

Looking for biomarkers of Hg exposure by transcriptome analysis in the aquatic plant *Elodea nuttallii*

N. Regier¹, L. Baerlocher², M. Münsterkötter³, L. Farinelli² and C. Cosio^{1*}

¹ Institut F.-A. Forel, Geneva University, CH-1290 Versoix, Switzerland. Claudia.Cosio@unige.ch

² Fasteris SA, CH-1228 Plan-les-Ouates, Switzerland

³ IBIS, Helmholtz Centre Munich, 85764 Neuherberg, Germany

Abstract. Recently developed genomics tools have a promising potential to identify early biomarkers of exposure to toxicants. In the present work we used transcriptome analysis (RNA-seq) of *Elodea nuttallii* –an invasive rooted macrophyte that is able to accumulate large amounts of metals- to identify biomarkers of Hg exposure. RNA-seq allowed identification of genes affected by Hg exposure and also unraveled plant response to the toxic metal: a change in energy/reserve metabolism caused by the inhibition of photosynthesis, and an adaptation of homeostasis networks to control accumulation of Hg. Data were validated by RT-qPCR and selected genes were further tested as biomarkers. Samples exposed in the field and to natural contaminated sediments clustered well with samples exposed to low metal concentrations under laboratory conditions. Our data suggest that this plant and/or this approach could be useful to develop new tests for water and sediment quality assessment.

Key words: Toxic metals, macrophyte, Cd, RNA-seq, water column, sediments

Introduction

Contamination of surface waters with metals, mainly caused by human activities, is a topic of great concern. To gain insight into the impact of metal contamination on ecosystems, the potential of various tests on metal accumulating organisms is explored. Recent genomic tools such as gene expression profiling show promising potential to identify biomarkers of exposure for use in the field. However, gene expression should react specifically and in a dose dependent manner to be used as biomarker. Rooted macrophytes are well-suited to be used in ecotoxicological tests because they are exposed to both water and sediments. *Elodea nuttallii* is a rooted macrophyte native to Northern America, but as a neophyte, has spread throughout the world. It is able to accumulate large amounts of metals, including Hg and Cd, and is therefore a good candidate for the aim of this study (Regier et al., 2012).

In the present work, we intended to identify biomarkers suitable to reveal Hg exposure in the field.

Materials and Methods

Plant exposure

Exposure of 10 cm long shoots collected in our culture (*Elodea nuttallii* from Lake Geneva established in microcosms in 2008) took place in 1 L beakers containing 1.2 µm filtered water from Lake Geneva under standardized conditions (20°C, 16/8 h light/dark) for 24 h. For RNAseq and RT-qPCR exposures were: no spike (control), 500 µg/L and 5 mg/L CdCl₂, 200 ng/L, 80 µg/L and 1 mg/L HgCl₂. For nCounter: control, 10, 50 and 200 ng/L HgCl₂, 200 µg/L HgCl₂, 30 ng/L CH₃HgCl, 200 ng/L HgCl₂ + 30 ng/L CH₃HgCl, 65 ng/L and 325 µg/L CuCl₂, 200 ng/L HgCl₂ + 65 ng/L CuCl₂, 500 µg/L CdCl₂ and 200 ng/L HgCl₂ + 500 µg/L CdCl₂, and contaminated natural sediments (Hg concentration 1.8±0.03 mg/kg). Other parameters tested were salinity (4 g/L NaCl), cold (10°C) and darkness. Additionally, plants were exposed in the field 24h in the Hg contaminated Babeni reservoir and the upstream Valcea reservoir (Romania) as a reference (Bravo, 2010) at 1, 4 and 6 m depth, corresponding to 3.63±0.12, 2.9±0.96, 3.09±0.61 and 0.30±0.05 ng/L Hg respectively.

Total RNA isolation

Plants were quickly rinsed in MilliQ water, dried with tissue and immediately flash-frozen in liquid nitrogen. Total RNA was extracted from shoots using TRI-reagent (Sigma) and further purified using the RNeasy Plant Mini kit (Qiagen).

Illumina mRNA sequencing

Samples were prepared using the mRNA-Seq sample prep kit (Illumina). Briefly, the mRNA was purified from total RNA. Double-stranded cDNA was synthesized and bar-coded adapters ligated to both ends of the cDNA fragments. The cDNA fraction (200±25 bp) was excised from a 2% agarose gel and enriched by PCR using adapter-specific primers. The cDNA libraries were sequenced using Illumina Solexa flow cell and the Chrysalis 36 cycles sequencing kit for 2x54 cycles (control, 200 ng/L HgCl₂, 500 µg/L CdCl₂, 5 mg/L CdCl₂) or for 75 cycles (control, 80 µg/L HgCl₂, 1 mg/L HgCl₂).

De novo transcriptome assembly

De novo assembly of the transcriptome was achieved with the VELVET program and its extension OASES. The average insert size was set to 200 bp and the hash values to 31 to obtain the best assembly size as well as accuracy.

Mapping of sequence reads onto transcripts and quantification of gene expression

The MAQ program was used to map the reads of each sample on the assembled contigs to quantify gene expression. Transcripts with at least a 2-fold expression difference between two samples and regulated in a dose-dependent manner to Hg and Cd were identified.

Transcript annotation and functional categorization

Contigs were annotated first with BLASTX searches against the NCBI non-redundant protein database and assigned to GO terms using the Blast2GO program. The transcriptome was further analyzed using the Functional Catalogue developed by the Munich Information Center for Protein Sequences (MIPS). The differentially expressed subsets of transcripts were then compared to the whole transcriptome to identify functional categories that were significantly enriched. To correct for multiple testing, the Bonferroni correction as well as the false discovery rate (FDR) were calculated. Functional categories with a p-value < 0.05 and a FDR < 0.05 were considered as significantly enriched.

RT-qPCR

Expression of 8 genes revealed as differentially expressed by RNA-seq was determined relative to 3 stably expressed genes by RT-qPCR. RNA was reverse

transcribed using the PrimeScript RT reagent Kit (Takara Bio Europe).

NanoString nCounter analysis

A subset of 79 genes, found to be differentially expressed in a dose dependent manner, was chosen for a global expression analysis by nCounter (NanoString). Hybridisation was done on 200 ng purified RNA. Fold-change in gene expression of exposed samples was calculated relative to the negative and control samples and averaged over the biological replicates. Data were analyzed with the Genesis software by hierarchical clustering with average linkage.

Results and Discussion

Sequencing and de novo transcriptome assembly

RNA-Seq resulted in between 1'677'801 and 15'995'086 reads per sample, and in a total number of 50'187'798 54 bp paired-end reads and 14'418'361 75 bp single-end reads. Sequence quality was checked by using the phiX control, which revealed an error rate of 0.08%. The raw sequence data generated in this study has been deposited in the ArrayExpress archive at EMBL (number E-MTAB-892). The assembly yielded a total number of 63'596 contigs with an N50 contig size of 713 bp, an average length of 437 bp and a maximal length of 9'738 bp. 62.9% of the contigs were longer than 200 bp and 9.6% of the contigs were longer than 1000 bp. Total length of the assembly was 27'844'112 bp. A quality control of the assembly was performed by aligning the 8 partial mRNA sequences of *E. nuttallii*, available in the NCBI nucleotide database, vs. the contigs obtained in our study. All sequences matched at least one contig with 99-100 % sequence homology over the whole overlapping length.

Mapping of reads onto contigs and analysis of gene expression

Between 29.1 and 54.2% of the reads of each sample could be mapped. We determined the coverage of each transcript by reads from each individual sample and normalized the values to the sample with the lowest total number of reads. Using a cut-off of 2-fold change in gene expression, we identified transcripts regulated in a dose-dependent manner in response to Hg and Cd. By applying these criteria, we found a total number of 212 differentially expressed genes in response to Cd (54 up- and 158 down-regulated), and 170 genes in response to Hg exposure (84 up- and 86 down-regulated). None of the genes were up-regulated by both toxic metals, but for down-regulated genes, we found that 13 of them responded to Cd and to Hg. Furthermore, we did not identify any gene that was up-regulated by one toxic metal and down-regulated by the other toxic metal dose-dependently.

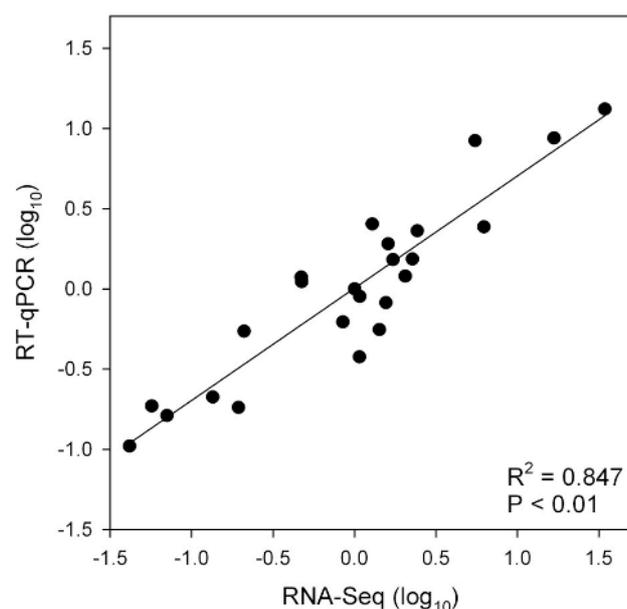
Validation of RNA-Seq by RT-qPCR

We selected 8 genes that showed distinct expression patterns in response to Hg or Cd, according to RNA-Seq. RT-qPCR analyses were performed on RNA samples obtained from an exposure independent from those used in RNA-Seq. We found a significant correlation between RNA-Seq and RT-qPCR data ($R^2 = 0.847$; $P < 0.01$; Figure 1).

Annotation and functional categorization by GO

BLASTX searches revealed sequence homologies ($E\text{-value} \leq 10^{-3}$) for 32'095 of the contigs (50.5%). Numerous contigs showed homologies to unknown, putative, hypothetical or expressed proteins. The sequences revealing significant BLAST hits were further annotated to GO terms. A total of 27'669 sequences (43.6%) could be assigned to at least one GO term. In the main category biological process, the most strongly represented processes were related to cellular process (58.5%), metabolic process (56.7%) and response to stimulus (21.0%). Genes involved in other important biological processes such as biological regulation, localization, developmental process and cellular component organization were also found to be well represented in the transcriptome. In the category of molecular function we found binding (56.2%) and catalytic activity (49.5%) to be most strongly represented. Transporter activity, structural molecule activity and molecular transducer activity were other important functions identified in the transcriptome.

Fig. 1. Correlation between RNA-Seq and RT-qPCR



data.

Several recent studies have also proven the use of very short sequences generated by RNAseq to be suitable for *de novo* assembly of the transcriptome of non-model

organisms, where no or scarce sequence information is available (Garg et al., 2011). In the ecotoxicology field many organisms are unsequenced, but results suggest that this kind of approach can now be reliably used with all organisms.

Functional categorization by FunCat

In addition to GO classification, we used the functional catalogue of MIPS to identify which biological functions were affected by Hg and Cd exposure. All sequences of the transcriptome that could be identified by BLASTX were classified. In addition, the genes represented in the differentially expressed subsets were classified to identify the functional categories that were significantly enriched in the subsets, and therefore considered as being related to toxic metal response. Concerning genes that were up-regulated in response to Hg contamination: 'cell cycle and DNA processing', 'interaction with the environment' and 'biogenesis of cellular components' were significantly enriched. Other important enriched categories were related to stress response, carbohydrate and protein metabolism and cellular signaling.

In the set of down-regulated genes in response to Hg, significantly enriched categories were 'cellular transport, transport facilities and transport routes' and 'interaction with the environment'. Within the category 'interaction with the environment', we found that the most strongly enriched sub-categories were related to 'homeostasis of cations' as well as 'homeostasis of metal ions' for down-regulated genes. Analysis of sub-categories of 'cellular transport, transport facilities and transport routes', revealed the sub-level category 'metal ion transport' to be strongly enriched. Genes related to detoxification were also over-represented in the set of down-regulated genes.

Concerning the response to Cd, for up-regulated genes, the most strongly enriched category was 'metabolism of energy reserves', followed by ion transport and transport ATPases, as well as signal transduction. For down-regulated genes we found the strongest enrichments related to homeostasis, cellular import and transport facilities and detoxification.

These results can be interpreted such that *E. nuttallii* responded to Hg and Cd treatment by the up-regulation of proteins (e.g. chaperones) known for their stress response function. The modification of reserve metabolism, notably sugar-catabolizing proteins might be caused by inhibited production of energy reserves by photosynthesis. Down-regulation of metal transporters and genes related to homeostasis appeared to, most probably, control and reduce accumulation of toxic metals.

Cluster analysis of gene expression

We further performed a global expression analysis using the Nanostring nCounter for a subset of genes, found to be differentially expressed by RNA-Seq, to evaluate

whether gene expression is metal-specific, dose-dependent and able to predict metal exposure in a complex environment. Data were analyzed by hierarchical clustering to assess relationships between the gene expression in shoots exposed to various concentrations of Hg, Cd, Cu, exposed to the Babeni reservoir (Hg and herbicide contamination), and exposure to variations of other abiotic parameters i.e. salinity, cold and darkness.

We found two main clusters (Figure 2): samples exposed to high ($\mu\text{g/L}$ range) metal concentration formed one group, irrespective of the metal used for exposure. The second cluster consisted of samples exposed to low (ng/L range) metal concentrations.

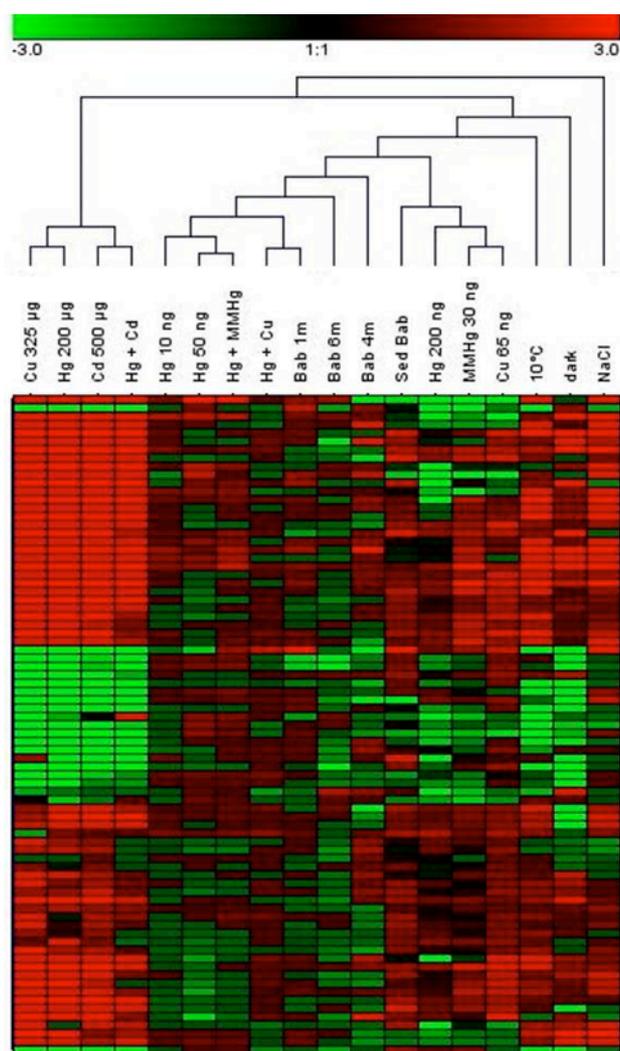


Fig. 2. Clustering of differentially expressed genes of plants exposed to various metal concentrations and in the field as revealed by nCounter analysis.

Globally, exposure to high metal concentrations led to more marked changes in the level of gene expression than exposure to low concentrations, as shown by color intensities in Figure 2. Samples exposed in the field and to natural contaminated sediments clustered well with samples exposed to low metal concentrations under laboratory conditions. Plants exposed to darkness, cold or salinity also showed a response, although different from the metal exposed samples. Genes identified in the laboratory after exposure to high concentrations were successfully used as biomarkers of Hg exposure in the field in a complex environment. On the other hand, the biomarkers also responded to high concentrations of Cu and Cd in the laboratory, suggesting that more genes would be needed to discriminate between the different metals. Moreover, most of the genes found to be differentially expressed by metals also responded to variation of the environmental factors examined, stressing that robust control is necessary to be able to apply the genomic tools in the field.

Conclusion

Data suggest that a genomic approach could be used to develop water and sediment quality tests with unsequenced organisms and confirmed the interest of rooted macrophytes for this aim.

Acknowledgements

The authors thank the financial supports provided from SER (contract IZLRZ3_128305) and by La Société Académique de Genève. Help in the fieldwork of the National Administration 'Apele Române'–Olt and Viorel Gheorghe Ungureanu (Bucarest University) are gratefully acknowledged

References

- Bravo AG. Mercury methylation and trophic transfer in contaminated freshwater systems. Geneva (Switzerland), 2010.
- Garg R, Patel RK, Tyagi AK, Jain M. *De Novo* Assembly of Chickpea Transcriptome Using Short Reads for Gene Discovery and Marker Identification. *DNA Research* 2011; 18: 53-63.
- Regier N, Larras F, Bravo AG, Ungureanu VG, Cosio C. Hg bioaccumulation in the macrophyte *Elodea nuttallii* in the field and in microcosm: Hg in shoots accumulated from the water might involve Cu transporters. *Chemosphere* 2012; submitted.