Transcriptome analysis of the quantitative distribution of the *Cyrtotrachelus buqueti* population in two cities in China

Running title: *Cyrtotrachelus buqueti*

Chaobing Luo[1,2,*], Anxuan Liu [1,2, 3], Wencong Long[1,2, 3], Hong Liao[1,2, 3], Yaojun Yang[1,#]

[1]Bamboo Diseases and Pests Control and Resources Development Key Laboratory of Sichuan Province, Leshan Normal University, Leshan 614000, Sichuan, China.

[2]College of Life Science, Leshan Normal University, Leshan 614000, Sichuan, China.

[3]College of Food and Bioengineering, Xihua University, Chengdu 6110039, Sichuan, China.

*** First author**

#Corresponding author

Yaojun Yang

Bamboo Diseases and Pests Control and Resources Development Key Laboratory of Sichuan Province, Leshan Normal University, Binhe road, No. 778, Leshan 614000, Sichuan, China.

TEL: +86-0833-2276293

Email: y8868g@163.com

Abstract

Background*: Cyrtotrachelus buqueti* is a forest pest that severely damages bamboo shoots. Reducing the population of this insect involves complex mechanisms and is dependent on diverse gene expression influenced by environmental factors.

Methods: In this study, samples from two regions of China, Muchuan in Sichuan Province and Chishui in Guizhou Province, were investigated through RNA-seq to explore the causes and molecular mechanisms underlying the population reduction of this species. Environmental factors, such as temperature, heavy metal content, and pH, may affect the reduced population of *C. buqueti* in Chishui.

Results*:* Approximately 44 million high-quality reads were generated, and 94.2% of the data were mapped to the transcriptome. A total of 15,641 out of the 29,406 identified genes were predicted. Moreover, 348 genes were differentially expressed between the two groups of imagoes and included 77 upregulated and 271 downregulated UniGenes. The functional analysis showed that these genes were significantly enriched in ribosome and metabolic pathway categories. The candidate genes, which contributed to *C. Buqueti* reduction, included 41 genes involved in the ribosome constitution category, five genes in the one-carbon pool pathway by folate category, and five heat shock protein genes.

Conclusions: Downregulation of these candidate genes seems to have impaired metabolic processes, such as protein, DNA, RNA, and purine synthesis, as well as carbon and folate metabolism, and finally resulted in the observed reduced population of *C. buqueti*. Furthermore, temperature, heavy metal content, and pH might influence the population by altering the expressions of genes involved in these metabolic processes.

Key words: *Cyrtotrachelus buqueti;* folate metabolism; heat shock protein; quantitative distribution; Transcriptome analysis

1. Introduction

Cyrtotrachelus buqueti belongs to the class of insects in the order Coleopterai, the family Curculionoidea, and the genus *Cyrtotrachelus*. It is found in southern China and several countries of Southeast Asia, such as the Socialist Republic of Vietnam, Thailand, and the Union of Myanmar (Rui-Ting, 2005). *C. buqueti* is a pest in cluster bamboo forests and seriously limits the development of bamboo products, adversely affecting the income of bamboo farmers. This species damages bamboo shoots; specifically, the larvae bore into cluster bamboo shoots, such as *Phyllostachys pubescens*, *Neosinocalamus affinis*, *Bambusa textilis*, and *Dendrocalamus farinosus*. For this reason, the insect has been listed as a dangerous forestry pest by the State Forestry Administration of China since 2003 (Yang et al., 2009).

C. buqueti produces one generation per year in Sichuan and overwinters inside puparia in the soil. The adult emerges from the soil from June to October, which is almost the same time period as bamboo shooting. The activity period of adults lasts for around one month. The adult starts to mate and oviposit after 2 days of feeding. Female adults lay eggs in bamboo shoots, using their long, hard mouths. The larva causes damage to bamboo shoots from July to September and pupates sometime between the last 10 days of July to the last 10 days of October (Wang et al., 2005, Nie, 2010). The high content of special volatiles, primary benzaldehyde, and the special proportion of components on the top of the tufting bamboo shoot could produce an important odor signal that attracts *C. buqueti* toward the plant (Yang et al., 2010). Feeding and oviposition behaviors enhance the difficulty forecasting population density and injury percentage of bamboo shoots. There are various methods to control the *C. buqueti*, for example, sapling quarantine, removing the puparia from the soil and the larva from the bamboo shoot manually, pesticides, and biological control, such as benzaldehyde (Nie, 2010). However, none of the methods are perfect, and are either inefficient or costly.

Based on our investigation, the *C. buqueti* hazard rate is more than 40% across the 67,000-hectare bamboo cluster forests at Muchuan and Mabian, Sichuan province. However, in Chishui, Guizhou province and its surrounding areas, such as Hejiang in the Sichuan province, only slight damage was observed and the hazard rate was less than 1% in 82,000-hectare bamboo clusters, which is less than 300 kilometers away from the first site (Yang, 2011). The reasons for and effects of this phenomenon

72 are not fully understood. At present, many forecast studies have been launched on insects like
73 *Drosophila* and mosquitoes (Guruprasad et al., 2010, Wegbreit and Reisen, 2000) in order to study their
74 population change and biological damage driven by agriculture against the background of global
75 environmental change (Clark et al., 2001). In addition, next generation high-throughput DNA
76 sequencing techniques, such as Illumina sequencing, provide the opportunity for research in life
77 science with considerable cost savings and efficiency (Ansorge, 2009). *C. buqueti* has been
78 investigated in several studies, but the molecular regulation mechanisms are poorly understood and
79 relevant genomic resources are scarce. Therefore, we sequenced and annotated the transcriptome of *C.*
80 *buqueti* from the two study areas to screen for any genes that may be responsible for the reduction in
81 the quantitative distribution of the insect, using Illumina HiSeq 2500 sequencing, with the goal of
82 discovering safer prevention methods.
83
84 **2. Materials and Methods**
85 *2.1 Insect material preparation*
86 *C. buqueti* Guerin, from Muchuan (N103°98′, E28°96′) and Chishui (N105°69′, E28°57′), were used
87 for transcriptome analysis. Six male adults were sampled. The digestive, reproductive, and musculature
88 systems of *C. buqueti* were removed and mixed. The mixtures were transferred immediately to liquid
89 nitrogen and stored subsequently at -80 ℃ until RNA extraction. The digestive, reproductive, and
90 musculature systems of *C. buqueti* were sampled from three comparable insects from same place using
91 three biological replications. The Muchuan and Chishui samples were used to construct six libraries,
92 which were named C2, C3, C4, D2, D3, and D4.
93
94 *2.2 Detection of temperature, heavy metal concentration, and pH*
95 The annual mean temperature, monthly mean maximum temperature, and monthly mean minimum
96 temperature were derived from the TIANQI network (www.tianqi.com).
97 Inductively coupled plasma mass spectrometry and potentiometry were performed to detect heavy
98 metal concentration and pH, respectively.
99
100 *2.3 RNA isolation, library construction, and sequencing*
101 Total RNA was extracted using the RNAprep pure Tissue Kit (DP431, TianGen Biotechnology, Beijing,

China) and was treated with RNase-free DNase I (TianGen Biotechnology, Beijing, China) to remove genomic DNA contamination. Briefly, 1.0% agarose gel stained with Gel-Red was evaluated to determine the RNA integrity. A K5500 spectrophotometer (KAIAO, Beijing, China) and an Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA) were used to assess RNA quantity and quality. The RNA integrity number (RIN) was greater than 8.0 for all samples (Clerico et al., 2015). RNA samples from the three systems for each group were pooled together in equal amounts, in order to generate one mixed sample. These six mixed RNA samples were used to construct the cDNA library and Illumina sequencing, which was managed by Beijing ANOROAD Bioinformatics Technology Co., Ltd.

The oligo-dT beads were used to isolate mRNA, which were then fragmented into short sections and added into the fragmentation buffer. The short mRNA fragments were used as templates, with random hexamer primers, to synthesize first-strand cDNA. Next, dNTPs, DNA polymerase I, and response buffer were used to synthesize the second-strand cDNA. The double-stranded cDNAs were purified using the QIAQuick PCR kit, eluted with ethidium bromide, and then used for "A" base addition and end-reparation. The cDNAs were finally ligated with sequencing adapters and a 1.0% agarose gel stained with G was used to fragment recovery.

2.4 Sequence tag preprocessing and mapping

The sequence tag was preprocessed according to a previously described protocol (Li et al., 2009, Langmead et al., 2009). Raw reads were cleaned by removing reads with adaptors, low quality (≤19%), or Ns (>5%). Clean reads were mapped to transcript sequences after assembly with Bowtie 2 software, allowing for a maximum of 2 nucleotide mismatches (Tatusov et al., 1997).

2.5 Gene expression calculation and pathway analysis

Gene expression was calculated using RPKM (Mortazavi et al., 2008). The DEseq package (ver. 2.1.0) was used to detect DEGs between the two sample groups (Anders and Huber, 2010, Wang et al., 2010). DEGs were selected based on |log2 Ratio|≥1 and a Padj value < 0.05. FDR was used to determine the P-value threshold in multiple tests. The absolute value of the log 2 (fold change) with RPKM ≥ 1 and an FDR ≤0.005 were used as the thresholds in the study to determine significant differences in gene expression.

2.6 Functional analysis of DEGs

Functional enrichment analyses, such as GO and KEGG, were performed in order to identify whether DEGs were significantly enriched in GO terms or metabolic pathways. Blast2GO was performed to implement the GO enrichment analysis of DEGs. The corrected P value of GO terms ($P< 0.05$) was considered significantly enriched by DEGs. The statistical enrichment analyses of differential expression genes were calculated using KOBAS software for the KEGG pathways. The GENES database for KEGG contained the whole genome, whether complete or incomplete, the PATHWAY database included gene functional statistics, and the LIGAND Database contained the enzyme database. Pathways with an FDR value ≤0.05, which were defined as those with genes, showed significant levels of differential expression.

2.7 Quantitative real-time PCR validation of RNA-Seq data

Twenty DEGs were chosen for validation with quantitative real-time PCR (RT-qPCR). The primers, which were designed with the Primer 3.0 software (http://biotools.umassmed.edu/bioapps/primer3_www.cgi), are listed in Table 1. RT-qPCR reactions were analyzed in the ABI StepOneTM Plus Real-Time PCR System with SYBR Green PCR Master Mix (TianGen, Beijing, China), and amplified with 1μL of the cDNA template, 10μL 2×SupperReal PriMix Plus, 2μL of 50×ROX Reference Dye, 0.5μL of each primer (20 μmol/μL), and a final volume of 20μL was achieved by adding water. The amplification program consisted of one cycle of 95 ℃ for 15 min, followed by 30 cycles of 95 ℃ for 5 s, 60 ℃ for 20 s and 72 ℃ for 20 s. Fluorescent products were detected in the last step of each cycle. The melting curve analysis was performed at the end of 30 cycles, in order to ensure proper amplification of target fragments. All RT-qPCR for each gene was performed in three biological replicates, with three technical repeats per experiment. Relative gene expressions were normalized by comparison with the expression of lotus β-actin (c28453_g1_i3), and were analyzed using the $2^{-\Delta\Delta C}$T method. The data are indicated as mean ±SE (n=9). Statistical analysis of the RT-qPCR data was conducted using the ANOVA procedure in SAS 8.1 (SAS Institute, Cary, NC, USA).

3 Results

3.1 Mortality and living environment analysis of C. buqueti in Chishui and Muchuan

We detected 10 pupal cells with *C. buqueti* adults or larvae in each area. The reason for the small sample size is that few pupal cells were found in Chishui. Eight adult and larvae died or decayed in the Chishui pupal cells (Figs. 1A–C), but only a few adults were found dead in those obtained from Muchuan (2/10) (Fig. 1D–F).

The quantitative distribution of *C. buqueti* in Muchuan was much higher than in Chishui. We also found several differences in the living environments of *C. buqueti* between the two regions. Chishui had higher heavy metal concentrations in the soil and pupal cells than Muchuan, with metals such as Hg and Se (Fig. 2A & B). The pH of the soil and pupal cells was higher in Muchuan than in Chishui (Fig. 2C). Moreover, the annual mean temperature in Chishui was 18.1 ℃ and could reach 28 ℃ in July. By contrast, the temperature in Muchuan was more than 1 ℃ lower, averaging at 17.0 ℃ (Fig. 2D).

3.2 Sequencing, de novo assembly, and functional annotations

Six sequencing libraries from the two distinct areas were prepared and sequenced with the Illumina HiSeq platform to investigate the population size variation of *C. buqueti*. A total of 52 million short reads were generated from the six libraries, with 44 million high-quality ($Q \geq 20$) 125-bp reads, which could be subdivided into 29,406 UniGenes with an average length of 747 bp (Table 2). The lengths of the assembled UniGenes primarily ranged from 201 to 20,229 bp, with a mean length of approximately 748 bp (Fig. 3A). Additionally, a BLAST comparison with the Uniprot protein database revealed that up to 15,641 UniGenes in the *C. Buqueti* transcriptome contained putative homologs in other databases (Fig. 3B). Among the 15,641 UniGenes, 15,703 were successfully annotated by Gene Ontology (GO) assignments and classified into three functional categories, namely, molecular function, biological process, and cellular component (Fig. 3D). The matched unique sequences assigned to molecular function were clustered into 22 classification bins with the largest subcategory being binding (62 UniGenes) and the second largest subcategory being catalytic activity (46 UniGenes). The unique sequences were classified into 23 bins with the most abundant comprising transcripts involved in the cellular process (65 UniGenes), metabolism (54 UniGenes), and single-organism process (54 unigenes) categories. The unique sequences in cellular components were divided into 20 classifications with the most abundant being cell part (75 UniGenes) and organelle (37 UniGenes) (Fig. 3D).

The matched unique sequences were divided into 24 categories using Clusters of Orthologous Groups (COGs) of proteins (Fig. 3C). The dominant category included general functional prediction, amino

acid transport and metabolism, translation, ribosomal structure and biogenesis, post-translational modification, protein turnover, and chaperones (Fig. 3D).

However, the effect of this diverse gene expression on the decreased population density of *C. buqueti* in Chishui remains unknown. First, we performed a hierarchical clustering of the six samples using the Euclidean distance method associated with complete linkage. C2 and C4 were close to each other, similar to D3 and D4 (Fig. 4A), suggesting the use of C2, C3, and C4 or D2, D3, and D4 as three biological repeats. We summarized the expression level of each gene with HT-seq by reads per kilobase million mapped reads (RPKM). The two groups were compared using DEGseq[127] to identify differentially expressed genes (DEGs). We found 348 DEGs by using a false discovery rate (FDR), 0.05, and a fold change of 2 as significance cutoffs. Among these DEGS, 77 genes were significantly upregulated and 271 were significantly downregulated in the D-group, compared with the C-group (Fig. 4C).

We also performed hierarchical clustering of all DEGs by using the R Project for Statistical Computing (Fig. 4B). Nine clusters were plotted with expression patterns (Fig. 4D). The S1 cluster included 18 upregulated genes from both samples, and the expression levels of these genes were similar in both the C- and D-groups. Thirty-two of these genes in the S2 cluster were upregulated in both groups, and the expression levels in the C-group were higher than those in the D-group, and 39 genes had opposite expression patterns. The 35 genes in S3 were downregulated in the D-group but showed no significant changes in the C-group. Interestingly, genes in the S4 cluster had expression patterns contrary to those in S3. The genes in cluster S5 possessed the most DEGs, but gene expression was not significantly different between the two groups. Notably, the gene expression pattern in C4 was distinct from the others. The S6 cluster was composed of 13 genes, the expression of which was downregulated in the C-group compared with the D-group. Genes in the S7–S9 clusters were downregulated in both groups (Fig. 4D).

3.3 Functional classification of DEGs between the two groups

We used GO assignments to classify the functions of DEGs in pairwise comparisons of cDNA libraries between the different groups (Fig. 5). Comparative results of the C and D groups showed that many GO terms were significantly enriched in the three categories. The translation and cellular protein metabolic process pathways were significantly enriched in the biological process category ($P < 0.05$) (Fig. 5A). The ribosome, non-membrane-bounded organelle, intracellular non-membrane-bounded

organelle, and ribonucleoprotein component GO terms were significantly enriched in the cellular component category (Fig. 5C). The structural molecule activity and structural constituents of ribosomes terms were significantly enriched in the molecular function category (Fig. 5B).

We then performed a KEGG analysis of the DEGs. Most of the genes were assigned to the ribosome, one-carbon pool by folate, biosynthesis of amino acids, microbial metabolism in diverse environments, purine metabolism, metabolic pathways, and biosynthesis of secondary metabolite processes categories (Fig. 5D).

3.4 Candidate genes involved in one-carbon metabolism, purine synthesis, and folate metabolism

Genes with expressions that were highly correlated with one-carbon pool by folate are potential candidates involved in *C. buqueti* reproduction, growth, and development. We suggested five candidate genes, namely, *AIRS* (c14771_g1_i2), *GARS* (c14771_g1_i1), *SHMT1* (c3484_g1_i2), *SHMT2* (c3484_g1_i1), and *THFD* (c28372_g1_i4), for one-carbon pool by folate. The expression of these genes in the C group significantly differed from their expression in in the D group. The expression levels of candidate genes encoding for *AIRS* and *GARS*, which are key enzymes in IMP synthesis (Ulrich et al., 2003) (Supplementary Fig. 1), were downregulated in the D group (Fig. 6). This suggests that IMP synthesis would be disordered in the D group. Moreover, *SHMT1* and *SHMT2*, which are serine hydroxymethyltransferases and are involved in one-carbon metabolism (Supplementary Fig 1), also had lower expression in the D group than in the C group (Fig. 6). *MTHFR* in the D group, which participates in folate metabolism (Supplementary Fig. 1), was significantly downregulated by twofold (log 2) compared with the C group (Fig. 6). Thus, folate metabolism might be destroyed in the D group. Functional analysis of these candidate genes may be beneficial for determining the molecular mechanisms underlying the disunity of *C. buqueti* quantitative distribution.

3.5 Genes related to retinal determination gene network (RDGN)

Genes involved in RDGN have been reported to also be involved in eye development. The genes related to RDGN belong to the PAX6, eyes absent (EYA), Sine oculis(SO), and dachshund (DAC) families (Pappu and Mardon, 2002). In this study, we found 10 PAX6 homologous genes, 10 EYA genes, three SO genes, and four DACH genes (Supplementary Table 1). The expression of these genes in the C group was not significantly different from that in the D group (Supplementary Fig. 2).

252 However, analyzing these RDGN genes may be useful for the determination of the molecular
253 mechanisms for genetic engineering or the marker-assisted selection of *C. buqueti* feeding, which could
254 alter the quantitative distribution of *C. buqueti*.
255
256 *3.6 DEGs involved in the ribosome category*
257 The molecular mechanism involved in the decreased population numbers in Chishui was the
258 downregulation of ribosomal protein genes, resulting in decreased ribosome numbers, and enlarged
259 ribosome size (Korostelev et al., 2006, Zhang et al., 2014). In our study, we found that 41
260 downregulated ribosomal protein genes were significantly enriched in the ribosome category of the
261 KEGG pathway analysis (Figs. 7A and B).
262
263 *3.7 Heat shock protein (HSP) relevant gene*
264 Temperature gradients associated with global warming threaten the survival of wild animals, but the
265 expression of HSPs can improve tolerance to heat shock. Five putative homologs of the HSP gene,
266 namely, *HSP1* (c30948_g1_i1), *HSP2* (c30174_g1_i1), *HSP3* (c24612_g1_i2), *HSP4* (c24612_g1_i1),
267 and *HSP5* (c28234_g1_i1), were identified in the transcriptomic data. These genes exhibited lower
268 expression levels in the D group than that in the C group (Fig. 8).
269
270 *3.8 RT-qPCR validation*
271 The transcriptional regulation revealed by RNA-seq was confirmed in three independent biological
272 experiments using RT-qPCR. A total of 20 genes were selected to design gene-specific primers; 10
273 were involved in the ribosome pathway, five in the one carbon pool by folate pathway, and five coded
274 HSP (Table 1). The RT-qPCR results for all genes were tested statistically and most genes showed
275 significantly different expression levels (P=0.05, Fig. 9). Moreover, 20 genes showed significant
276 correlations (P=0.05) between the RT-qPCR data and the RNA-seq results, which indicated good
277 reproducibility between the transcript abundance assayed by RNA-seq and the expression profile
278 revealed by RT-qPCR (Fig. 9).
279
280 **4. Discussion**
281 We identified and studied two separate regions with different population sizes of *C. buqueti*. Few were

observed in Chishui in Guizhou province, when compared to the abundant population in Muchuan in Sichuan province. The reduced *C. buqueti* numbers are the result of complex factors and are dependent on diverse gene expression controlled by environmental factors. We investigated pupal cells with *C. buqueti* adults or larvae. The results showed that the adults and larvae died in the pupal cells in Chishui, but only a few adults were found dead inside the pupal cells in Muchuan, as predicted. The potential reasons might be that the living environment altered, inducing the formation of several physical and physiological defects in the insects or causing bamboo shoots to become more toxic to the insect. These characteristics shortened the lives of both larvae and adults and even caused death.

Heat is known to cause endogenous (species-native) proteins to misfold into aggregation-prone species, the toxicity of which is mitigated and reversed by chaperones (De, 1995, Vabulas et al., 2010, Verghese et al., 2012). Newly synthesized proteins are particularly susceptible to heat-induced misfolding and aggregation and are apparently major triggers of the heat-shock response, as well as its main beneficiaries (Vabulas et al., 2010). At lower temperatures, more ribosomes are engaged in protein synthesis (Yun et al., 1996). The structures and functions of DNA and RNA are altered at high temperatures, and thus stable heredity can be adversely affected (Liao et al., 2015). We analyzed the temperature patterns in the two regions and found that Chishui had a higher average temperature than Muchuan (Fig. 2D), especially during July and August, with maximum temperatures of over 40 ℃. Ironically, Nanning, in Guangxi Province, China, which has a higher annual mean temperature than Chishui (data not shown), reported serious damage caused by *C. buqueti*. These results indicate that high temperature only partially, rather than primarily, affects the *C. buqueti* proliferation. In addition, our results showed that Hg, Se, and As existed in higher concentrations in the soil and pupal cells in Chishui, and that the Cd concentration was higher in the pupal cell in Chishui, than was found in Muchuan (Figs. 2A and B). Plants and animals may accumulate heavy metals like Cd, Se, Hg, Cr, and Pb from the soil, which negatively affects protein synthesis, ribosome structure and function, and RNA and DNA synthesis (Ahmad et al., 2014, Wang et al., 2004) A preliminary hypothesis on the mode of action and likely proximal site for organic mercurial inhibition of synaptosome fraction protein synthesis are provided (Cheung and Verity, 1981) in Supplementary Fig 3. Therefore, we assume that a high heavy metal concentration might play a dominant role in reducing the population of *C. buqueti* in Chishui. Moreover, the bioavailability and toxicity of heavy metals is partially dependent on pH. Heavy

312 metals generally became more available to organisms at increased pH levels (Sijm et al., 2000). The
313 results showed that pH levels below 5 in Chishui and up to pH 6 in Muchuan, especially up to pH 7.5
314 in pupal cells (Fig. 2C), might affect the ribosome-catalyzed peptidyl transfer and the bioavailability
315 and toxicity of heavy metals. In summary, the combined effects of high-temperature, high
316 concentration of heavy metals, and high pH values might combine and adversely affect the diverse
317 metabolism processes of *C. buqueti* by affecting protein, DNA, and RNA synthesis in Chishui samples,
318 which might reduce the quantitative population of *C. buqueti*, and even cause local extinction of the
319 insect population.
320
321 The global analysis of transcriptomes could facilitate the identification of systemic gene expression and
322 regulatory mechanisms, in order to successfully analyze the transcriptomes of several species (Xie et
323 al., 2012, Sweetman et al., 2012). In our study, we sequenced and annotated the transcriptome of *C.*
324 *buqueti* samples from the two study areas to screen for genes responsible for the observed variety of
325 quantitative distribution. A total of 52 million pair-end reads were generated. An average of 94.2% of
326 the data were mapped to the transcriptome (Table 2). Among the 29,406 genes identified, 15,641 were
327 previously predicted in the reference. Moreover, 348 were differentially expressed between the insects
328 in the two different areas, with 77 upregulated and 271 downregulated UniGenes. The functional
329 analysis identified that these genes were significantly enriched in ribosome constitution and metabolic
330 pathways. Candidate genes included 41 involved in ribosome constitution, five in pathway one-carbon
331 pool by folate, and five HSP genes. A comparative transcriptomic analysis detected several DEGs and
332 potential candidate genes. Adding to the currently available expressed sequences would help further the
333 comprehensive understanding of the transcription profiles of the reduced quantitative distribution *C.*
334 *buqueti*. Transcriptome-wide gene expression profiles were compared between the libraries to identify
335 corresponding genes associated with the variety of *C. buqueti* distribution. DEG expression was
336 remarkably downregulated, and only 77 genes were upregulated in the D group that were not
337 upregulated in the C group (Fig. 4C). These variations may be consistent with the morphological and
338 physiological changes in *C. buqueti*, which were probably due to changes in the living environment of
339 this species. Moreover, nine clusters were plotted with expression patterns that indicated that certain
340 genes could be responsible for the reduced distribution. The GO enrichment analysis revealed that
341 translation, ribosome, and structural constituents of ribosomes were overrepresented terms for DEGs

(Figs. 5A–C). These may contribute to the regulation of ribosome protein gene expression, thus affecting ribosome quantity and enlarging ribosome size. KEGG analysis also revealed that most genes were assigned to the ribosome, one-carbon pool by folate, biosynthesis of amino acids, purine metabolism, metabolic pathways, and biosynthesis of secondary metabolite processes categories (Fig. 5D). Thus, these processes play an important role in ribosome quantity, enlargement of ribosome size, and enhancement of enzyme activity.

Folate is a cofactor that transfers single-carbon units (e.g., CH_3 and CHO) in numerous reactions. One-carbon metabolism includes nucleotide synthesis (purine ring and thymidine from uridine; Supplementary Fig 4 and 1) (Blatch et al., 2015). Thus, insufficient folate would result in the inadequate production of thymidine and possible misincorporations of uracil in DNA (Blount et al., 1997). Insects also appear to require folate for nucleotide synthesis (Blatch et al., 2015). Folate deficiencies decrease SAM levels and DNA methylation levels, which then disrupt gene regulation in mice (Friso and Choi, 2002). Many insects, such as *Drosophila melanogaster*, have methylated DNA, although mechanisms of epigenetic regulation in some insects, including *D. melanogaster*, appear to be quite different from those observed in mammals (Lyko et al., 2000). Studies on folate in insects have provided conflicting results. Folate has been reported to stimulate growth, to have no effect, or even to be toxic. However, insects cannot use folate directly without transforming into THF. 5,10-Methylenetetrahydrofolate reductase (MTHFR) reduces folate into available THF, which then goes on to participate in DNA synthesis and amino acid transformation. Downregulated *MTHFR* expression causes the accumulation of homocysteine (Hyc) and disordered biochemical processes, such as cell cycle regulation, DNA replication, and DNA and protein methylation (Christensen et al., 1999). We found the following five genes in this study: *AIRS* and *GARS* were the key enzymes of IMP synthesis; *SHMT1* and *SHMT2* were serine hydroxymethyltransferases and are involved in the one-carbon pathway; and the *MTHFR* gene participates in folate metabolism. These genes were downregulated in the D-group (Fig. 6), which can explain the reduced quantitative distribution of *C. buqueti* in Chishui.

Ribosomes are integral parts of any given cell, playing an important role in linking genotypes to phenotypes by manufacturing the proteome. Ribosome construction is essential to maintain cellular

372 homeostasis and is also the single most expensive metabolic process for a cell (Raska et al., 2004). A
373 dividing mammalian cell requires 10 million ribosomes (Purves, 2000), and a rapidly growing bacterial
374 cell requires as many as 20,000 ribosomes (Berg et al., 2006). Therefore, the downregulation of
375 ribosomal protein genes results in the corresponding reduction of ribosome proteins, thus decreasing
376 the quantity of ribosomes in a cell (Korostelev et al., 2006). The enlargement of ribosomes would
377 induce an alteration of biological ribosomal functions and decrease ribosome biogenesis in cells. A
378 previous study showed that cell growth and proliferation could be derived from ribosome biogenesis
379 (Zhang et al., 2014). Thus, ribosome biogenesis is tightly linked with cell growth and proliferation
380 (Trainor and Merrill, 2014, Raiser et al., 2014, Bolze and Casanova, 2013). Our study found that 32
381 downregulated ribosome protein genes were significantly enriched in the ribosome category of the
382 KEGG pathway analysis (Figs. 7A and B). Therefore, downregulation of ribosomal protein genes may
383 cause alteration in *C. buqueti* cells, such as reduced cell number, cell size, and quantity of cilia. These
384 alterations may have induced an alteration in the quantity of *C. buqueti* in Chishui. However, further
385 studies are needed to elucidate the mechanisms underlying the downregulation of ribosome protein
386 genes and the enlarged size of the ribosomes.
387
388 HSPs function as molecular chaperones, and are key in the maintenance of proper protein folding and
389 overall proteostasis, preventing and reversing deleterious protein misfolding (Mayer and Bukau, 2005)
390 (Supplementary Table 2) . Hsp70 assumes a critical role in helping refold aggregated or misfolded
391 proteins, as well as in the folding of newly synthesized proteins, among other functions (Mayer and
392 Bukau, 2005, Höhfeld et al., 2001) (Supplementary Fig 5). It is a crucial component of protein quality
393 control systems, and specialized forms interact with a large continuum of substrates (Höhfeld et al.,
394 2001). Interestingly, five HSP genes were downregulated in the D group, which were the homologous
395 genes of Hsp70 in humans (Fig.8). Moreover, Chishui had a higher average temperature than Muchuan
396 (Fig. 2D). Accordingly, the possibility of protein misfolding might increase, and the overall proteostasis
397 was probably threatened because of the downregulation of HSP genes. These characteristics reduced
398 protein quality and quantity in the *C. buqueti* pupal cells in Chishui. Furthermore, lower expression
399 levels of HSP genes caused lower thermotolerance, whereas the higher temperature in Chishui required
400 a higher thermotolerance of *C. buqueti*. Therefore, *C. buqueti* in Muchuan might respond better to
401 extreme weather, especially to heat. These phenomena indicated that downregulation of HSP genes

402 may decrease thermotolerance and probably result in the observed population size differences between
403 the two regions. However, the mechanisms for the downregulation of HSP genes and amounts of *C.*
404 *buqueti* remain unknown.
405
406 **Conclusions**
407 We explored an effective and safe method to further the molecular understanding of *C. buqueti*.
408 Muchuan and Chishui, which had different *C. buqueti* population sizes, were investigated using
409 transcriptome analysis. Several possible environmental factors, namely temperature, heavy metals and
410 pH value, were found to be responsible for the death of *C. buqueti* adults and larvae in the pupal cells
411 in Chishui, due to hindered metabolic processes, such as protein synthesis, DNA synthesis, RNA
412 synthesis, one-carbon metabolism, purine synthesis, and folate metabolism. Furthermore, we analyzed
413 the expression of RDGE and some HSP genes. The data suggest that the expression of RDGN genes
414 was unchanged, but HSP genes were downregulated in the D group. However, this molecular
415 mechanism did not offer a safer or more effective method for *C. buqueti* prevention, but enriched the
416 molecular knowledge of this pest. Further studies on the molecular mechanisms of feeding behavior,
417 signal molecules, and flight, among others, might help to achieve this goal.
418

Competing interests

These authors have no conflict of interest to declare

Funding

Supported by the National Natural Science Foundation of China (31470655)

Authors' contributions

Chaobing Luo and Yaojun Yang analyzed and interpreted the data. Anxuan Liu ,Wencong Long and Hong Liao performed the experiment, and was a major contributor in writing the manuscript. All authors read and approved the final manuscript

Acknowledgements

We would like to thank ANOROAD (Beijing Biotechnology Corporation) for its assistance in origin data processing and related bioinformatics analysis. We also thank Xiang Nong and members of the laboratory for suggestions and discussion of this work and manuscript revision.

References

AHMAD, M. M., ALI, A., KHAN, M. A. & ABDIN, M. Z. 2014. Biomolecular Characteristics of Aspergillus niger Under Cadmium Metal Stress. *Environmental Processes,* 2**,** 241-250.

ANDERS, S. & HUBER, W. 2010. Differential expression analysis for sequence count data. *Genome Biology,* 11**,** R106.

ANSORGE, W. J. 2009. Next-generation DNA sequencing techniques. *New Biotechnology,* 25**,** 195-203.

BERG, J. M., TYMOCZKO, J. L. & STRYER, L. 2006. Biochemistry. 5th edition. *National Center for Biotechnology Information õs Bookshelf***,** 93-126.

BLATCH, S. A., STABLER, S. P. & HARRISON, J. F. 2015. The effects of folate intake on DNA and single-carbon pathway metabolism in the fruit fly Drosophila melanogaster compared to mammals. *Comparative Biochemistry & Physiology Part B Biochemistry & Molecular Biology,* 189**,** 34-39.

BLOUNT, B. C., MACK, M. M., WEHR, C. M., MACGREGOR, J. T., HIATT, R. A., WANG, G., WICKRAMASINGHE, S. N., EVERSON, R. B. & AMES, B. N. 1997. Folate deficiency causes uracil misincorporation into human DNA and chromosome breakage: implications for cancer and neuronal damage. *Proceedings of the National Academy of Sciences,* 94**,** 3290-3295.

BOLZE, A. & CASANOVA, J. L. 2013. Ribosomal protein SA haploinsufficiency in humans with isolated congenital asplenia. *Science,* 340**,** 976-8.

CHEUNG, M. & VERITY, M. A. 1981. Methyl mercury inhibition of synaptosome protein synthesis: Role of mitochondrial dysfunction ☆. *Environmental Research,* 24**,** 286-298.

CHRISTENSEN, B., ARBOUR, L., TRAN, P., LECLERC, D., SABBAGHIAN, N., PLATT, R., GILFIX, B. M., ROSENBLATT, D. S., GRAVEL, R. A. & FORBES, P. 1999. Genetic polymorphisms in methylenetetrahydrofolate reductase and methionine synthase, folate levels in red blood cells, and risk of neural tube defects. *American Journal of Medical Genetics,* 84**,** 151-157.

CLARK, J. S., CARPENTER, S. R., BARBER, M., COLLINS, S., DOBSON, A., FOLEY, J. A., LODGE, D. M., PASCUAL, M., JR, P. R. & PIZER, W. 2001. Ecological forecasts: an emerging imperative. *Science,* 293**,** 657-660.

462 CLERICO, E. M., TILITSKY, J. M., MENG, W. & GIERASCH, L. M. 2015. How hsp70 molecular
463 machines interact with their substrates to mediate diverse physiological functions. *Journal of*
464 *Molecular Biology,* 427**,** 1575-88.
465 DE, M. A. 1995. The heat-shock response. *New Horizons,* 3**,** 198.
466 FRISO, S. & CHOI, S. W. 2002. Gene-nutrient interactions and DNA methylation. *Journal of Nutrition,*
467 132**,** 2382S.
468 GURUPRASAD, B. R., HEGDE, S. N. & KRISHNA, M. S. 2010. Seasonal and Altitudinal Changes in
469 Population Density of 20 Species of Drosophila in Chamundi Hill. *Journal of Insect Science,*
470 10**,** 123.
471 H HFELD, J., CYR, D. M. & PATTERSON, C. 2001. From the cradle to the grave: molecular
472 chaperones that may choose between folding and degradation. *Embo Reports,* 2**,** 885.
473 KOROSTELEV, A., TRAKHANOV, S., LAURBERG, M. & NOLLER, H. F. 2006. Crystal Structure
474 of a 70S Ribosome-tRNA Complex Reveals Functional Interactions and Rearrangements. *Cell,*
475 126**,** 1065-77.
476 LANGMEAD, B., TRAPNELL, C., POP, M. & SALZBERG, S. L. 2009. Ultrafast and
477 memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology,*
478 10**,** R25.
479 LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G.,
480 ABECASIS, G. & DURBIN, R. 2009. The Sequence Alignment/Map format and SAMtools.
481 *Bioinformatics,* 25**,** : 2078–2079.
482 LIAO, J. L., ZHOU, H. W., PENG, Q., ZHONG, P. A., ZHANG, H. Y., HE, C. & HUANG, Y. J. 2015.
483 Transcriptome changes in rice (Oryza sativa L.) in response to high night temperature stress at
484 the early milky stage. *BMC Genomics,* 16**,** 18.
485 LYKO, F., RAMSAHOYE, B. H. & JAENISCH, R. 2000. DNA methylation in Drosophila
486 melanogaster. *Nature,* 408**,** 538-40.
487 MAYER, M. P. & BUKAU, B. 2005. Hsp70 chaperones: Cellular functions and molecular mechanism.
488 *Cellular & Molecular Life Sciences Cmls,* 62**,** 670-84.
489 MORTAZAVI, A., WILLIAMS, B. A., MCCUE, K., SCHAEFFER, L. & WOLD, B. 2008. Mapping
490 and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods,* 5**,** 621.
491 NIE, X. W. 2010. Preliminary Report on Experiment of Biological Features of Cyrtotrachelus buqueti

and Its Prevention and Control. *Forest Inventory & Planning*.

PAPPU, K. & MARDON, G. 2002. *Retinal Specification and Determination in Drosophila*, Springer Berlin Heidelberg.

PURVES 2000. *The Cell - A Molecular Approach*, ASM.

RAISER, D. M., NARLA, A. & EBERT, B. L. 2014. The emerging importance of ribosomal dysfunction in the pathogenesis of hematologic disorders. *Leukemia & Lymphoma,* 55**,** 491-500.

RASKA, I., KOBERNA, K., MAL NSK , J., FIDLEROV , H. & MASATA, M. 2004. The nucleolus and transcription of ribosomal genes. *Biology of the Cell,* 96**,** 579-594.

RUI-TING, J. U. 2005. Cyrtotrachelus buqueti in Shanghai. *Forest Pest & Disease*.

SIJM, D., KRAAIJ, R. & BELFROID, A. 2000. Bioavailability in soil or sediment: exposure of different organisms and approaches to study it. *Environmental Pollution,* 108**,** 113.

SWEETMAN, C., WONG, D. C., FORD, C. M. & DREW, D. P. 2012. Transcriptome analysis at four developmental stages of grape berry (Vitis vinifera cv. Shiraz) provides insights into regulated and coordinated gene expression. *BMC Genomics,* 13**,** 691.

TATUSOV, R. L., KOONIN, E. V. & LIPMAN, D. J. 1997. A genomic perspective on protein families. *Science,* 278**,** págs. 631-637.

TRAINOR, P. A. & MERRILL, A. E. 2014. Ribosome biogenesis in skeletal development and the pathogenesis of skeletal disorders. *Biochimica Et Biophysica Acta,* 1842**,** 769-78.

ULRICH, C. M., ROBIEN, K. & MCLEOD, H. L. 2003. Cancer pharmacogenetics: polymorphisms, pathways and beyond. *Nature Reviews Cancer,* 3**,** 912-20.

VABULAS, R. M., RAYCHAUDHURI, S., HAYERHARTL, M. & HARTL, F. U. 2010. Protein Folding in the Cytoplasm and the Heat Shock Response. *Cold Spring Harbor Perspectives in Biology,* 2**,** a004390.

VERGHESE, J., ABRAMS, J., WANG, Y. & MORANO, K. A. 2012. Biology of the Heat Shock Response and Protein Chaperones: Budding Yeast (Saccharomyces cerevisiae) as a Model System. *Microbiology & Molecular Biology Reviews Mmbr,* 76**,** 115.

WANG, J., MAO, Z. & ZHAO, X. 2004. Response of Saccharomyces cerevisiae to chromium stress. *Process Biochemistry,* 39**,** 1231-1235.

WANG, L., FENG, Z., WANG, X. & ZHANG, X. 2010. DEGseq: an R package for identifying

differentially expressed genes from RNA-seq data. *Bioinformatics,* 26**,** 136-8.

WANG, W. D., CHEN, F. Z. & WANG, X. Q. 2005. Reproductive Behavior of Cyrtotrachelus bugueti. *Sichuan Journal of Zoology,* 51**,** 1615-29.

WEGBREIT, J. & REISEN, W. K. 2000. Relationships among weather, mosquito abundance, and encephalitis virus activity in California: Kern County 1990-98. *Journal of the American Mosquito Control Association,* 16**,** 22-7.

XIE, F., BURKLEW, C. E., YANG, Y., LIU, M., XIAO, P., ZHANG, B. & QIU, D. 2012. De novo sequencing and a comprehensive analysis of purple sweet potato (Impomoea batatas L.) transcriptome. *Planta,* 236**,** 101-113.

YANG, Y. 2011. Larvae Population Dynamics of Cyrtatrachelus buqueti and the Forecasting Models with Climate Factors. *Scientia Silvae Sinicae,* 47**,** 82-87.

YANG, Y. J., QIN, H., WANG, S. F., WANG, Y. P., LIAO, H., LIU, C. & LI, S. G. 2010. Antennal ultrastructure and electroantennogram responses of Cyrtotrachelus buqueti Guerin-Meneville (Coleoptera: Curculionidae) to volatiles of bamboo shoot. *Acta Entomologica Sinica,* 27**,** 353-4.

YANG, Y. J., WANG, S. F., GONG, J. W., LIU, C., MU, C. & QIN, H. 2009. [Relationships among Cyrtotrachelus buqueti larval density and wormhole number and bamboo shoot damage degree]. *Ying yong sheng tai xue bao = The journal of applied ecology / Zhongguo sheng tai xue xue hui, Zhongguo ke xue yuan Shenyang ying yong sheng tai yan jiu suo zhu ban,* 20**,** 1980-1985.

YUN, H. S., HONG, J. & LIM, H. C. 1996. Regulation of ribosome synthesis in Escherichia coli: Effects of temperature and dilution rate changes. *Biotechnology & Bioengineering,* 52**,** 615-24.

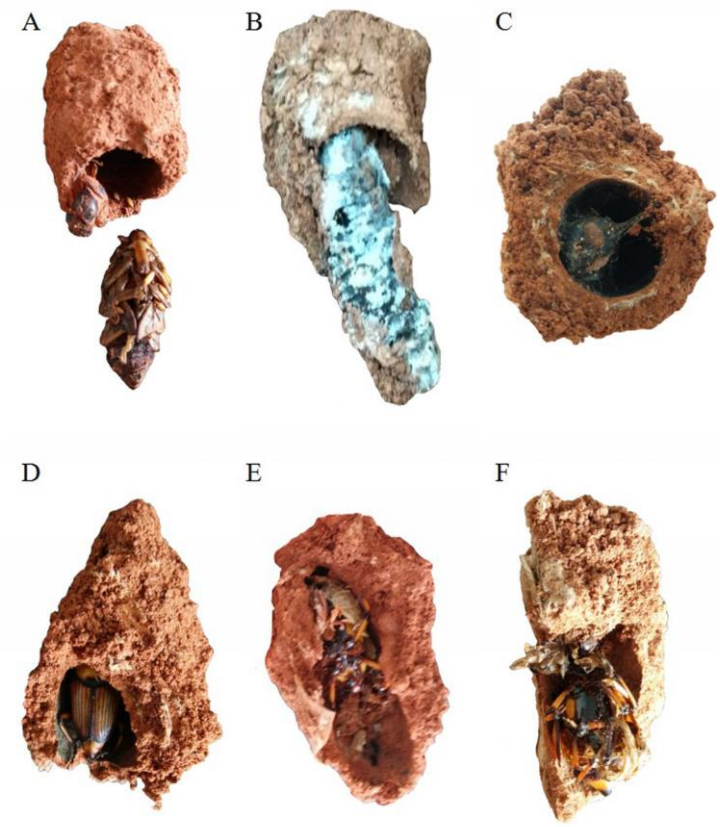
ZHANG, Q., SHALABY, N. A. & BUSZCZAK, M. 2014. Changes in rRNA transcription influence proliferation and cell fate within a stem cell lineage. *Science,* 343**,** 298-301.

550 **Figure legends**

551 Figure 1. Pupal cells with *Cyrtotrachelus buqueti* adults or larvae in Chishui, Guizhou Province and

552 Muchuan, Sichuan Province. Pupal cells with *C. buqueti* (A–C) adults or larvae from Chishui and (D–F)

553 adults from Muchuan.



554

555

Figure 2. Living environment of *C. buqueti*. Heavy metal content in (A) soil (Se, Hg, and As) and (B) pupal cells (Se, Cd, Hg, and As); (C) pH levels of soil and pupal cells; (D) Temperatures in Muchuan and Chishui, including annual mean temperatures, monthly maximum average temperatures, and monthly minimum mean temperatures.

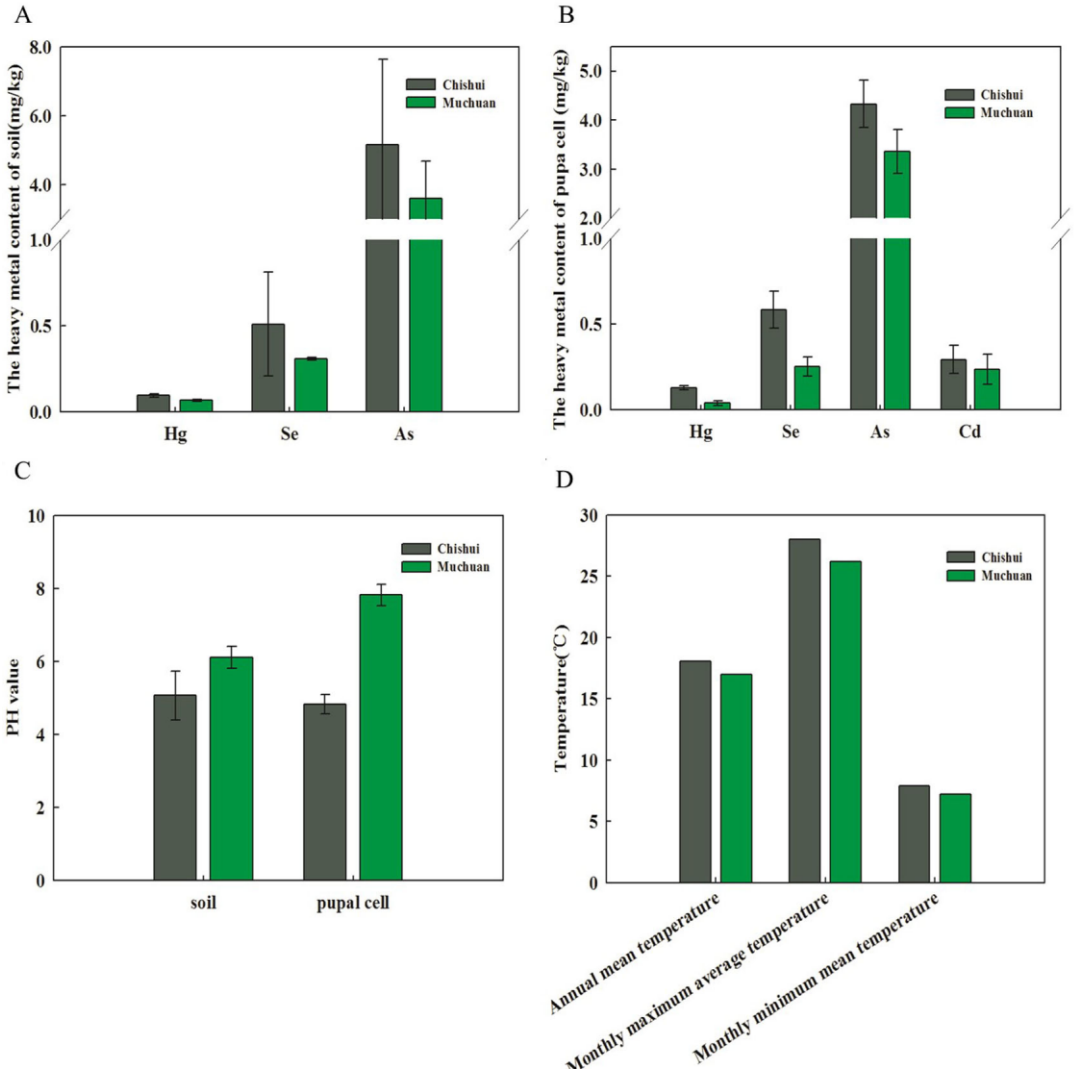
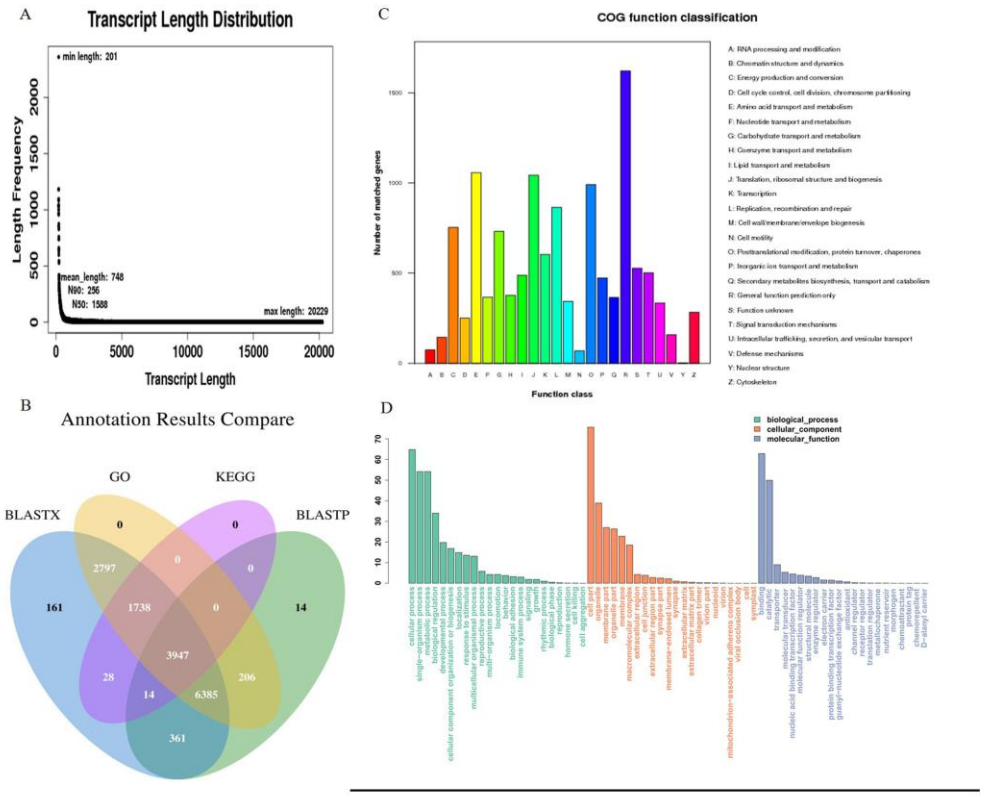
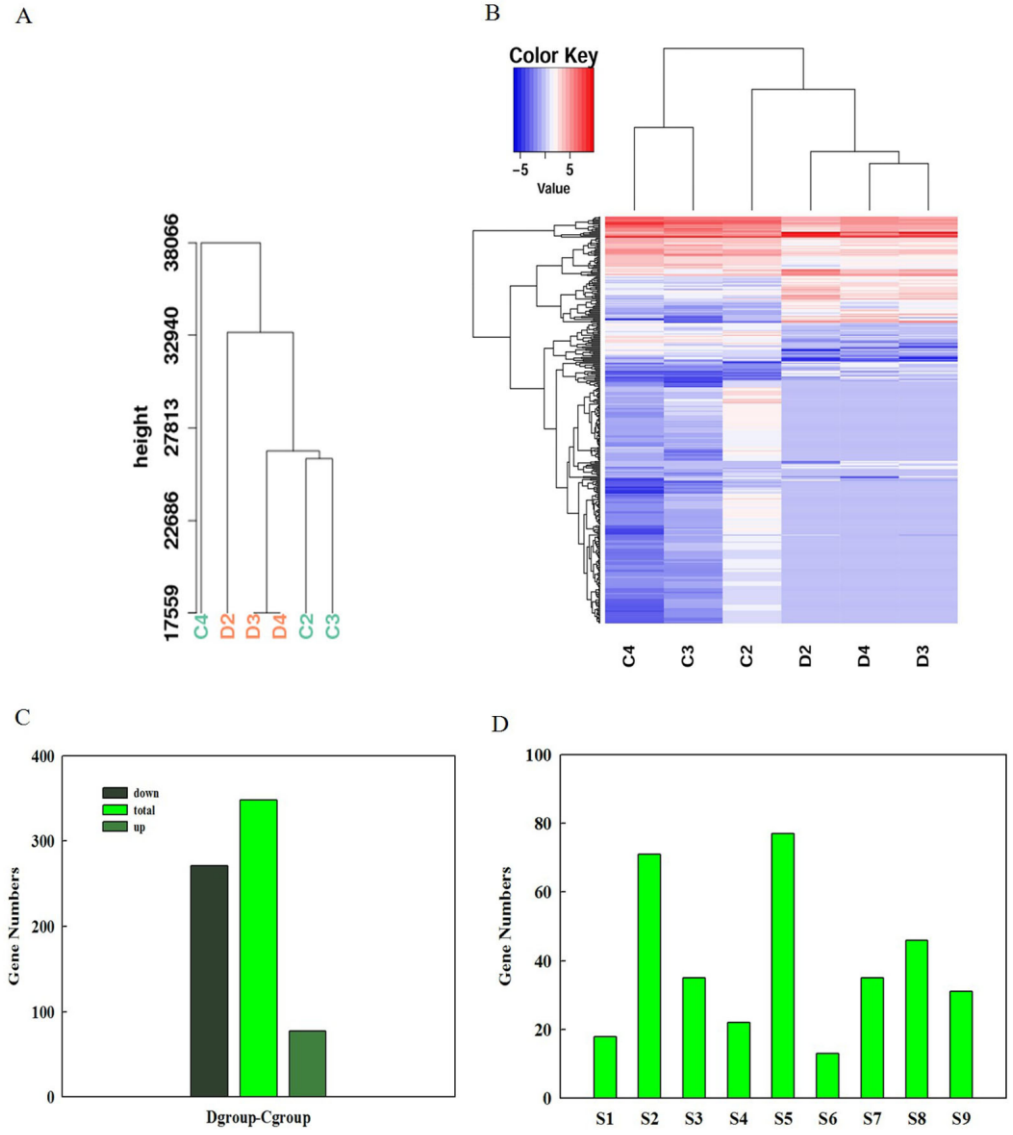


Figure 3. Functional annotations of UniGenes. (A) Length distribution of UniGenes. A total of 15,641 UniGeneswere assembled. (B) Venn Diagram: summary of annotation results. (C) Clusters of Orthologous Groups functional categories in the *C. buqueti* transcriptome. A total of 15,071 UniGenes were subcategorized into 24 categories. The y-axis represented the number of UniGenes, and the x-axis represented categories. (D) Gene Ontology functional categories in the *C. buqueti* transcriptome, including molecular function, cellular component, and biological process.

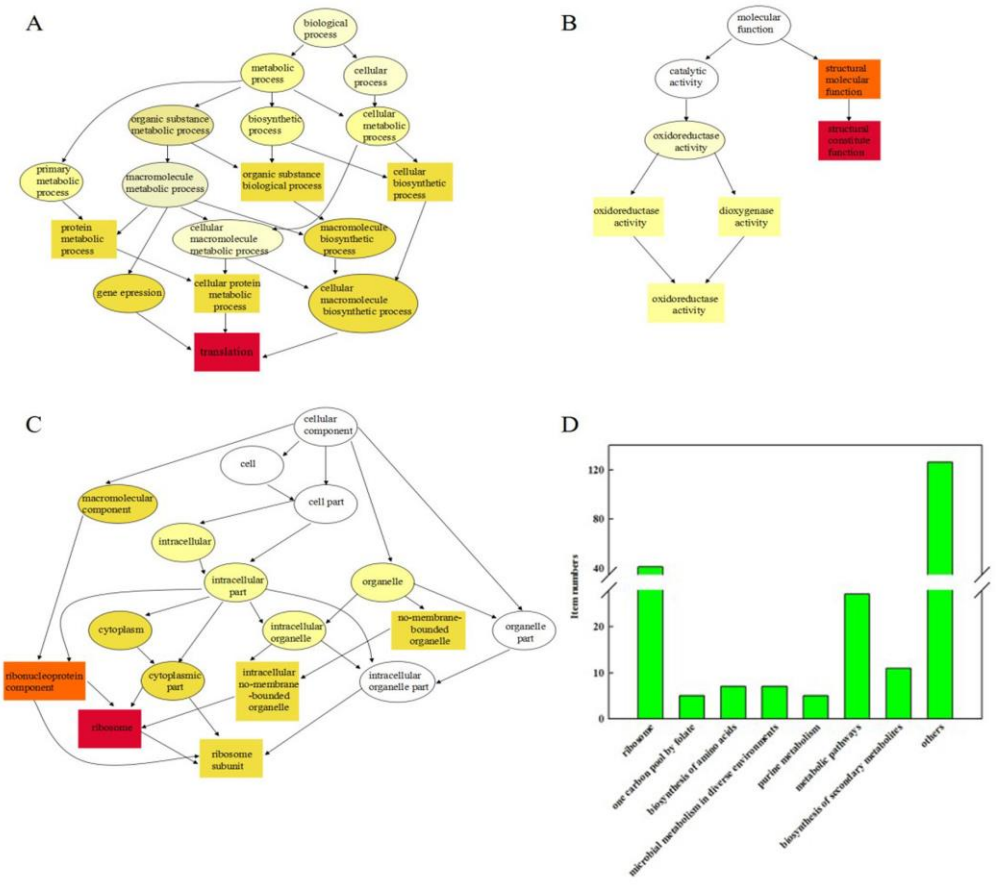


570 Figure 4. Overview of serial analysis of DEGs identified by pairwise comparisons of the six *C. buqueti*
571 transcriptomes, namely C2, C3, C4, D2, D3, and D4. (A) Cluster of all samples. (B) Heatmap of DEGs
572 across the six *C. buqueti* transcriptomes. The expression values of six libraries are presented as RPKM
573 normalized log 2 transformed counts. Red and blue colors indicate upregulated and downregulated
574 transcripts, respectively. The nine main clusters are shown. (D) Counting the number of each cluster.



575
576

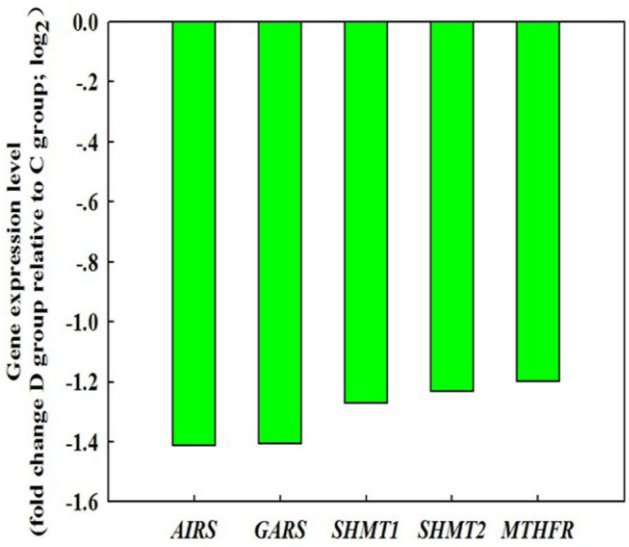
577 Figure 5. Functional classification of DEGs between the two groups. Directed acyclic graphs of (A)
578 biological process, (B) molecular function, and (C) cellular component categories. (D) Main KEGG
579 Orthology (KO) classifications of DEGs. A total of 348 DEGs were mapped to eight categories; a:
580 ribosome; b: one-carbon pool by folate; c: biosynthesis of amino acids; d: microbial metabolism in
581 diverse environments; e: purine metabolism; f: metabolic pathways; g: biosynthesis of secondary
582 metabolites; and h: others.



583

584

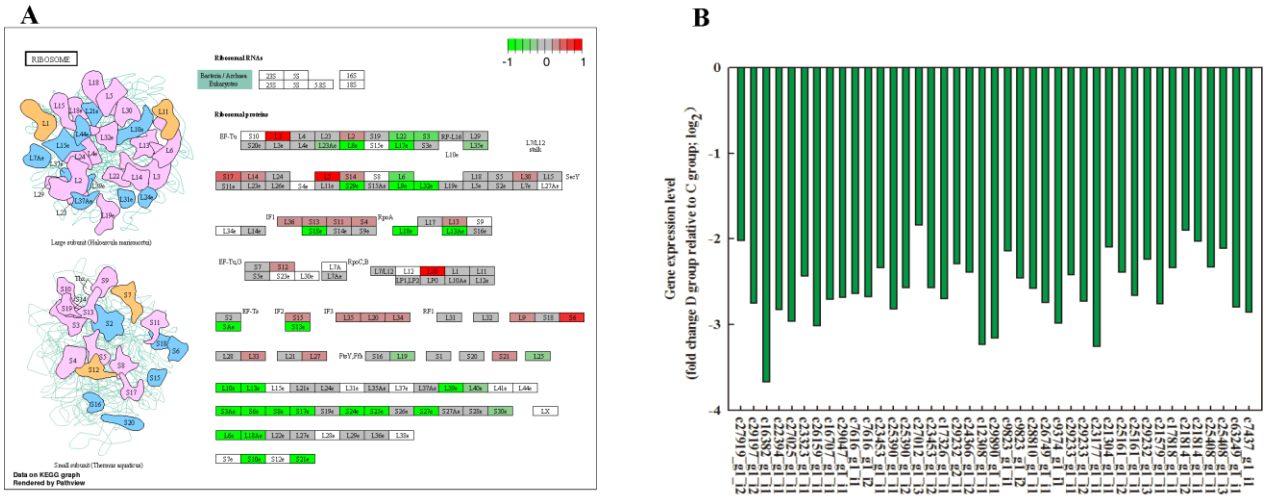
585 Figure 6. Gene expression levels of *AIRS* (c14771_g1_i2), *GARS* (c14771_g1_i1), *SHMT1*
586 (c3484_g1_i2), *SHMT2* (c3484_g1_i1), and *THFD* (c28372_g1_i4). Twofold (log 2) change of D
587 group relative to C group



588

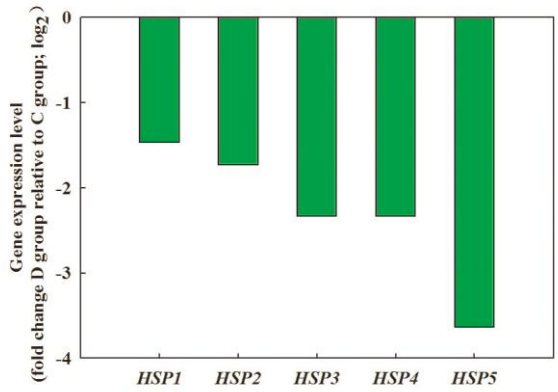
589

590 Figure 7. Gene expression levels of ribosomal protein genes. (A) Data on the KEGG graph rendered by
591 pathview. (B) Gene expression level of ribosome protein genes, including 41 genes: c27919_g1_i2,
592 c29197_g1_i2, c16382_g1_i1, c22394_g1_i1, c27025_g1_i1, c23323_g1_i1, c26159_g1_i1,
593 c16707_g1_i1, c29047_g1_i1, c7616_g1_i1, c7616_g1_i2, c23453_g1_i1, c25390_g1_i1,
594 c25390_g1_i2, c27012_g1_i3, c23453_g1_i2, c17326_g1_i1, c29232_g2_i1, c24366_g1_i2,
595 c12308_g1_i1, c29890_g1_i1, c9823_g1_i1, c9823_g1_i2, c28810_g1_i1, c26749_g1_i1,
596 c9374_g1_i1, c29233_g1_i1, c29233_g1_i2, c23177_g1_i1, c21304_g1_i1, c25161_g1_i2,
597 c25161_g1_i1, c29232_g1_i3, c21579_g1_i1, c17818_g1_i1, c21814_g1_i2, c21814_g1_i1,
598 c25408_g1_i1, c25408_g1_i3, c63249_g1_i1, and c7437_g1_i1. Twofold (log 2) change D group
599 relative to C group



600
601

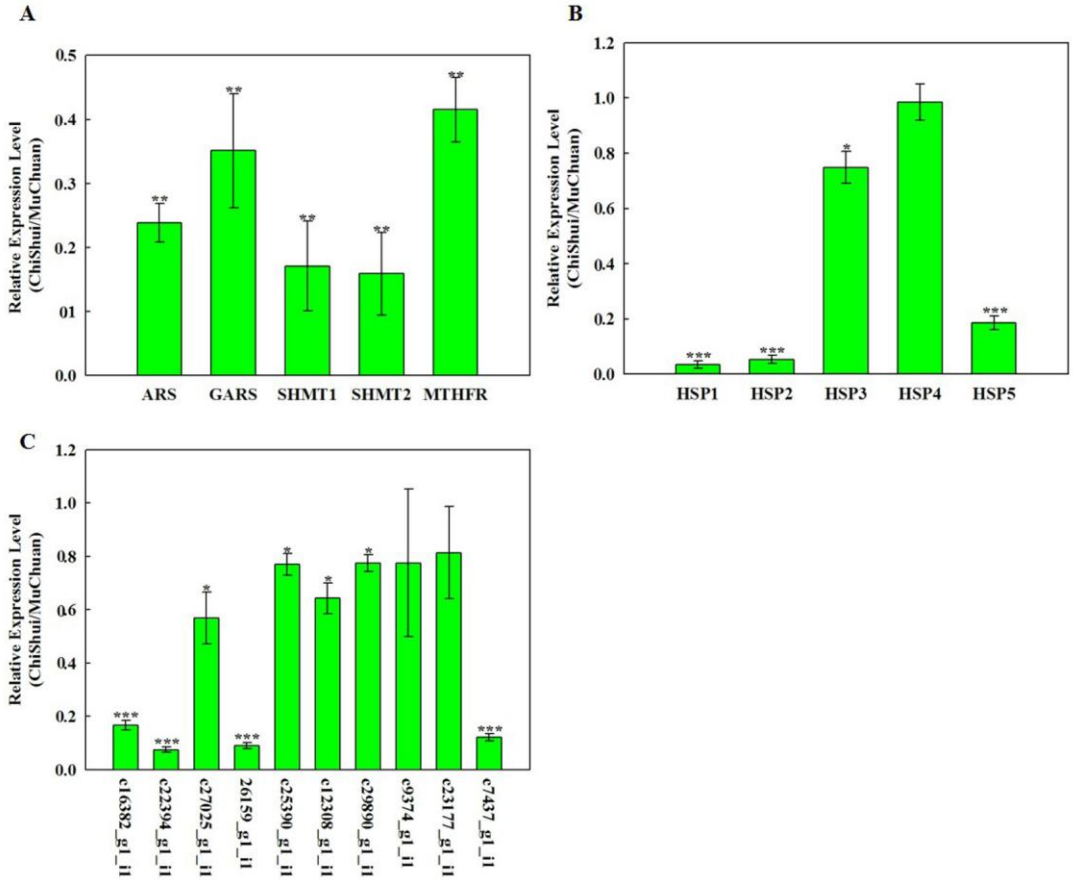
602 Figure 8. Gene expression level of five HSPs. Five genes, *HSP1* (c30948_g1_i1), *HSP2*
603 (c30174_g1_i1), *HSP3* (c30174_g1_i1), *HSP4* (c30174_g1_i1), and *HSP5* (c30174_g1_i1). Twofold
604 (log 2) change in D group relative to C group.



605

606

Figure 9. Real-time quantitative RT-qPCR confirmation of 20 candidate genes. The left y-axis indicates relative gene expression levels determined by RT-qPCR. Relative gene expressions were normalized by comparison with the expression of lotus β-actin (c28453_g1_i3), and were then analyzed using the $2^{-\Delta\Delta C}$T method. The expression values were adjusted by setting the expression of MuChuan samples to be 1 per gene. All RT-qPCRs for each gene used three biological replicates, with three technical replicates per experiment; the error bars indicate SE. Different "*"(*-***) indicates the significant difference at P=0.05. (A) The expression levels of candidate genes involved in one-carbon metabolism, purine synthesis, and folate metabolism. (B) The expression levels of HSPs. (C) The expression levels of candidate genes involved in the ribosome category.



618 Table 1.Primers used in RT-qPCR

Gene ID	RT-qPCR primer	
	Forward primer (5'-3')	Reverse primer (5'-3')
c28453_g1_i3	CCCGAGTTTTCGCCCATAGA	TGTTTATCTATCGGCTTGCTTGC
c14771_g1_i2	GAACTGACAAAAGCGGGCGT	TGTCTTTGAGCATGGGTTCCCT
c14771_g1_i1	ATGGGGTTGCTGAAGGTTGC	TCCTACTGCGAATCCTGCCA
c3484_g1_i2	AGTGCAGATGAGTGGGGTGT	GCAGATCCAAACCCATGATCCT
c3484_g1_i1	AGTGCAGATGAGTGGGGTGT	GCAGATCCAAACCCATGATCCT
c28372_g1_i4	GTCAAACCTGGCGCTTGTGT	ACTGGAGTTATGAGACCCGCT
c30948_g1_i1	ATCGCATGTTGTCGGAAGCG	CGCAGTCTTGTGTGGCTTGT
c30174_g1_i1	ATGCCGCCAAGAACCAGGTA	GCCAATGTTTCAGATCGCCCT
c24612_g1_i2	CGGCGTGATGACCAAGATGA	AACTTGGTCGGGATGGTCGT
c24612_g1_i1	ACGGAGATCAACCTGCCCTT	AATGCCCTCGAACTTGCTGC
c28234_g1_i1	ACAAGAAGTTCGCGTGCCTG	TCGCCTTGCACAGCTTTTCG
c16382_g1_i1	GCAGTTCTTCACCAGCGTCT	ACTTTAAGATCCGTGCGCTGC
c22394_g1_i1	TGAAGCTGGTGAAGAAGTACGC	TGGGGTCGTTCTTCTTCTCCT

c27025_g1_i1	TGAACTTCACACGCACCATAC	CAAGAAGGTCAAGCGGTGT
c26159_g1_i1	TGAGGTGCTGAAGAAGTTTGGC	CCATTCGTTTCTGCGTTCAACC
c25390_g1_i1	GCCTTCTTGCTCTTCTCGGT	TGGCTTCAACAAGGTCGTGT
c12308_g1_i1	TGCAAGATGTGTTGTCTGCC	GGTGCATGTTCAAACCTGGT
c29890_g1_i1	CCGCATGAAGAAGAAGCTCC	CTCCATCTGAAAAGCCGTCG
c9374_g1_i1	CGGTGTCTCCCTCAAGAAGT	AGTACCACACACGACCTCTG
c23177_g1_i1	CGGACTATGAGGCAATGCTG	AGACACACTCCCAAGCTACC
c7437_g1_i1	GGCCTGATCATCAAGGTCCT	ATGAGGGAGAGCATGTCGTT

621 Table 2. Summary statistics from Illumina sequencing of the *C. buqueti* transcriptome.

Sequencing	
Total number of reads	52,439,226
Number of clean reads	44,036,144
Number of contigs	97,721
Total length of contigs (bp)	73,064,686
Contig N50	1,588
Average length of contigs (bp)	330
Number of primary unigenes	73,210
Number of final unigenes	29,406
Total length of final unigenes (bp)	231,348,394
Average length of final unigene (bp)	1245.5740
GC percentage (%)	38.4200

622

623