *A. thaliana* while 33 genes had GO terms relating to stress and stimulus response.

We also examined the over-representation of GO terms that were differentially expressed for specific pairwise comparisons between treatments. We compared CW versus HD and CW versus MWbs for each species as these comparisons had among the

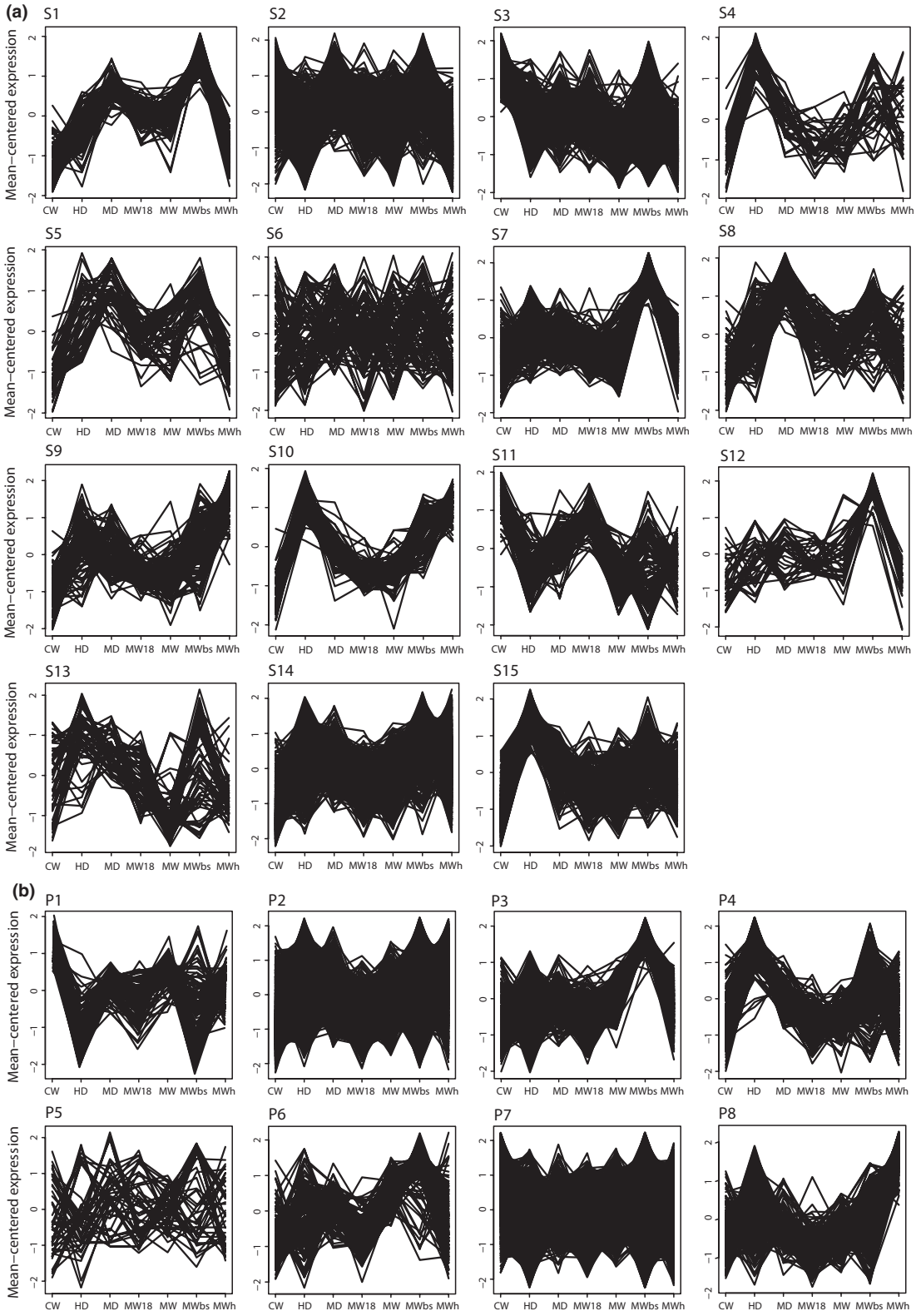

Fig. 6 Gene expression clusters from weighted gene co-expression network analysis (WGCNA) for spruce (a) and pine (b). Expression values are normalized by their standard deviation and mean-centered to show similarity in their patterns more clearly. Experimental treatments are indicated by: CW, cold wet; HD, hot dry; MD, mild dry; MW18, mild wet 18 h; MW, mild wet; MWbs, mild wet budset; MWh, mild wet heat.

greatest differences in gene expression. In addition, CW and HD had the most contrasting temperature and moisture conditions. We found several GO terms overrepresented in the CW/HD comparison for spruce (Table S3) including generation of precursor metabolites and energy and carbohydrate metabolic processes. In addition, 53 of the top 100 differentially expressed genes were

Table 3 The results of a Fisher's exact test examining the number of genes associated with gene ontology (GO) slim terms, showing GO terms over-represented in the set of genes with a significant treatment effect compared with the remaining expressed genes in (a) spruce and (b) pine, and (c) GO terms over-represented in the set of genes with a significant interaction between treatment and species

GO term	Name	Type	FDR	*P*-value	Test group (no.)	Reference group (no.)	Un- annotated test (no.)	Un- annotated reference group (no.)
(a)								
GO:0005975	Carbohydrate metabolic process	P	2.80E-06	1.60E-08	300	557	3288	9314
GO:0006091	Generation of precursor metabolites and energy	P	2.80E-06	2.60E-08	141	212	3447	9659
GO:0015979	Photosynthesis	P	2.20E-04	3.10E-06	59	71	3529	9800
GO:0005576	Extracellular region	C	2.70E-03	5.00E-05	86	136	3502	9735
GO:0044710	Single-organism metabolic process	P	2.70E-03	6.50E-05	288	604	3300	9267
GO:0006950	Response to stress	P	3.90E-03	1.10E-04	410	912	3178	8959
GO:0009579	Thylakoid	C	3.90E-03	1.40E-04	110	194	3478	9677
GO:0019748	Secondary metabolic process	P	3.90E-03	1.50E-04	122	221	3466	9650
GO:0009628	Response to abiotic stimulus	P	5.80E-03	2.50E-04	248	522	3340	9349
GO:0003824	Catalytic activity	F	9.10E-03	4.30E-04	1823	4693	1765	5178
GO:0006629	Lipid metabolic process	P	3.50E-02	1.80E-03	212	458	3376	9413
GO:0050896	Response to stimulus	P	3.90E-02	2.20E-03	656	1597	2932	8274
GO:0003700	Sequence-specific DNA-binding transcription factor activity	F	4.20E-02	2.80E-03	132	269	3456	9602
GO:0001071	Nucleic acid-binding transcription factor activity	F	4.20E-02	2.80E-03	132	269	3456	9602
(b)								
GO:0005975	Carbohydrate metabolic process	P	7.20E-09	3.40E-11	560	326	6677	6186
GO:0044710	Single-organism metabolic process	P	2.30E-03	2.20E-05	536	369	6701	6143
GO:0006629	Lipid metabolic process	P	5.90E-03	1.00E-04	416	283	6821	6229
GO:0003700	Sequence-specific DNA-binding transcription factor activity	F	5.90E-03	1.40E-04	254	159	6983	6353
GO:0001071	Nucleic acid-binding transcription factor activity	F	5.90E-03	1.40E-04	254	159	6983	6353
GO:0005794	Golgi apparatus	C	7.10E-03	2.00E-04	145	80	7092	6432
GO:0009719	Response to endogenous stimulus	P	1.60E-02	5.30E-04	291	194	6946	6318
GO:0005576	Extracellular region	C	1.90E-02	7.20E-04	154	91	7083	6421
GO:0006810	Transport	P	3.40E-02	1.80E-03	808	627	6429	5885
GO:0051234	Establishment of localization	P	3.40E-02	1.80E-03	808	627	6429	5885
GO:0051179	Localization	P	3.40E-02	1.80E-03	808	627	6429	5885
GO:0009056	Catabolic process	P	3.40E-02	2.00E-03	714	549	6523	5963
GO:0009607	Response to biotic stimulus	P	3.70E-02	2.30E-03	204	134	7033	6378
GO:0005654	Nucleoplasm	C	3.80E-02	2.50E-03	311	219	6926	6293
(c)								
GO:0003824	Catalytic activity	F	3.5E-3	1.7E-5	800	3238	843	4287
GO:0005975	Carbohydrate metabolic process	P	2.6E-2	2.5E-4	116	364	1527	7161

P, biological process; M, molecular function; C, cellular component.
The analysis was conducted using Blast2GO and only significant results ($\alpha = 0.05$, false discovery rate (FDR) corrected) are shown.

annotated and five were identified as heat shock proteins, while 17 had GO terms relating to stress and stimulus response. Similarly for pine, 79 of the top 100 differentially expressed genes were annotated and six were identified as heat shock proteins, while 17 genes had GO terms relating to stress and stimulus response. For the CW/MWbs comparison, we found several terms related to reproductive development in pine over-represented (Table S4); 69 of the top 100 differentially expressed genes were annotated and 17 had GO terms relating to stress and stimulus response, six were genes related to carbohydrate metabolism and one term was related to reproduction. For spruce, 74 of the top 100 differentially expressed genes were annotated. Twenty had GO terms relating to stress and stimulus response, nine were related to carbohydrate metabolism and seven were related to reproduction.

We also examined the annotations of DEGs and CEGs, comparing them to annotations in the orthologs that did not have either diverged or conserved patterns of expression to control for any biases in gene function that the method of ortholog detection would introduce. The top functional category over-represented in the CEGs was sequence-specific DNA-binding transcription factor activity; however, it was not significant after correcting for multiple comparisons. The top over-represented biological processes for the DEGs were translation catalytic activity and carbohydrate metabolic processes (Table 3c).

We used these annotations to examine the over-representation of GO terms for clusters that we identified using the co-expression network analysis (Fig. 6). In spruce, several clusters exhibited strong upregulation in MWbs (S1, S7, and S12 clusters). Although there were no significantly over-represented GO slim

terms in these clusters, there were several genes relating to reproductive and leaf development in the S7 cluster. There were many clusters with coordinated expression in the HD and MWbs treatments (e.g. S4, S13, and S15 clusters), including the S14 cluster, which had an over-representation of genes relating to stress response and nitrogen compound metabolic process (Table S3). The S9 and S10 clusters were upregulated in HD, MWbs, and MD, and the S10 cluster had genes relating to stress response and stimulus response over-represented. The S8 and S5 clusters were composed of genes upregulated in HD, MWbs and MD, but downregulated in CW. The expression of genes in the S6 cluster was influenced by all treatments and had an over-representation of genes involved in lipid metabolism, photosynthesis and secondary metabolic processes. The large S2 cluster was comprised of genes upregulated in CW, MD and MWbs but downregulated in HD and MWh, and had genes relating to metabolism and photosynthesis. The S3 cluster, which had high gene expression in the CW treatment, had an excess of genes relating to carbohydrate metabolism and cell wall biogenesis. Finally, the S11 group had genes highly expressed in CW and MW18 that were related to lipid and secondary metabolic processes.

For pine, the P1 cluster was downregulated in the MWbs and HD treatments and had an over-representation of genes related to stress and stimulus response (Table S4; Fig. 6). The P4 and P8 clusters also contained genes relating to stress and stimulus response and were upregulated in the HD and MWbs treatments, as were genes in the P6 group. The P2 cluster had genes associated with the regulation of transcription and their expression was strongly influenced by all treatments. Genes relating to lipid and carbohydrate metabolism, photosynthesis and response to stimulus were over-represented in the P7 cluster. Finally, similar to the S1, S12, and S7 clusters in spruce, the P3 cluster in pine consisted of genes upregulated in the MWbs treatment that were associated with reproduction and biotic stimulus.

Data access

Unfiltered reads for the 91 libraries are available in the SRA (PRJNA193174; Sequence and Read Archive, NCBI). Fasta files for the reference assemblies of interior spruce and lodgepole pine and the expression count data are archived on Dryad (doi: 10.5061/dryad.1p888) and Treegenes. Blast2GO.dat files for the expressed genes containing the BLAST results, annotations, InterProScan results, and enzyme codes in both species are available in the Dryad archive.

Discussion

Conservation and divergence of plasticity in gene expression

The lineages leading to lodgepole pine and interior spruce diverged over 140 million yr ago (Savard *et al.*, 1994; Wang *et al.*, 2000), yet our results suggest that expression patterns are conserved in almost three-quarters of identified orthologs with climate-induced differential expression. Of all 14 691 identified orthologs, we found many more genes with patterns of environment-induced expression that were conserved in both species (4298 CEGs) than genes with patterns that differed between species (2428 DEGs; Fig. 4). This suggests that selection has favored conservation of the pattern of plastic responses in a large number of genes, despite the long divergence time. However, the 2428 DEGs represent almost 20% of the identified orthologs, indicating that a significant fraction of the genome has evolved plastic expression in response to environment. We also found that half of all identified orthologs differed in their average amounts of expression between the species (7418 orthologs with significant species term; Fig. 4). This suggests that evolution leads to broad changes in gene expression levels more often than plastic changes of expression in response to environment, at least for the climatic treatments and plant organs that we examined. If expression divergence is correlated with sequence divergence (as found by Guan *et al.*, 2013), successful identification of orthologs may be reduced for genes that have undergone rapid sequence divergence, which would cause us to underestimate the number of DEGs. Alternatively, some of the putative cases of diverged expression could be due to ortholog misidentification. It is also worth noting that many of the apparent cases of divergence in expression could be caused by differences in the timing of expression responses in each species, as a result of either developmental program or experimental conditions. This could be especially relevant for differences in expression within the MWbs treatment, as these individuals were sampled 2 months after those in the other treatments.

While we cannot make any direct inferences about whether the evolution of expression divergence was gradual, punctuated, or more rapid in one species than the other, our results are consistent with an average rate of 0.059% of all genes diverging in their pattern of expression plasticity per million years (2428 DEGs/14 691 orthologs/280 million yr). This rate of expression divergence is > 4 times higher than that found in a recent study of tissue-dependent expression in spruce (xylem versus phelloderm), where only 60 out of 5407 differentially expressed genes also had diverged in their patterns of expression (Raherison *et al.*, 2013). We cannot rule out the possibility that the apparently higher rate of expression divergence reported in the present study is driven by a difference in power between the studies. Nonetheless, it is clear that expression divergence occurs slowly in conifers relative to angiosperms (e.g. Jiao *et al.*, 2005; Ma *et al.*, 2005; Walia *et al.*, 2009). It would be interesting to extend this study to other conifers to better evaluate rates of expression divergence across the phylogeny.

Although patterns of gene expression can evolve by selection or drift, it is thought that most expression divergence is neutral rather than adaptive (Tirosh & Barkai, 2011). Here, we found that DEGs were more likely to be differentially expressed in certain treatments (Fig. 5), and to be members of certain co-expression clusters (Table S2). In spruce, we found more upregulation of DEGs in MWbs and more downregulation of DEGs in MWh, with patterns that tended toward the opposite of this in pine (although less strongly and consistently so). In contrast, CEGs were more likely to be differentially expressed between CW and

MWh or HD (Fig. 5). These nonrandom associations suggest that natural selection has played an active role in the divergence of expression in response to the environmental cues associated with budset and heat stress. Unfortunately, it is not possible to use these data to evaluate whether the differences between species arose in response to positive selection or the relaxation of purifying selection. The observed differences between species in response to hot wet environments are consistent with physiological evidence that shows that black spruce (*Picea mariana*) and jack pine (*Pinus banksiana*) have very different mechanisms for coping with heat stress under well-watered conditions (Ewers *et al.*, 2005).

Differential expression among treatments and gene function

We found a substantial number of genes that showed plasticity in their expression profiles among treatments within each species (49% in pine and 26% in spruce). This large-scale remodeling of expression in response to environmental conditions is a common finding of microarray and RNAseq studies (e.g. Kreps *et al.*, 2002; Qin *et al.*, 2008). As expected, many of the genes responding to the treatments are known to be involved in stress responses in other species. We used co-expression network analysis to explore these patterns in more detail and identify which types of genes tended to be involved in response to specific combinations of environmental conditions (Fig. 6). Within species, expression patterns of particular clusters were in some cases driven by a single treatment (e.g. the P3 pine cluster), while in others, co-expression was evident across treatments (e.g. the S6 spruce cluster). Some of the best defined clusters of co-expressed genes were related to the MWbs treatment, which reflects the extensive transcriptional remodeling that is associated with seasonal developmental processes (Ruttink *et al.*, 2007; Holliday *et al.*, 2008; El Kayal *et al.*, 2011). For spruce, the S7 cluster showed distinct upregulation that was restricted to the budset treatment (Fig. 6). A diverse array of genes were found in this cluster, including 30 annotated as transcription factors. Many of the homologs of these genes are related to phytohormone signaling (e.g. comp43628_c0, ethylene response factor), and floral and leaf development in *A. thaliana* (e.g. comp68536_c0, NO APICAL MERISTEM; comp457_c1, LEUNIG; comp13440_c0, LEAFY; comp7287_c1, BAM3; comp76292_c0_seq1, ELF8). The S7 cluster was also over-represented for DEGs (Table S2), suggesting possible divergence in the budset gene expression program between the two species. While such differences may be a result of changes in the timing of budset, they may also be partly explained by the large differences in bud morphology and architecture between pine and spruce. The P3 cluster in pine appeared to be involved in processes similar to S7, as these genes were also upregulated in the budset treatment, with many assigned GO categories relating to development and transcription. For example, 34 genes corresponded to transcription factors in *A. thaliana* and several are involved in meristem identity (e.g. comp56289_c0, UNUSUAL FLORAL ORGANS; comp11703_c0, NO APICAL MERISTEM; comp7928_c0 and comp13032_c0, SHOOT MERISTEMLESS; comp7584_c0 PHABULOSA 1D; comp39095_c0, LEAFY; and comp2425_c0, KNOTTED-LIKE). These clusters (S7 and P3) are probably related to a gene expression program involved in the formation of the leaf primordial within overwintering buds. Out of the 212 orthologs found in S7, 33 were also found in P3, suggesting limited conservation in budset gene expression profiles between the species.

Several clusters in both spruce and pine stood out as a consequence of their coordinated expression patterns for both the heat stress and budset treatments (Fig. 6). This suggests possible overlap in stress response pathways for these treatments or, alternatively, heat-induced cessation of growth (though there was no evidence of budset in the heat treatment). For spruce, the S4, S13 and S15 clusters were strongly upregulated in the HD and MWbs treatments. The larger S14 cluster and the S10 cluster contained many genes upregulated in the HD, MWh and MWbs treatments. In both of these clusters, the GO term response to stress was over-represented, and many of the genes (34 in S14 and 22 in S10) encoded heat shock proteins – so-named because of their expression under heat stress, but which are expressed under a variety of stresses in plants (Sun *et al.*, 2002; Wang *et al.*, 2003; Swindell *et al.*, 2007). The small P4 cluster in pine showed a similar response and function with a significant number of genes related to stress and stimulus response, including 11 putative heat shock proteins. The P7 and the P8 clusters possessed several ethylene-response element binding factors (ERFs), including a homolog to DEHYDRATION-RESPONSIVE ELEMENT BINDING PROTEIN 2 (DREB2, comp67986_c0) found in the P7 treatment. Closer inspection of this particular gene model shows that it was only upregulated in the HD treatment, which is consistent with its role in heat and drought stress in *A. thaliana* (Liu *et al.*, 1998; Schramm *et al.*, 2008). Another gene that stood out in the P7 cluster was PHYTOCHROME A (PHYA, comp8631_c0), which was strongly upregulated in response to both HD and MWbs treatments. This is an enigmatic result as PHYA is a photoreceptor involved in photoperiodic developmental processes (including budset), but may reflect a transition to dormancy in the heat treatment as noted above. For pine, the P2 cluster had a pattern of coordinated heat stress/budset expression.

Future directions

Understanding the genomic basis of adaptation requires the integration of evidence from studies of gene expression, protein and sequence divergence, and population and quantitative trait differentiation. The work on gene expression described here is the first in a series of studies exploring the genomic basis of local adaptation to climate in lodgepole pine and interior spruce. Here, we characterized variation in gene expression within and among species in response to climatic treatments, identifying genes with highly conserved or diverged patterns of differential expression. Work in progress will characterize patterns of sequence and protein divergence in these species and explore correlations with expression divergence to further investigate whether selection has shaped expression patterns. Using exome re-sequencing and

association genetic approaches at the landscape scale, we plan to characterize the genomic basis of local adaptation to climate within each species. This will allow us to identify whether the same genes tend to be involved in plastic responses to climate and adaptive divergence among populations and species. By studying how populations adapt to climatic variations in space, we aim to predict how populations will respond to temporal climatic changes, and help inform forest management practices.

Acknowledgements

The authors would like to thank the Aitken and Rieseberg lab groups for helpful suggestions in the preparation of this manuscript. We thank G. O'Neill of the BC Ministry of Forests, Lands and Natural Resources Operations for providing seedlings, and P. Smets, R. Belvas and C. Fitzpatrick for cultivating seedlings and maintaining treatments. This work is part of the AdapTree Project funded by the Genome Canada Large Scale Applied Research Project program, with co-funding from Genome BC and the Forest Genetics Council of BC (co-Project Leaders S. N. Aitken and A. Hamann).

References

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S. 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications* **1**: 95–111.

Alberto F, Aitken S, Alia R, González-Martínez S, Hanninen H, Kremer A, Lefevre F, Lenormand T, Yeaman S, Whetten R *et al.* 2013. Potential for evolutionary response to climate change – evidence from tree populations. *Global Change Biology* **19**: 1645–1661.

Allen CD, Macalady AK, Chenchouni H, Bachelet D, Mcdowell N, Vennetier M, Kitzberger T, Rigling A, Breshears DD, Hogg EHT *et al.* 2010. A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *Forest Ecology and Management* **259**: 660–684.

Beaulieu J, Giguère I, Deslauriers M, Boyle B, MacKay J. 2013. Differential gene expression patterns in white spruce newly formed tissue on board the International Space Station. *Advances in Space Research* **52**: 760–772.

Birol I, Raymond A, Jackman SD, Pleasance S, Coope R, Taylor GA, Yuen MM, Keeling CI, Brand D, Vandervalk BP *et al.* 2013. Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. *Bioinformatics* **29**: 1492–1497.

Bonan GB. 2008. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *Science* **320**: 1444–1449.

De La Torre AR, Wang T, Jaquish B, Aitken SN. 2014. Adaptation and exogenous selection in a *Picea glauca* × *Picea engelmannii* hybrid zone: implications for forest management under climate change. *New Phytologist* **201**: 687–699.

Eckert AJ, Shahi H, Datwyler SL, Neale DB. 2012. Spatially variable natural selection and the divergence between parapatric subspecies of lodgepole pine (*Pinus contorta*, Pinaceae). *American Journal of Botany* **99**: 1323–1334.

Ekblom R, Galindo J. 2010. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **107**: 1–15.

El Kayal W, Allen CCG, Ju CJT, Adams E, King-Jones S, Zaharia LI, Abrams SR, Cooke JE. 2011. Molecular events of apical bud formation in white spruce, *Picea glauca*. *Plant, Cell & Environment* **34**: 480–500.

Ewers BE, Gower ST, Bond-Lamberty B, Wang CK. 2005. Effects of stand age and tree species on canopy transpiration and average stomatal conductance of boreal forests. *Plant, Cell & Environment* **28**: 660–678.

Franks S, Webber J, Aitken SN. 2013. Evolutionary and plastic responses to climate change in terrestrial plant populations. *Evolutionary Applications* **7**: 123–139.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q *et al.* 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnology* **29**: 644–652.

Guan Y, Dunham MJ, Troyanskaya OG, Caudy AA. 2013. Comparative gene expression between two yeast species. *BMC Genomics* **14**: 33.

Hamann A, Aitken SN. 2013. Conservation planning under climate change: accounting for adaptive potential and migration capacity in species distribution models. *Diversity and Distributions* **19**: 268–280.

Hodgins KA, Lai Z, Oliveira LO, Still DW, Scascitelli M, Barker MS, Kane NC, Dempewolf H, Kozik A, Kesseli RV *et al.* 2014. Genomics of Compositae crops: reference transcriptome assemblies and evidence of hybridization with wild relatives. *Molecular Ecology Resources* **14**: 166–177.

Hogg EH, Brandt JP, Michaelian M. 2008. Impacts of a regional drought on the productivity, dieback, and biomass of western Canadian aspen forests. *Canadian Journal of Forest Research* **38**: 1373–1384.

Holliday JA, Ralph SG, White R, Bohlmann J, Aitken SN. 2008. Global monitoring of autumn gene expression within and among phenotypically divergent populations of Sitka spruce (*Picea sitchensis*). *New Phytologist* **178**: 103–122.

IPCC. 2007. *Climate change 2007: the physical science basis. Contribution of Working Group I to the fourth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge, UK and New York, NY, USA: Cambridge University Press.

Jiao Y, Ma L, Strickland E, Deng XW. 2005. Conservation and divergence of light-regulated genome expression patterns during seedling development in rice and Arabidopsis. *Plant Cell* **17**: 3239–3256.

Joosen RVL, Lammers M, Balk PA, Bronnum P, Konings M, Perks M, Stattin E, van Wordragen MF, van der Geest AL. 2006. Correlating gene expression to physiological parameters and environmental conditions during cold acclimation of *Pinus sylvestris*, identification of molecular markers using cDNA microarrays. *Tree Physiology* **26**: 1297–1313.

Kolosova N, Miller B, Ralph S, Ellis BE, Douglas C, Ritland K, Bohlmann J. 2004. Isolation of high-quality RNA from gymnosperm and angiosperm trees. *BioTechniques* **36**: 821–824.

Kreps JA, Wu Y, Chang H, Zhu T, Wang X, Harper JF, Mesa T, Row M, Diego S, California JAK *et al.* 2002. Transcriptome changes for Arabidopsis in response to salt, osmotic, and cold stress. *Plant Physiology* **130**: 2129–2141.

Lai Z, Kane NC, Kozik A, Hodgins KA, Dlugosch KM, Barker MS, Matvienko M, Yu Q, Turner KG, Pearl SA *et al.* 2012. Genomics of Compositae weeds: EST libraries, microarrays, and evidence of introgression. *American Journal of Botany* **99**: 1–10.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**: 559.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**: 323.

Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research* **13**: 2178–2189.

Liu Q, Kasuga M, Sakuma Y, Abe H, Miura S, Yamaguchi-Shinozaki K, Shinozaki K. 1998. Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in Arabidopsis. *Plant Cell* **10**: 1391–1406.

Ma L, Chen C, Liu X, Jiao Y, Su N, Li L, Wang X, Cao M, Sun N, Zhang X *et al.* 2005. A microarray analysis of the rice transcriptome and its comparison to *Arabidopsis*. *Genome Research* **15**: 1274–1283.

Mbogga MS, Hamann A, Wang T. 2009. Historical and projected climate data for natural resource management in western Canada. *Agricultural and Forest Meteorology* **149**: 881–890.

Movahedi S, Van Bel M, Heyndrickx KS, Vandepoele K. 2012. Comparative co-expression analysis in plant biology. *Plant, Cell & Environment* **35**: 1787–1798.

Namroud M-C, Beaulieu J, Juge N, Laroche J, Bousquet J. 2008. Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Molecular Ecology* **17**: 3599–3613.

Parchman TL, Gompert Z, Mudge J, Schilkey FD, Benkman CW, Buerkle CA. 2012. Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology* **21**: 2991–3005.

Porth I, White R, Jaquish B, Alfaro R, Ritland C, Ritland K. 2012. Genetical genomics identifies the genetic architecture for growth and weevil resistance in spruce. *PLoS One* **7**: e44397.

Qin D, Wu H, Peng H, Yao Y, Ni Z, Li Z, Zhou C, Sun Q. 2008. Heat stress-responsive transcriptome analysis in heat susceptible and tolerant wheat (*Triticum aestivum* L.) by using Wheat Genome Array. *BMC Genomics* **9**: 432.

Raherison E, Rigault P, Caron S, Poulin P-L, Boyle B, Verta J-P, Giguère I, Bomal C, Bohlmann J, MacKay J. 2013. Transcriptome profiling in conifers and the PiceaGenExpress database show patterns of diversification within gene families and interspecific conservation in vascular gene expression. *BMC Genomics* **13**: 434.

Rehfeldt GE, Wykoff WR, Ying CC. 2001. Physiologic plasticity, evolution, and impacts of a changing climate on *Pinus contorta*. *Climatic Change* **50**: 355–376.

Rehfeldt GE, Ying CC, Spittlehouse DL, Hamilton DA Jr. 1999. Genetic responses to climate in *Pinus contorta*: niche breadth, climate change, and reforestation. *Ecological Monographs* **59**: 375–407.

Reid KE, Holliday JA, Yuen M, Nguyen A, Aitken SN, Bohlmann J. 2013. Sequencing of Sitka spruce (*Picea sitchensis*) cDNA libraries constructed from autumn buds and foliage reveals autumn-specific spruce transcripts. *Tree Genetics & Genomes* **9**: 683–691.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

Roche L. 1969. A genecological study of the genus *Picea* in British Columbia. *New Phytologist* **68**: 505–554.

Ruttink T, Arend M, Morreel K, Storme V, Rombauts S *et al.* 2007. A molecular timetable for apical bud formation and dormancy induction in poplar. *Plant Cell* **19**: 2370–2390.

Rweyongeza DM, Yang R-C, Dhir NK, Barnhardt LK, Hansen C. 2007. Genetic variation and climatic impacts on survival and growth of White spruce in Alberta, Canada. *Silvae Genetica* **56**: 3–4.

Savard L, Li P, Strauss SH, Chase MW, Michaud M, Bousquet J. 1994. Chloroplast and nuclear gene sequences indicate Late Pennsylvanian time for the last common ancestor of extant seed plants. *Proceedings of the National Academy of Sciences, USA* **91**: 5163–5167.

Schramm F, Larkindale J, Kiehlmann E, Ganguli A, Englich G, Vierling E, von Koskull-Doring P. 2008. A cascade of transcription factor DREB2A and heat stress transcription factor HsfA3 regulates the heat stress response of Arabidopsis. *Plant Journal* **53**: 264–274.

Snell-Rood EC, Van Dyken JD, Cruickshank T, Wade MJ, Moczek AP. 2010. Toward a population genetic framework of developmental evolution: the costs, limits, and consequences of phenotypic plasticity. *BioEssays* **32**: 71–81.

Sork VL, Aitken SN, Dyer RJ, Eckert AJ, Legendre P, Neal DB. 2013. Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes* **9**: 901–911.

Sun WN, Van Montagu M, Verbruggen N. 2002. Small heat shock proteins and stress tolerance in plants. *Biochimica Et Biophysica Acta* **1577**: 1–9.

Swindell WR, Huebner M, Weber AP. 2007. Transcriptional profiling of Arabidopsis heat shock proteins and transcription factors reveals extensive overlap between heat and non-heat stress response pathways. *BMC Genomics* **8**: 125.

Tirosh I, Barkai N. 2011. Inferring regulatory mechanisms from patterns of evolutionary divergence. *Molecular Systems Biology* **7**: 530.

Ukrainetz NK, O'Neill GA, Jaquish B. 2011. Comparison of fixed and focal point seed transfer systems for reforestation and assisted migration: a case study for interior spruce in British Columbia. *Canadian Journal of Forestry Research* **41**: 1452–1464.

Walia H, Wilson C, Ismail AM, Close TJ, Cui X. 2009. Comparing genomic expression patterns across plant species reveals highly diverged transcriptional dynamics in response to salt stress. *BMC Genomics* **10**: 398.

Wall PK, Leebens-Mack J, Chanderbali AS, Barakat A, Wolcott E, Liang H, Landherr L, Tomsho LP, Hu Y, Carlson JE *et al.* 2009. Comparison of next generation sequencing technologies for transcriptome characterization. *BMC Genomics* **10**: 347.

Wang T, Campbell EM, O'Neill GA, Aitken SN. 2012. Projecting future distributions of ecosystem climate niches: uncertainties and management applications. *Forest Ecology and Management* **279**: 128–140.

Wang T, Hamann A, Yanchuk A, O'Neill GA, Aitken SN. 2006. Use of response functions in selecting lodgepole pine populations for future climate. *Global Change Biology* **12**: 2404–2416.

Wang WX, Vinocur B, Altman A. 2003. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta* **218**: 1–14.

Wang X-Q, Tank DC, Sang T. 2000. Phylogeny and divergence times in Pinaceae: evidence from three genomes. *Molecular Biology and Evolution* **17**: 773–781.

Wegrzyn JL, Liechty JD, Stevens KA, Wu LS, Loopstra CA, Vasquez-Gross H, Dougherty WM, Lin BY, Zieve JJ, Martinez-Garcia PJ *et al.* 2014. Unique features of the lobololly pine (*Pinus taeda* L.) megagenome revealed through sequence annotation. *Genetics* **196**: 891–909.

Woods A, Coates KD, Hamann A. 2005. Is an unprecedented Dothistroma needle blight epidemic related to climate change? *BioScience* **55**: 761.

Wu TD, Watanabe CK. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**: 1859–1875.

Wu HX, Ying CC. 2004. Geographic pattern of local optimality in natural populations of lodgepole pine. *Forest Ecology and Management* **194**: 177–198.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 The number and taxonomic origin of contaminating sequences of the interior spruce and lodgepole pine reference assemblies based on BLASTX to the NR protein database.

Fig. S2 Histograms showing the number of contigs of different sizes yielded by the assemblies built with reads combined from *x* individuals from different treatments versus from only the MD treatment.

Table S1 The number of individuals grown under the experimental conditions and included in the expression analysis for each treatment in pine and spruce

Table S2 Number of conserved expression genes (CEGs) and diverged expression genes (DEGs) in co-expression clusters identified by WGCNA in spruce and pine, along with the expected counts based on a χ^2 contingency test

Table S3 The GO slim terms over-represented in spruce for differentially expressed genes among treatments, and for clustered genes

Table S4 The GO slim terms over-represented in pine for differentially expressed genes among treatments, and for clustered genes

Methods S1 Further documentation of methods, including sections on 'Experimental design and sampling', 'Reference assembly', 'Annotation', and 'Assessment of assembly strategy'.

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.