Construction of subtracted EST and normalised cDNA libraries from liver of chemical-exposed three-spined stickleback (Gasterosteus aculeatus) containing pollutant-responsive genes as a resource for transcriptome analysis

Brown, MM; Williams, Timothy; Chipman, James; Katsiadaki, I; Sanders, M; Craft, JA

DOI:
10.1016/j.marenvres.2008.02.043

License:
Creative Commons: Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

Document Version
Peer reviewed version

Citation for published version (Harvard):

Link to publication on Research at Birmingham portal

Publisher Rights Statement:

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

• Users may freely distribute the URL that is used to identify this publication.
• Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
• Users may not further distribute the material nor use it for the purposes of commercial gain.
• Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 19. Mar. 2021
Accepted Manuscript

Construction of subtracted EST and normalised cDNA libraries from liver of chemical-exposed three-spined stickleback (Gasterosteus aculeatus) containing pollutant responsive genes as a resource for transcriptome analysis

Margaret M. Brown, Timothy D. Williams, J. Kevin Chipman, Ioanna Katsiadaki, Matthew Sanders, John A. Craft

PII: S0141-1136(08)00048-2
DOI: 10.1016/j.marenvres.2008.02.043
Reference: MERE 3200

To appear in: Marine Environmental Research

Please cite this article as: Brown, M.M., Williams, T.D., Chipman, J.K., Katsiadaki, I., Sanders, M., Craft, J.A., Construction of subtracted EST and normalised cDNA libraries from liver of chemical-exposed three-spined stickleback (Gasterosteus aculeatus) containing pollutant responsive genes as a resource for transcriptome analysis, Marine Environmental Research (2008), doi: 10.1016/j.marenvres.2008.02.043

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Construction of subtracted EST and normalised cDNA libraries from liver of chemical-exposed three-spined stickleback (*Gasterosteus aculeatus*) containing pollutant responsive genes as a resource for transcriptome analysis

Margaret M. Brown a,*, Timothy D. Williams b, J. Kevin Chipman b, Ioanna Katsiadaki c, Matthew Sanders c, John A. Craft a

a Department of Biological and Biomedical Sciences, Glasgow Caledonian University, UK; b School of Biosciences, The University of Birmingham, UK; c CEFAS, Weymouth laboratory, UK.

Abstract

The three-spined stickleback (*Gasterosteus aculeatus*) is ideally suited to laboratory studies, while its wide distribution in the northern hemisphere gives it great potential as a sentinel organism. In the setting of a UK-wide collaboration (Fish Toxicogenomics) we have developed a microarray for transcriptomic analysis of chemical responses in populations of *G. aculeatus* under laboratory and field conditions. Although several EST libraries are available for this species none are from chemical-exposed fish and thus unlikely to include a full set of pollutant-responsive genes. To harvest such transcripts cDNA libraries were produced from liver of chemical-exposed mature males. Two normalised full-length libraries were generated
by different methods: 1) partial subtraction of polyA⁺RNA against solid-phase cDNA using magnetic bead technology; 2) degradation of double stranded cDNA formed by abundant transcripts. To enrich for pollutant-responsive genes a subtracted EST library was also generated. For each library ~1.5K clones were sequenced and characterised using Blast2GO. All libraries contained pollutant-responsive transcripts not previously available while additionally the subtracted library was generally enriched ~1.2 -10 fold for transcripts expected to be induced in response to the pollutants.

Keywords: Stickleback; cDNA; Subtractive suppression hybridisation; Microarray; normalisation

*Corresponding author: m.a.brown@gcal.ac.uk (M.A. Brown)

The three-spined stickleback (Gasterosteus aculeatus) is a small euryhaline species with wide distribution. Its prevalence in freshwater systems in Northern Europe and suitability to laboratory conditions (short life cycle; easily cultivated; readily induced breeding cycle) make it highly informative in ecotoxicological studies, in particular of endocrine disruption (Wibe et al., 2002; Hahlbeck et al. 2004a; Hahlbeck et al., 2004b; Katsiadaki et al., 2006). Biology, behaviour and physiology have been studied extensively in G. aculeatus however, functional genomics knowledge is absent although as a major research organism for evolutionary biologists (Peichel, 2005) the genome is sequenced though not annotated. In the setting of a UK-wide collaboration (Fish Toxicogenomics -
www.biosciences.bham.ac.uk/fishtoxicogenomics) an expression microarray for *G. aculeatus* has been developed for analysis of chemical responses under lab and field conditions. To be useful in this context, an array must include pollutant-responsive cDNAs. Available EST libraries (Kingsley et al., 2004), derived from multiple tissues of non-exposed fish, are unlikely to contain full suites of pollutant-responsive liver transcripts. These transcripts must be chemically induced prior to generation of normalised libraries; or alternatively selected by subtraction of constitutively expressed transcripts from the transcriptome of chemical-exposed fish. To provide resources for our array both approaches were adopted to generate full-length cDNA and EST libraries.

To induce oestrogen-, heavy metal- and PAH-responsive genes, fish were exposed to ethynyl oestradiol (EE2), copper (Cu) and di-benz(a,h)anthracene (DbA) respectively. For library construction polyA⁺ RNA was isolated from liver of mature male fish exposed (24 or 48 hours) to either Cu, 10 or 100 µg/l; EE2, 10 or 100 ng/l; DbA, 1 or 10 µg/l or water. Full-length cDNA libraries were produced from pooled polyA⁺ RNA (chemical-exposed and control samples (n=2) equal quantities from each time point/concentration) by alternate methods of normalisation: 1) partial-subtraction of polyA⁺ RNA against solid-phase cDNA using magnetic bead technology (Ga_NmlG); 2) degradation of double stranded cDNA formed by abundant transcripts (Ga_NmlY). To remove abundant transcripts in Method 1, first-strand cDNA was synthesised directly on oligo(dT) cellulose magnetic beads, RNA template eluted and the solid-phase cDNA library hybridised with fresh pooled polyA⁺ RNA. Resulting cDNA/RNA hybrids were collected by magnet and partially subtracted polyA⁺ RNA remaining in solution subjected to two further subtractions prior to synthesis and linear amplification of cDNA (SMART™PCR cDNA Synthesis Kit, Clontech) (Zhu et
This generated adequate material for construction of a phagemid library (SMART™ cDNA Library Construction Kit, Clontech). In Method 2, duplex-specific nuclease technology (DSN) (Zhulidov et al., 2004) was used for normalisation (DNA Trimmer Kit, Evrogen) and a plasmid library synthesised by long-distance PCR (Creator™ SMART™ cDNA Library Construction Kit, Clontech). For the subtracted EST library (Ga_S1TG), pooled polyA⁺ RNA from liver of chemical-exposed fish (equal quantities from each time-point/concentration) was subtracted against pooled control (equal quantities from 24 and 48 hours) by suppression subtractive hybridisation (SSH) as described (Brown et al., 2004).

Randomly selected clones (~1.5K from each library) were subjected to one-pass sequencing and scored for quality using PHRED20, minimum 100 consecutive bases as criteria. Sequencing quality varied (Table 1), consistently good results were obtained with plasmid vector (Ga_NmlY) but 3’ polyT structures within the phage vector (Ga_NmlG) and 3’ polyA and the 5’ GC rich region created by the SMART™ system within the ESTs (Ga_S1TG) caused a number to fail. All libraries contained pollutant-responsive transcripts not previously available in stickleback, however, the subtracted library was generally enriched ~1.2-10-fold for transcripts expected to be induced by the model chemicals.

Of the readable sequences 1817/3568 did not form contigs with those from the Kingsley library; of these 833 were singlets the remainder forming contigs within the libraries. Taking into account redundancy levels (Table 1) about 49% of the 1817 sequences represent unique clones not previously available however, several genes may be represented although not contiguous. It was not expected to infer which genes were altered specific to the pollutants, this will be discerned downstream by back-screening of the samples against the arrayed libraries.
Gene Ontology based data mining (Blast2GO (Conesa et al., 2005)) showed the libraries covered a wide range of biological functions. A small number of transcripts were universally over-represented, mainly ribosomal or those involved in protein synthesis, but generally gene identity profiles differed between libraries. In the normalised libraries this could be due to the relatively small numbers sequenced however it is noteworthy there was less redundancy in that normalised by partial subtraction (Method 1). The subtracted library was more enriched for chemical-responsive transcripts both known (e.g.s CYP1A, metallothionein, vitellogenin) and novel to this species (e.g.s DbA: fatty acid binding proteins, EE2: insulin-like growth factor binding protein; Cu: betaine homocysteine methyltransferase) (Table 2). Blast2GO, a powerful annotation tool for high-throughput processing of sequence data, has been shown to reach 60-70% accuracy with datasets where annotation/functionality is known. With ESTs from non-model species lower accuracy would be expected and more detailed manual annotation of predicted full length cDNAs extracted from the genomic database (http://www.ensembl.org) would be desirable but labour intensive.

Provision of resources for expression arrays must ideally be comprehensive for detection of gene events specific to chemical exposure. The methods used here for library construction sought to maximise transcriptome coverage accordingly. Starting material was pooled to produce single libraries containing transcripts responsive to several pollutants, an economy on time and resources. The resulting libraries were complementary and provided coverage of the transcriptome with the subtracted EST library more highly enriched for pollutant-responsive transcripts. The combined use of complementary cDNA libraries is therefore advantageous to isolate tissue or treatment specific transcripts.
References


Table 1.
Comparison of sequencing quality and Blast2GO annotation of the normalised and subtracted libraries.

<table>
<thead>
<tr>
<th>LIBRARY</th>
<th>Ga_NmlG (directional cloning)</th>
<th>Ga_NmlY (directional cloning)</th>
<th>Ga_S1TG (non-directional cloning)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Norm/Method 1</td>
<td>Norm/Method 2</td>
<td>Subtracted SSH</td>
</tr>
<tr>
<td>QUALITY SEQUENCE</td>
<td>751/1191 (63%)</td>
<td>1275/1440 (89%)</td>
<td>1542/2016 (77%)</td>
</tr>
<tr>
<td>NO HOMOLOG</td>
<td>127 (17%)</td>
<td>69 (5%)</td>
<td>73 (23%)</td>
</tr>
<tr>
<td>SIMILAR TO UNKNOWN</td>
<td>232 (31%)</td>
<td>221 (17%)</td>
<td>686 (45%)</td>
</tr>
<tr>
<td>ANNOTATED TRANSCRIPTS</td>
<td>392 (52%)</td>
<td>985 (78%)</td>
<td>783 (32%)</td>
</tr>
<tr>
<td>UNIQUE IDS</td>
<td>262/392 (67%)</td>
<td>394/985 (40%)</td>
<td>323/783 (40%)</td>
</tr>
<tr>
<td>GO ASSOCIATIONS</td>
<td>121</td>
<td>182</td>
<td>177</td>
</tr>
<tr>
<td>(BIOLOGICAL PROCESS/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOLECULAR FUNCTION)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2.

ESTs from transcripts up-regulated in response to chemical exposures (EE2, dibenzoanthracene, copper) were isolated from liver of *G. aculeatus* by suppression subtractive hybridisation (treated minus control). Randomly selected clones were subjected to one-pass sequencing and annotated using Blast2GO. All sequences are available at NCBI, dbEST accession numbers EG588073 – EG591696.
<table>
<thead>
<tr>
<th>Process</th>
<th>Accession</th>
<th>Most similar to</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase I metabolism</td>
<td>EG590952</td>
<td>cytochrome P450 1A</td>
</tr>
<tr>
<td></td>
<td>EG591310</td>
<td>Cytochrome P450 3K1 (CYP3K1) (P450 LMC2)</td>
</tr>
<tr>
<td></td>
<td>EG590829</td>
<td>cytochrome P450 2K5</td>
</tr>
<tr>
<td></td>
<td>EG591338</td>
<td>cytochrome P450 2N1</td>
</tr>
<tr>
<td></td>
<td>EG590214</td>
<td>cytochrome P450 2P</td>
</tr>
<tr>
<td></td>
<td>EG590878</td>
<td>cytochrome P450 3A49</td>
</tr>
<tr>
<td></td>
<td>EG590659</td>
<td>cytochrome P450 monoxygenase CYP2K6</td>
</tr>
<tr>
<td></td>
<td>EG590599</td>
<td>cytochrome P450, family 2, subfamily 1, polypeptide 2, B (CYP22B)</td>
</tr>
<tr>
<td></td>
<td>EG590223</td>
<td>cytochrome P450, family 4, subfamily a, polypeptide 1 (CYP4A1)</td>
</tr>
<tr>
<td></td>
<td>EG591178</td>
<td>cytochrome P450, (CYP2) (naphthalene-1-epoxide)</td>
</tr>
<tr>
<td></td>
<td>EG590478</td>
<td>similar to Cytochrome P450 7A1 (Cholesterol 7-alpha-hydroxylase)</td>
</tr>
<tr>
<td></td>
<td>EG591390</td>
<td>similar to family 4 cytochrome P450 70 isoform 1</td>
</tr>
<tr>
<td>Phase II metabolism</td>
<td>EG088877</td>
<td>glutathione S-transferase</td>
</tr>
<tr>
<td></td>
<td>EG590578</td>
<td>PREDICTED similar to UDP-glucuronosyltransferase 2A1 precursor, microsomal</td>
</tr>
<tr>
<td></td>
<td>EG590874</td>
<td>similar to sulfotransferase family 3A, member 1</td>
</tr>
<tr>
<td></td>
<td>EG591152</td>
<td>UDP-glucuronosyltransferase (UDPGT) UGT3</td>
</tr>
<tr>
<td>Oxidative stress</td>
<td>EG590483</td>
<td>14kDa apolipoprotein</td>
</tr>
<tr>
<td></td>
<td>EG590549</td>
<td>17-beta-hydroxysterol dehydrogenase type 3</td>
</tr>
<tr>
<td></td>
<td>EG590666</td>
<td>3-xeno-5-alpha-monoo-4 alpha-hydroxyn酸 2</td>
</tr>
<tr>
<td></td>
<td>EG590205</td>
<td>B-cadherin</td>
</tr>
<tr>
<td></td>
<td>EG590665</td>
<td>beta globin mRNA, complete cds</td>
</tr>
<tr>
<td></td>
<td>EG590420</td>
<td>beta tubulin</td>
</tr>
<tr>
<td></td>
<td>EG591549</td>
<td>beta-2 microglobulin precursor</td>
</tr>
<tr>
<td></td>
<td>EG591437</td>
<td>calreticulin 1, epithelial</td>
</tr>
<tr>
<td></td>
<td>EG590927</td>
<td>chitinase</td>
</tr>
<tr>
<td></td>
<td>EG590252</td>
<td>cholesteryl-ester transfer protein</td>
</tr>
<tr>
<td></td>
<td>EG591502</td>
<td>cholesterol II</td>
</tr>
<tr>
<td></td>
<td>EG590704</td>
<td>chorogettin L</td>
</tr>
<tr>
<td></td>
<td>EG590274</td>
<td>chiron protein</td>
</tr>
<tr>
<td></td>
<td>EG590642</td>
<td>eukaryotic binding inhibitor (GABA receptor modulator, acyl-Coenzyme A binding protein)</td>
</tr>
<tr>
<td></td>
<td>EG590598</td>
<td>dmp1</td>
</tr>
<tr>
<td></td>
<td>EG590861</td>
<td>similar to insulin-like growth factor 2 precursor (LOC532929) mRNA</td>
</tr>
<tr>
<td></td>
<td>EG590963</td>
<td>VgC mRNA for phospholipid vesicles, complete cbs</td>
</tr>
<tr>
<td></td>
<td>EG590342</td>
<td>violutoprotein</td>
</tr>
<tr>
<td></td>
<td>EG592123</td>
<td>zona-pellucida protein</td>
</tr>
<tr>
<td></td>
<td>EG590468</td>
<td>PREDICTED similar to calmodulin</td>
</tr>
<tr>
<td>Cholesterol metabolism</td>
<td>EG590481</td>
<td>15 kDa seleniumprotein</td>
</tr>
<tr>
<td></td>
<td>EG590673</td>
<td>acetyl-Coenzyme A acyltransferase 2 (acetocholesteryl-Coenzyme A dehydrogenase)</td>
</tr>
<tr>
<td></td>
<td>EG590254</td>
<td>acyl-CoA thioesterase 1</td>
</tr>
<tr>
<td></td>
<td>EG590993</td>
<td>adenosine monophosphate reductase translocator</td>
</tr>
<tr>
<td></td>
<td>EG590934</td>
<td>aluminio glycerole aminotransferase</td>
</tr>
<tr>
<td></td>
<td>EG590172</td>
<td>aldolase B</td>
</tr>
<tr>
<td></td>
<td>EG591220</td>
<td>Annexin A5</td>
</tr>
<tr>
<td></td>
<td>EG590467</td>
<td>bile acid-Coenzyme A: amino acid Coenzyme A transferase</td>
</tr>
<tr>
<td></td>
<td>EG591607</td>
<td>clineC232 ubiquin mRNA, partial cbs</td>
</tr>
<tr>
<td></td>
<td>EG591448</td>
<td>cholester-2-pyrophosphate</td>
</tr>
<tr>
<td></td>
<td>EG590882</td>
<td>glutathione Peroxidasse</td>
</tr>
<tr>
<td></td>
<td>EG590546</td>
<td>peroxiredoxin 4</td>
</tr>
<tr>
<td></td>
<td>EG591402</td>
<td>peroxiredoxin 2, mitochondria isoform 2</td>
</tr>
<tr>
<td></td>
<td>EG590575</td>
<td>phospholipid hydroperoxide glutathione peroxydase A</td>
</tr>
<tr>
<td></td>
<td>EG591340</td>
<td>PREDICTED similar to ubiquitin isoform 1</td>
</tr>
<tr>
<td>Copper responsive proteins</td>
<td>EG591322</td>
<td>similar to 80-beta-spectrin-binding protein (FLAP) (MLK-886-binding protein)</td>
</tr>
<tr>
<td></td>
<td>EG591574</td>
<td>ubiquitin-binding protein homolog</td>
</tr>
<tr>
<td></td>
<td>EG591609</td>
<td>ubiquitin-conjugating enzyme E2-like</td>
</tr>
<tr>
<td></td>
<td>EG591131</td>
<td>thioredoxin-like 4 (tia4), mRNA</td>
</tr>
<tr>
<td>Cholesterol metabolism</td>
<td>EG590419</td>
<td>COMM domain containing 3</td>
</tr>
<tr>
<td></td>
<td>EG588881</td>
<td>betaine-homocysteine methyltransferase</td>
</tr>
<tr>
<td></td>
<td>EG590588</td>
<td>cytochrome c oxidase subunit 1</td>
</tr>
<tr>
<td></td>
<td>EG591098</td>
<td>cytochrome c oxidase subunit II</td>
</tr>
<tr>
<td></td>
<td>EG590623</td>
<td>cytochrome c oxidase subunit III</td>
</tr>
<tr>
<td></td>
<td>EG590661</td>
<td>cytochrome c oxidase, subunits VIa</td>
</tr>
<tr>
<td></td>
<td>EG591694</td>
<td>PREDICTED similar to retinol binding protein 7, cellular</td>
</tr>
<tr>
<td>fatty acid metabolism</td>
<td>EG591416</td>
<td>fatty acid binding protein Hb-isoenzyme</td>
</tr>
<tr>
<td></td>
<td>EG590708</td>
<td>fatty acid binding protein</td>
</tr>
<tr>
<td></td>
<td>EG591611</td>
<td>Hydroxycyt-8-Coenzyme A dehydrogenase/se-7-ketocor-5-Coenzyme A thiolesterase</td>
</tr>
<tr>
<td></td>
<td>EG590737</td>
<td>luecic fatty acid binding protein</td>
</tr>
<tr>
<td></td>
<td>EG590662</td>
<td>acetyl-Coenzyme A acyltransferase 2 (acetocholesteryl-Coenzyme A thiolesterase)</td>
</tr>
<tr>
<td></td>
<td>EG590299</td>
<td>THT-binding protein mRNA, partial cbs</td>
</tr>
<tr>
<td>Cholesterol metabolism</td>
<td>EG590393</td>
<td>transmembrane 7 superfamily member 2</td>
</tr>
<tr>
<td></td>
<td>EG590455</td>
<td>isoprostane-diphosphate delta isoenzyme</td>
</tr>
<tr>
<td></td>
<td>EG592723</td>
<td>fatty acid diphosphate synthase (dihydroxyacetone-1-phosphate synthase)</td>
</tr>
<tr>
<td></td>
<td>EG590478</td>
<td>similar to Cytochrome P450 7A1 (Cholesterol 7-alpha-hydroxylase)</td>
</tr>
<tr>
<td></td>
<td>EG590252</td>
<td>cholesteryl-ester transfer protein</td>
</tr>
<tr>
<td>transcriptional repression</td>
<td>EG591387</td>
<td>RNA-binding protein VgRP71</td>
</tr>
<tr>
<td>Apoptosis</td>
<td>EG591110</td>
<td>PREDICTED similar to Ras GTPase-activating-like protein IQGAP1 (P195)</td>
</tr>
<tr>
<td></td>
<td>EG591612</td>
<td>cdk death-regulatory protein GRIM9</td>
</tr>
<tr>
<td></td>
<td>EG590808</td>
<td>PEST-containing nuclear protein</td>
</tr>
<tr>
<td></td>
<td>EG590888</td>
<td>PEST-2</td>
</tr>
</tbody>
</table>

**NOTES:**
- **Phase I metabolism:** Processes related to the initial activation of xenobiotics and endogenous compounds, often involving cytochrome P450 enzymes.
- **Phase II metabolism:** Processes that conjugate xenobiotics and endogenous compounds, often involving glutathione S-transferases and UDP-glucuronosyltransferases.
- **Oxidative stress:** Processes that respond to cellular damage due to reactive oxygen species.
- **Cholesterol metabolism:** Processes involved in the synthesis, catabolism, and transport of cholesterol.
- **Fatty acid metabolism:** Processes involved in the synthesis and degradation of fatty acids.
- **Transcriptional repression:** Processes that suppress gene expression.
- **Apoptosis:** Processes involved in programmed cell death.