RESEARCH ARTICLE | FEBRUARY 28 2019

Transcriptome-wide gene expression alteration of sloe bug (Insecta: Hemiptera: Dolycoris baccarum) to divergent-altitude environments FREE

Qi-Lin Zhang ✉; Ming-Long Yuan; Xian-Yu Deng; Jun Guo; Feng Wang; Lian-Bing Lin ✉

AIP Conf. Proc. 2079, 020020 (2019)
https://doi.org/10.1063/1.5092398

Transcriptome-Wide Gene Expression Alteration of Sloe Bug (Insecta: Hemiptera: Dolycoris baccarum) to Divergent-Altitude Environments

Qi-Lin Zhang[1, a, b)], Ming-Long Yuan[2, b)], Xian-Yu Deng[1], Jun Guo[1], Feng Wang[1], Lian-Bing Lin[1, c)]

[1]*Faculty of Life Science and Technology, Kunming University of Science and Technology, Kunming 650500, China*
[2]*State Key Laboratory of Grassland Agro-Ecosystems, College of Pastoral Agricultural Science and Technology, Lanzhou University, Lanzhou 730020, China.*

[a)]Corresponding author: zhangqilin88888@126.com
[b)]These authors contributed equally to this work.
[c)]Corresponding author: linlb@kmust.edu.cn

Abstract. Animals can remodel their gene expression to compensate for the effects of different ecological factors, which can confer resistance to extreme environment alteration, such as altitude changes. Understanding response of gene expression to divergent-altitude habitats, such as Tibetan Plateau (TP) and lower altitude area, in natural populations remains one of the greatest challenges. Previous studies extensively focused on vertebrates, while the current lack of understanding about transcriptomic gene expression alteration in divergent-altitude adaptation of invertebrates, especially the insects. Here, we downloaded raw data/reads of *Dolycoris baccarum* distributed in 1,300 m a.s.l (above sea level) and its corresponding TP strain (3,200 m a.s.l), and performed a mixed assembly of these two transcriptomes to construct unigene set of *D. baccarum*. Subsequently, we analyzed the gene expression changes between two populations, and validate the sequencing data using real time quantitative PCR. A total of 4,816 differentially expressed genes were largely enriched to GO involved in DNA repair, hypoxia response, energy production, suggesting that substantial plasticity in gene expressions was mainly associated with the ability to resist UV radiation, hypoxia, and cold. Our study firstly provides novel insights into gene expression variations of insects in divergent-altitude adaptation.

INTRODUCTION

Identifying the gene expression of adaptive traits to environmental stressor is a major mission in evolutionary biology research [1]. The environment conditions on the Tibetan Plateau (TP), with low temperature, hypoxia and strong ultraviolet radiation (UV), which create a metabolic challenge within which constant energy production must be maintained in the face of these stress [2]. Vertebrate adaptations to high-altitude environments are classic examples of genomic plasticity under natural selection of extreme environment [2]. By comparing with the correspondingly low-altitude populations, several animal strains living in the TP are well adapted to these harsh environmental conditions [3, 4], indicating the genomic plasticity in animals increases adaptation of the TP animals to high-altitude stressor. In particular, field manipulation experiments offer the power to distinguish transcriptional variation that is plastic over physiological timescales. Therefore, alteration of abundance in divergent-altitude environments is key pattern of adaptive evolution.

Studies of the transcriptome response to environmental stressors can provide important insights into the genes (i.e. expression) involved in physiological acclimation and evolutionary adaptation [1]. Yet, group-specific genetic basis for high-altitude adaptation were also detected in some TP animals. Tang *et al.* identified differentially expressed genes (DEGs) among domestic goats living in different altitudes, and most of them were not enriched to gene ontology (GO) terms commonly associated with hypoxia in mammals, but were ribosome and protein synthesis [5]. Insects

2018 International Conference on Biotechnology and Bioengineering (8th ICBB)
AIP Conf. Proc. 2079, 020020-1–020020-6; https://doi.org/10.1063/1.5092398
Published by AIP Publishing. 978-0-7354-1807-3/$30.00

have specific exchange and respiratory system (the tracheal system) and oxygen is directly delivered to tissues where cellular respiration takes place by air-filled tubes and tracheoles, which is significantly different from vertebrate gas exchange system including lungs and a circulatory system with an oxygen carrier (haemoglobin) [6]. In addition, insects are ectothermic invertebrates, so environmental heat is a key power to promote living, development and reproduction of insects. These large physiological differences of insects with vertebrates may have evolved unique mechanisms under long-time adaptive evolution to high altitude environments. For example, mitochondrial genes expression of insects is opposite to that of vertebrates under low oxygen [7]. These observations imply that mechanisms for high-altitude adaption of insects may have both convergent and divergent adaptive strategies with vertebrate.

D. baccarum (Hemiptera: Pentatomidae) is a polyphagous pest on various crops and one of the major pests of alfalfa in China. This pest is widely distributed in China, from eastern region (low altitude) to the TP (high altitude). Due to its broad geographic distribution that spans heterogeneous landscapes, there is potential for local adaptation. Moreover, populations occurring at different altitudes differ in physiological parameters that are likely to be adaptive. Previously, we determined uplift of the TP contributed to the speciation of the grassland caterpillars (Lepidoptera: Lymantriidae: *Gynaephora*) [8], indicating that high-altitude adaption is main power in the process of divergent evolution of insects to different altitudes. Therefore, a pair of *D. baccarum* strains living in different-altitude environments provides an ideal model system to study genetic basis of insects to adapt to high-altitude habitats.

MATERIALS AND METHODS

One of two *D. baccarum* strains was collected from the Loess Plateau with relatively low altitude (DBQY, 1300 m above sea level (a.s.l.)), 86% of oxygen content at sea level (percentage of oxygen content at sea level is calculated at http://www.altitude.org/air_pressure.php). Another strain was collected from alpine meadow of the TP (DBHC, 3200 m a.s.l., 69% of oxygen content at sea level). Clean reads of two *D. baccarum* strains were downloaded from NCBI Sequence Read Archive (SRA) repository [Accession Number: SRX2207346 and SRX2207347]. The *de novo* transcriptome assembly for these filtered reads was performed with Trinity software under default parameters [9]. We kept only contigs longer than 200 bp for subsequent analyses. We then used CD-HIT-EST with threshold 0.90 to cluster sequence and eliminated redundancy in the final assembly.

Based on the Trinity assembly results, the blastx tool with E-value of 10^{-5} was used to annotate unigenes to NCBI non-redundant (Nr) protein database, Swiss-Prot, and GO. We predicted open reading frames (ORFs) for each unigenes by Transdecoder program in the Trinity package.[10] To remove redundancy transcripts and keep more sequence information, the longest ORF of each reference genes was extracted to represent unique gene. Subsequently, each of ORFs was hit to Swisspprot and Pfam database using blastp and Hmmscan software based upon their corresponding protein sequences, to further guide extraction of coding sequence (CDS) by Transdecoder.Predict software [10]. Moreover, we annotated all predicted CDS according to Nr protein database to perform further GO enrichment.

All reads of each of DBQY and DBHC were mapped to referenced all-unigenes set of *D. baccarum* using Bowtie software [11], and then we calculated expression level of each genes by using RSEM (RNA-Seq by Expectation Maximization) software for each of two populations based on reads per kilobases per million reads (RPKM) value . Subsequently, DEGs and their significance level were detected with EBSeq software based upon RPKM [12]. Then we corrected significance by control the FDR according to the above mentioned BH method. Finally, If fold change of unigenes expression ≥ 2 and FDR < 0.001, these DEGs were considered as reliable targets in further analyses. GO enrichment for DEGs was conducted using Blast2GO pipeline [13]. Significance levels of all GO terms were corrected according to the methods of Benjamini and Hochberg (BH) by using the *p.adjust* module in the R package to control the false discovery rate (FDR) in multiple pairwise comparisons. The GO terms with corrected *p*-values of < 0.05 and ES values of > 1.3, were considered as significantly enriched target categories.

To ensure validation of the RNA-Seq results, a portion of the pooled total RNA was used for qRT-PCR analysis as follows. We synthesized first-strand cDNA using RevertAid™ First Strand cDNA Synthesis Kit (Thermo Scientific, USA) according to product manuals. The cDNA products were diluted to 100 ng/μl using free-RNase water. We designed gene-specific primers for the qRT-PCR using Beacon Designer 7 (Table 1). The experiments were performed on the ABI PRISM 7500 Fast Real-Time PCR System (Applied Biosystems, USA) with SYBR® Premix Ex Taq II Kit (Takara, Japan). We selected *Ef1a* gene as the endogenous control, and each qRT-PCR analysis was performed in triplicate. Calculation of the relative expression level for every gene was performed based on the $2^{-\Delta\Delta CT}$ method [14].

TABLE 1. Primers used in qRT-PCRs.

Unigene name	Primer	Sequence (5'-3')	Size (bp)
Unigene21429_All	DBP1F	ATTTGCGTGAAAGGAAAG	139
	DBP1R	CTCAAGTTAATCAAGATGGAA	
Unigene18939_All	DBP2F	TCAGTATAATAGCCTTCTTG	134
	DBP2R	GTCCATTCCACTTAACTTA	
Unigene19176_All	DBP3F	CTGCCTTGTAAGTTAAGA	103
	DBP3R	GCTCTGTGGTATTAGTAG	
Unigene35465_All	DBP4F	CTCACTTCTTGGCTACAG	77
	DBP4R	GAGGTCGGTAGAGAGTAT	
Unigene5640_All	DBP5F	TCACCGAAGAAGAAGTAG	113
	DBP5R	ACGATAGAAAGACGACAT	

RESULTS

By assembling clean reads downloaded from SRA database, 126,494 reconstructed contigs were generated using Trinity in DBQY, with a median length of 388 bp and an N50 of 781 bp. We also obtained 114,298 contigs with a median length of 409 bp and an N50 of 1018 bp from DBHC. We further clustered contigs to produce non-redundant unigene sets by using CD-HIT-EST for each of DBQY and DBHC. Total 84,926 unigenes were obtained with an N50 length of 1,477 bps and a mean length of 717 bps in DBQY. For DBHC, 72,923 unigenes were obtained with an N50 length of 2,133 bps and a mean length of 929 bp. We constructed a reference dataset by clustering mixed assemblies from two *D. baccarum* strains to analyze DEGs.

A total of 4,816 unigenes were differentially expressed between two populations, of which up-regulated genes were 3,052 and down-regulated genes were 1,764, and former distribution evidently exhibited more dispersive and wide state (Fig. 1). We also performed quantitative real-time PCR for randomly selecting 5 DEGs to further confirmed transcriptomic results. The results from qRT-PCR were consistent with the sequencing data (Pearson correlation coefficients = 0.885, $P < 0.05$). All of DEGs were significantly enriched to 48 GO terms that belong to three classifications, of which biological process occupied 28 GO categories, 13 Go terms were included in cellular function and 7 GO terms involved in molecular component. Within biological process, the GO terms mainly associated with ubiquitination regulation, nucleus regulation, response to stress and DNA damage, mitochondrial function, oxidation-reduction process and histone regulation" (Table 2). The GO terms belong to cellular component were mainly involved in mitochondrial structure and function, ATP production and metabolism, cellular part and structure. Among molecular component, DEGs were significantly enriched to minimum GO terms, which mainly are related to ubiquitination function, ATP production, compound and protein binding and nucleic acid binding.

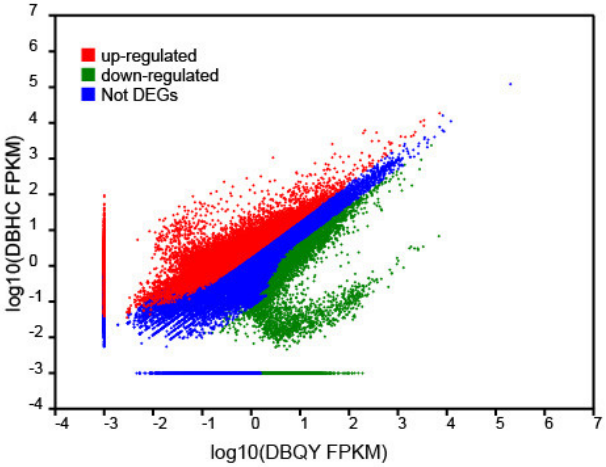

FIGURE 1. Expression level of two *D. baccarum* populations. The genes are classified into three classes. Red dot indicates up-regulated genes of DBHC to DBQY. Green dots are down-regulated genes. Blue dot indicates genes without different

expressions. The horizontal coordinate is the expression level of DBQY and the vertical coordinate is the expression level of DBHC.

TABLE 2. List of the enriched GO terms of differentially expressed genes between the two *Dolycoris baccarum* populations. Note: Enrichment score (ES) values of all listed terms were more than 1.3.

Go terms (Id)	FDR	Subcategory
mitotic cell cycle (GO:0000278)	1.41E-09	Biological process
negative regulation of protein modification process (GO:0031400)	5.13E-08	Biological process
regulation of catenin import into nucleus (GO:0035412)	7.46E-08	Biological process
regulation of protein import into nucleus, translocation (GO:0033158)	6.19E-07	Biological process
mitotic spindle organization (GO:0007052)	1.02E-06	Biological process
negative regulation of histone ubiquitination (GO:0033183)	1.08E-06	Biological process
negative regulation of DNA repair (GO:0045738)	1.12E-06	Biological process
regulation of histone H2A K63-linked ubiquitination (GO:1901314)	1.43E-06	Biological process
negative regulation of double-strand break repair (GO:2000780)	2.08E-06	Biological process
negative regulation of protein ubiquitination (GO:0031397)	3.67E-06	Biological process
mitochondrial electron transport, NADH to ubiquinone (GO:0006120);	4.88E-06	Biological process
negative regulation of cellular protein metabolic process (GO:0032269)	5.91E-06	Biological process
negative regulation of response to DNA damage stimulus (GO:2001021)	6.73E-06	Biological process
intracellular steroid hormone receptor signaling pathway (GO:0030518)	8.19E-06	Biological process
positive regulation of cellular process (GO:0048522)	1.20E-05	Biological process
mitotic chromosome condensation (GO:0007076);	1.20E-04	Biological process
oxidation-reduction process (GO:0055114);	1.50E-04	Biological process
regulation of histone modification (GO:0031056)	6.90E-04	Biological process
chromosome organization (GO:0051276)	1.36E-03	Biological process
negative regulation of chromosome organization (GO:2001251)	1.53E-03	Biological process
positive regulation of protein transport (GO:0051222)	2.50E-03	Biological process
mitochondrion organization (GO:0007005);	3.10E-03	Biological process
response to stress (GO:0006950)	4.23E-03	Biological process
mitotic DNA damage checkpoint (GO:0044773)	1.33E-02	Biological process
multicellular organismal process (GO:0032501)	2.91E-02	Biological process
gland morphogenesis (GO:0022612)	3.36E-02	Biological process
M phase of mitotic cell cycle (GO:0000087)	4.23E-02	Biological process
negative regulation of DNA metabolic process (GO:0051053)	4.81E-02	Biological process
cytosol (GO:0005829);	1.29E-17	Cellular function
mitochondrial respiratory chain (GO:0005746);	2.32E-09	Cellular function
mitochondrial part (GO:0044429);	2.43E-07	Cellular function
protein complex (GO:0043234);	3.01E-07	Cellular function
ATP-binding cassette (ABC) transporter complex (GO:0043190);	4.72E-07	Cellular function
cytoplasm (GO:0005737);	9.09E-06	Cellular function
mitochondrion (GO:0005739);	1.68E-05	Cellular function
nuclear part (GO:0044428)	1.43E-03	Cellular function
cell part (GO:0044464)	2.37E-03	Cellular function
nucleus (GO:0005634)	1.33E-02	Cellular function
cytoplasmic part (GO:0044444)	1.59E-02	Cellular function
microtubule associated complex (GO:0005875)	3.51E-02	Cellular function
respiratory chain complex IV (GO:0045277)	4.92E-02	Cellular function
ATP binding (GO:0005524)	2.37E-09	Molecular component
ubiquitin-ubiquitin ligase activity (GO:0034450)	1.24E-08	Molecular component
protein binding (GO:0005515)	4.24E-05	Molecular component
organic cyclic compound binding (GO:0097159)	2.37E-03	Molecular component
ribonucleoside binding (GO:0032549)	3.42E-02	Molecular component
antioxidant activity (GO:0016209)	4.58E-02	Molecular component
RNA binding (GO:0003723)	4.64E-02	Molecular component

DISCUSSION

Previous studies extensively focus on response of sequence variation to high-altitude, but exploration for genetic basis of divergent-altitude adaptation based on gene expression was limited, such as only in goats and migratory locust (*Locusta migratoria*) [4, 5]. The DEGs between DBHC and DBQY were mainly enriched to histone ubiquitination and modification, mitotic function and DNA repair, mitochondrial electron transport (Table 2). Here, the results firstly showed GO categories involved in histone ubiquitination and modification occupied the largest proportion in all enriched terms of DEGs. Previous evidence showed that histone ubiquitination and modification play a key role in regulation of DNA damage response [15]. Secondly, several GO terms were related to mitotic function and DNA repair. According to previous studies, DNA repair could be regulated throughout the cell cycle [16], and we enriched related-mitotic function terms mainly participated in response to DNA damage, including "mitotic DNA damage checkpoint", "mitotic DNA integrity checkpoint" and even "mitotic spindle organization (function of DNA repair)" [17]. Overall, a surprising overrepresentation from our enriched GO list was categories associated with DNA repair. This observation may be explained by several reasons as follows: with hypoxia different, UV is an intermittent stress in the TP (only in the daylight), and is generally easier to mitigate by moving to the shelter or sometime receiving sufficient heat from microhabitats. Additionally, a number of candidate GO categories involved in DNA repair were also found in Tibetan locusts [4]. Therefore, intermittent and essential exposure of *D. baccarum* in UV triggered expression changes of genes associated with DNA repair to confer protection for genome stability in high-altitude adaptation. Furthermore, this phenomenon indicates adaptive evolution to non-sustainable challenge may be mainly reflected on gene expression in *D. baccarum*, as well as strong UV is also one of the main environmental stressors for insects at high altitude (it was even not inferior to hypoxia). Certainly, we found that GO category associated with ATP production and mitochondrion under hypoxia, suggesting their potential involvement in hypoxia adaptation in *D. baccarum* living in the TP. Similar GO terms were also reported in endothermic and ectothermic vertebrates living in high altitude environments, including goats [5]. These results demonstrate the complexity of high-altitude adaptation in insects.

CONCLUSION

The high degree of plasticity in gene expression patterns over the long-term acclimation period is striking. DEGs showed several GO terms that may be involved in changes in oxygen content, UV radiation, and temperature, mainly includes oxidoreductase, energy metabolism, mitochondrion, histone modification, DNA repair, and response to DNA damage. These findings updated understanding of insect adaptation to divergent-altitude environments.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (31201520 and 31760042).

REFERENCES

1. Z. A. Cheviron, A. Whitehead, R. T. Brumfield, Mol. Ecol. **17**, 4556-4569 (2008).
2. Y. Wang, L. Yang, K. Zhou, Y. Zhang, Z. Song, S. He, Genome Biol. Evol. **7**, 2970-2982 (2015).
3. Y. Luo, X. Yang, Y. Gao, Mitochondr. DNA **24**, 313-319 (2013).
4. D. Zhao, Z. Zhang, A. Cease, J. Harrison, L. Kang, BMC Genomics **14**, 631-631 (2013).
5. Q. Tang, W. Huang, J. Guan, L. Jin, T. Che, Y. Fu, *et al.*, Gene **567**, 208-216 (2015).
6. S. M. Simelane, S. Abelman, F. D. Duncan, Acta Biotheoretica **64**, 161-196 (2016).
7. D. C. McMullen, K. B. Storey, Insect Biochem. Molec. Biol. **38**, 367-373 (2008).
8. M. L. Yuan, Q. L. Zhang, Z. F. Wang, Z. L. Guo, G. S. Bao, PloS ONE **10**, e0127257 (2015).
9. M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, *et al.*, Nature biotechnology **29**, 644-652 (2011).
10. B. J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, *et al.*, Nat. Protocols **8**, 1494-1512 (2013).
11. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Genome Biol. **10**, R25-R25 (2009).
12. N. Leng, J. A. Dawson, J. A. Thomson, V. Ruotti, A. I. Rissman, B. M. G. Smits, *et al.*, Bioinformatics **29**, 1035-1043 (2013).

13. A. Conesa, S. Götz, J. M. García-Gómez, J. Terol, M. Talón, M. Robles, Bioinformatics **21**, 3674-3676 (2005).
14. K. J. Livak, T. D. Schmittgen, Methods **25**, 402-408 (2001).
15. S. Jackson, D. Durocher, Molecular Cell **49**, 795–807 (2013).
16. D. Branzei, M. Foiani, Nat. Rev. Mol. Cell Biol. **9**, 297-308 (2008).
17. A. A. Guerrero, M. C. Gamero, V. Trachana, A. Fütterer, C. Paciosbras, N. P. Díazconcha*, et al.*, Prod. Natl. Acad. Sci. USA **107**, 4159-4164 (2010).