De novo assembly and characterization of Muscovy duck liver transcriptome and analysis of differentially regulated genes in response to heat stress

Tao Zeng · Liping Zhang · Jinjun Li · Deqian Wang ·
Yong Tian · Lizhi Lu

Received: 16 November 2014 / Revised: 28 December 2014 / Accepted: 14 January 2015 / Published online: 9 February 2015
© Cell Stress Society International 2015

Abstract High temperature is a major abiotic stress limiting animal growth and productivity worldwide. The Muscovy duck (*Cairina moschata*), sometimes called the Barbary drake, is a type of duck with a fairly unusual domestication history. In Southeast Asia, duck meat is one of the top meats consumed, and as such, the production of the meat is an important topic of research. The transcriptomic and genomic data presently available are insufficient to understanding the molecular mechanism underlying the heat tolerance of Muscovy ducks. Thus, transcriptome and expression profiling data for this species are required as important resource for identifying genes and developing molecular marker. In this study, de novo transcriptome assembly and gene expression analysis using Illumina sequencing technology were performed. More than 225 million clean reads were generated and assembled into 36,903 unique transcripts with an average length of 1,135 bp. A total of 21,221 (57.50 %) unigenes were annotated. Gene Ontology (GO) analysis of the annotated unigenes revealed that the majority of sequenced genes were associated with transcription, signal transduction, and apoptosis. We also performed gene expression profiling analysis upon heat treatment in Muscovy ducks and identified 470 heat-response unique transcripts. GO term enrichment showed that protein folding and chaperone binding were significant enrichment, whereas KEGG pathway analyses showed that Ras and MAPKs were

activated after heat stress in Muscovy ducks. Our research enriched sequences information of Muscovy duck, provided novel insights into responses to heat stress in these ducks, and serve as candidate genes or markers that can be used to guide future efforts to breed heat-tolerant duck strains.

Keywords Muscovy duck · Heat stress · Gene expression · RNA-seq · Transcriptome

Introduction

The Muscovy duck (*Cairina moschata*), sometimes called the Barbary drake, is a type of duck with a fairly unusual domestication history. Muscovy ducks are native to the tropical climates of Central and South America, Mexico, and extreme southern Texas in the USA. Muscovy ducks were intentionally released in the USA by private individuals, businesses, and governmental organizations as an ornamental species believed to enhance the aesthetic appeal of urban parks and lakes. In Florida, Muscovy ducks have been reported in nearly all 67 counties. These large ducks are extremely common in urban parks and near lakes, ponds, and streams because people often feed them at these sites. In Southeast Asia, duck meat is one of the top meats consumed, and as such, the production of the meat is an important topic of research (Aronal et al. 2012). In China, Muscovy breeds are popular because they have stronger tasting meat than the usual domestic ducks, which are descendants of the Mallard (*Anas platyrhynchos*). Additionally, the meat is lean when compared to the fatty meat of mallard-derived ducks; its leanness and tenderness are often comparable to that of veal (Saez et al. 2009). The carcass of a

Electronic supplementary material The online version of this article (doi:10.1007/s12192-015-0573-4) contains supplementary material, which is available to authorized users.

T. Zeng · L. Zhang · J. Li · D. Wang · Y. Tian · L. Lu (✉)
Institute of Animal Husbandry and Veterinary Medicine, Zhejiang
Academy of Agricultural Sciences, Hangzhou 310021, China
e-mail: lulizhibox@163.com

L. Zhang
College of Chemistry and Life Sciences, Zhejiang Normal
University, Jinhua 321004, China

Muscovy duck is also much heavier than most other domestic ducks, which makes it an ideal food source.

Despite the economic and commercial importance of Muscovy duck, there have been few genomics studies on this bird species. There are only 1,158 nucleotide sequences available in the NCBI GenBank as of 21 September 2014 (<http://www.ncbi.nlm.nih.gov/GenBank/>). The lack of sequence information for species of duck family has greatly delayed their research at the molecular level.

Warming of the climate system is unequivocal, and since the 1950s, many of the observed changes are unprecedented over decades to millennia. Until now, there are about 30,000 papers in PubMed focusing on heat stress response. Many questions regarding the function of molecular chaperones and heat-shock response in general have been answered in the last few decades. Still, the accurate regulation of heat-shock response is not comprehensively understood. High temperature can retard animal or plant growth, development, and yield; therefore, agriculture will be seriously affected by global warming in the future (Sohail et al. 2012; Li et al. 2013). The Muscovy duck was mostly bred in the southern China, where the temperature was much higher than the northern part. Knowledge of heat-responsive genes and proteins is therefore critical for further understanding the molecular mechanisms of stress tolerance in this species.

Massively parallel sequencing including next generation sequencing (NGS) techniques like 454 pyrosequencing/Roche, Illumina/Solexa GAIIX, ABI/SOLiD, Pac Biosciences/PacBioRS, and Helicos Biosciences/tSMS, and DRS have been widely applied for transcriptome analysis (Ma et al. 2012; Fang et al. 2012; Santos et al. 2012). These efficient, sensitive, reliable, and cost-effective sequences technologies do not require prior knowledge of genomic sequence and have been used to characterize the transcriptomes of animal for gene discovery, marker development, and understanding gene regulatory networks of important biological processes (Martin et al. 2013; Edwards et al. 2012; Bhardwaj et al. 2013; Frey et al. 2014; Tan et al. 2013).

In this study, we conducted transcriptome analysis using Illumina sequencing technology for the purposes of (1) enriching the genomics resource of Muscovy ducks and (2) providing a deeper understanding of the molecular mechanisms regarding Muscovy duck responses to heat stress. More than 60 million reads from normal and heat-treated Muscovy ducks were obtained and assembled into 36,903 unique sequences. We then compared global expression profiles of Muscovy duck liver between normal and heat stress conditions and identified a number of heat stress-responsive genes. This study will provide novel insights into the molecular mechanism that underlies the heat stress response as well as candidate genes or markers that can be used to guide future efforts to breed heat-resistant duck strains.

Methods

Animal and stress treatment

The animal care and used protocol was approved by the Institutional Animal Care and Use Committee of Nanjing Agricultural University (Nanjing, China). Muscovy ducks were raised in Yuyao, Zhejiang Province, China. Animal stress was performed as described previously (Zeng et al. 2013). The animals were kept in a temperature-controlled room with flowing air and sufficient water and fed a standard diet three times daily during the experimental period. For the hyperthermia challenge, ducks were maintained at 20 °C for 2 weeks before the experiments. The temperature was increased at 10 °C h⁻¹ from 20 to 39±0.5 °C. Experiment group ducks were kept at 39±0.5 °C for 1 h while the control group ducks were kept at 20 °C for the same time. Each group has two replicates. Livers were collected from each individual in all groups and stored at -80 °C until analyzed.

RNA isolation and library construction for transcriptome analysis

Total RNA extraction, mRNA purification, and cDNA library construction were conducted by LC Sciences (Houston, TX, USA). Briefly, total RNA was extracted using Trizol reagent (Invitrogen, Carlsbad, CA, USA) following the manufacturer's procedure. Total RNA quantity and purity were analyzed using Bioanalyzer 2100 and RNA 6000 Nano LabChip Kit (Agilent, Palo Alto, CA, USA) with RIN >8.0. Approximately 10 µg of total RNA of each liver sample was used to isolate Poly(A) mRNA with poly-T oligo-attached magnetic beads (Invitrogen). Following purification, the mRNA is fragmented into small pieces using divalent cations under elevated temperature. Then the cleaved RNA fragments were reverse-transcribed to create the final cDNA library in accordance with the protocol for the mRNA-Seq Sample Preparation Kit (Illumina, San Diego, CA, USA), and the average insert size for the paired-end libraries was 300 bp (±50 bp). Subsequently, we performed the paired-end sequencing (100 bp) on an Illumina HiSeq 2500 at the LC-BIO following the vendor's recommended protocol.

De novo assembly and annotation

Before assembly, high-quality clean reads were generated from the raw reads by removing adapter sequences, duplicated sequences, low-quality reads with ambiguous bases ("N" >5 %) and reads with more than 10 % of *Q* value <20 bases. All subsequent analyses were based on clean reads. The resulting high-quality cleaned reads were assembled de novo into contigs using Trinity with strand-specific option "-SS_lib_type" set to "F" and "min_kmer_cov" set to 2

(Grabherr et al. 2011). The overlapped reads were first connected to form contigs, and then mapped back to the contigs so that the different contigs and their distances from the same transcript were estimated and assembled to form the longer sequence, which were unigenes. The contigs were again mapped back to the unigenes to obtain the final unigenes that could not be extended on either end.

BLASTX alignment was conducted with a cutoff E value of 10^{-5} between unigenes and Nr (non-redundant), Swiss-Prot (a manually annotated and reviewed protein sequence database), Pfam (protein family), KEGG (Kyoto Encyclopedia of Genes and Genomes), and COG (Clusters of Ortholog Groups). GO terms were assigned to the Muscovy duck assembled transcript based on the GO terms annotated to their corresponding homologues. To predict and classify possible functions, unigenes were also compared against the COG database. Biochemical pathways were predicted from the Muscovy duck transcripts using the Pathway Tools.

Protein coding region prediction (CDS) and EST-SSRs marker identification

The nucleic acid sequences and amino acid sequences of protein-coding region (CDS) were predicted by using Genescan. MISA (<http://pgsc.ipk-gatersleben.de/misa/>) was used to identify the potential SSR makers in all of the unique sequences. More than six di-nucleotides repeats and more than five tri-nucleotide, tetra-nucleotide, penta-nucleotide, and hexa-nucleotide repeats were used as the search criteria for SSRs in MISA script (Zhou et al. 2012).

Quantitative real-time PCR analysis

A portion (1 μ g) of the total RNA obtained from each extraction was reverse-transcribed in a 20- μ L reaction volume by using the TransScript First-Strand cDNA Synthesis SuperMix (TransGen, Beijing, China) following the manufacturer's instruction. The qPCR was performed on an ABI 7300 (Applied Biosystems, Foster City, CA, USA). Reactions were performed in a 20- μ L reaction mixture containing 2 μ L cDNA template, 0.4 μ M forward/reverse primer, 10 μ L 2 $\times$ SYBR qPCR Mix, and 0.4 μ L ROX reference dye (Takara, Osaka, Japan). The cycling protocol included an initial step at 94 $^{\circ}$ C for 3 min, followed by 40 cycles of denaturation at 94 $^{\circ}$ C for 10 s, and annealing at 60 $^{\circ}$ C for 30 s. Experiments for the detection of all the genes, including β -actin, were performed in triplicates. The relative expression levels of the genes tested were calculated using the $2^{-\Delta\Delta C_t}$ method. The data are expressed as means $\pm$ standard error. Results were considered statistically significant at $P < 0.05$. All of the primers used in this study are listed in Additional file 1.

Results and discussions

Sequencing, de novo assembly, and homology search

Strand-specific RNA-seq libraries were prepared from livers of Muscovy ducks (*C. moschata*) under normal and heat stress conditions. For each condition, two replicates were performed. Additionally, each RNA-seq library was sequenced on an Illumina HiSeq 2500 system. Due to the absence of reference genomic sequences, de novo assembly was applied to construct transcripts from these RNA-seq reads. We used Trinity software for the de novo assembly of the Illumina reads, which has been demonstrated to be efficient for de novo reconstruction of transcriptomes from RNA-seq data (Iyer and Chinnaiyan 2011; Grabherr et al. 2011). After removing low quality, cutting adapter, and filtering junk, a total of 225.74 million clean reads were obtained, corresponding to 22.75 Gb clean data. De novo assembly of these high-quality cleaned reads generated 67,880 unique transcripts with an average length of 1,135 bp and the total length of 25,288,989 bp, as well as 36,903 unigenes with an average length of 1,011 bp and the total length of 15,378,495 bp (Table 1). Of the 36,903 unigenes, 17,219 unigenes (46.67 %) were between 200 and 500 bp; 8,238 unigenes (22.32 %) ranged from 501 to 1,000 bp; and 11,446 unigenes (31.01 %) were larger than 1,000 bp. The length distribution of the assembled unigenes is shown in Fig. 1.

Homology search for the sequences obtained after clustering was conducted using BLASTX against protein sequences Nr databases at NCBI. We determined the distribution of top hit species and found that the majority of the annotated sequences corresponded to known nucleotide sequences of bird species, with *A. platyrhynchos* (44.0 %) followed by *Gallus gallus* (18.2 %), *Ficedula albicollis* (5.4 %), and *Melopsittacus undulatus* (4.9 %) (Fig. 2). Normally, almost all sequences of Muscovy duck are similar with *A. platyrhynchos*, but the sequences information of *A. platyrhynchos* are not integral in the NCBI.

Functional annotation and classification

Protein functions can be predicted from annotation of the most similar proteins in Nr, Swiss-Prot, Pfam, KEGG, COG, and GO databases. We matched unigene sequences against three protein databases, Nr, Swiss-Prot, and Pfam, and obtained 21,221, 19,143, and 11,362 unigenes, respectively (Table 2). Distinct gene sequences were first searched using BLASTX against the Nr database using a cutoff E value of 10^{-5} (Altschul et al. 1990; Jiang et al. 2013).

GO functional annotation was according to the Nr annotation. Of the 21,221 annotated unigenes, 16,596 sequences (44.97 %) were assigned with one or more terms. These 16,596 unigenes were categorized into 44 GO functional groups,

Table 1 Summary for Muscovy duck transcriptome

	All	Min length	Median length	Mean length	N50	Max length	Total length
Transcript	67,880	201	640	1,135	2,023	17,938	77,045,990
Unigenes	36,903	201	544	1,011	1,844	17,938	37,329,376

which are distributed under the three main categories: biological process (8,467), cell components (14,927), molecular function (12,563) (Additional file 2). The top 10 groups in the three main categories are shown in Fig. 3. Within the biological process category, transcription, signal transduction, and apoptosis were the most enriched. Under the cellular components category, the nucleus, integral to membrane, and cytoplasm were the most highly represented GO terms. In the category of molecular function, the most abundant groups included ATP binding, zinc ion binding, and protein binding, and other appealing groups included DNA binding, metal ion binding, and RNA binding. It has been suggested that the genes showing high representation for these processes are from metabolically active developing tissues and are different in function. These processes could be involved in gene regulation and cell defense response via stress signaling pathways, which might be abiotic stress activated (Xu et al. 2013; Bhardwaj et al. 2013; Edwards et al. 2012; Utsumi et al. 2012).

In addition to GO analysis, COG analysis was used to further evaluate the function of the assembled unigenes. A total of 18,284 unigenes with COG annotations were grouped into 24 functional categories (Additional file 3). Among the 24 COG categories, the cluster for “signal transduction mechanisms” (15.50 %) represented the largest group, followed by

“General function prediction only” (15.27 %) and “Posttranslational modification, protein turnover, chaperones” (7.48 %). Only a small portion of the unigenes were assigned to “Cell motility” (0.42 %) or “Cell wall/membrane/envelope biogenesis” (0.53 %) (Fig. 4). The most abundant cluster in the heat-treated groups was “Signal transduction mechanisms,” indicating that these genes play a vital regulation role in the Muscovy duck senescence and stress response.

KEGG is thought to provide a basic platform for systematic analysis of gene function in terms of the networks of genes products (Kanehisa et al. 2012). To further predict biochemical pathways from the assembled unigenes, KEGG analysis was also conducted. A total of 20,402 annotated unigenes had significant matches with 8,162 hits in the KEGG database, and the top 25 KEGG pathways are shown in Fig. 5. The pathways were classified into five main categories: metabolic, genetic information processing, environment information processing, cellular processes, and organismal systems (Additional file 4). As shown in Fig. 5, focal adhesion, endocytosis, and MAPK signaling pathway were the most represented pathways. These results provide a valuable resource for investigating specific processes, functions, and pathways as well as facilitate the identification of novel genes involved in heat stress tolerance in the liver of Muscovy ducks.

Fig. 1 Size distribution of the sequencing reads used for assembly

Fig. 2 Species distribution of the top BLAST hits. Homology search for the sequences obtained after clustering was done using BLASTX against protein sequences in the NCBI Nr databases. Maximum homology of Muscovy duck transcripts were observed with *A. platyrhynchos*

Frequency and distribution of SSRs in the Muscovy liver transcriptome

SSRs or microsatellites are stretches of short nucleotide motifs ranging from one to six nucleotides in length. These are repeated in tandem and are evenly spread across prokaryotic and eukaryotic genome. Compared with other types of molecular markers, SSRs have many advantages, such as simplicity, effectiveness, abundance, reproducibility, and extensive genomic coverage (Powell et al. 1996). Thus, SSRs serve as important molecular marker discovery centers for studying linkage maps of animals, genetic analysis for economically important quantitative traits, plant evolution, and breeding studies (Victoria et al. 2011; Sonah et al. 2011). Using MISA software, after screening, SSRs in the 36,903 unigenes sequences and 7,755 SSRs distributed in 7,326 sequences were identified. Based on the repeat motifs, all SSRs loci were divided into mono-nucleotide, di-nucleotide, tri-nucleotide, quad-nucleotide, penta-nucleotide, and hexa-nucleotide. The most prevalent SSR type was mono-nucleotides (5,093, 65.67 %), followed by tri-nucleotides (1,695, 21.86 %), then di-nucleotides (669, 8.63 %), penta-nucleotide (158, 2.04 %),

quad-nucleotide (99, 1.28 %), and hexa-nucleotide (41, 0.53 %) (Table 3).

Comparative analysis for identification of heat responsive genes

To generate a resource for heat stress responsive genes, representative unigene transcripts showing twofold or greater differential expression were analyzed under control and heat stress conditions in Muscovy ducks. A rigorous algorithm to identify differentially expressed genes was developed based on the method of Audic and Claverie (1997). For group A (control A vs heat A), 2,935 unigenes were differentially expressed (1,483 up-regulated and 1,452 down-regulated), while for group B (control B vs heat B), 3,897 unigenes were differentially expressed (2,266 up-regulated and 1,631 down-regulated). Among these differentially expressed genes, 470 genes were differentially expressed among the two groups (240 up-regulated and 230 down-regulated). The full list of differentially expressed transcript tags can be seen in Additional file 5. We also show a heatmap of 50 differentially expressed genes in Fig. 6.

To investigate the biological significance of the differentially expressed genes regulated by heat stress in Muscovy ducks, it is important to have the GO descriptions, i.e., detailed annotations of gene molecular function, biological process it is involved in, and cellular component of the gene product. Figure 7 illustrates the GO terms of transcripts tags with differential expression. Among the heat stress-induced transcripts, Gos associated with protein folding and chaperone binding were significant enrichment. The ability for an organism encountering stress to regulate cellular processes by transcriptional control can allow it to cope with stress-induced damage of macromolecular components and to avoid extreme cellular damage that can lead to cell death (Lockwood et al.

Table 2 Summary of function annotation of assembled unigenes

Public protein database	Number of unigene hits	Percentage (%) ^a
NCBI NR	21,221	57.50
Swiss-Prot	19,143	51.87
Pfam	11,362	30.79
KEGG	20,402	55.29
COG	18,284	49.55
GO	16,596	44.97

^a Proportion of the 36,903 assembled unigenes

Fig. 3 Function classifications of Gene Ontology (GO) terms of all Muscovy duck transcripts. The results are summarized in three main categories: biological process, cellular component, and molecular function. The x-axis indicates the subcategories, and the y-axis indicates the number of genes in the same category

2010). It is well known heat shock proteins (HSPs) and other chaperones are induced by various stress. The increased expression of molecular chaperones is a primary component of the cellular stress response and a key indicator of environmental stress (Kultz 2005; Dahloff 2004; Tomanek and Zuzow

2010). Thus, the enrichment of the GO associated with protein folding and chaperone binding in Muscovy ducks is not surprising. In the present study, two HSPs (HSP70 and DnaJ protein) and other chaperones were strongly induced by heat stress. Upon heat shock, HSPs are activated by a change in

Fig. 4 Histogram of Clusters of Orthologous Groups (COG) classification. All unigenes were aligned to the COG database to predict and classify possible functions. Out of 21,221 hits in the NCBI Nr database, 18,284 unigenes were annotated and separated into 24 functional categories

Fig. 5 Pathway assignment based on the Kyoto Encyclopedia of Genes and Genomes (KEGG). A total of 20,402 annotated unigenes had significant matches with 8,162 hits in the KEGG database. The pathways were classified into five main categories: metabolic, genetic information processing, environment information processing, cellular processes, and organismal systems

their oligomerization state and through phosphorylation; subsequently, they bind to denaturing proteins and prevent their aggregation (Haslbeck et al. 2005). They exert an anti-apoptotic activity and, additionally, enhance the cell's ability to combat oxidative stress (Concannon et al. 2003). Although there is still uncertainty about the causality versus correlation of these events, HSPs are key players in the collaborative effort of protecting the cell from the stress of heat.

Fifteen genes under the Go category of metal ion binding were enriched in response to heat stress in Muscovy ducks. Three zinc ion-binding genes (GLIS2, ZBTB7A, and TRAF4) were highly expressed in heat stress group than control group. GLIS2 can act either as a transcriptional repressor or as a transcriptional activator, depending on the cell context. It also can repress transcriptional activation by CTNNB1 in the Wnt signaling pathway (Kim et al. 2007). ZBTB7A plays a key

Table 3 Frequency of EST-SSRs in Muscovy duck

Motif length	Repeat numbers								Total (%)	
	4	5	6	7	8	9	10	>10		
Mono-	—	—	—	—	—	—	—	5,093	5,093	65.67
Di-	—	—	346	145	46	42	49	41	669	8.63
Tri-	—	997	457	221	40	—	—	—	1,695	21.86
Quad-	—	81	18	—	—	—	—	—	99	1.28
Penta-	155	3	—	—	—	—	—	—	158	2.04
Hexa-	41	—	—	—	—	—	—	—	41	0.53
Total	196	1,061	821	366	86	42	49	5,134	7,755

Fig. 6 Clustering analysis of 50 differentially expressed genes. Heatmap of Pearson's correlation across 50 differentially expressed genes. A dendrogram of correlation between genes is shown to the left of the heatmap. Each column represents an experimental sample and each row represents a gene. Expression differences are shown in different colors. Red means high expression and green indicates low expression

role in the instruction of early lymphoid progenitors to develop into B lineage by repressing T cell instructive Notch signals. TRAF4 is an adapter protein and signal transducer that links members of the tumor necrosis factor receptor (TNFR) family to different signaling pathways. It plays a role in the activation of NF-kappa-B and JNK, regulation of cell survival and apoptosis, and regulation of NF-kappa-B in response to signaling through Toll-like receptors (Abell and Johnson 2005). Genes from these families have been reported to play significant roles in plant response to various environmental stresses (Xu et al. 2013). In animals, especially waterfowls, there is little transcriptome information available regarding these genes. Therefore, our research may have an important value in understanding the mechanism against heat stress in Muscovy ducks.

Ras superfamily GTPases are membrane-bound small GTP-binding proteins that play a vital role(s) in diverse cell physiology, including cell cycle progression, cell division, regulation of cell morphology and motility, and intracellular trafficking of molecular and organelles (Bar-Sagi and Hall 2000; Ridley and Hall 1992). It has been shown that heat stress can activate Ras, major components of Ras signaling pathways, as demonstrated by pull-down assay using the Ras binding domain of Raf (Han et al. 2002). In our study, levels of Ras GTPase-activating protein 1 (RASA1) were elevated at heat stress group in Muscovy ducks, indicating that Ras molecular may be directly activated by heat stress. Among the many signaling pathways that respond to mitogens and stresses, MAPK family members are crucial for maintenance of cells through regulating the activities of nuclear transcription factors (Yang et al. 2003; Rubinfeld and Seger 2004). After heat stress, mitogen-activated protein kinase kinase 6 (MAP3K6) was up-regulated in Muscovy ducks and encodes a MAPK, which is a family of proteins that has been shown to be induced in response to salinity stress in marine mussel (Lockwood and Somero 2011). In this context, MAPKs are likely to be important regulators of the cell cycle (Kultz and Burg 1998). Consistent with previous studies, the gene p38 MAP kinase (p38MAPK) was also up-regulated. In rat liver in vivo, heat stress promotes activation of SAPK/JNK and p38MAPK and their upstream kinases (Maroni et al. 2000). p38MAPK is known to play a role(s) in triggering the apoptotic process in response to various stresses. In addition, p38MAPK activated by a low dosage of oxidative stress is involved in mitotic arrest (Kurata 2000). Currently, the consequences of p38MAPK activation by heat stress are not clear since mechanism of action is very complicated and diverse.

In order to validate differential gene expression obtained through RNA-seq, a total of six genes were selected out of the transcripts having twofold or greater differential expression under heat stress group, and their qRT-PCR analysis was performed and the results are shown in Fig. 8. Among those transcripts, we identified genes encoding HSPA1_8

Fig. 7 Function classifications of Gene Ontology terms of the differentially expressed genes regulated by heat stress. We divided the sets into the three major Gene Ontology (GO) domains: biological process, cellular component, and molecular function, and the subcategories (y-axis) and percentage (x-axis) of genes were calculated

Fig. 8 qRT-PCR verification of six differentially expressed genes. Genes up-regulated and down-regulated by heat treatment. β -Actin was used as a reference gene for normalization of gene-expression data. comp26270_c0_seq1: heat shock 70 kDa protein 1/8; comp10854_c0_seq1: p38 MAP kinase; comp160809_c0_seq1: Ras GTPase-activating protein 1; comp25753_c0_seq1: mitogen-activated protein kinase kinase kinase 6; comp62407_c0_seq1: insulin-like growth factor 1; comp26345_c1_seq1: cytochrome P450, family 7, subfamily A

(comp26270_c0_seq1), p38MAPK (comp10854_c0_seq1), RASA1 (comp160809_c0_seq1), MAP3K6 (comp25753_c0_seq1), IGF1 (comp62407_c0_seq1), and CYP7A1 (comp26345_c1_seq1). The expression pattern of these six genes obtained through qRT-PCR data largely corroborated with RNA-seq data. The qRT-PCR analysis confirms that RNA-seq approach has provided reliable data regarding differential gene expression of Muscovy ducks under heat stress.

Conclusions

In this report, we present the sequencing, de novo assembly, and analysis of heat-stressed Muscovy ducks using the Illumina sequencing technology. The transcriptome is described in details in the “Results and discussion” section, with an emphasis on annotation using the Nr, Swiss-Prot, Pfam, KEGG, and COG databases. We generated more than 225 million clean reads and assembled 36,903 unigenes. We also performed gene expression profiling analysis upon heat treatment in Muscovy ducks and identified 470 heat-response unique transcripts. GO term enrichment showed that protein folding and chaperone binding were significant enrichment. KEGG pathway analyses showed that Ras and MAPKS were activated after heat stress in Muscovy ducks. Our research enriched sequences information of Muscovy duck, provided novel insights into responses to heat stress in these ducks, and served as candidate genes or markers that can be used to guide future efforts to breed heat-tolerant duck strains.

Acknowledgments This work was sponsored by the earmarked fund for National Waterfowl-Industry Technology Research System (CARS-43-2) and Zhejiang Technology Innovation Team Project for Animal Health Breeding (2010R50027).

Conflict of interest The authors declare that they have no competing interests.

Authors' contributions LL and TZ conceived and designed the experimental plan. LZ, JL, and TZ performed experiments. GL, YT, and TZ analyzed and interpreted the sequence data. TZ drafted the manuscript. All authors read and approved the final manuscript.

References

- Abell AN, Johnson GL (2005) MEKK4 is an effector of the embryonic TRAF4 for JNK activation. *J Biol Chem* 280(43):35793–35796
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410
- Aronal AP, Huda N, Ahmad R (2012) Amino acid and fatty acid profiles of Peking and Muscovy duck meat. *Int J Poult Sci* 11:229–236
- Audic S, Claverie JM (1997) Detection of eukaryotic promoters using Markov transition matrices. *Comput Chem* 21:223–227

- Bar-Sagi D, Hall A (2000) Ras and Rho GTPases: a family reunion. *Cell* 103(2):227–238
- Bhardwaj J, Chauhan R, Swarnkar MK, Chahota RK, Singh AK, Shankar R, Yadav SK (2013) Comprehensive transcriptomic study on horse gram (*Macrotyloma uniflorum*): de novo assembly, functional characterization and comparative analysis in relation to drought stress. *BMC Genomics* 14:647
- Concannon CG, Gorman AM, Samali A (2003) On the role of Hsp27 in regulating apoptosis. *Apoptosis* 8(1):61–70
- Dahlhoff EP (2004) Biochemical indicators of stress and metabolism: applications for marine ecological studies. *Annu Rev Physiol* 66:183–207
- Edwards CE, Parchman TL, Weekley CW (2012) Assembly, gene annotation and marker development using 454 floral transcriptome sequences in *Ziziphus celata* (Rhamnaceae), a highly endangered, Florida endemic plant. *DNA Res* 19(1):1–9
- Fang G, Munera D, Friedman DI, Mandlik A, Chao MC, Banerjee O, Feng Z, Losic B, Mahajan MC, Jabado OJ, Deikus G, Clark TA, Luong K, Murray IA, Davis BM, Keren-Paz A, Chess A, Roberts RJ, Korlach J, Turner SW, Kumar V, Waldor MK, Schadt EE (2012) Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. *Nat Biotechnol* 30(12):1232–1239
- Frey KG, Herrera-Galeano JE, Redden CL, Luu TV, Servetas SL, Mateczun AJ, Mokashi VP, Bishop-Lilly KA (2014) Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. *BMC Genomics* 15:96
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29(7):644–652
- Han SI, Oh SY, Jeon WJ, Kim JM, Lee JH, Chung HY, Choi YH, Yoo MA, Kim HD, Kang HS (2002) Mild heat shock induces cyclin D1 synthesis through multiple Ras signal pathways. *FEBS Lett* 515(1–3):141–145
- Haslbeck M, Franzmann T, Weinfurter D, Buchner J (2005) Some like it hot: the structure and function of small heat-shock proteins. *Nat Struct Mol Biol* 12(10):842–846
- Iyer MK, Chinnaiyan AM (2011) RNA-Seq unleashed. *Nat Biotechnol* 29(7):599–600
- Jiang X, Zeng T, Zhang S, Zhang Y (2013) Comparative proteomic and bioinformatic analysis of the effects of a high-grain diet on the hepatic metabolism in lactating dairy goats. *PLoS One* 8(11):e80698
- Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40(Database issue):D109–D114
- Kim YS, Kang HS, Jetten AM (2007) The Kruppel-like zinc finger protein Glis2 functions as a negative modulator of the Wnt/beta-catenin signaling pathway. *FEBS Lett* 581(5):858–864
- Kultz D (2005) Molecular and evolutionary basis of the cellular stress response. *Annu Rev Physiol* 67:225–257
- Kultz D, Burg M (1998) Evolution of osmotic stress signaling via MAP kinase cascades. *J Exp Biol* 201(Pt 22):3015–3021
- Kurata S (2000) Selective activation of p38 MAPK cascade and mitotic arrest caused by low level oxidative stress. *J Biol Chem* 275(31):23413–23416
- Li YF, Wang Y, Tang Y, Kakani VG, Mahalingam R (2013) Transcriptome analysis of heat stress response in switchgrass (*Panicum virgatum* L.). *BMC Plant Biol* 13:153
- Lockwood BL, Somero GN (2011) Transcriptomic responses to salinity stress in invasive and native blue mussels (genus *Mytilus*). *Mol Ecol* 20(3):517–529

- Lockwood BL, Sanders JG, Somero GN (2010) Transcriptomic responses to heat stress in invasive and native blue mussels (genus *Mytilus*): molecular correlates of invasive success. *J Exp Biol* 213(Pt 20):3548–3558
- Ma PF, Guo ZH, Li DZ (2012) Rapid sequencing of the bamboo mitochondrial genome using Illumina technology and parallel episodic evolution of organelle genomes in grasses. *PLoS One* 7(1):e30297
- Maroni P, Bendinelli P, Zuccorononno C, Schiaffonati L, Piccoletti R (2000) Cellular signalling after in vivo heat shock in the liver. *Cell Biol Int* 24(3):145–152
- Martin LB, Fei Z, Giovannoni JJ, Rose JK (2013) Catalyzing plant science research with RNA-seq. *Front Plant Sci* 4:66
- Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol Breed* 2:225–238
- Ridley AJ, Hall A (1992) The small GTP-binding protein rho regulates the assembly of focal adhesions and actin stress fibers in response to growth factors. *Cell* 70(3):389–399
- Rubinfeld H, Seger R (2004) The ERK cascade as a prototype of MAPK signaling pathways. *Methods Mol Biol* 250:1–28
- Saez G, Davail S, Gentes G, Hocquette JF, Jourdan T, Degraze P, Baeza E (2009) Gene expression and protein content in relation to intramuscular fat content in Muscovy and Pekin ducks. *Poult Sci* 88(11):2382–2391
- Santos CS, Pinheiro M, Silva AI, Egas C, Vasconcelos MW (2012) Searching for resistance genes to *Bursaphelenchus xylophilus* using high throughput screening. *BMC Genomics* 13:599
- Sohail MU, Hume ME, Byrd JA, Nisbet DJ, Ijaz A, Sohail A, Shabbir MZ, Rehman H (2012) Effect of supplementation of prebiotic mannan-oligosaccharides and probiotic mixture on growth performance of broilers subjected to chronic heat stress. *Poult Sci* 91(9):2235–2240
- Sonah H, Deshmukh RK, Sharma A, Singh VP, Gupta DK, Gacche RN, Rana JC, Singh NK, Sharma TR (2011) Genome-wide distribution and organization of microsatellites in plants: an insight into marker development in *Brachypodium*. *PLoS One* 6(6):e21298
- Tan Y, Liu J, Chen X, Zheng H, Li F (2013) RNA-seq-based comparative transcriptome analysis of the syngas-utilizing bacterium *Clostridium ljungdahlii* DSM 13528 grown autotrophically and heterotrophically. *Mol Biosyst* 9(11):2775–2784
- Tomanek L, Zuzow MJ (2010) The proteomic response of the mussel congeners *Mytilus galloprovincialis* and *M. trossulus* to acute heat stress: implications for thermal tolerance limits and metabolic costs of thermal stress. *J Exp Biol* 213(Pt 20):3559–3574
- Utsumi Y, Tanaka M, Morosawa T, Kurotani A, Yoshida T, Mochida K, Matsui A, Umemura Y, Ishitani M, Shinozaki K, Sakurai T, Seki M (2012) Transcriptome analysis using a high-density oligomicroarray under drought stress in various genotypes of cassava: an important tropical crop. *DNA Res* 19(4):335–345
- Victoria FC, da Maia LC, de Oliveira AC (2011) In silico comparative analysis of SSR markers in plants. *BMC Plant Biol* 11:15
- Xu Y, Gao S, Yang Y, Huang M, Cheng L, Wei Q, Fei Z, Gao J, Hong B (2013) Transcriptome sequencing and whole genome expression profiling of chrysanthemum under dehydration stress. *BMC Genomics* 14:662
- Yang SH, Sharrocks AD, Whitmarsh AJ (2003) Transcriptional regulation by the MAP kinase signaling cascades. *Gene* 320:3–21
- Zeng T, Jiang X, Li J, Wang D, Li G, Lu L, Wang G (2013) Comparative proteomic analysis of the hepatic response to heat stress in Muscovy and Pekin ducks: insight into thermal tolerance related to energy metabolism. *PLoS One* 8(10):e76917
- Zhou Y, Gao F, Liu R, Feng J, Li H (2012) De novo sequencing and analysis of root transcriptome using 454 pyrosequencing to discover putative genes associated with drought tolerance in *Ammopiptanthus mongolicus*. *BMC Genomics* 13:266