Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/etap

Short communication

Generally detected genes in comparative transcriptomics in bivalves: Toward the identification of molecular markers of cellular stress response

Jingjing Miao, Luping Chi, Luqing Pan, Ying Song*

The Key Laboratory of Mariculture, Ministry of Education, Ocean University of China, Qingdao 266003, PR China

ARTICLE INFO

Article history:
Received 3 September 2014
Received in revised form
12 December 2014
Accepted 13 December 2014
Available online 7 January 2015

Keywords:
Transcriptomic
Differential expression
Bivalve
Meta-analysis
Cellular stress response
Detection frequency

ABSTRACT

The specificity and representativeness of protein-coding genes identified by transcriptomics as biomarkers for environmental toxicological stress is crucial. We extracted the differential gene expression profile data from 49 published comparative transcriptomic studies of bivalves from January 2004 till November 2014 performed in 15 different bivalve species. Among the studies, 77 protein-coding genes were frequently detected when we use threefold of the average detection frequency as cut-off. Cellular organization and communication, protein and energy metabolism, stress response are the main functional classes of these proteins. We consider if these protein-coding genes represent common cellular stress responses of bivalves.

1. Introduction

Coastal ecosystems have been subject to increased contamination from both inorganic, e.g. heavy metals, and organic compounds, e.g. polycyclic aromatic hydrocarbons (PAHs) and organochlorine pesticides (OCPs) in the past decades. Pollution caused by these contaminants represents one of the major factors of environmental stress in aquatic environments, where the term "environmental stress" defines all the conditions in which physiological processes of a living system are altered by external factors (Viarengo and Canesi, 1991; Dondero et al., 2006). Bivalve molluscs play a fundamental role in the functioning of the marine ecosystem, constitute very valuable commercial resources in aquaculture, and have been widely used as sentinel organisms in the biomonitoring of marine pollution. Molecular, biochemical, cytological, immunological and physiological techniques have been extensively studied

* *Corresponding author at*: Laboratory of Environmental Physiology of Aquatic Animal, Fisheries College, Ocean University of China, Yushan Road 5, Qingdao 266003, PR China. Tel.: +86 532 82032963; fax: +86 532 82032963.
E-mail address: panlq@ouc.edu.cn (L. Pan).

http://dx.doi.org/10.1016/j.etap.2014.12.007
1382-6689/

in bivalves and developed for application in monitoring and assessing deleterious effects in biological systems (Galloway et al., 2002, 2004; Handy et al., 2003; Lam and Gray, 2003; Bolognesi et al., 2004).

General physiological stress is manifested in many ways, e.g. decrease in lysosomal membrane stability, indicators of oxidative stress (lipid peroxidation and antioxidant enzyme activities) and genotoxicity (DNA integrity) (Moore et al., 2004). These biological effects are non-specific in regard to different groups of contaminants and highly sensitive to chemical stressors (Lehtonen and Schiedek, 2006; van der Oost et al., 2003). With corresponding to the general biological effects, responses to stress stimuli are supported by a more or less important reprogram of the gene expression, where a specific gene can be alternatively induced or suppressed depending on its physiological role, as demonstrated for yeast cells challenged with a set of stress agents (Gasch et al., 2000). Not only in yeast but evolutionarily conserved in all cellular organisms, cellular stress response (CSR) is a universal mechanism and not stressor specific (Kültz, 2003). Proteins involved in key aspects of the CSR are conserved in all organisms (Kültz, 2005). With this context comes the imperative to improve our understanding of the molecular biomarkers in bivalves indicating the CRS, particularly those that may serve as early warning signals of the health status in bivalves, a pre-requirement when following the precautionary principle.

The last 10 years have seen the introduction and wide spread adoption of transcriptomic techniques used in comparative mRNA profiling studies for the identification of biomarkers for specific conditions of bivalves or components of biological processes and pathways (Li et al., 2013; Suárez-Ulloa et al., 2013). On the basis of these transcriptomic studies, we performed a meta-analysis with current comparative transcriptional data of bivalves, in order to explore the generally detected protein-coding genes and discuss the association of the genes with CRS in bivalves.

2. Methods

The differential gene expression profile data of bivalves were extracted from 21 published transcriptomic studies deal with toxic effects of heavy metals or organic contaminants that used the subtractive (SSH), high-throughput deep sequencing or microarray technique and were performed in 8 different bivalve species. We also analyzed data from 12 studies addressed transcriptional response of bivalves to pathogens and 16 studies addressed other biological questions, in order to make a comparison with the frequently detected protein-coding genes from toxic effect studies. The total dataset contained all the mRNA identifications presented in the 49 articles was shown in Table S1. These original research studies were from journals via PUBMED (latest access on 25 November 2014).

Supplementary Table S1 related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.etap.2014.12.007.

From each paper, we extracted the differentially expression genes into our database manually. When the same gene appeared more than once in the same study due to either different isoforms or detection under different conditions, we extracted it only once. We did not consider whether the protein was up- or down-regulated. The "detection frequency" was defined as the chance that a particular gene can be detected in one experiment, which is equal to the gene counts divided by the number of source papers. Threefold of the average detection frequency was defined as the threshold for "the frequently detected protein". This threshold was equal to 95 percentiles in general. By this definition, we picked out the list of frequently detected protein-coding genes in the total 49 studies and also in the 21 toxic studies and 12 phathogenic stress studies.

The functional interpretation of the frequently detected protein-coding genes was based on Gene Ontology. On the basis of literature information, mainly from NCBI, the generally detected proteins were divided into functional classes using GOC at Database (2013) (http://eagl.unige.ch/GOCat/) (Gobeill et al., 2013). Each protein-coding gene was assigned to one major function. The genome ontology identities were determined using CateGOrizer (http://www.animalgenome.org/bioinfo/tools/countgo/) (Hu et al., 2008).

3. Results and discussion

Among the protein-coding gene identifications from the 21 toxic effect studies, the average detection frequency of one particular protein to be found in one study is 0.05. However, results of our survey showed that genes code for some protein or protein families are much more often detected in general, with a maximum detection frequency of 0.67, regardless of the exposed chemical type in different bivalve species and tissues. We used threefold of the average (0.14) as the cut-off ($p < 0.05$) and found that 77 protein-coding genes were frequently detected over all gene identifications (Table 1).

We also analyzed the frequency of protein-coding genes in pathogenic stress studies and other biological stress studies separately. The detected frequency of 21 protein-coding genes is over threefold of the average (0.25) in the 12 pathogenic stress experiments, while in the 16 other environmental stress studies, there are 56 protein-coding genes with detected frequency over threefold of the average (0.19) (Table 1). For the total protein-coding gene identifications in the 49 published studies, we listed the top 77 frequently detected protein-coding genes. The detected frequencies of these 77 genes were over fivefold of the average detected frequency (0.1) (Table 1). Among the top 77 frequently detected genes from the total 49 studies, 53 genes were the same with the 77 frequently detected genes from the toxic studies, and 24 genes were different (shown in boldface in Table 1).

The genes at the top of the list code for cytoskeleton proteins (tubulin, actin and myosin), substance metabolism proteins (cathepsin, ubiquitin, serine protease, CYP450 and glutathione S-transferase), cellular organization and biogenesis (actin, tubulin, bromodomain-containing protein, collagen, etc.), cellular growth, cycle and death (elongation factor, ferritin, cyclin, etc.), nucleotide and nucleic acid metabolism (rab GTPase-activating protein, myc-homolog) and response to stress and stimuli such as defensin, glutathione peroxidase (GPx), superoxide dismutase (SOD), heat shock protein