Genome-wide analysis of the potato Hsp20 gene family: identification, genomic organization and expression profiles in response to heat stress

Peng Zhao1†, Dongdong Wang1†, Ruoqiu Wang1, Nana Kong1, Chao Zhang1, Chenghui Yang1, Wentao Wu1,2, Haoli Ma1* and Qin Chen1*

Abstract

Background: Heat shock proteins (Hsps) are essential components in plant tolerance mechanism under various abiotic stresses. Hsp20 is the major family of heat shock proteins, but little of Hsp20 family is known in potato (Solanum tuberosum), which is an important vegetable crop that is thermosensitive.

Results: To reveal the mechanisms of potato Hsp20s coping with abiotic stresses, analyses of the potato Hsp20 gene family were conducted using bioinformatics-based methods. In total, 48 putative potato Hsp20 genes (StHsp20s) were identified and named according to their chromosomal locations. A sequence analysis revealed that most StHsp20 genes (89.6%) possessed no, or only one, intron. A phylogenetic analysis indicated that all of the StHsp20 genes, except 10, were grouped into 12 subfamilies. The 48 StHsp20 genes were randomly distributed on 12 chromosomes. Nineteen tandem duplicated StHsp20s and one pair of segmental duplicated genes (StHsp20-15 and StHsp20-48) were identified.

A cis-element analysis inferred that StHsp20s, except for StHsp20-41, possessed at least one stress response cis-element. A heatmap of the StHsp20 gene family showed that the genes, except for StHsp20-2 and StHsp20-45, were expressed in various tissues and organs. Real-time quantitative PCR was used to detect the expression level of StHsp20 genes and demonstrated that the genes responded to multiple abiotic stresses, such as heat, salt or drought stress. The relative expression levels of 14 StHsp20 genes (StHsp20-4, 6, 7, 9, 20, 21, 33, 34, 35, 37, 41, 43, 44 and 46) were significantly up-regulated (more than 100-fold) under heat stress.

Conclusions: These results provide valuable information for clarifying the evolutionary relationship of the StHsp20 family and in aiding functional characterization of StHsp20 genes in further research.

Keywords: Genome-wide, Potato, Hsp20 gene family, Heat stress
called as small Hsp [9]. Hsp20 is the major family of heat shock proteins induced by elevated temperature-associated stress in plants [10, 11]. Hsp20 is encoded by a multigene family and is considered the most produced protein under heat stress conditions in many higher plants [12, 13].

Hsp20s are ATP-independent molecular chaperones and can form oligomeric protein complexes of 200–800 kDa, which consist of 9 to 50 subunits [14, 15]. Hsp20 can avert protein denaturation, and thus maintain the stability and normal functions of proteins in both eukaryotic and prokaryotic cells [6, 16]. The existing evidence suggests that Hsp20 plays an important role in plant heat tolerance. Hsp20s possess a conserved structure, consisting of a variable N-terminal region, a more conserved C-terminal region and a C-terminal extension [6]. The more conserved C-terminal region is usually named as the alpha-crystallin domain (ACD), which contains approximately 80 to 100 amino acid residues. The three different regions possess varied functions. The ACD functions in substrate interactions, while the N-terminal region participates in substrate binding and the C-terminal extension is responsible for homo-oligomerization [17–20]. The ACD contains two conserved regions, one in the N-terminal consensus region and the other is connected through a hydrophobic β6-loop at the C-terminal common region. The two conserved regions consist of 4 anti-parallel sheets and 3 β-strands respectively [16, 21]. Furthermore, unlike other Hsp families, the Hsp20 gene family exhibits extensive sequence variability and evolutionary divergence [22].

The number of plant Hsp20 genes is approximately four times greater than that of animals [10]. The Hsp20 gene family members have been investigated in many plants, such as Arabidopsis, rice, soybean, pepper and tomato. There are 19 Hsp20 genes in Arabidopsis [23], 39 in rice [24], 51 in soybean [25], 35 in pepper [26] and 42 in tomato [27]. Following maize, wheat and rice, potato is the fourth-largest food crop in the world. Potatoes are formed from underground stems through a process known as tuberization, but high temperatures inhibit the process and decrease the amount of photosynthetic product transported into the tubers, causing a large yield loss [28]. To date, the potato Hsp20 gene family members have not been identified and their functions under heat stress conditions remain to be elucidated. With the availability of the whole-genome sequence of potato, it is now possible to more fully study the potato Hsp20 gene family.

Here, we used bioinformatics methods to identify Hsp20 genes from potato genome, and analyze the sequence features, chromosomal locations, phylogenetic relationships, cis-elements, tissue-specific expression levels and dynamic expression patterns in response to different abiotic stresses, including heat stress. The results provide useful information for further functional investigations of the StHsp20 gene family.

**Methods**

**Identification of the Hsp20 family members in potato genome**

The whole potato protein sequence was downloaded from the Potato Genome Sequencing Consortium (PGSC, http://potato.plantbiology.msu.edu/integrated_searches.shtml). To identify potato Hsp20 candidates, the Hidden Markov Model (HMM) analysis was used for the search. We downloaded HMM profile of Hsp20 (PF00011) from Pfam protein family database (http://pfam.xfam.org/) and used it as the query (P<0.001) to search the potato protein sequence data [29]. To avoid missing probable Hsp20 members because of incomplete ACD domains, a BLASTP-algorithm based search using Arabidopsis Hsp20 amino acid sequences as queries was conducted with an e-value ≤ 1e⁻³. Additionally, keywords "Hsp20" and "small heat shock protein" were employed to search against PGSC database. After removing all of the redundant sequences, the output putative Hsp20 protein sequences were submitted to CDD (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi), Pfam and SMART (http://smart.embl-heidelberg.de/) to confirm the conserved Hsp20 domain. The predicted protein sequences lacking the Hsp20 domain or with a molecular weight outside of the 15–42-kDa range were excluded. All of the non-redundant and high-confidence genes were assigned as potato Hsp20s (StHsp20s). These StHsp20 genes were named on the basis of their positions on pseudomolecules [24].

**Sequence analysis and structural characterization**

All of the high-confidence Hsp20 sequences were submitted to ExPASy (http://web.expasy.org/protparam/) to calculate the number of amino acids, molecular weights and theoretical isoelectric points (pl). The chromosomal locations and intron numbers of StHsp20s were acquired through the PGSC. The MEME program (version 4.11.2, http://alternate.meme-suite.org/tools/meme) was used to identify the conserved motifs in the StHsp20s sequences, with the following parameters: any number of repetitions, maximum of 10 misfits and an optimum motif width of 6 - 200 amino acid residues. The exon–intron structures of the StHsp20 genes were identified on the Gene Structure Display Server (GSDS, http://gds.cbi.pku.edu.cn/) [30].

**Chromosomal localization and gene duplication**

The chromosomal positions of the StHsp20 genes were acquired from the potato genome browser at the PGSC. MapChart software [31] was used for the mapping of
\textit{StHsp20} genes’ chromosomal positions and relative distances. The \textit{StHsp20} gene duplication was confirmed based on two criteria: (a) the length of the shorter aligned sequence covered > 70\% of the longer sequence; and (b) the similarity of the two aligned sequences were > 70\% [32, 33]. Two genes separated by five or fewer genes in 100-kb chromosome fragment were considered as tandem duplicated genes [34]. The segmental duplicated genes of \textit{StHsp20} were identified by searching the segmental genome duplications of potato at the Plant Genome Duplication Database (PGDD, http://chibba.agnet.uga.edu/duplication/).

**Phylogenetic analysis and classification of potato \textit{Hsp20} genes**

The full-length amino acid sequences of Hsp20s (Additional file 1: Table S1) derived from \textit{Arabidopsis} [35], soybean [25], rice [24] and \textit{Populus} [36] combined with newly identified \textit{StHsp20}s were used for phylogenetic analysis. All of the acquired sequences were first aligned by ClustalX (version 1.83) software [37] with the default parameters. An unrooted neighbor-joining phylogenetic tree was constructed using MEGA6 software [38] with bootstrap test of 1000 times. The potato \textit{Hsp20} genes were classified into different groups according to the topology of phylogenetic tree and the classifications of \textit{Hsp20}s in four other species.

**Analysis of \textit{Cis}-acting element in \textit{StHsp20} genes’ promoters**

The upstream sequences (1.5 kb) of the \textit{StHsp20}-coding sequences were retrieved from the PGSC and then submitted to PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/; [39]) to identify six regulatory elements, abscisic acid (ABA)-responsive elements, involved in the ABA responsiveness; dehydration-responsive elements (DREs), involved in dehydration, low-temperature and salt stresses; heat stress elements (HSEs), involved in heat stress response; low temperature responsive elements (LTRE), involved in low-temperature response; TC-rich repeats, involved in defense and stress response; and W-boxes, binding site of WRKY transcription factor in defense responses.

**Plant materials and abiotic stress treatments**

The doubled monoploid (DM) potato was used in this study. All of the lines were cultured in Murashige and Skoog (MS) medium [40] containing 3\% sucrose and 0.8\% agar at pH 5.9. The plant material was sustained in an artificial climate chamber with 16 h light/8 h dark photoperiod and temperature of 22 ± 1 °C. The four-week-old plantlets were then transferred into cuvettes containing 1/2 MS liquid medium and maintained in an artificial growth chamber at 22 ± 1 °C (16 h light/8 h dark period) for a week before being subjected to an abiotic stress. For heat stress, the plantlets were exposed to 35 °C; for salt stress, the plantlets were incubated with 150 mM NaCl; and for drought stress, the plantlets were treated with 260 mM mannitol. Under these different stress conditions, the aboveground of whole plants were collected at 0, 3 and 24 h after treatments. All of the collected samples were froze in liquid nitrogen rapidly and stored at – 80 °C refrigerator before RNA extraction.

**RNA-sequencing (RNA-seq) data analysis of \textit{StHsp20} genes**

The Illumina RNA-seq data were downloaded from the PGSC to study the expression patterns of \textit{StHsp20} genes. The RNA-seq data (Additional file 2: Table S2) included various developmental stages, tissues and stress treatments. To render the data suitable for cluster displays, absolute FPKM values were divided by the mean of all of the values, and the ratios were transformed by log2. HemI [41] software was used to generate the heatmap.

**Total RNA extractions and expression analyses of potato \textit{Hsp20} genes**

Primer Premier 5 was used to design primers specific to the \textit{StHsp20} genes (Additional file 3: Table S3). Total RNA was extracted using an RNAsimple Total RNA Kit (BioTeke, Beijing, China). The cDNA was reverse-transcribed by First Strand cDNA Synthesis Kit, ReverTra Ace-α (TOYOBO, Shanghai, China). All of the operational procedures followed the manufacturer’s protocols. Before the qRT-PCR analysis, 1 μl cDNA was diluted with 4 μl nuclease-free water.

qRT-PCR was carried out using the KAPA SYBR FAST qPCR Kit Master Mix (2x) Universal (KAPA BIOSYSTEMS, Boston, United States) on a Bio-Rad CFX96 Real Time PCR System. Each PCR reaction was conducted in a 20-μl reaction volume containing 10 μl KAPA SYBR, 0.5 μl 10 μM solution of each primer, 1 μl diluted cDNA and 8 μl ddH₂O. The PCR program was set as follow: 95 °C for 2 min and 40 cycles of 95 °C for 5 s and 60 °C for 30 s. The melt curve was analyzed from 65 °C to 95 °C with increments of 0.5 °C every 5 s. For each sample, three biological repeats, with two technical replicates each, were performed to acquire reliable results. The housekeeping gene \textit{ef1a} was used as the internal reference gene. The synthetic cDNA was diluted to 3-, 9-, 27- and 81-fold to establish the standard curve for each \textit{StHsp20} gene and \textit{ef1a}. The relative expression levels of the \textit{StHsp20} genes were calculated using the standard curve and normalized by the control’s expression. The results were displayed by means ± standard deviation (SD).
Results

Identification and analysis of StHsp20 genes
A total of 58 Hsp20s were obtained by HMM analysis, 52 sequences were found by local BLASTP, and 35 sequences were acquired by keyword search against the PGSC database. After removing the repetitive sequences, 65 sequences were reserved and submitted to CDD, Pfam and SMART to confirm the ACD domain. Sequences without a typical ACD domain and with a molecular weight outside of the 15–42-kDa range were excluded. Finally, 48 sequences were confirmed as potato Hsp20 genes and named based on their chromosomal locations. Gene names, gene IDs, chromosomal locations, open reading frame lengths, exon numbers, amino acid numbers, molecular weights and pIs were listed in Table 1. The lengths of the StHsp20 proteins ranged from 133 (StHsp20-36) to 303 amino acids (StHsp20-15). The molecular weights of StHsp20s were between 15.3 kDa (StHsp20-36) and 34.0 kDa (StHsp20-15). StHsp20 genes were distributed on 12 potato chromosomes. The predicted pl values of StHsp20 ranged from 4.91 (StHsp20-5) to 9.88 (StHsp20-39).

StHsp20 gene structure
Structures and phases of introns/exons were determined by the alignment of genomic DNA with full-length cDNA of StHsp20s. Among the StHsp20 genes, nearly half (20, 41.7%) were intronless, 23 (47.9%) had one intron, and only 5 genes (10.4%), StHsp20-15 (12 introns), StHsp20-22 (5 introns), StHsp20-15 (12 introns), StHsp20-45 (8 introns) and StHsp20-48 (5 introns), had two or more introns (Fig. 1). Interestingly, all of the tandem duplicated genes were intronless and the pair of segmentally duplicated genes, StHsp15 and StHsp48, had multiple introns. StHsp48 was shorter than StHsp15 in sequence length, but shared a highly conserved region with StHsp15. The conserved region possessed the same motif patterns and intron pattern existed. According to previous research, three patterns were proposed. Pattern 1 means no intron, Pattern 2 means one intron, and Pattern 3 means more than one intron [24]. Most StHsp20 members of the CI subfamily lacked introns, and the CII and ER subfamilies had no introns. However, all of the members of the CV, CVI, CVII, Po, MI and MII subfamilies had one intron, which indicated a close phylogenetic relationship (Fig. 1; Table 1). In addition, three genes (StHsp15, StHsp45 and StHsp48) belonging to the CIII subfamily had 12, 8 and 5 introns, respectively (Fig. 1; Table 1). The presence of multiple introns indicated a particular phylogenetic status.

Phylogenetic analysis of StHsp20 genes
To analyze the evolutionary relationships of Hsp20 genes in potato, Arabidopsis, soybean, rice and Populus, an unrooted phylogenetic tree was constructed using full-length amino acid sequences. In total, 19 sequences from Arabidopsis, 22 sequences from rice, 47 sequences from potato, 46 sequences from soybean and 25 sequences from Populus were assessed in the phylogenetic tree (Fig. 2). The potato Hsp20 family member StHsp20-29 was excluded from the phylogenetic tree because it was too divergent to be aligned with other sequences. The 159 Hsp20s were classified into 12 distinct subfamilies, 71 cytosol Is (CIs), 13 CIIIs, 3 CIVs, 5 CVs, 3 CVIs, 3 CVIIIs, 5 mitochondria Is (MIs), 6 MIIs, 12 plastids (Ps), 6 peroxisomes (Pos) and 11 endoplasmic reticulum (ERs). However, the remaining 10 potato Hsp20s could not be clustered into any subfamily. Except for the unclassified StHsp20s, 37 StHsp20s existed in 11 subfamilies, except for the CIV subfamily. Most of the Hsp20s, including 29 StHsp20s, were classified into CI–CVII, which indicated that cytosol might be the main functional area for plant Hsp20s. Remarkably, StHsp20 members were more closely related to those in the same subfamily from different species than to the other Hsp20s from the same species, which implied a relatively high synteny between the same Hsp20 subfamily across various species. It was interesting that the P and M (MI and MII) subfamily members had a close relationship with each other, which indicated that the M subfamily evolved from the P subfamily once again [6]. No Hsp20 protein of monocotyledon (rice) was found in CIV subfamily. According to previous study [35], CIV subfamily of Hsp20s existed only in dicotyledon.

A close relationship between the phylogenetic classification and intron pattern existed. According to previous research, three patterns were proposed. Pattern 1 means no intron, Pattern 2 means one intron, and Pattern 3 means more than one intron [24]. Most StHsp20 members of the CI subfamily lacked introns, and the CII and ER subfamilies had no introns. However, all of the members of the CV, CVI, CVII, Po, MI and MII subfamilies had one intron, which indicated a close phylogenetic relationship (Fig. 1; Table 1). In addition, three genes (StHsp15, StHsp45 and StHsp48) belonging to the CIII subfamily had 12, 8 and 5 introns, respectively (Fig. 1; Table 1). The presence of multiple introns indicated a particular phylogenetic status.
| Name       | Gene ID            | Chr. | Genomic Location | ORF | Exon | AA  | MW (kDa) | pI   |
|------------|--------------------|------|------------------|-----|------|-----|--------|-----|
| StHsp20-1  | PGSC0003DMG400008713 | 1    | 6195006 - 6196992 | 732 | 2    | 243 | 26.8   | 5.80|
| StHsp20-2  | PGSC0003DMG400008714 | 1    | 6199128 - 6200118 | 669 | 2    | 222 | 24.5   | 8.91|
| StHsp20-3  | PGSC0003DMG400008715 | 1    | 6205818 - 6207478 | 639 | 2    | 212 | 23.8   | 9.30|
| StHsp20-4  | PGSC0003DMG400020718 | 1    | 3702817 - 3703750 | 576 | 1    | 191 | 21.8   | 6.87|
| StHsp20-5  | PGSC0003DMG400016460 | 2    | 3622323 - 36224237 | 414 | 2    | 137 | 15.7   | 4.91|
| StHsp20-6  | PGSC0003DMG400012619 | 2    | 4759303 - 47596399 | 426 | 2    | 141 | 16.3   | 8.32|
| StHsp20-7  | PGSC0003DMG40003219 | 3    | 46254502 - 46257461 | 702 | 2    | 233 | 25.9   | 6.98|
| StHsp20-8  | PGSC0003DMG400025350 | 3    | 53138385 - 53139192 | 651 | 2    | 216 | 24.5   | 4.98|
| StHsp20-9  | PGSC0003DMG400024476 | 3    | 53878612 - 53879610 | 571 | 1    | 191 | 21.9   | 6.87|

Table 1: Features of StHsp20 genes identified in potato
Chromosomal location and gene duplication of StHsp20s

The 48 StHsp20 genes were distributed on 12 potato chromosomes randomly (Fig. 3). The majority of StHsp20 genes were located on the proximate or the distal ends of the chromosomes. The maximum number of nine predicted StHsp20 genes, scattered in two clusters, were present on chromosome 9, and only one gene existed on chromosome 5.

During the progress of evolution, both tandem duplication and segmental duplication contribute to the generation of gene family [42]. Thus, we analyzed the duplication events of StHsp20 genes. Based on the

| Name     | Gene ID                | Chr. | Genomic Location    | ORF  | Exon | AA   | MW (kDa) | pI   |
|----------|------------------------|------|---------------------|------|------|------|----------|------|
| StHsp20-45 | PGSC0003DMG400047019 | 12   | 19504973 - 19513520 | 909  | 9    | 302  | 33.8     | 8.24 |
| StHsp20-46 | PGSC0003DMG400028624 | 12   | 49808899 - 49810200 | 690  | 2    | 229  | 26.2     | 8.44 |
| StHsp20-47 | PGSC0003DMG402010796 | 12   | 54506438 - 54507366 | 750  | 2    | 249  | 28.2     | 8.23 |
| StHsp20-48 | PGSC0003DMG400029311 | 12   | 57570831 - 57572324 | 423  | 6    | 140  | 15.5     | 6.19 |

Zhao et al. BMC Genomics (2018) 19:61
Page 6 of 13
defined criteria, 19 genes (39.6%) were confirmed to be tandem duplicated genes. Two separate pairs of tandem duplicated genes located on chromosome 10 and chromosome 12. Two groups of three tandem duplicated genes located on chromosome 1 and 8. Five and four tandem duplicated genes located on chromosome 6 and 9, separately. Additionally, two genes (StHsp20-15 and StHsp20-48) were segmentally duplicated genes, and the length of segmentally duplicated chromosome was 625 kb. Segmental duplication only accounted for 4.2% of the StHsp20 genes. Based on above results, it could be inferred that tandem duplication and segmental duplication contribute to the expansion of StHsp20 family together, but the former played a predominant role.

**Stress-related cis-elements in StHsp20 promoters**

To further study the potential regulatory mechanisms of StHsp20 during abiotic stress responses, the 1.5-kb upstream sequences from the translation start sites of StHsp20 genes (promoter regions of StHsp20-2, StHsp20-11, StHsp20-15 and StHsp20-32 were absent)

| Motif | Width | Best possible match |
|-------|-------|---------------------|
| 1     | 113   | FPPSSRETSAFANTRIDKET PEAKVVDVPGILKHEEKYVE VEEDRVAQISGERSREKEKND KWIHRVSSGFMRFRLPEN AKMDQIKASMENGVTYVTVP KEIEKCP |
| 2     | 20    | DVDIKKAMENGVLTVTVPK |
| 3     | 19    | ADLPGKEDVKVQVEDNG |
| 4     | 15    | SNFDPSLDVFDFP |
| 5     | 11    | MSLIPRFGGGR |
| 6     | 45    | TDDATGTASVLAGIKMPRKM YMNLTVALLVLGILVSNKMK |
| 7     | 15    | RSYGKFSTSNFLPEN |
| 8     | 21    | VYEDPVSTELQEDASYTL |
| 9     | 8     | VKSIDSIG |
| 10    | 41    | PSHEYLETYTSNAPSLSFPHVP QYMAQIEYKTEPEAHIF |

**Fig. 2** Phylogenetic tree of Hsp20s from Arabidopsis, Populus, soybean, rice and potato. The phylogenetic tree was constructed using the NJ (Neighbor-joining) method with 1000 bootstrap replications. The 12 subfamilies were distinguished in different colors, and the unclassified StHsp20s were colored in purple.
were submitted into PlantCARE to detect the cis-elements. Six abiotic stress response elements, ABA-responsive elements, DRE, HSE, LTRE, TC-rich repeat and W-box, were analyzed and displayed in Fig. 4. Except for StHsp20-23 and StHsp20-41, the other StHsp20s possessed at least 1 stress-response-related cis-element, which indicated that the expressions of StHsp20s were associated with these abiotic stresses. In total, 32 StHsp20s (72.8%) had one or more HSEs, suggesting a potential heat-stress response under high temperature conditions. One to two LTREs existed in 11 StHsp20s, and 1 DRE was found in StHsp20-33. TC-rich repeats and W-boxes were located in 34 and 13 StHsp20s, respectively. Anyhow, the cis-element analysis illustrated that StHsp20 genes could respond to different abiotic stresses.

Expression patterns of StHsp20 genes in different tissues

Using the RNA-seq data, a heatmap of 48 StHsp20 genes, represented by FPKM values in different tissues and organs, was established by HemI (Fig. 5). Most of StHsp20s were expressed in one tissue at least, except for StHsp20-2 and StHsp20-45, which were barely expressed in any tissue or organ. Six genes including StHsp20-18, StHsp20-24, StHsp20-25, StHsp20-26, StHsp20-29 and StHsp20-30, were highly expressed in all of the tissues. Some StHsp20 genes showed similar expression patterns in various tissues. StHsp20-1, StHsp20-3, StHsp20-12, StHsp20-40 and StHsp20-48 showed relatively high expression levels in vegetative organs, such as shoots, stolons and petioles, but undetectable levels in leaves, sepals, stamens, flowers and petals. StHsp20-6, StHsp20-9, StHsp20-19, StHsp20-33, StHsp20-34, StHsp20-35, StHsp20-41 and StHsp20-43 were highly expressed in callus. StHsp20-7, StHsp20-8, StHsp20-11, StHsp20-27 and StHsp20-32 exhibited high expression levels in shoots and callus.

Expression profiles of StHsp20s under abiotic stress

To further explore the expression changes in the StHsp20 genes under various abiotic stresses including heat, salt and drought, qRT-PCR was used to investigate the transcript levels of each StHsp20 gene with 3 biological repetitions and 2 technical repetitions. Generally, the relative expression level of the StHsp20 genes under all stress conditions fluctuated during the 24-h treatments (Fig. 6). The relative expression level of StHsp20-45 was not shown because the non-specific primers may lead to unreliable results. Most of the StHsp20 genes
were sensitive to heat stress, and none of the genes were down-regulated, but StHsp20-29 and StHsp20-30 showed no differences after being treated for 3 h and 24 h under heat stress. The expression levels of StHsp20-10 and StHsp20-13 were up-regulated only after a 24-h heat treatment. The relative expression levels of 14 StHsp20 genes (StHsp20-2, StHsp20-11, StHsp20-15 and StHsp20-32 were absent) were analyzed by PlantCARE. The upstream length to the translation start site can be inferred according to the scale at the bottom.

Although the Hsp20 family is generally induced by heat stress, we also determined whether the family is involved in responses to salt and drought stresses. The expression levels of StHsp20 genes under salt and drought stresses varied among the 47 members. The expression pattern of each StHsp20 was different from that under heat stress. Nearly half of the StHsp20 (40.4%) genes were down-regulated after being treated for 3 h or 24 h. Six genes (StHsp20-11, 14, 15, 23, 30 and 40) and 10 genes (StHsp20-4, 6, 9, 10, 11, 14, 30, 36, 44 and 46) were not sensitive to salt and drought stresses, respectively. The remaining StHsp20s were up-regulated under salt and drought stresses, but the changes were not as extreme as that under heat stress. The differential expression patterns compared with those under heat stress indicated there were different response and regulatory mechanisms of the StHsp20 family under various abiotic stress conditions.

RNA-seq data of StHsp20 under abiotic stress after treated for 24 h was collected from PGSC and processed to compare the expression abundance with that of qRT-PCR. The relative expression level was represented by stress/control (Additional file 4:...
However, two sections of the results were not completely in accordance with each other. Under heat stress, only 10 of the StHsp20 genes showed high expression level while 20 StHsp20s showed low expression level. High expression levels were confirmed in nearly half of StHsp20s under salt and drought stresses, meanwhile 3 and 2 StHsp20s exhibited low expression level respectively. In a word, compared the two set of results from RNA-seq and qRT-PCR, 4 genes (StHsp20-5, 10, 13 and 22) showed a similar expression pattern under 3 abiotic stresses, and 5 genes (StHsp20-7, 15, 29, 42 and 47) had similar expression pattern under salt and drought stress.

Discussion

Hsp20s, as molecular chaperone, inhibit the irreversible aggregation of denaturing proteins, thus enhance the thermotolerance of plant [16]. With the availabilities of the whole genome sequence of many plants, several Hsp20 families have been identified, such as Arabidopsis, rice, Populus, pepper and tomato [23, 26, 27, 36]. However, little is known about Hsp20 family in potato.

The current study identified 48 StHsp20 genes, and analyzed their structure, chromosomal location, phylogeny, gene duplication, stress-related cis-elements and expression patterns in different tissues and abiotic stresses. The study provides comprehensive information.
on the StHsp20 gene family and will aid in understanding the functional divergence of Hsp20 genes in potato.

Previous research identified 19, 39, 35 and 42 Hsp20 genes in Arabidopsis, rice, pepper and tomato, respectively [24, 26, 27, 35]. The low number of Hsp20 genes in Arabidopsis is related to its small genome. Forty-eight Hsp20 genes were identified in potato, which was close to the numbers found in pepper and tomato, which also belong to Solanaceae.

Gene organization plays a vital role in the evolution of multiple gene families [43]. In this study the percentage of intronless StHsp20 genes is similar to that of pepper (45.71%) [26] and tomato (30.95%) [27]. Additionally, StHsp20 genes of the CII and ER subfamilies, as well as most StHsp20 genes of the CI subfamily, were intronless (Fig. 2, Table 1). Members of the CV, CVI, MI, MII, P and Po subfamilies had only one intron. The results are also in accordance with that in pepper and tomato. Additionally, similar motif arrangements were found in the same subfamily members (Figs. 1c, 2). This correlation between intron numbers and motif arrangement further confirmed the classifications of the StHsp20 genes. In some studies, genes with few or no introns were considered to have enhanced expression levels in plants [44, 45]. To response to various stresses timely, genes must be rapidly activated, which would be assisted by a compact gene structure with less introns [46]. Most of the StHsp20 genes were highly induced under heat stress (Fig. 6), which may approve the above standpoints in other research.

In earlier studies, Arabidopsis Hsp20 genes were classified into seven subfamilies (CI, CII, CIII, M, P, ER and Po), and five genes could not be clustered into any subfamily [23]. Subsequently, four new nucleocytoplasmic subfamilies (CIV, CV, CVI and CVII) and a mitochondrial subfamily (MII) were identified [35]. In our study, the phylogenetic tree showed that Hsp20 genes were classified into 12 distinct subfamilies. The StHsp20 genes existed in 11 of the 12 subfamilies. There was no Hsp20 gene of potato in the CIV subfamily, which may be the result of gene loss during evolution.

Most of the StHsp20 genes (61.7%) were grouped into a nucleocytoplasmic subfamily, which was also illustrated in Arabidopsis, pepper and tomato [23, 26, 27]. Among these subfamilies, CI was the largest subfamily, containing 18 StHsp20 genes. Based on these results, we inferred that, because proteins are mainly synthesized in the cytoplasm, this could be the primary place for Hsp20 proteins to interact with denatured proteins, preventing inappropriate aggregation and degradation. Furthermore, the Hsp20 genes in the same subfamily from different species were more similar than those of the same species but belonging to various subfamilies. The finding indicated that synteny might exist in Arabidopsis, Populus, rice and soybean Hsp20 proteins, and that Hsp20 subfamilies diversified before the divergence within these species.

The expansions of gene families and genome evolutionary mechanisms mainly depend on gene duplication events [47]. The major duplication patterns are tandem duplication and segmental duplication [48]. In this research, 48 StHsp20 genes were located unevenly on 12 potato chromosomes, and most of the StHsp20 genes were located on the terminal regions of the chromosomes. Although the genome size of potato is almost 7 times that of Arabidopsis, the number of Hsp20 genes in potato (48 genes) is only 2.5 times that in Arabidopsis (19 genes). This could be the result of different whole genome duplication events in Arabidopsis and potato. A total of 21 StHsp20 duplicated genes were detected in potato, including one pair of segmentally duplicated genes (StHsp20-15 and StHsp20-48) and four tandem duplicated gene groups (Fig. 3), which revealed that both tandem and segmental duplications contributed to the evolution of Hsp20 genes in potato. Similar expression patterns under various abiotic stresses were found within the tandem duplicated gene groups (Fig. 6). The similar expression patterns indicated the analogous functions and structures of tandem duplicated StHsp20 genes. The redundancies of functions and similarities of structures may reflect shared induction mechanisms.

The expression patterns of Hsp20 genes in different tissues have been described in many species, such as Arabidopsis, rice, pepper and tomato [24, 26, 27, 35]. There is no uniform gene expression pattern for plant Hsp20 genes. According to the RNA-seq data of potato, several StHsp20 genes such as StHsp20-22 and StHsp20-41, exhibited incongruous expression patterns in various tissues, indicating that different StHsp20 proteins may have diverse functions. Three genes, StHsp20-18, StHsp20-26 and StHsp20-30, were highly and indiscriminately expressed in all of the investigated tissues under normal condition. Similar with several Hsp20 genes in soybean, the three StHsp20s showed specific housekeeping expression activity [25].

qRT-PCR was used to investigate the transcript levels of each StHsp20 under different abiotic stresses. The two genes (StHsp20-29 and StHsp20-30) with distinctive expression patterns were highly expressed in all of the investigated tissues, but no induction was observed under heat stress. Thus, we may assume that the two genes are lacking of chaperone activities. The results confirmed the association of potato Hsp20 proteins with thermostolerance; however, the existence of numerous Hsp20s may lead to functional redundancy [6]. In addition, similar expression patterns in StHsp20 genes may be caused by shared induction mechanisms. Because the heat shock response network involves heat shock proteins and heat
stress-related transcription factors (Hsfs), the expression levels of Hsp20 genes rely heavily on the activation of Hsfs under heat stress. During a 24-h heat treatment, the StHsp20 genes showed different transcript accumulation levels. It was reported that the same set of Hsps could be regulated by different Hsfs on transcription level [49, 50], which indicated that StHsp20 genes are specifically controlled by various Hsfs. The differences in transcription levels of StHsp20s may be the reflection of different upstream regulating genes of Hsfs.

Based on qRT-PCR, all of the StHsp20 genes responded to salt and drought stress; however, the expression level of several StHsp20s was down-regulated (Fig. 6). Under heat stress, Hsfs are activated and bound to HSEs in the Hsp20 gene promoters to regulate the expressions of downstream genes. Nevertheless, various cis-elements were found in promoter regions of StHsp20s (Fig. 4), and these are involved in the responses of StHsp20 genes to other abiotic stresses. Thus, StHsp20 genes could be induced by both heat stress and other abiotic stresses. The multiple abiotic stress responses of StHsp20 genes reflected an interconnected induction mechanism involving Hsf transcription factors.

Compared with expression pattern represented by RNA-seq data, the expression profile generated by qRT-PCR was not completely equal to that. The difference of expression pattern may be caused by multiple reasons. Although the same plant material (DM) was used for research, only aboveground part of plant was collected in our research, while the whole plant was sampled for RNA sequencing. Specific to heat stress, the plant was treated for 24-h in normal photoperiod of 16 h light/8 h dark in our study, but the plant for RNA sequencing was treated in the dark. The potato RNA-seq data used in our research was presented as FPKM. Compared with raw read counts, FPKM value can better reduce sample differences. However, the FPKM value could be significantly changed due to highly expressed genes [51]. The bias of FPKM value leads to different expression compared with qRT-PCR.

Conclusions
Here, a genome-wide analysis of potato Hsp20 family was performed, and 48 StHsp20 genes were confirmed. Subsequently, analyses of StHsp20 genes on gene structures, phylogeny, chromosomal location, gene duplication, stress-related cis-elements, expression patterns in different tissues and abiotic stresses, were conducted based on bioinformatics and qRT-PCR methods. Most of StHsp20 genes were sensitive to heat stress and were up-regulated rapidly, indicating that StHsp20 genes play important roles in the acquired thermostolerance of potato. The study provides comprehensive information on the StHsp20 gene family in potato and will aid in determining the StHsp20 gene functions.
References

1. Giorno F, Wolters-Arts M, Grillo S, Scharf KD, Vriezen WH, Mariani C. Developmental and heat stress-regulated expression of HspA2 and small heat shock proteins in tomato anthers. J Exp Bot. 2009;60(12):453–62.

2. Wang W, Vinicur B, Altman A. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. Planta. 2003;218(1):11–14.

3. Mittler R. Abiotic stress, the field environment and stress combination. Trends Plant Sci. 2006;11(15):15–9.

4. Herman DJ, Knowles LO, Knowles NR. Heat stress affects carbohydrate metabolism during cold-induced sweetening of potato (Solanum tuberosum L.). J. Plant. 2017;2(5):563–82.

5. De MA. Heat shock proteins: facts, thoughts, and dreams. Shock. 1999;1(1):1.

6. Waters ER. The evolution, function, structure, and expression of the plant shsps. J Exp Bot. 2013;64(2):391.

7. Mogk A, Bukaú B. Role of shsps in organizing cytosolic protein aggregation and disaggregation. Cell Stress Chaperon. 2017;22(4):495–502.

8. Wang W, Vinicur B, Shoyeov O, Altman A. Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends Plant Sci. 2004;9(5):244–52.

9. Sarkan NK, Kim YK, Grover A. Rice shs proteins: genomic organization and expression profiling under stress and development. BMC Genomics. 2005;10(1):393.

10. Waters ER, Lee GJ, Vierling E. Evolution, structure and function of the small heat shock proteins in plants. J Exp Bot. 1996;47(3):325–38.

11. Sung DY, Kaplan F, Lee KJ, Guy CL. Acquired tolerance to temperature extremes. Trends Plant Sci. 2003;8(4):179–87.

12. Vierling E. The roles of heat shock proteins in plants. Annu Rev Plant Biol. 2003;54(1):579–620.

13. Chang YI, Liu HC, Liu NY, Hsu FC, Ko SS. (2006). Arabidopsis Hsas2, a novel heat shock protein, is essential for acquired thermotolerance during long recovery after acclimation. Plant Physiol. 2006;140(4):307–305.

14. Lee GJ, Vierling E. A small heat shock protein cooperates with heat shock protein 70 systems to reactivate a heat-denatured protein. Plant Physiol. 2000;122(1):189–98.

15. Cashikar AG, Duennwald M, Lindquist SL. A chaperone pathway in protein disaggregation. Hsp26 alters the nature of protein aggregates to facilitate reactivation by Hsp104. J Biol Chem. 2005;280(25):23869–75.

16. Haslebeck M, Vierling E. A first line of stress defense: small heat shock proteins and their function in protein homeostasis. J Mol Biol. 2015;427(1):1537–48.

17. Kirschner M, Winkeleau S, Thierfelder JM, Nover L. Transient expression and heat-stress-induced co-aggregation of endogenous and heterologous small heat-stress proteins in tobacco protoplasts. Plant J. 2002;30(2):417–24.

18. Lee GJ, Vierling E. A small heat shock protein cooperates with heat shock protein 70 systems to reanimate a heat-denatured protein. Plant Physiol. 2000;122(1):189–98.

19. Cashikar AG, Duenwald M, Lindquist SL. A chaperone pathway in protein disaggregation. Hsp26 alters the nature of protein aggregates to facilitate reactivation by Hsp104. J Biol Chem. 2005;280(25):23869–75.

20. Haslebeck M, Vierling E. A first line of stress defense: small heat shock proteins and their function in protein homeostasis. J Mol Biol. 2015;427(1):1537–48.

21. Kirschner M, Winkeleau S, Thierfelder JM, Nover L. Transient expression and heat-stress-induced co-aggregation of endogenous and heterologous small heat-stress proteins in tobacco protoplasts. Plant J. 2002;30(2):417–24.

22. Lee GJ, Vierling E. A small heat shock protein cooperates with heat shock protein 70 systems to reanimate a heat-denatured protein. Plant Physiol. 2000;122(1):189–98.

23. Kirschner M, Winkeleau S, Thierfelder JM, Nover L. Transient expression and heat-stress-induced co-aggregation of endogenous and heterologous small heat-stress proteins in tobacco protoplasts. Plant J. 2002;30(2):417–24.