BMC
Genomics

Hsf and *Hsp* gene families in *Populus*: genome-wide identification, organization and correlated expression during development and in stress responses

Jin Zhang[1,2], Bobin Liu[1,3], Jianbo Li[1], Li Zhang[1], Yan Wang[4], Huanquan Zheng[5], Mengzhu Lu[1,2*] and Jun Chen[1*]

Abstract

Background: Heat shock proteins (Hsps) are molecular chaperones that are involved in many normal cellular processes and stress responses, and heat shock factors (Hsfs) are the transcriptional activators of *Hsps*. *Hsfs* and *Hsps* are widely coordinated in various biological processes. Although the roles of *Hsfs* and *Hsps* in stress responses have been well characterized in *Arabidopsis*, their roles in perennial woody species undergoing various environmental stresses remain unclear.

Results: Here, a comprehensive identification and analysis of *Hsf* and *Hsp* families in poplars is presented. In *Populus trichocarpa*, we identified 42 paralogous pairs, 66.7% resulting from a whole genome duplication. The gene structure and motif composition are relatively conserved in each subfamily. Microarray and quantitative real-time RT-PCR analyses showed that most of the *Populus Hsf* and *Hsp* genes are differentially expressed upon exposure to various stresses. A coexpression network between *Populus Hsf* and *Hsp* genes was generated based on their expression. Coordinated relationships were validated by transient overexpression and subsequent qPCR analyses.

Conclusions: The comprehensive analysis indicates that different sets of *PtHsps* are downstream of particular *PtHsfs* and provides a basis for functional studies aimed at revealing the roles of these families in poplar development and stress responses.

Keywords: Coexpression, Expression analysis, Gene family, Heat shock factor (Hsf), Heat shock protein (Hsp), *Populus*

Background

During their growth, plants are subjected not only to abiotic stresses, such as irradiation, temperature, salinity, and drought, but also biotic stresses, such as herbivore and pathogen attacks. These stress factors can simultaneously act on the plants causing cell injury and producing secondary stresses [1,2]. As sessile organisms, plants cannot move to avoid these stresses, and thus have developed mechanisms, such as morphological adaptation, to tolerate these stresses [3].

Along with other stresses, heat stress can trigger the expression of certain genes that were not expressed under

"normal" conditions [4-6]. Heat shock proteins (Hsps) accumulate when the expression of their genes is triggered by heat, as well as other stresses [7-9]. Hsps are molecular chaperones that regulate the folding, localization, accumulation, and degradation of protein molecules in both plant and animal species [10]. The expression of *Hsps* is controlled and regulated by specific types of transcription factors called heat shock factors (Hsfs), which normally exist as inactive proteins [11].

Plant Hsps are classified into five families based on their approximate molecular weights: small Hsp (sHsp), Hsp60, Hsp70, Hsp90, and Hsp100. Genes encoding Hsfs and Hsps have been well characterized in some model plants, such as *Arabidopsis* and rice [12,13]. In *Arabidopsis*, 21, 27, 18, 18, 7, and 4 genes have been identified as *Hsf*, *sHsp*,

* Correspondence: lumz@caf.ac.cn; chenjun@caf.ac.cn
[1]State Key Laboratory of Tree Genetics and Breeding, Key Laboratory of Tree Breeding and Cultivation of the State Forestry Administration, Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China
Full list of author information is available at the end of the article

Zhang *et al. BMC Genomics* (2015) 16:181

Page 2 of 19

Hsp60, *Hsp70*, *Hsp90*, and *Hsp100* family members, respectively [13-19].

To date, various stress responses and functions of Hsf and Hsp members have been reported. In *Arabidopsis*, HsfA1a, HsfA1b, and HsfA1d act as the main positive regulators of heat shock response [20]. *Arabidopsis* HsfA2 was significantly increased under high light and heat stress or by H_2O_2 treatment. It is the key regulator in the induction of the defense system under several types of environmental stresses [21,22]. The cytosolic class I *sHsp* in *Rosa chinensis*, *RcHsp17.8*, was induced by heat, cold, salt, drought, osmotic, and oxidative stresses [23]. A plastid nucleoid-localized sHsp (Hsp21) interacts with the plastid nucleoid protein pTAC5 and is essential for chloroplast development in *Arabidopsis* under heat stress [24]. Hsp60 proteins as chaperones participate in the folding and aggregation of many proteins that are transported to organelles, such as chloroplasts and mitochondria [25,26]. *Arabidopsis hsp60* mutants showed defects in chloroplast, embryo, and seedling development and also increased cell death [27,28]. Unlike other family members that are mainly expressed when the organism is subjected to environmental assaults, Hsp70s play essential functions in facilitating refolding and proteolytic degradation of abnormal proteins under both normal and stress conditions [29,30]. Hsp100 proteins belong to the Caseinolytic protease (Clp) family, which forms an ATP-dependent protease complex that is able to hydrolyze casein [19,31-33]. Tomatoes that have an antisense suppression of *Hsp100/ClpB* are more heat sensitive [34]. Moreover, Lee *et al.* [19] noted that *Arabidopsis* mutant plants containing a chloroplast-localized *Hsp100/ClpB-p* knockout turned yellow after heat treatment. Hsps were not only transcriptionally regulated by Hsfs, but could also regulate the activity of Hsfs through a feedback loop caused by their physical interaction. In tomato, Hsp70 and Hsp90 regulate the Hsfs' function by directly interacting with Hsfs. Hsp70 represses the activity of HsfA1 and HsfB1, while the DNA binding activity of HsfB1 is stimulated by Hsp90 in tomato [35]. Above all, Hsfs and Hsps play crucial roles in plant development and various stress tolerances.

As perennial species, poplars undergo seasonal variations and various environmental stresses frequently. The completion of the *Populus trichocarpa* genome sequencing project in 2006 makes it an ideal genetic model for studying tree development and physiology [36]. To date, 28 *Populus* Hsfs have been reported based on the V2.2 *P. trichocarpa* genome database [11]. In our previous study, we have reported the subcellular localization and expression, under various abiotic stresses, of 10 *Hsp90* genes identified from the V3.0 *P. trichocarpa* genome database [37]. However, the other Hsp families in poplar remain

unclear, and little is known about the transcriptional regulatory relationships between *Populus* Hsfs and Hsps. Here, we provide a comprehensive analysis of the gene organization and expression of *Populus* Hsfs and other *Hsp* genes, including *sHsp*, *Hsp60*, *Hsp70* and *Hsp100*, under different abiotic stresses. A complex transcriptional regulatory network between *Populus* Hsfs and Hsps has been generated based on their transcription patterns in poplar.

Results
Identification and phylogenetic analysis of the *Hsf* and *Hsp* gene families in poplar

After automated database searching and a manual review, 118 genes were identified as members of the *Hsf* and *Hsp* families, including *sHsp*, *Hsp60*, *Hsp70*, and *Hsp100*, of *P. trichocarpa*. The *Hsf* and *Hsp* gene families, *PtHsf* and *PtHsp*, respectively, in poplar were relatively large compared with those in *Arabidopsis* and rice. The numbers of identified genes in the *Hsf*, *sHsp*, *Hsp60*, *Hsp70*, and *Hsp100* families of *P. trichocarpa* were 28, 37, 28, 20, and 5, respectively (Table 1). The subcellular localization predictions suggested that the PtHsfs are targeted to the nucleus, while PtHsps are localized to various cytosolic organelles. Detailed information on the *Hsf* and *Hsp* genes in *P. trichocarpa*, including their gene IDs and the characteristics of their encoded proteins, are listed in Additional file 1: Table S1 and Additional file 2: Table S2.

To evaluate the evolutionary relationship of the Hsf and Hsp proteins, a phylogenetic analysis of each family was performed based on the full-length amino acid sequences from both *P. trichocarpa* and *Arabidopsis* (Figures 1 and 2, left panel). Each family could be classified into different subfamilies. The *PtHsf* family contains three subfamilies: type A (17 genes), type B (10 genes), and type C (1 gene). However the subfamilies in each of the *PtHsp* families could be assigned based on the proteins' predicted subcellular localization. The *sHsp* family was classified into cytosolic, endoplasmic reticulum (ER), peroxisome (PX), chloroplast (CP), and mitochondrial (MT) subfamilies in *P. trichocarpa*. There are six groups of cytosolic *sHsp* genes, C-I, C-II, C-III, C-IV, C-V, and C-VI, and two

Table 1 Numbers of *Hsf* and *Hsp* genes in *Arabidopsis*, *Populus* and rice

	A. thaliana	O. sativa	P. trichocarpa
Hsf	21	25	28
sHsp	27	29	37
Hsp60	18	20	28
Hsp70	18	26	20
Hsp90	7	9	10
Hsp100	4	5	5

Zhang et al. BMC Genomics (2015) 16:181

Page 3 of 19
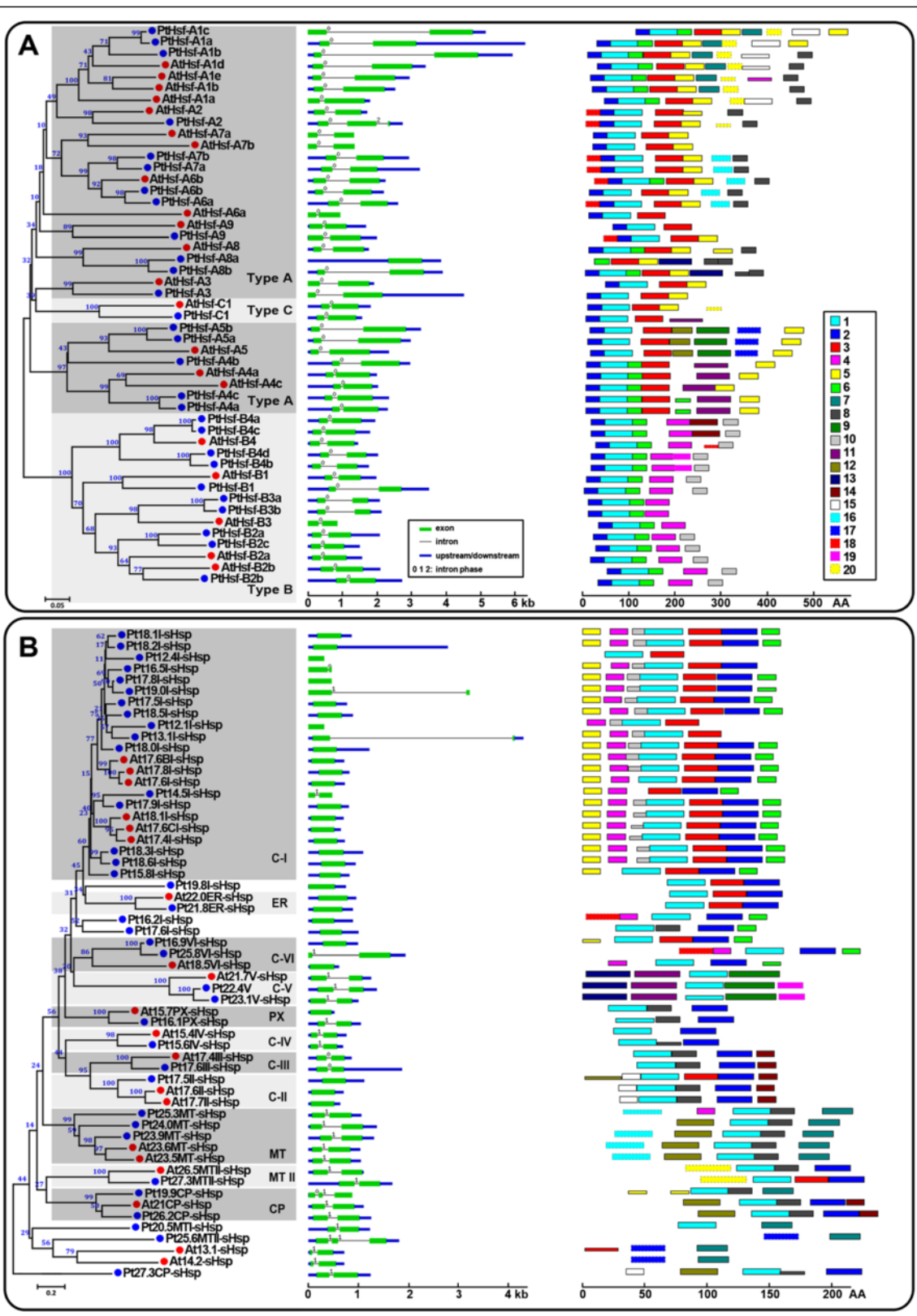
Figure 1 (See legend on next page.)

Zhang *et al. BMC Genomics* (2015) 16:181

Page 4 of 19

Figure 1 Phylogenetic relationships, gene structures and motif compositions of *Hsf* and *sHsp* family members in *A. thaliana* (At) and *P. trichocarpa* (Pt). Multiple alignment of Hsf **(A)** and sHsp **(B)** proteins from *A. thaliana* (At) and *P. trichocarpa* (Pt) was performed using MEGA 5.0 by the Neighbor-Joining (NJ) method with 1000 bootstrap replicates (left panel). Exon/intron structures of the *Hsf* and *sHsp* genes are shown in the middle panel. Green boxes represent exons and black lines represent introns. The numbers indicate the splicing phases of the *Hsf* and *sHsp* genes: 0, phase 0; 1, phase 1; and 2, phase 2. A schematic representation of conserved motifs (obtained using MEME) in Hsf and sHsp proteins is displayed in the right panel. Different motifs are represented by different colored boxes. Details of the individual motifs are in the Additional file 4: Table S4 and Additional file 5: Table S5.

groups of mitochondrial *sHsp* genes, MT and MT II, in *P. trichocarpa*. Notably, the C-I *sHsp* group in the genome of *P. trichocarpa* is large, containing 19 genes compared with 6 in *Arabidopsis* (Figure 1). The *Hsp60* family was divided into four subfamilies in *P. trichocarpa*: cytosol-localized *Cpn60* (18 genes), mitochondrion-localized *Hsp60* (3 genes), and chloroplast-localized *Cpn60-a* (4 genes) and *Cpn60-b* (3 genes). The *Hsp70* family contains genes encoding 10 cytosolic Hsp70s, 4 binding proteins (*BIPs*, *Hsp70* homologs in the ER), 2 plastid Hsp70s (cpHsc70s), 2 mitochondrial Hsp70s (mtHsc70s), and 2 truncated Hsp70s (Hsp70ts). The *Hsp100* family can be divided into three classes in *P. trichocarpa*, cytoplasmic (Cyt, 2 genes), chloroplastic (CP, 2 genes), and mitochondrial (MT, 1 gene) (Figure 2).

Structure of *Hsf* and *Hsp* genes and conserved motifs of Hsf and Hsp proteins in poplar

To gain further insights into the structural diversity of *Hsf* and *Hsp* genes in *P. trichocarpa*, we compared the exon/intron organization in the coding sequences between individual *Hsf* and *Hsp* genes of *P. trichocarpa* and *Arabidopsis* (Figures 1 and 2, middle panel). Most closely related members in the same *Hsf* or *Hsp* subfamilies shared similar intron numbers or exon length. In the *PtHsp70* family, cytosolic *Hsp70s* have zero or one intron, ER-localized *BIPs* have six introns, mitochondrion-localized *mtHsc70s* have five introns and chloroplast-localized *cpHsc70s* have seven introns, while truncated *Hsp70ts* have no introns (Figure 2B). Interestingly, two cytosol-localized *Hsp60* members, *PtCpn60-8.1* and *PtCpn60-8.2*, have no introns in their coding regions while the other *P. trichocarpa Hsp60s* contain several introns (8–16) (Figure 2A). We then compared the intron phases with respect to the codons. The intron phases were remarkably well conserved within the same subfamilies in the detected *PtHsf* and *PtHsp* families.

We further detected the exon/intron structure of 42 paralogous pairs of the *PtHsf* and *PtHsp* genes (10, 9, 12, 9, and 2 pairs in *P. trichocarpa Hsf*, *sHsp*, *Hsp60*, *Hsp70*, and *Hsp100* gene families, respectively, Additional file 3: Table S3). Although 36 paralogous pairs showed conserved intron numbers and gene lengths, one *Hsf* gene pair (*PtHsfA8a/PtHsfA8b*), four *sHsp* pairs (*Pt17.8I-*

sHsp/Pt19.0I-sHsp, *Pt12.2I-sHsp/Pt13.1I-sHsp*, *Pt14.5I-sHsp/Pt17.9I-sHsp*, and *Pt16.9VI-sHsp/Pt25.8VI-sHspI*), and one *Hsp60* pair (*PtCpn60-a3/PtCpn60-a4*) exhibited certain degrees of variation (Figures 1 and 2).

We then used the Multiple Expectation Maximization for Motif Elicitation (MEME) [38] to predict the conserved motifs shared among the related proteins within these families. In each family, 20 putative motifs were identified. The details of these motifs are listed in Additional file 4: Tables S4; Additional file 5: Tables S5; Additional file 6: Tables S6; Additional file 7: Tables S7 and Additional file 8: Tables S8. Most of the closely related members in the phylogenetic tree shared common motif compositions with each other (Figures 1 and 2, right panel).

Chromosomal location and duplication of *Hsf* and *Hsp* genes in poplar

To investigate the expansion of *Hsf* and *Hsp* genes in *P. trichocarpa*, the identified *PtHsf* and *PtHsp* genes were plotted on the chromosomes. Of 118 *PtHsf* and *PtHsp* genes, 114 were distributed on 19 chromosomes, while only four genes (*PtHsf-C1*, *Pt16.2I-sHsp*, *PtCpn60-2.2*, and *PtCpn60-4.2*) localized to unassembled genomic sequence scaffolds (Figure 3). The distributions of *Hsf* and *Hsp* genes among the chromosomes appeared to be uneven in *P. trichocarpa*: chromosome (chr) V, VII, XVI, XVIII, and XIX contain only one or two *Hsf* and *Hsp* genes, while relatively high densities of *Hsf* and *Hsp* genes were discovered on chr I, III, VI, VIII, IX, and X. In particular, *Hsfs* and *Hsps* were clustered on the duplicated fragments of chr VIII and X in *P. trichocarpa*.

The genome of *P. trichocarpa* has experienced at least two whole genome duplication (WGD) events, followed by a series of chromosomal reorganizations involving reciprocal tandem/terminal fusions and translocations [36]. Approximately 84.7% (100 of 118) *Hsf* and *Hsp* genes were located in the replicated region, while 44 genes lacked copies on the corresponding duplicated blocks. A chromosome region containing two or more genes within 200 kb can be defined as a gene cluster [39]. The gene clusters were distributed unevenly among *Hsf* and *Hsp* gene families. The *sHsp*, *Hsp60*, *Hsp70*, and *Hsp100* families contain five, one, three, and one clusters, respectively (Figure 3 and Additional file 3: Table S3). The smallest tandem duplication clusters consisted of only two
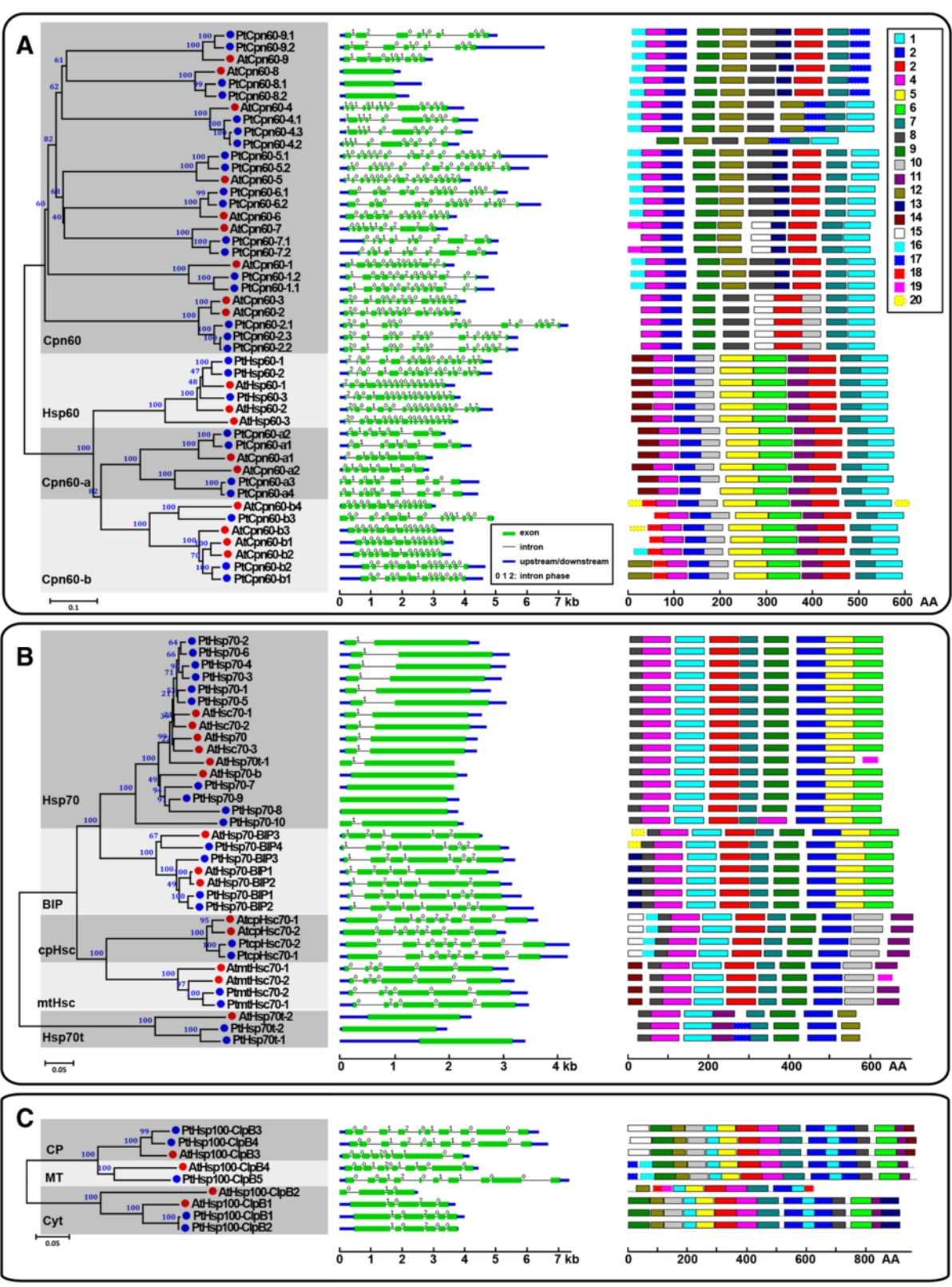
Zhang *et al. BMC Genomics* (2015) 16:181

Page 5 of 19

Figure 2 (See legend on next page.)

Zhang et al. BMC Genomics (2015) 16:181

Page 6 of 19

(See figure on previous page.)

Figure 2 Phylogenetic relationships, gene structures and motif compositions of *Hsp60*, *Hsp70* and *Hsp100* family members in *A. thaliana* (At) and *P. trichocarpa* (Pt). Multiple alignment of Hsp60 **(A)**, Hsp70 **(B)** and Hsp100 **(C)** proteins from *A. thaliana* (At) and *P. trichocarpa* (Pt) was performed using MEGA 5.0 by the Neighbor-Joining (NJ) method with 1000 bootstrap replicates (left panel). Exon/intron structures of the *Hsp60* **(A)**, *Hsp70* **(B)** and *Hsp100* **(C)** genes are shown in the middle panel. Green boxes represent exons and black lines represent introns. The numbers indicate the splicing phases of the *Hsp60* **(A)**, *Hsp70* **(B)** and *Hsp100* **(C)** genes: 0, phase 0; 1, phase 1; and 2, phase 2. A schematic representation of conserved motifs (obtained using MEME) in Hsp60 **(A)**, Hsp70 **(B)** and Hsp100 **(C)** proteins is displayed in the right panel. Different motifs are represented by different colored boxes. Details of the individual motifs are in the Additional file 6: Table S6, Additional file 7: Table S7 and Additional file 8: Table S8.
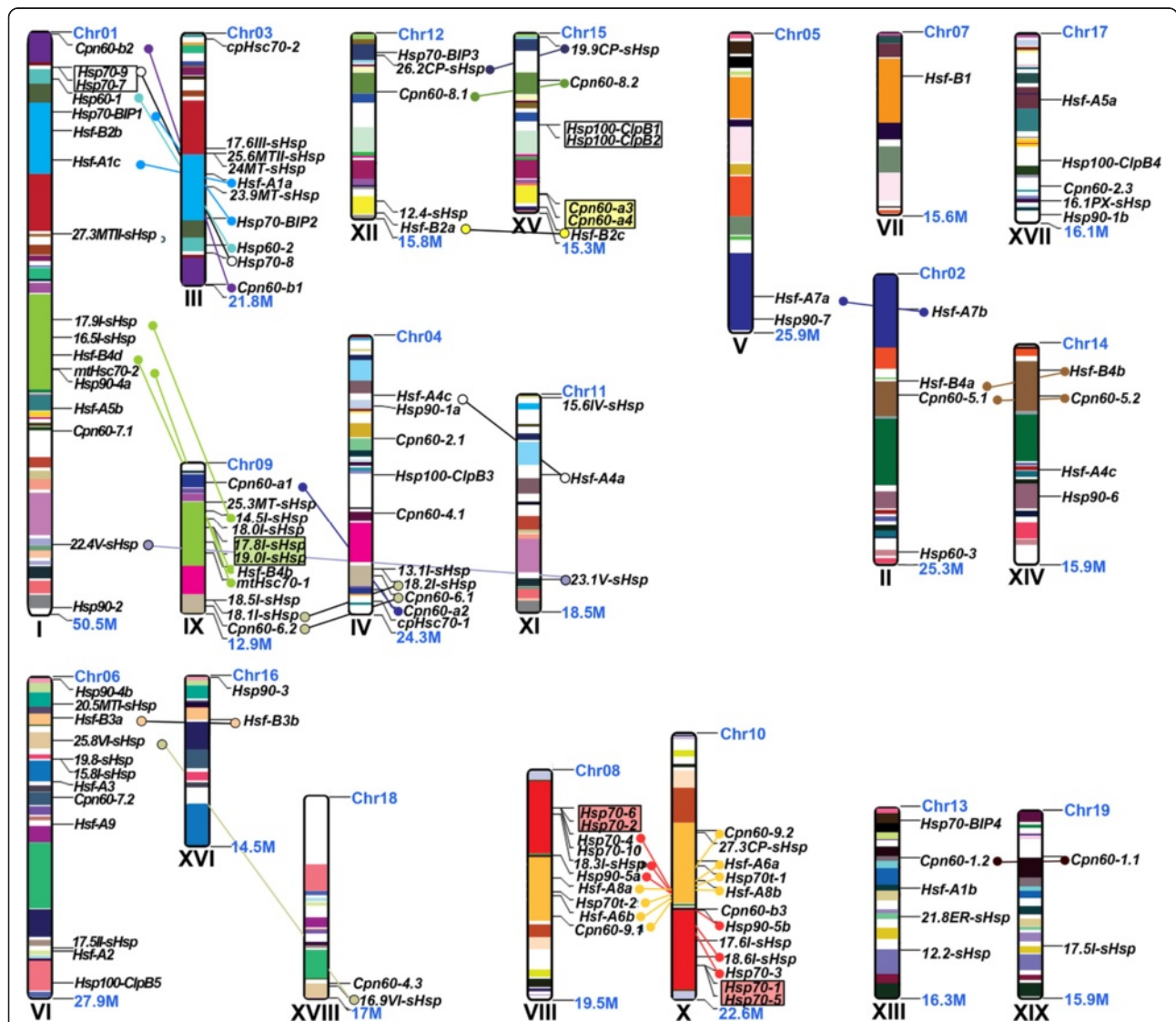

Figure 3 Chromosomal locations of *Populus* Hsfs and Hsps. 114 of 118 *Hsf* and *Hsp* genes are mapped to 19 chromosomes. The schematic diagram of *Populus* genome-wide chromosome organization arisen from the whole genome duplication event was adapted from Tuskan *et al.* [36] and Chai *et al.* [54]. Homologous blocks derived from segmental duplication are indicated using the same colors. Small circles connected by colored line indicate corresponding sister gene pairs.

Zhang *et al. BMC Genomics* (2015) 16:181

Page 7 of 19

genes and the largest cluster, in the *Hsp70* family, had five genes. Interestingly, none of the *Hsf* genes were represented in tandem clusters.

Among the 42 paralogous pairs of the *Populus Hsf* and *Hsp* families, 66.7% gene pairs (28 of 42 pairs) were generated by whole genome duplication and 19% (8 of 42 pairs) by tandem duplication (Additional file 3: Table S3). To verify whether Darwinian positive selection was involved in the *Hsf* and *Hsp* gene divergence after duplication, the nonsynonymous (Ka) versus synonymous (Ks) substitution rate ratios were calculated for the 42 paralogous pairs [40]. A Ka/Ks ratio significantly lower than 0.5 suggests a purifying selection for both duplicates [41]. The summary of Ka/Ks for the 42 *Hsf* and *Hsp* gene paralogous pairs is shown in Additional file 3: Table S3.

Expression patterns of *Populus Hsf* and *Hsp* genes

Publicly available Expressed Sequence Tags (ESTs) provide a useful tool to survey gene expression profiles using a digital northern blot [42]. We conducted a preliminary expression analysis of *Hsf* and *Hsp* genes by counting the frequencies of ESTs obtained from different tissues and under various growth conditions in different *Populus* cDNA libraries (Additional file 9: Figure S1). A complete search of the digital expression profiles from PopGenIE (http://popgenie.org/) [43] yielded 77 *Populus Hsf* and *Hsp* genes in the cDNA libraries. The frequencies of these ESTs were relatively low, and most *Hsf* and *Hsp* genes were represented by only one EST in the cDNA libraries. Nevertheless, these expression profiles suggested that most of the *PtHsf* and *PtHsp* genes had broad expression patterns across different tissues.

We then investigated the global expression profiles of *Hsf* and *Hsp* genes by examining previously published microarray data in poplars. At first, Affymetrix (GSE13990) [44] and a Nimblegen (GSE13043) [45] microarray data from Gene Expression Omnibus [46] were used to analyze the expression patterns of *Hsf* and *Hsp* genes in different tissues. Most *Populus Hsf* and *Hsp* genes were detected in the two different platforms. The majority of *Hsf* and *Hsp* genes showed a tissue-specific expression pattern (Figure 4). Four *sHsps* (*Pt16.2I-sHsp*, *Pt18.3I-sHsp*, *Pt18.5I-sHsp*, and *Pt21.8ER-sHsp*) and two *Hsp60s* (*PtCpn60-5.1* and *PtCpn60-7.2*) had high transcript levels in the differentiating xylem. Two *Hsfs* (*PtHsf-A3* and *PtHsf-B3a*), one *sHsp* (*Pt16.5I-sHsp*) and one *Hsp70* (*PtHsp70-BIP3*) were preferentially expressed in male and female catkins, but almost all of the *Hsp60s* had low transcript levels in catkins (Figure 4). During stem development, 9 *Hsfs,* 13 *sHsps*, and 2 *Hsp100s* had high levels in the basal stem undergoing the secondary growth (internode 9). In comparison, most *Hsp60* and *Hsp70* genes showed high expression levels in the upper stem (internode 2 and 3) (Figure 4).

To explore the possible roles of *Populus Hsf* and *Hsp* genes in response to various abiotic stresses, we then analyzed their expression patterns under heat, drought, low nitrogen level, mechanical wounding, and methyl jasmonate (MeJ) treatment. Four *Hsf* genes (*PtHsf-A2*, *PtHsf-A6b PtHsf-B2c*, and *PtHsp-C1*), three *sHsps* (*Pt18.0I-sHsp*, *Pt19.8I-sHsp*, and *Pt23.9MT-sHsp*), one *Hsp70* (*PtHsp70-5*), and one *Hsp100* (*PtHsp100-ClpB2*) were up-regulated under nitrogen deprivation in both genotypes 1979 and 3200. Mechanical wounding caused the up-regulation of four *Populus Hsfs* in expanding leaves at 90 h after wounding, followed by a down-regulation at 1 week in young leaves and expanding leaves. In cell culture, the addition of MeJ led to the down-regulation of most *sHsp* and *Hsp70* genes (Figure 5). Notably, C-I *sHsp* genes were significantly up-regulated under drought stress in the two genotypes, while the other *Hsf* and *Hsp* genes were not significantly changed (Figure 5).

In a previous study, the physiological conditions of poplars at temperatures between 22°C and 42°C were divided into four states based on photosynthetic activity: the baseline (22°C, the growth temperature), optimum (31.75°C, temperature producing the maximum net CO_2 assimilation rate), 20% inhibition of optimum (38.4°C), and 30% inhibition of optimum (40.5°C) [47]. Most *sHsp* and *Hsp70* genes were induced when the temperature was increased to the optimum, and then continued to be induced when the photosynthesis was inhibited under heat stress, as shown using dataset GSE26199 in Figure 6 [47]. Because of the restrictions of the microarray probe sets, many *Hsf* and *Hsp* genes were not detected in GSE26199. To gather more information on the *Populus Hsf* and *Hsp* genes' expression profiles under high temperatures, RNA-seq data from a hybrid poplar (*P. alba* × *P. glandulosa*) stem (internode 5) under the heat treatments (our unpublished data, Additional file 10: Table S9) were used to analyze the expression of *Populus Hsf* and *Hsp* genes. In the hybrid poplar, nine *Hsfs*, seven *Hsp60s*, and almost all of the *sHsps*, *Hsp70s*, and *Hsp100s* were induced 2 h after treated at 37°C (Figure 6). After a 2-h recovery period following the 37°C treatment, the expression levels of the induced *Hsfs*, *Hsp60s*, *Hsp70s*, and *Hsp100s* decreased, but not those of the *sHsps* (Figure 6).

Confirmation of *Populus Hsf* and *Hsp* gene expression levels by qRT-PCR

To confirm the expression profiles of *Populus Hsf* and *Hsp* genes obtained from the microarray and RNA-seq data, a qRT-PCR analysis of selected *Hsf* and *Hsp* genes (seven *Hsf*, three *sHsp*, three *Hsp60*, three *Hsp70*, and three *Hsp100*) was performed on five different tissues (YL - young leaves, ML - mature leaves, PS - primary stem, SS - secondary stem, and R - root) of hybrid poplar.
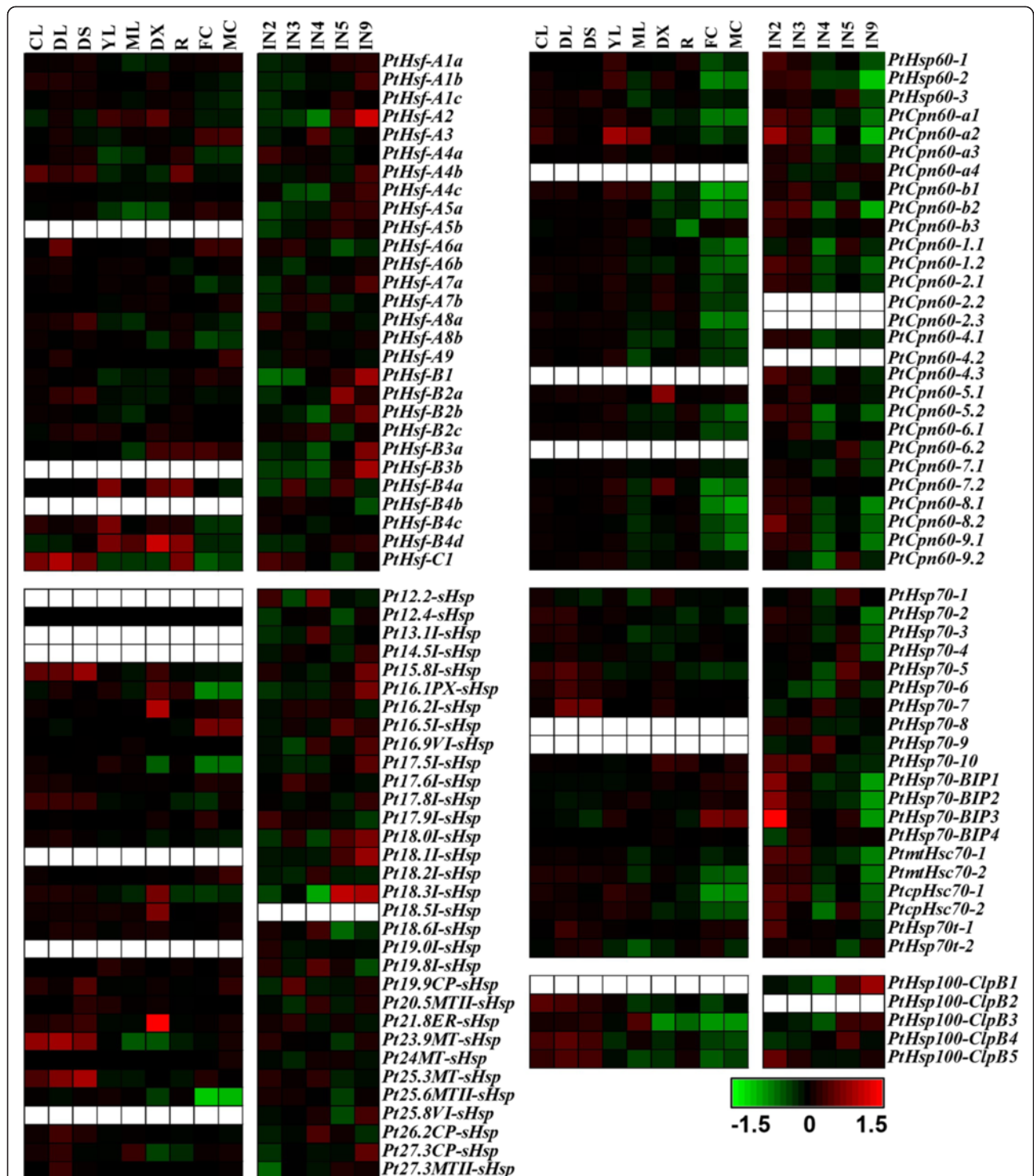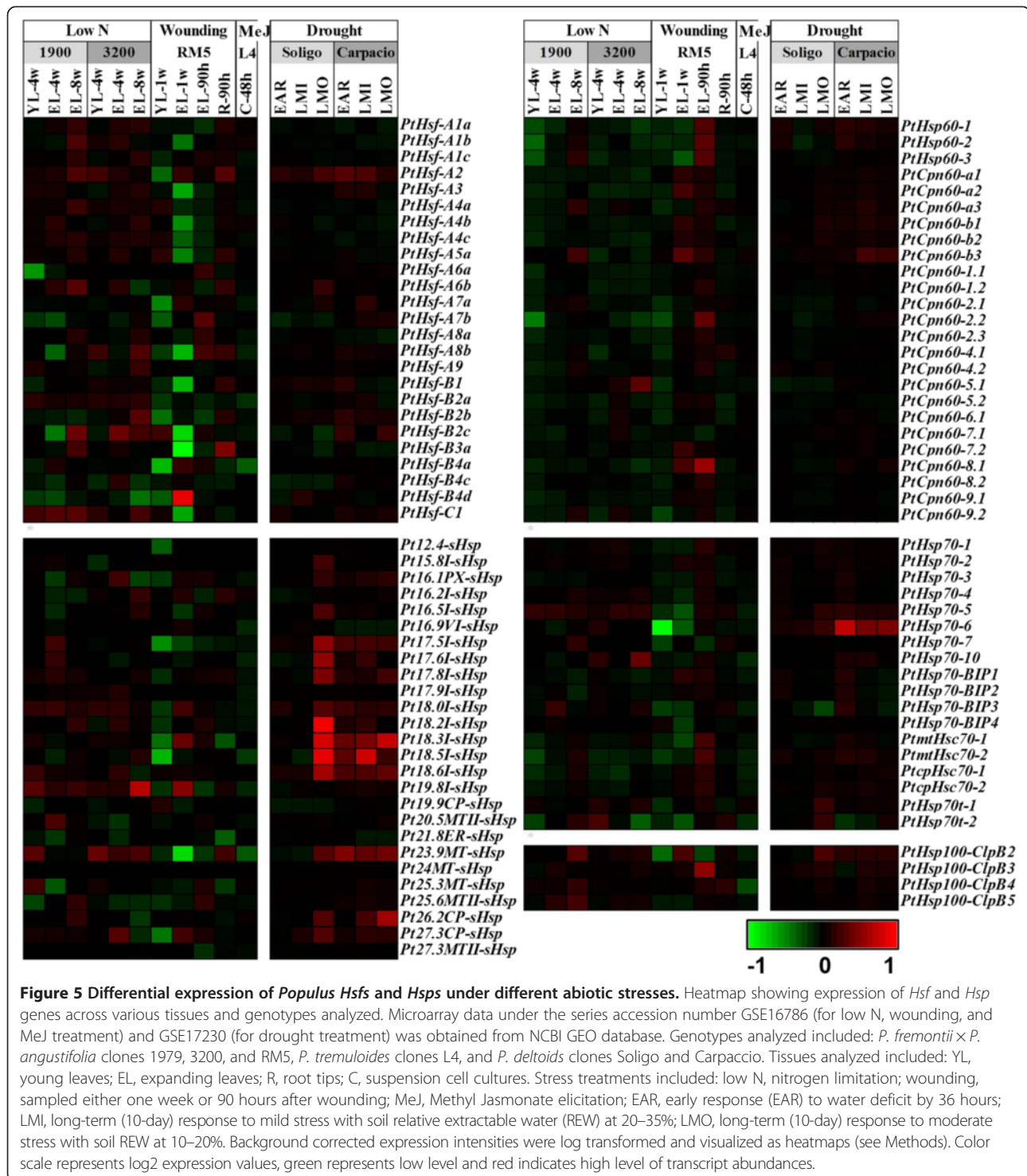
Zhang *et al. BMC Genomics* (2015) 16:181

Page 8 of 19

Figure 4 Expression profiles of *Populus* Hsfs and Hsps across different tissues. Heatmap showing expression of *Hsf* and *Hsp* genes across various tissues and different stem development/growth stages. The Affymetrix microarray data (GSE13990) and the NimbleGen microarray data (GSE13043) were obtained from NCBI Gene Expression Omnibus (GEO) database. CL, continuous light-grown seedling; DL, etiolated dark-grown seedling transferred to light for 3 h; DS, dark-grown seedlings; YL, young leaf; ML, mature leaf; R, root; DX, differentiating xylem; FC, female catkins; MC, male catkins; IN2-IN5, and IN9, stem internode 2 to internode 5, and internode 9. Background corrected expression intensities were log transformed and visualized as heatmaps (see Methods). Color scale represents log2 expression values, green represents low level and red indicates high level of transcript abundance. Blank represents a gene has no corresponding probe sets in the microarray data.

Zhang *et al. BMC Genomics* (2015) 16:181

Page 9 of 19

Figure 5 Differential expression of *Populus Hsfs* and *Hsps* under different abiotic stresses. Heatmap showing expression of *Hsf* and *Hsp* genes across various tissues and genotypes analyzed. Microarray data under the series accession number GSE16786 (for low N, wounding, and MeJ treatment) and GSE17230 (for drought treatment) was obtained from NCBI GEO database. Genotypes analyzed included: *P. fremontii × P. angustifolia* clones 1979, 3200, and RM5, *P. tremuloides* clones L4, and *P. deltoids* clones Soligo and Carpaccio. Tissues analyzed included: YL, young leaves; EL, expanding leaves; R, root tips; C, suspension cell cultures. Stress treatments included: low N, nitrogen limitation; wounding, sampled either one week or 90 hours after wounding; MeJ, Methyl Jasmonate elicitation; EAR, early response (EAR) to water deficit by 36 hours; LMI, long-term (10-day) response to mild stress with soil relative extractable water (REW) at 20–35%; LMO, long-term (10-day) response to moderate stress with soil REW at 10–20%. Background corrected expression intensities were log transformed and visualized as heatmaps (see Methods). Color scale represents log2 expression values, green represents low level and red indicates high level of transcript abundances.

The gene expression patterns were mostly consistent with the microarray data (Figures 5 and 7).

We also analyzed the response of the selected *Hsf* and *Hsp* genes to high temperatures by qRT-PCR. A heat treatment composed a 37°C pretreatment for 3 h, a subsequent 45°C treatment for 3 h, and a 2-h recovery interval was performed (Figure 8) [37]. All of the selected *Hsf* and *Hsp* genes were induced immediately after 30 min in the 37°C pretreatment and 45°C treatment. The expression levels of the detected genes after the 45°C treatment were relatively high compared with that of the 37°C pretreatment. After a 2-h recovery from
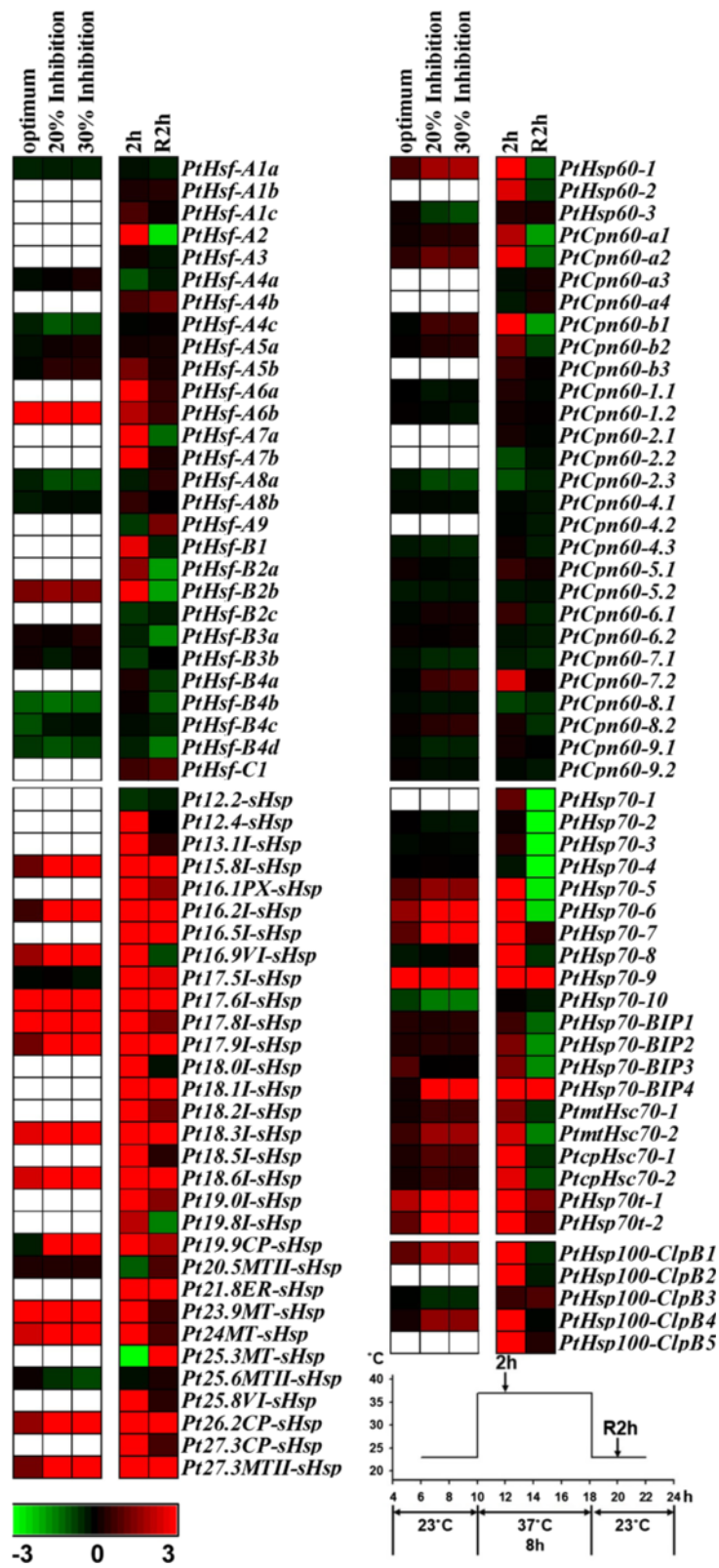
Zhang et al. BMC Genomics (2015) 16:181

Page 10 of 19

Figure 6 (See legend on next page.)

Zhang *et al. BMC Genomics* (2015) 16:181

Page 11 of 19

(See figure on previous page.)

Figure 6 Expression profiles of *Populus* Hsfs and Hsps under heat stress. Heatmap showing expression of *Hsf* and *Hsp* genes under heat stress. Microarray data (GSE26199) obtained from NCBI GEO database and our RNA-seq data (unpublished, see Methods) were used to be analyzed. In GSE26199, the expression changes of *Hsf* and *Hsp* genes under photosynthetic optimum (31.75℃, temperature producing the maximum net CO_2 assimilation rate), 20% inhibition of optimum (38.4℃) and 30% inhibition of optimum (40.5℃) relative to baseline (22℃, the growth temperature) were analyzed in *P. trichocarpa* leaves. In our RNA-seq data, the expression changes of *Hsf* and *Hsp* genes during heat stress (2 h, 2 hours after heat treatment at 37℃; R2h, 2 hours after recovery from 37℃) were analyzed in the 5th internode of *P. alba* × *P. glandulosa*. The expression level of genes was determined based on the value of FPKM (Fragments Per Kilobase of exon per Million fragments mapped). Details of the FPKM are shown in Additional file 10: Table S9. Color scale represents log2 expression values, green represents low level and red indicates high level of transcript abundance. Blank represents the gene has no corresponding probe sets in the microarray data.

the 37°C pretreatment, the induced transcription levels of the selected *Hsf*, *Hsp60*, *Hsp70*, and *Hsp100* genes were low; however, the *sHsps* maintained their induced transcription levels. These gene expression patterns were consistent with the RNA-seq data (Figure 6). Interestingly, in the 2-h plant recovery period after the 45°C treatment, the increased transcription levels of almost all of the selected *Hsf* and *Hsp* genes did not decline (Figure 8).

Populus Hsfs transcriptionally regulate the expression of different sets of *Hsp* genes

In *Arabidopsis*, some Hsfs have been shown to promote rapid *Hsp* expression by binding to *cis*-acting sequences, designated as heat shock elements (HSEs), in the promoter of *Hsps* [21,48]. To explore the potential regulatory network between *PtHsfs* and their downstream *PtHsps*, we constructed a coexpression network between *PtHsfs* and *PtHsps* using WGCNA [49]. As shown in
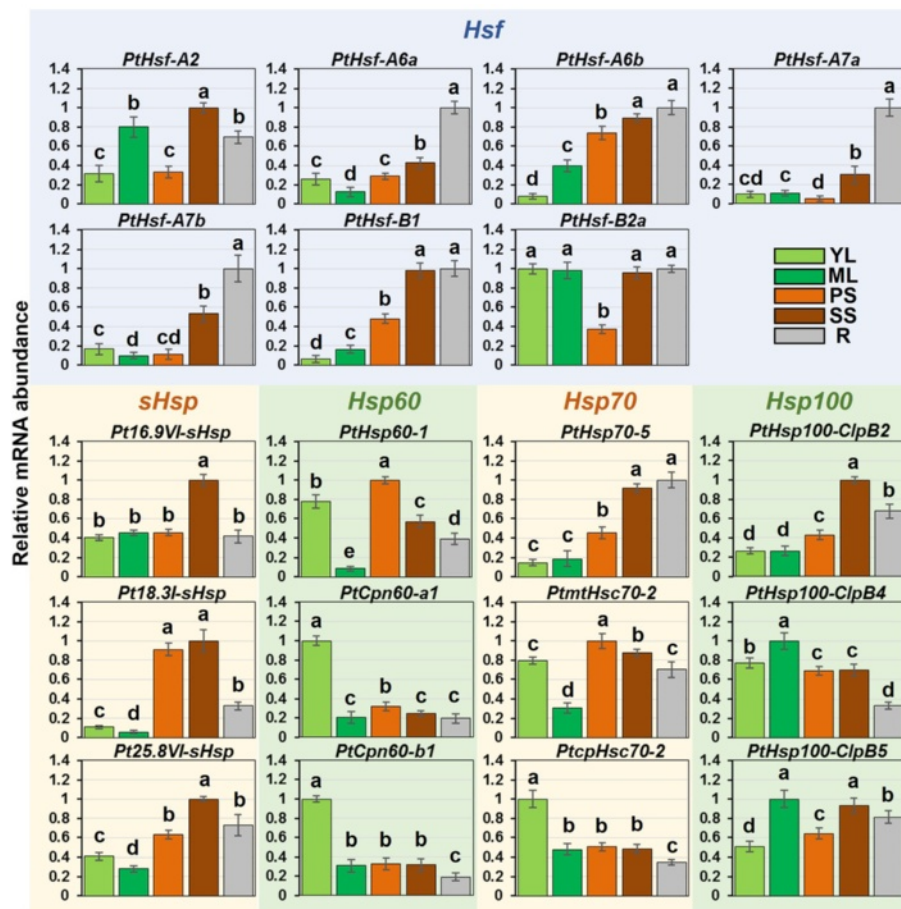

Figure 7 qRT-PCR analysis of selected *Populus* Hsfs and Hsps in different tissues. The relative mRNA abundance of selected seven *Hsfs*, three *sHsps*, three *Hsp60s*, three *Hsp70s* and three *Hsp100s* was normalized with respect to two reference genes *PtActin* and *PtTubulin* in five different tissues. Bars represent standard deviations (SD) of three technical replicates. YL, young leaves; ML, mature leaves; PS, primary stem; SS, secondary stem; R, roots. Bars with the same letter are not significantly different according to Duncan test and Fisher's protected LSD test ($P < 0.05$).
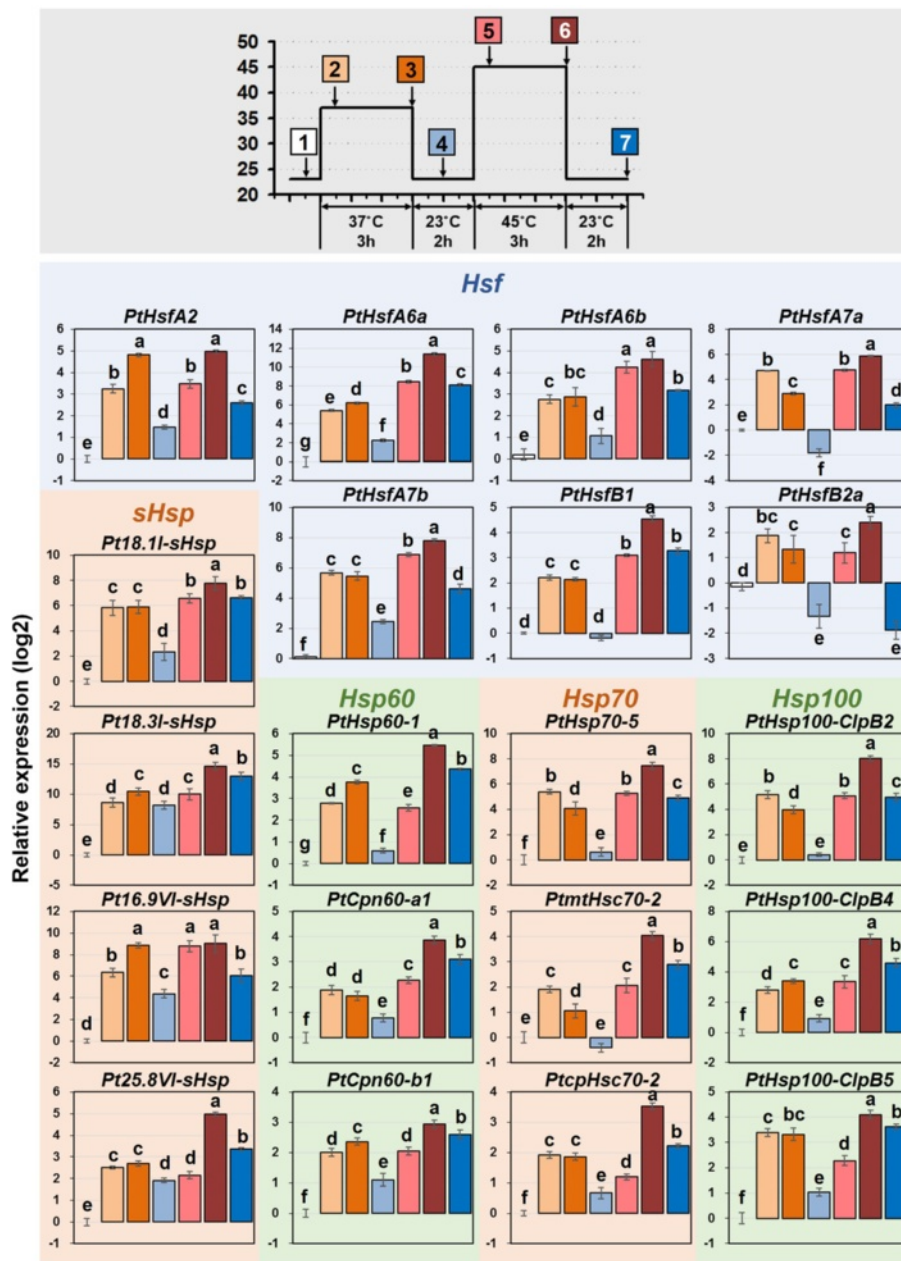
Zhang *et al. BMC Genomics* (2015) 16:181

Page 12 of 19

Figure 8 qRT-PCR analysis of selected *Populus* Hsfs and Hsps under heat stress. The relative mRNA abundance of selected seven *Hsfs*, four *sHsps*, three *Hsp60s*, three *Hsp70s*, and three *Hsp100s* during heat stress were quantified. Seedlings were heated to 37°C for 3 h (pretreatment), return to 23°C for 2 h, heated to 45°C for 3 h (treatment), and then allowed to 23°C for 2 h. 1, control; 2, 30 min after pretreatment at 37°C; 3, 2 h after pretreatment at 37°C; 4, 1 h after recovery at 23°C; 5, 30 min after treatment at 45°C; 6, 2 h after treatment at 45°C; 7, 2 h after recovery at 23°C. Relative expression represents log2 expression values. Bars with the same letter are not significantly different according to Duncan's test and Fisher's protected LSD test (*P* < 0.05).

Figure 9, there are many coexpression relationships between *Populus* Hsfs and Hsps, and one *Hsf* could be coexpressed with several *Hsps*. Moreover, several *PtHsfs*, including *PtHsfA2*, *PtHsfA6a*, and *PtHsfB2a*, showed high coexpression levels when bound to *PtHsps*. To validate the possible regulatory roles of *PtHsfs*, we transiently overexpressed *PtHsfs* in *P. trichocarpa* to detect the

transcriptional regulation of the downstream *PtHsps* using qRT-PCR. Two type A *Hsfs* (*PtHsfA2* and *PtHsfA6a*) and three type B *Hsfs* (*PtHsfB1*, *PtHsfB2a*, and *PtHsfB2b*) were selected for infiltration (see Methods). Three days after the infiltration of *PtHsfs*, the expression levels of selected *Hsps* (three *sHsp*, three *Hsp60*, three *Hsp70*, and three *Hsp100*) were examined to evaluate the transcriptional
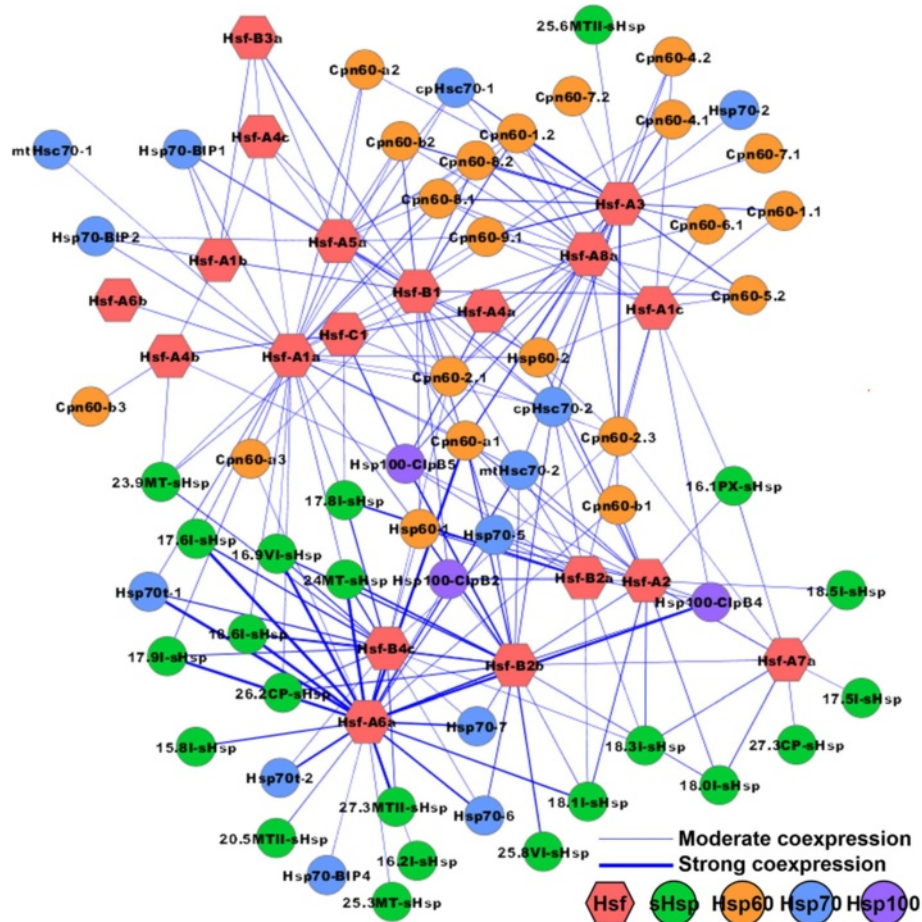
Figure 9 Coexpression network of *Hsfs* and *Hsps* in *Populus*. Nodes represent *Hsfs* and *Hsps* in *Populus*, edges indicate pairwise correlation constructed by WGCNA. Node color codes represent different gene families. Red hexagons indicate *Hsfs*, solid circles with green, orange, blue, and purple indicate *sHsps*, *Hsp60s*, *Hsp70s*, and *Hsp100s*, respectively. Thin edges indicate moderate coexpressions and thick edges indicate strong coexpressions between the two nodes. The network was created using Cytoscape (see Methods).

regulatory relationships between selected *PtHsf* and selected *PtHsps* (Figure 10).We revealed that, the expression levels of all of the detected *Hsps*, except *25.8VI-sHsp*, were induced by the overexpression of *PtHsfA2* (Figure 10A). One *sHsps* (*18.1I-sHsp*), two *Hsp60s* (*Hsp60-1* and *Cpn60-a1*), three *Hsp70s* (*Hsp70-5*, *mtHsc70-2*, and *cpHsc70-2*), and two *Hsp100s* (*Hsp100-ClpB2* and *Hsp100-ClpB4*) were significantly induced by the transient overexpression of *PtHsfA6a*, *PtHsfB2a*, and *PtHsfB2b* (Figure 10B, D, and E). However, only a few selected *Hsps* were induced by the overexpression of *PtHsfB1* (Figure 10C).

Discussion

Duplications play major roles in the diversification of the *Populus Hsf* and *Hsp* gene families

Gene duplications represent the major mechanism for gene family expansion through either WGD or tandem duplications during evolution [50]. Like most gene families, *Hsf* and *Hsp* appear to have undergone complicated

evolutionary processes [11,13]. The abundance of *Hsf* and *Hsp* genes in *P. trichocarpa* may be the result of multiple gene duplication events, represented by a whole genome duplication following multiple segmental and tandem duplications [36]. However, 44 of the 118 *Hsf* and *Hsp* genes lack copies in the corresponding duplicated blocks in the genome of *P. trichocarpa* (Figure 3). This suggested that dynamic rearrangements or segmental losses may have occurred following the segmental duplication. In plants, the exon/intron diversification of gene family members has played an important role in the evolution of multiple gene families [51]. Only 6 of 42 paralogous pairs of *Hsf* and *Hsp* genes exhibited certain degrees of variation in *P. trichocarpa* (Figures 1 and 2). The differences might be derived from the single intron addition or subtraction events during the structural evolution of the *Hsf* or *Hsp* genes.

In *P. trichocarpa*, ~33.4% of predicted genes originated from whole genome duplication and 15.6% from tandem
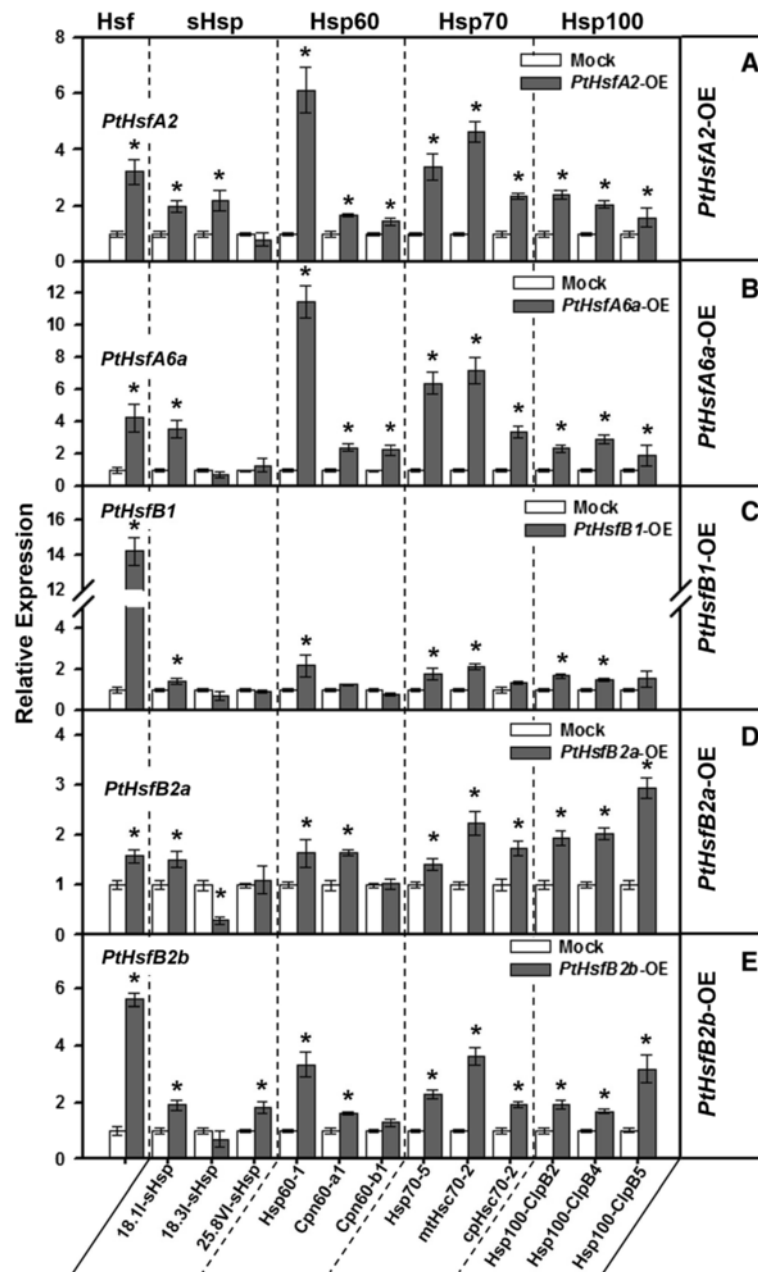
Zhang *et al. BMC Genomics* (2015) 16:181

Page 14 of 19

Figure 10 qRT-PCR validation of transcriptional expression of *Hsps* regulated by transient overexpression of *PtHsfs*. The relative mRNA abundance of selected three *sHsps*, three *Hsp60s*, three *Hsp70s* and three *Hsp100s* were quantified in transient overexpression leaves of *PtHsfA2* **(A)**, *PtHsfA6b* **(B)**, *PtHsfB1* **(C)**, *PtHsfB2a* **(D)**, and *PtHsfB2b* **(E)**. The relative expressions represent log2 expression values. Asterisk, significant differences ($P < 0.05$) between overexpression and mock control.

duplications [36]. Our study indicates that the *Hsf* and *Hsp* gene families possess a higher whole genome duplication ratio (47.5%, 56 of 118 genes) and lower tandem duplication ratio (13.5%, 16 of 118 genes) in *P. trichocarpa*, which are dramatically different from the averages. This high retention rate of whole genome duplications and low retention rate of tandem duplications are also inconsistent with the previous studies on other families [52-56]. The low *K*a/

*K*s ratio indicates that all gene pairs, except for two (*PtHsf-A6a/PtHsf-A6b* and *Pt19.9CP-sHsp/Pt26.2CP-sHsp*), might have evolved under the influence of purifying selection.

Populus *Hsf* and *Hsp* genes are vital players in plant development and in response to abiotic stresses

Plant *Hsf* and *Hsp* genes are implicated in a variety of biological processes [6,12,13,57]. The study on the *Populus*

Zhang *et al. BMC Genomics* (2015) 16:181

Page 15 of 19

Hsf and *Hsp* gene families showed that these genes were differentially expressed in various analyzed tissues. Most *Hsp60* and *Hsp70* genes showed high expression levels in the upper stem (internode 2 and 3), indicating their putative roles in the primary stem growth of *Populus* (Figure 4). Our qRT-PCR results showed that all selected *Populus Hsf* and *Hsp* genes exhibited diverse expression levels in the five examined tissues (Figure 7), indicating that these *Hsf* and *Hsp* genes often participate in plant development. It is worth noting that the expression of *Pt18.3I-sHsp* was obviously higher in stem than in leaves and roots, indicating that it may have potential roles in stem development.

Hsfs, as transcriptional activators of *Hsps*, cooperated with *Hsps* to form a network responding to various stresses. They play a broad role in the tolerance to multiple environmental stress treatments apart from heat stress [6,35]. The response of Hsfs and Hsps to heat and other abiotic stresses overlaps extensively, indicating that *Hsfs* and *Hsps* are important in the cross-talk of multiple environmental stress response pathways [12,13]. Interestingly, under drought stress, most members of the *Populus Hsf*, *Hsp60*, *Hsp70*, and *Hsp100* families were not regulated, while only members of the *sHsp* family were significantly up-regulated (especially *Populus* C-I *sHsp* genes) in the hybrid poplar (Figure 6). The specific response to drought stress of *Populus* C-I *sHsp* might be related to the member expansion of the *Populus* C-I *sHsp* subfamily, but the mechanisms need to be further examined. The expression of *Hsps* could be down-regulated to their initial levels when plants recovered from high temperatures [37]. However, the *Populus sHsps* remained at a high expression level 2 h after the plants' recovery from 37°C (Figure 8). This indicated that *Populus sHsps* play an irreplaceable role compared with other *Hsps* (*Hsp60*, *Hsp70*, and *Hsp100* genes) in response to heat and drought stress.

Complex transcriptional regulatory network between *Populus* Hsfs and Hsps

Several *PtHsfs*, including *PtHsfA2*, *PtHsfA6a*, and *PtHsfB2a*, showed high coexpression levels with *PtHsps*, indicating that these *PtHsfs* might be the key regulators among the numerous *PtHsfs*. Nishizawa *et al.* [21] showed that *Arabidopsis HsfA2* is a key regulator in the induction of the defense system under several types of environmental stresses. In our study, the overexpression of *PtHsfA2* induced the expression of 11 of 12 *Hsps* (Figure 10A), indicating that *PtHsfA2* is also an activator of the downstream *Hsps*. The expression of *PtHsfA2* was induced under nitrogen deprivation, drought, and heat stresses (Figures 5 and 6), implying that *PtHsfA2* is a key regulator under various stresses in poplars. A set of *PtHsps* could be simultaneously induced by the transient overexpression of *PtHsfA6a*,

PtHsfB2a, or *PtHsfB2b*, indicating that these *Hsps* were generally regulated by the three *Hsfs* (Figure 10B, D, and E). Recently, Ikeda *et al.* [58] reported that *HsfB1* and *HspB2b* suppress the general heat shock response under non-heat-stress conditions and in the attenuating period. However, *HsfB1* and *HsfB2b* appear to be necessary for the expression of heat stress-inducible *Hsps* under heat stress conditions, which is necessary for acquired thermotolerance. In this study, most detected *Hsps* could not be activated by the overexpression of *PtHsfB1* (Figure 10C), while they could be activated by overexpressing *PtHsfB2a* or *PtHsfB2b* (Figure 10D and E). These results indicated that *PtHsfB1* is not a direct activator of these *Hsps*. Under heat stress, *PtHsfB1* was also induced dramatically, indicating that although *PtHsfB1* could not regulate the downstream *Hsps* directly, it might be involved in the heat stress response in some indirect manner. Tomato (*Solanum lycopersicum*) HsfB proteins may be coactivators of HsfAs [59]. In *Arabidopsis*, the heat stress response is finely regulated by activation and repression activities of *Hsfs* [58]. The expression of a particular *Populus Hsfs* can induce the expression of a different set of *Hsps*, suggesting a complex transcriptional regulatory network between *Populus* Hsfs and *Hsps*. Additionally, the *Hsps* might also be regulated by both the activation and the repression mechanisms during the heat stress response in *Populus*. However, the precise regulatory mechanisms among *Hsfs* and *Hsps* of woody plants during development and stress responses require further investigation.

Conclusions
In this study, 118 members of *Populus Hsf* and *Hsp* (including *sHsp*, *Hsp60*, *Hsp70* and *Hsp100*) gene families were identified. A comprehensive analysis of these genes, including phylogeny, chromosomal location, gene structure, expression profiling and heat stress responses, was performed. In the phylogenetic tree, the majority of subfamilies contained members from both *Populus* and *Arabidopsis*, which suggested that the functions of most of *Hsfs* and *Hsps* were conserved during evolution. In addition, a gene duplication analysis revealed that the class I cytosolic *sHsp* subfamily had expanded more than that of *Arabidopsis*, showing both the conservation and divergence of gene families during the evolution. An extensive expression analysis indicated that *Populus Hsf* and *Hsp* genes may play various conserved roles in different biological processes in plants. Moreover, the coexpression network and transient overexpression of *PtHsfs* imply that there is a complex transcriptional regulatory network between *Populus* Hsfs and *Hsps*. The present study established a solid foundation for functional research on the *Populus Hsf* and *Hsp* gene families and

Zhang *et al. BMC Genomics* (2015) 16:181

Page 16 of 19

has improved our understanding of the functions of *Hsf* and *Hsp* genes in woody plants.

Methods

Sequence retrieval and phylogenetic reconstruction

Published *Arabidopsis Hsf* and *Hsp* sequences [13] were retrieved and were used as queries in BLAST searches against the *Populus trichocarpa* genome database (http://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Ptrichocarpa, release 3.0) to identify potential *Populus Hsfs* and *Hsps*. WoLF PSORT (http://psort.hgc.jp/) was used to predict their protein subcellular localizations [60]. The isoelectric points and molecular weights were estimated using the Compute pI/Mw tool from ExPASy (http://web.expasy.org/compute_pi). The phylogenetic trees were constructed using the neighbor-joining method [61] in the MEGA package V5.1 [62] with bootstrap values from 1,000 replicates indicated at each node. To predict putative orthologous genes, a reciprocal Best Blast Hit search was used. The gene names of *Populus Hsfs* are according to Scharf et al. [63] and the gene names of *Populus sHsps* were revised according to their molecular weights in the *P. trichocarpa* genome V3.0 based on Water et al. [64].

Chromosomal locations, gene structures, and a conserved motif analysis

The chromosomal locations of the *Hsf* and *Hsp* genes were determined using the *Populus* genome browser (http://www.phytozome.net/poplar). Homologous chromosome segments resulting from whole-genome duplication events were identified as described previously [37]. The exon and intron structures were illustrated using Gene Structure Display Server (GSDS, http://gsds.cbi.pku.edu.cn) [65] by aligning the cDNA sequences with the corresponding genomic DNA sequences from Phytozome (http://www.phytozome.net). The conserved motifs in the encoded proteins were analyzed by MEME online program (http://meme.sdsc.edu, v4.9.0) [38]. MEME was run locally with the following parameters: number of repetitions = any, maximum number of motifs = 20, and optimum motif width = 30 to 70 residues for Hsf, Hsp60, Hsp70 and Hsp100. Because of the short sHsp protein length, we set an optimum motif width of between 10 and 40.

Publicly available microarray data analyses and coexpression network generation

For abiotic and hormonal treatments, the Affymetrix microarray data available in the NCBI GEO database under the series accession numbers GSE26199 (heat stress), GSE17230 (drought stress), and GSE16786 were analyzed [47,66,67]. GSE16786 is composed of the following five subsets: GSE14893 (nitrogen limitation, genotype 1979), GSE14515 (nitrogen limitation, genotype 3200), GSE16783 (1 week after leaf wounding), GSE16785 (90 h after leaf wounding), and GSE16773 (methyl jasmonate-elicited suspension cell cultures). Probe sets corresponding to *Populus Hsf* and *Hsp* genes were identified using the online Probe Match tool POParray (http://aspendb.uga.edu/poparray). For genes with more than one probe sets, the median of expression values was considered. The expression data were gene-wise normalized. The data was normalized using the Gene Chip Robust Multiarray Analysis (GCRMA) algorithm followed by log transformation and average calculations. After normalization and log transformation of data for all the *Populus* genes present on the chip, the log signal intensity values for *Populus* probe set IDs corresponding to *Hsf* and *Hsp* genes were extracted for further analyses. The expression profiles were performed in MultiExperiment Viewer (MeV) v4.7.4 [68].

The WGCNA package [49] provides a robust set of R functions for constructing weighted coexpression networks. The data from the microarray data and our RNA-seq data in Figures 4, 5, 6 were used for the coexpression analysis. The similarity matrix was transformed into an adjacency matrix using a method that employs a power function. The network picture was created using Cytoscape [69].

Plant material and growth conditions

One-year-old *P. trichocarpa* and hybrid poplar (*P. alba* × *P. glandulosa*) clones (84 K) were grown in a growth chamber under long-day conditions (16 h light/8 h dark) at 23°C. Plant materials for the qRT-PCR of different tissues and under heat treatment conditions were collected from 84 K. For the heat treatment used in RNA-seq, 1-year-old seedlings of *P. alba* × *P. glandulosa* were treated at 37°C for 8 h (10:00 a.m.-18:00 p.m.) and then recovered to 23°C. During the heat treatment, three time points (0 h, 2 h, and R2 h) were selected to sample collection for RNA-seq. At each time point, the 5th internodes of seedlings under both heat treatment and control conditions were collected separately. For the heat treatment used in the qRT-PCR, the chamber was heated to 37°C for 3 h (pretreatment), returned to 23°C for 2 h, heated to 45°C for 3 h (treatment), and then allowed to recover for 2 h. Leaves from three different plants were harvested at seven selected time points during heat stress treatments. For the transient overexpression of 35S::PtHsfs, *P. trichocarpa* leaves were infiltrated by *Agrobacterium tumefaciens*. Samples were harvested, frozen immediately in liquid nitrogen, and stored at −80°C for further analysis. Three biological replicates were performed.

Zhang *et al. BMC Genomics* (2015) 16:181

Page 17 of 19

Plasmid construction and transient overexpression in *P. trichocarpa*

The coding sequences of *PtHsfA2, A6a, B1, B2a,* and *B2b* were amplified from the cDNA of *P. trichocarpa,* and inserted into pMDC32 to produce the *35S::PtHsfs* constructs using the Gateway cloning system (Invitrogen, Carlsbad, USA). Each *35S::PtHsfs* construct was introduced into the *A. tumefaciens* GV3101 by electroporation. A sample of 1 mL from an overnight *A. tumefaciens* culture was centrifuged at $6,000 \times g$ for 1 min, the supernatant was discarded and the pellet was resuspended in infiltration media (10 mM $MgCl_2$ and 150 μM acetosyringone) ($OD_{600} = 0.1$). Then, the *A. tumefaciens* was infiltrated into *P. trichocarpa* epidermal cells. After 3 days, the infiltrated leaf regions were collected for RNA extraction.

RNA isolation and real-time qRT-PCR

Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen, Venlo, Netherlands) with an on-column RNase-free DNase I (Qiagen, Venlo, Netherlands) treatment to remove any contamination of genomic DNA. The first-strand cDNA synthesis was carried out with ~1 μg RNA using the SuperScript III reverse transcription kit (Invitrogen, Carlsbad, USA) and random primers according to the manufacturer's procedure. Primers with melting temperatures of 58–60°C and amplicon lengths of 150–250 bp were designed using Primer3 software (http://frodo.wi.mit.edu/primer3/input.htm). All primer sequences are listed in Additional file 11: Table S10. Real-time qRT-PCR was conducted on a 7500 Real Time PCR System (Applied Biosystems, CA, USA) using a SYBR Premix Ex Taq™ Kit (TaKaRa, Dalian, China) according to the manufacturer's instructions. Relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method [70]. The *PtActin* and *PtTubulin* genes were used as internal controls.

Statistical analysis

The statistical significance of differences in measured parameters was tested using the procedures of DPS (Zhejiang University, China). Differences between the means among different tissues or gene pairs were compared using Duncan's test and Fisher's protected least significant difference (LSD) test at 0.05 probability levels.

Additional files

Additional file 1: Table S1. *Hsf* and *Hsp* gene families in *Populus.*

Additional file 2: Table S2. List of all *Hsf, sHsp, Hsp60, Hsp70* and *Hsp100* gene sequences identified in *Populus.* The complete coding sequences and the corresponding amino acid sequences of *Hsf* and *Hsp* genes identified from *P. trichocarpa* were obtained from Phytozome (http://www.phytozome.net/poplar, release 3.0).

Additional file 3: Table S3. Divergence between paralogous *Hsf* and *Hsp* gene pairs in *Populus.*

Additional file 4: Table S4. Sequence logos for the conserved motifs of Hsf proteins in *Arabidopsis* and *Populus.*

Additional file 5: Table S5. Sequence logos for the conserved motifs of sHsp proteins in *Arabidopsis* and *Populus.*

Additional file 6: Table S6. Sequence logos for the conserved motifs of Hsp60 proteins in *Arabidopsis* and *Populus.*

Additional file 7: Table S7. Sequence logos for the conserved motifs of Hsp70 proteins in *Arabidopsis* and *Populus.*

Additional file 8: Table S8. Sequence logos for the conserved motifs of Hsp100 proteins in *Arabidopsis* and *Populus.*

Additional file 9: Figure S1. In silico EST analysis of *Populus Hsfs* and *Hsps.* EST frequency for each gene was calculated by evaluating its EST representation among 17 cDNA libraries available at PopGenIE (http://popgenie.org/) [43]. Color bar at bottom represents the frequencies of EST counts. P: petioles, IS: imbibed seeds, B: bark, AC: active cambium, FB: flower buds, TW: tension wood, MC: male catkins, DC: dormant cambium, AS: apical shoot, DB: dormant buds, FC: female catkins, CZ: cambial zone, CSL: cold stressed leaves, SL: senescing leaves, YL: young leaves, R: roots, SM: shoot meristem.

Additional file 10: Table S9. FPKM of *Populus Hsfs* and *Hsps* during heat treatment obtained from RNA-seq data. One-year-old seedlings of *P. alba × P. glandulosa* were treated at 37°C for 8 h (10:00 a.m. ~ 18:00 p.m.) and then recovered to 23°C. Control plants (CK) were grown under 23°C. During the heat treatment, materials collected at three time points were used for RNA-seq. 0 h, before heat treatment; 2 h, 2 h after 37°C heat treatment; R2 h, 2 h after recovered from 37°C to 23°C. At the time points of 2 h and R2 h, the 5th internode of four independent plants in heat treatment group and control group were collected for RNA extraction and RNA-seq. Differential expression was calculated by Cuffdiff using the FPKM.)

Additional file 11: Table S10. Primer sequences used in qRT-PCR analysis.

Author details

[1]State Key Laboratory of Tree Genetics and Breeding, Key Laboratory of Tree Breeding and Cultivation of the State Forestry Administration, Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China. [2]Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037, China. [3]College of Forestry, Fujian Agriculture and Forestry University, Fuzhou, Fujian 350002, China. [4]State Key Laboratory of Crop Biology, Shandong Key Laboratory of Crop Biology, College of Life Sciences, Shandong Agricultural University, Tai'an, Shandong 271018, China. [5]Department of Biology, McGill University, 1205 Dr Penfield Avenue, Montreal, Quebec H3A 1B1, Canada.

Zhang *et al. BMC Genomics* (2015) 16:181

Page 18 of 19

References

1. Levitt J. Responses of plants to environmental stresses. In: Water, radiation, salt, and other stresses, vol. II. New York: Academic Press; 1980.
2. Wang W, Vinocur B, Altman A. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. Planta. 2003;218(1):1–14.
3. Shao H-B, Guo Q-J, Chu L-Y, Zhao X-N, Su Z-L, Hu Y-C, et al. Understanding molecular mechanism of higher plant plasticity under abiotic stress. Colloids Surf B: Biointerfaces. 2007;54(1):37–45.
4. Morimoto RI. Cells in stress: transcriptional activation of heat shock genes. Science. 1993;259:1409–9.
5. Feder M. Integrative biology of stress: molecular actors, the ecological theater, and the evolutionary play. In: International Symposium on Environmental Factors, Cellular Stress and Evolution, Varanasi, India: 2006. 2006.
6. Wahid A, Gelani S, Ashraf M, Foolad MR. Heat tolerance in plants: an overview. Environ Exp Bot. 2007;61(3):199–223.
7. Lindquist S, Craig E. The heat-shock proteins. Annu Rev Genet. 1988;22 (1):631–77.
8. Morimoto R, Tissieres A, Georgopoulos C. Heat shock proteins: Structure, function and regulation. In: Cold Spring Harbor Lab. Press, Cold Spring Harbor, NY. 1994.
9. Gupta SC, Sharma A, Mishra M, Mishra RK, Chowdhuri DK. Heat shock proteins in toxicology: how close and how far? Life Sci. 2010;86(11):377–84.
10. Feder ME, Hofmann GE. Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. Annu Rev Physiol. 1999;61:243–82.
11. Wang F, Dong Q, Jiang H, Zhu S, Chen B, Xiang Y. Genome-wide analysis of the heat shock transcription factors in *Populus trichocarpa* and *Medicago truncatula*. Mol Biol Rep. 2012;39(2):1877–86.
12. Hu W, Hu G, Han B. Genome-wide survey and expression profiling of heat shock proteins and heat shock factors revealed overlapped and stress specific response under abiotic stresses in rice. Plant Sci. 2009;176(4):583–90.
13. Swindell WR, Huebner M, Weber AP. Transcriptional profiling of *Arabidopsis* heat shock proteins and transcription factors reveals extensive overlap between heat and non-heat stress response pathways. BMC Genomics. 2007;8(1):125.
14. Nover L, Bharti K, Döring P, Mishra SK, Ganguli A, Scharf K-D. *Arabidopsis* and the heat stress transcription factor world: how many heat stress transcription factors do we need? Cell Stress Chaperones. 2001;6(3):177.
15. Scharf KD, Siddique M, Vierling E. The expanding family of *Arabidopsis thaliana* small heat stress proteins and a new family of proteins containing alpha-crystallin domains (Acd proteins). Cell Stress Chaperones. 2001;6 (3):225–37.
16. Siddique M, Gernhard S, von Koskull-Doring P, Vierling E, Scharf KD. The plant sHSP superfamily: five new members in *Arabidopsis thaliana* with unexpected properties. Cell Stress Chaperones. 2008;13(2):183–97.
17. Lin BL, Wang JS, Liu HC, Chen RW, Meyer Y, Barakat A, et al. Genomic analysis of the Hsp70 superfamily in *Arabidopsis thaliana*. Cell Stress Chaperones. 2001;6(3):201–8.
18. Krishna P, Gloor G. The Hsp90 family of proteins in *Arabidopsis thaliana*. Cell Stress Chaperones. 2001;6(3):238–46.
19. Lee U, Rioflorido I, Hong SW, Larkindale J, Waters ER, Vierling E. The *Arabidopsis* ClpB/Hsp100 proteins: chaperones for stress and chloroplast development. Plant J. 2007;49(1):115–27.
20. Yoshida T, Ohama N, Nakajima J, Kidokoro S, Mizoi J, Nakashima K, et al. *Arabidopsis* HsfA1 transcription factors function as the main positive regulators in heat shock-responsive gene expression. Mol Genet Genomics. 2011;286(5–6):321–32.
21. Nishizawa A, Yabuta Y, Yoshida E, Maruta T, Yoshimura K, Shigeoka S. *Arabidopsis* heat shock transcription factor A2 as a key regulator in response to several types of environmental stress. Plant J. 2006;48(4):535–47.
22. Zhang L, Li Y, Xing D, Gao C. Characterization of mitochondrial dynamics and subcellular localization of ROS reveal that HsfA2 alleviates oxidative damage caused by heat stress in *Arabidopsis*. J Exp Bot. 2009;60(7):2073–91.
23. Jiang C, Xu J, Zhang H, Zhang X, Shi J, Li M, et al. A cytosolic class I small heat shock protein, RcHSP17. 8, of *Rosa chinensis* confers resistance to a variety of stresses to *Escherichia coli*, yeast and *Arabidopsis thaliana*. Plant Cell Environ. 2009;32(8):1046–59.
24. Zhong L, Zhou W, Wang H, Ding S, Lu Q, Wen X, et al. Chloroplast small heat shock protein HSP21 interacts with plastid nucleoid protein pTAC5 and is essential for chloroplast development in *Arabidopsis* under heat stress. Plant Cell. 2013;25:2925–43.
25. Lubben TH, Donaldson GK, Viitanen PV, Gatenby AA. Several proteins imported into chloroplasts form stable complexes with the GroEL-related chloroplast molecular chaperone. Plant Cell. 1989;1(12):1223–30.
26. Suzuki K, Nakanishi H, Bower J, Yoder DW, Osteryoung KW, Miyagishima S-Y. Plastid chaperonin proteins Cpn60α and Cpn60β are required for plastid division in *Arabidopsis thaliana*. BMC Plant Biol. 2009;9(1):38.
27. Apuya NR, Yadegari R, Fischer RL, Harada JJ, Zimmerman JL, Goldberg RB. The *Arabidopsis* embryo mutant schlepperless has a defect in the chaperonin-60α gene. Plant Physiol. 2001;126(2):717–30.
28. Ishikawa A, Tanaka H, Nakai M, Asahi T. Deletion of a chaperonin 60β gene leads to cell death in the *Arabidopsis* lesion initiation 1 mutant. Plant Cell Physiol. 2003;44(3):255–61.
29. Hartl FU, Bracher A, Hayer-Hartl M. Molecular chaperones in protein folding and proteostasis. Nature. 2011;475(7356):324–32.
30. Frydman J. Folding of newly translated proteins in vivo: the role of molecular chaperones. Annu Rev Biochem. 2001;70(1):603–47.
31. Batra G, Chauhan VS, Singh A, Sarkar NK, Grover A. Complexity of rice Hsp100 gene family: lessons from rice genome sequence data. J Bioscience. 2007;32(3):611–9.
32. Katiyar-Agarwal S, Agarwal M, Gallie DR, Grover A. Search for the cellular functions of plant Hsp100/Clp family proteins. Crit Rev Plant Sci. 2001;20 (3):277–95.
33. Hwang BJ, Park WJ, Chung CH, Goldberg AL. Escherichia coli contains a soluble ATP-dependent protease (Ti) distinct from protease La. Proc Natl Acad Sci U S A. 1987;84(16):5550–4.
34. Yang J-Y, Sun Y, Sun A-Q, Yi S-Y, Qin J, Li M-H, et al. The involvement of chloroplast HSP100/ClpB in the acquired thermotolerance in tomato. Plant Mol Biol. 2006;62(3):385–95.
35. Hahn A, Bublak D, Schleiff E, Scharf K-D. Crosstalk between Hsp90 and Hsp70 chaperones and heat stress transcription factors in tomato. Plant Cell. 2011;23(2):741–55.
36. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science. 2006;313(5793):1596–604.
37. Zhang J, Li J, Liu B, Zhang L, Chen J, Lu M. Genome-wide analysis of the *Populus* Hsp90 gene family reveals differential expression patterns, localization, and heat stress responses. BMC Genomics. 2013;14:532.
38. Bailey TL, Williams N, Misleh C, Li WW. MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. 2006;34 suppl 2: W369–73.
39. Holub EB. The arms race is ancient history in *Arabidopsis*, the wildflower. Nat Rev Genet. 2001;2(7):516–27.
40. Hurst L. The *Ka/Ks* ratio: diagnosing the form of sequence evolution. Trends Genet. 2002;18(9):486.
41. Yang X, Tuskan GA. Divergence of the Dof gene families in poplar, *Arabidopsis*, and rice suggests multiple modes of gene evolution after duplication. Plant Physiol. 2006;142(3):820–30.
42. Hu R, Chi X, Chai G, Kong Y, He G, Wang X, et al. Genome-wide identification, evolutionary expansion, and expression profile of homeodomain-leucine zipper gene family in poplar (*Populus trichocarpa*). PLoS One. 2012;7(2):e31149.
43. Sjödin A, Street NR, Sandberg G, Gustafsson P, Jansson S. The Populus Genome Integrative Explorer (PopGenIE): a new resource for exploring the Populus genome. New Phytol. 2009;182(4):1013–25.
44. Wilkins O, Nahal H, Foong J, Provart NJ, Campbell MM. Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors. Plant Physiol. 2009;149(2):981–93.
45. Dharmawardhana P, Brunner A, Strauss S. Genome-wide transcriptome analysis of the transition from primary to secondary stem development in *Populus trichocarpa*. BMC Genomics. 2010;11(1):150.
46. Barrett T, Edgar R. Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. Method Enzymol. 2006;411:352–69.
47. Weston DJ, Karve AA, Gunter LE, Jawdy SS, Yang X, Allen SM, et al. Comparative physiology and transcriptional networks underlying the heat shock response in *Populus trichocarpa*, *Arabidopsis thaliana* and *Glycine max*. Plant Cell Environ. 2011;34(9):1488–506.
48. Guo A, Chen S, Liu K, Liu Y, Ni L, Zhang K, et al. Isolation of heat shock factor HsfA1a-binding sites in vivo revealed variations of heat shock elements in *Arabidopsis thaliana*. Plant Cell Physiol. 2008;49(9):1306–15.
49. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9(1):559.

Zhang *et al. BMC Genomics* (2015) 16:181

Page 19 of 19

50. Cannon SB, Mitra A, Baumgarten A, Young ND, May G. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. BMC Plant Biol. 2004;4(1):10.

51. Xu G, Guo C, Shan H, Kong H. Divergence of duplicate genes in exon–intron structure. Proc Natl Acad Sci U S A. 2012;109(4):1187–92.

52. Lan T, Yang Z-L, Yang X, Liu Y-J, Wang X-R, Zeng Q-Y. Extensive functional diversification of the *Populus* glutathione S-transferase supergene family. Plant Cell. 2009;21(12):3749–66.

53. Hu R, Qi G, Kong Y, Kong D, Gao Q, Zhou G. Comprehensive analysis of NAC domain transcription factor gene family in *Populus trichocarpa*. BMC Plant Biol. 2010;10:145.

54. Chai G, Hu R, Zhang D, Qi G, Zuo R, Cao Y, et al. Comprehensive analysis of CCCH zinc finger family in poplar (*Populus trichocarpa*). BMC Genomics. 2012;13(1):253.

55. Kalluri UC, Difazio SP, Brunner AM, Tuskan GA. Genome-wide analysis of Aux/IAA and ARF gene families in *Populus trichocarpa*. BMC Plant Biol. 2007;7:59.

56. Barakat A, Bagniewska-Zadworna A, Choi A, Plakkat U, DiLoreto D, Yellanki P, et al. The cinnamyl alcohol dehydrogenase gene family in *Populus*: phylogeny, organization, and expression. BMC Plant Biol. 2009;9(1):26.

57. Kotak S, Larkindale J, Lee U, von Koskull-Doring P, Vierling E, Scharf KD. Complexity of the heat stress response in plants. Curr Opin Plant Biol. 2007;10(3):310–6.

58. Ikeda M, Mitsuda N, Ohme-Takagi M. *Arabidopsis* HsfB1 and HsfB2b act as repressors of the expression of heat-inducible Hsfs but positively regulate the acquired thermotolerance. Plant Physiol. 2011;157(3):1243–54.

59. Bharti K, von Koskull-Döring P, Bharti S, Kumar P, Tintschl-Körbitzer A, Treuter E, et al. Tomato heat stress transcription factor HsfB1 represents a novel type of general transcription coactivator with a histone-like motif interacting with the plant CREB binding protein ortholog HAC1. Plant Cell. 2004;16(6):1521–35.

60. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, et al. WoLF PSORT: protein localization predictor. Nucleic Acids Res. 2007;35 (Web Server issue):W585–7.

61. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4(4):406–25.

62. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011;28(10):2731–9.

63. Scharf K-D, Berberich T, Ebersberger I, Nover L. The plant heat stress transcription factor (Hsf) family: structure, function and evolution. Biochim Biophys Acta. 2012;1819(2):104–19.

64. Waters ER, Aevermann BD, Sanders-Reed Z. Comparative analysis of the small heat shock proteins in three angiosperm genomes identifies new subfamilies and reveals diverse evolutionary patterns. Cell Stress Chaperones. 2008;13(2):127–42.

65. Guo AY, Zhu QH, Chen X, Luo JC. GSDS: a gene structure display server. Yi Chuan. 2007;29(8):1023–6.

66. Cohen D, Bogeat-Triboulot M-B, Tisserant E, Balzergue S, Martin-Magniette M-L, Lelandais G, et al. Comparative transcriptomics of drought responses in *Populus*: a meta-analysis of genome-wide expression profiling in mature leaves and root apices across two genotypes. BMC Genomics. 2010;11(1):630.

67. Yuan Y, Chung JD, Fu X, Johnson VE, Ranjan P, Booth SL, et al. Alternative splicing and gene duplication differentially shaped the regulation of isochorismate synthase in *Populus* and *Arabidopsis*. Proc Natl Acad Sci U S A. 2009;106(51):22020–5.

68. Howe E, Holton K, Nair S, Schlauch D, Sinha R, Quackenbush J. MeV: multiExperiment viewer. In: Biomedical Informatics for Cancer Research. New York: Springer; 2010. p. 267–77.

69. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504.

70. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods. 2001;25(4):402–8.