

A cDNA Microarray for *Crassostrea virginica* and *C. gigas*

Matthew J. Jenny,^{1,2} Robert W. Chapman,^{1,3} Annalaura Mancina,^{1,4} Yian A. Chen,⁵ David J. McKillen,^{1,4} Hal Trent,^{1,4} Paul Lang,⁶ Jean-Michel Escoubas,⁷ Evelyne Bachere,⁸ Viviane Boulo,⁸ Z. John Liu,⁹ Paul S. Gross,^{1,4} Charles Cunningham,^{1,10} Pauline M. Cupit,^{1,10} Arnaud Tanguy,¹¹ Ximing Guo,¹² Dario Moraga,¹³ Isabelle Boutet,¹¹ Arnaud Huvet,¹⁴ Sylvain De Guise,¹⁵ Jonas S. Almeida,¹⁶ Gregory W. Warr^{1,4}

¹Hollings Marine Laboratory, Charleston, SC 29412, USA

²Department of Biology, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA

³South Carolina Department of Natural Resources, Marine Resources Research Institute, Charleston, SC 29412, USA

⁴Department of Biochemistry and Molecular Biology, Medical University of South Carolina, Charleston, SC 29425, USA

⁵Department of Statistics, Texas AM University, College Station, TX 77843, USA

⁶Department of Fisheries and Wildlife, Oregon State University, Corvallis, OR 97331, USA

⁷Écologie Microbienne des Insectes et Interactions Hôte-Pathogène, Université de Montpellier, 34095, Montpellier, France

⁸IFREMER-CNRS-Université de Montpellier II, UMR5171—Génome Populations Interactions Adaptation, F-34095, Montpellier cedex 5, France

⁹Department of Fisheries and Allied Aquacultures and Program of Cell and Molecular Biosciences, Auburn University, Auburn, AL 36849, USA

¹⁰Department of Biology, University of New Mexico, Albuquerque, NM 87131, USA

¹¹Evolution et Génétique des Populations Marines UMR 7144, Station Biologique, 29682, Roscoff, France

¹²Haskin Shellfish Research Laboratory, Institute of Marine and Coastal Sciences, Rutgers University, Port Norris, NJ 08901, USA

¹³Laboratoire des Sciences de l'Environnement Marin (LEMAR), UMR CNRS 6539, Institut Universitaire Européen de la Mer, Université de Bretagne Occidentale, 29280, Plouzané, France

¹⁴UMR Physiologie et Ecophysiologie des Mollusques Marins, IFREMER Centre de Brest, 29280, Plouzané, France

¹⁵Department of Pathobiology, University of Connecticut, Storrs, CT 06269, USA

¹⁶Program in Biomathematics and Biostatistics, University of Texas M.D. Anderson Cancer Center, Houston, TX 77030, USA

Received: 30 August 2006 / Accepted: 23 March 2007 / Published online: 1 August 2007

Abstract

The eastern oyster, *Crassostrea virginica*, and the Pacific oyster, *C. gigas*, are species of global economic significance as well as important components of estuarine ecosystems and models for genetic and environmental studies. To enhance the molecular tools available for oyster research, an international group of collaborators has constructed a 27,496-feature cDNA microarray containing 4460 sequences derived from *C. virginica*, 2320 from *C. gigas*, and 16 non-oyster DNAs serving as positive and negative controls. The performance of the array was assessed by gene expression profiling using gill and digestive gland RNA derived from both *C. gigas* and *C. virginica*, and digestive gland RNA from *C. ariakensis*. The utility of the microarray for detection of homologous genes by cross-hybridization between species was also assessed and the correlation between hybridization intensity and sequence homology for selected genes determined. The oyster

cDNA microarray is publicly available to the research community on a cost-recovery basis.

Keywords: bivalve — *Crassostrea gigas* — *Crassostrea virginica* — heterologous hybridization — microarray — oyster

Introduction

Oysters are members of the Lophotrochozoa, one of the three clades of Bilateria and a relatively understudied group that lacks well developed genetic models. The eastern oyster, *Crassostrea virginica*, and the Pacific oyster, *C. gigas*, are major components of a global aquaculture industry; *C. gigas* alone accounts for an annual production of greater than 4 million tons with a value in excess of \$3 billion. Regular disturbances in the production of *C. gigas* have been experienced for decades, however, with mortalities reported during summer periods in Japan (Koganezawa 1974), United States (Cheney et al. 2000), and France (Gouilletquer et al. 1998; Soletchnik et al. 1999). Oysters are important

Correspondence to: Matthew Jenny; E-mail: mjenny@whoi.edu

components of the estuarine marine ecosystem (Newell 1988), but overharvesting, loss of habitat, declines in environmental quality, and increased incidence of disease have led to a dramatic reduction in the health of entire ecosystems, making oysters an important indicator species of estuarine health. In addition, the oyster has proved a useful model for studying the population biology and genetics of a highly fecund organism (Hedgecock et al. 2005).

In acknowledgment of the oyster's commercial and ecological importance, as well as its underutilization in mainstream genetics research, the Oyster Genome Consortium was formed to advocate for and cooperate in the development of genetic and genomic tools for oysters. This international community has been instrumental in the construction of bacterial artificial chromosome (BAC) libraries for *C. gigas* and *C. virginica* (Cunningham et al. 2006) and in the adoption, by the Community Sequencing Program at the Joint Genome Institute, of *C. gigas* for a preliminary BAC and expressed sequence tag (EST) project. Further, in recent years the availability of oyster DNA sequence information has been growing, with several groups reporting EST collections for *C. gigas* and *C. virginica* (Jenny et al. 2002; Gueguen et al. 2003; Boutet et al. 2004; Huvet et al. 2004; Peatman et al. 2004; Tanguy et al. 2004, 2005; David et al. 2005). These collections have proven to be invaluable resources for oyster gene discovery, as they contain genes whose expression is regulated in response to environmental, chemical, and biological stressors that are thought to have contributed to the decline in health of oyster populations worldwide. To maximize the utility of these collections and to address the need for microarrays to facilitate functional genomic studies in oysters, the community of researchers in this area pooled their EST resources for construction of a publicly available cDNA microarray. Such a microarray would have practical uses including ecological monitoring, evaluation of oyster live-stock for selective breeding programs, and understanding of higher-order gene regulation in oysters.

In this report, we introduce the first publicly available oyster microarray. We describe the content and design of the array, and we report the degree of cross-species hybridization between the *C. gigas* and *C. virginica* features and include hybridization information about an oyster species not included in the array (*C. ariakensis*).

Materials and Methods

Libraries and EST Collections. Clones from 8 unmodified cDNA libraries and from 24 suppression

subtractive hybridization (SSH) generated libraries of *C. virginica* and *C. gigas* were used to construct the microarray (Table 1). The details of the construction of many of these libraries have been reported elsewhere (Jenny et al. 2002; Gueguen et al. 2003; Boutet et al. 2004; Huvet et al. 2004; Peatman et al. 2004; Tanguy et al. 2004, 2005; David et al. 2005). The construction of cDNA libraries not previously reported is described below. All of the previously unreported *C. gigas* and *C. virginica* SSH clone sequences are publicly accessible in the NCBI dbEST database or at <http://www.marinegenomics.org> (McKillen et al. 2005). Contig analysis and optimal microarray probe selection was performed on the *C. virginica* ESTs to minimize redundancy of the clones printed on the microarray (Chen et al. 2004).

Environmental Challenges and SSH Library Construction

Adult *C. virginica* were exposed to metal challenge, hypoxia followed by hyperthermia, and immune stimulation. For metal challenge, oysters were exposed to 0.44 μM cadmium (CdCl_2 salt) for 24 h. For the hypoxia/hyperthermia challenges, oysters were exposed to air for 30 h at room temperature (23°C) and then placed in a beaker of filtered natural seawater (25 ppt) at 32°C for 90 min. Immune challenge was by an injection of a suspension of heat-killed bacteria and fungi prepared from *Vibrio parahaemolyticus*, *Micrococcus luteus*, *Bacillus subtilis*, and *Fusarium oxysporum*. The final stock concentration of the bacterin/fungal suspension was created from a pool of 1×10^9 cfu/ml of each bacterium and 2.5×10^6 spores/ml of the *F. oxysporum*, which was heated in boiling water for 5 min before aliquoting and storage at -20°C. The oyster shells were notched opposite the adductor muscle and 100 μl of the suspension were injected using tuberculin needles into the adductor muscle. Samples were collected 4 and 24 h after challenge. Hemocytes and/or digestive gland were taken after each challenge and homogenized in RLT buffer containing 0.05% vol/vol β -mercaptoethanol and stored at -20°C until total RNA isolation by RNeasy[®] Miniprep kits (Qiagen, Valencia, CA). *C. virginica* SSH libraries were created using the Clontech PCR-Select[™] cDNA Subtraction Kit (Clontech Laboratories, Mountain View, CA) according to the manufacturer's protocols. For library construction, 2 μg of total RNA from control and challenged oysters, respectively, were used for first- and second-strand cDNA synthesis. Hybridization and subtraction steps were carried out in both the forward and reverse directions. PCR products were cloned into the pCR[®]2.1-TOPO vector as per the manufacturer's instructions

Table 1. Summary of libraries from which clones were derived

Tissue/cell type	Library type	Challenge/treatment	Species
Hemocyte (F & R)	SSH	4-h bacterin challenge	<i>C. virginica</i> ^a
Hemocyte (F & R)	SSH	24-h 50 ppb Cd exposure	<i>C. virginica</i> ^a
Hemocyte (F & R)	SSH	24-h hypoxia/hyperthermia	<i>C. virginica</i> ^a
Digestive gland (F)	SSH	4-h bacterin challenge	<i>C. virginica</i> ^a
Digestive gland (F)	SSH	24-h bacterin challenge	<i>C. virginica</i> ^a
Digestive gland (F)	SSH	24-h 50 ppb Cd exposure	<i>C. virginica</i> ^a
Digestive gland (F)	SSH	24-h hypoxia/hyperthermia	<i>C. virginica</i> ^a
Hemocyte (F & R)	SSH	<i>Perkinsus marinus</i> challenge	<i>C. virginica</i> ^b
Hemocyte (F & R)	SSH	<i>Perkinsus marinus</i> challenge	<i>C. gigas</i> ^b
Gill (F & R)	SSH	<i>Perkinsus marinus</i> challenge	<i>C. virginica</i> ^b
Gill (F & R)	SSH	<i>Perkinsus marinus</i> challenge	<i>C. gigas</i> ^b
Hemocyte (F & R)	SSH	Lipopolysaccharide injection	<i>C. gigas</i> ^c
Mantle/gonad (F)	SSH	"Summer mortality" resistance	<i>C. gigas</i> ^d
Digestive gland (F & R)	SSH	Hydrocarbon mixture	<i>C. gigas</i> ^e
Digestive gland (F & R)	SSH	Pesticide exposures	<i>C. gigas</i> ^f
Gill (F & R)	SSH	Pesticide exposures	<i>C. gigas</i> ^f
Digestive gland (F & R)	SSH	Hypoxia exposures	<i>C. gigas</i> ^g
Gill (F & R)	SSH	Hypoxia exposures	<i>C. gigas</i> ^g
Mantle (F & R)	SSH	Hypoxia exposures	<i>C. gigas</i> ^g
Hemocyte	Unmodified	Control, 25 ppt natural seawater	<i>C. virginica</i> ^h
Hemocyte	Unmodified	Control and bacterin injected	<i>C. virginica</i> ^a
Digestive gland	Unmodified	Multiple stressors	<i>C. virginica</i> ^a
D-veliger larvae	Unmodified	Control, 25 ppt natural seawater	<i>C. virginica</i> ^f
Gonadal tissue	Unmodified	Control	<i>C. virginica</i> ⁱ
Gill	Unmodified	Control	<i>C. virginica</i> ⁱ
Hemocyte	Unmodified	<i>Vibrio</i> spp. injection	<i>C. gigas</i> ^l
Mantle	Unmodified	Control	<i>C. gigas</i> ^k

F, Forward subtracted library; R, reverse subtracted library.

^aJenny M, <http://www.marinegenomic.org>.

^bTanguy et al. (2004) Gene 338, 121–131.

^cCunningham <http://www.marinegenomics.org>.

^dHuvet et al. (2004) Gene 343, 211–220.

^eBoutet et al. (2004) Gene 329, 147–157.

^fTanguy et al. (2005) FEBS J 272, 390–403.

^gDavid et al. (2005) FEBS J 272, 5635–5652.

^hJenny et al. (2002) Mar Biotechnol 4, 81–93.

ⁱPeatman et al. (2004) Proc Mar Biotechnol 6, S491–S496.

^jGueguen et al. (2003) Gene 303, 139–145.

^kCunningham C, unpublished data.

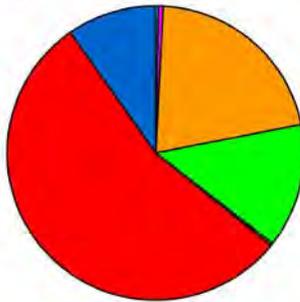
(TOPO TA Cloning Kit, Invitrogen, Carlsbad, CA). The pCR[®]2.1-TOPO constructs were transformed into chemically competent XL1 Blue MRF' *E. coli* cells. Plasmids were isolated using Qiagen[®] Turboprep 96 kits and the Qiagen[®] Biorobot 9600, according to the manufacturer's instructions (Qiagen). Plasmid inserts were sequenced using standard M13F and M13R primers.

cDNA Library Construction. The SMART[™] cDNA Library Construction Kit (Clontech Laboratories) was used for construction of the libraries. cDNA was ligated into a λTriplEx2 vector and packaged using the Gigapack[®] Gold λ (Stratagene, LaJolla, CA) packaging system. The cDNA libraries were mass transformed by infecting the BM25.8 strain of *E. coli* with the packaged phage at a 1:10 ratio of phage to bacteria at 31°C for 90 min. Transformed bacteria were selected on LB plates with carbenicillin

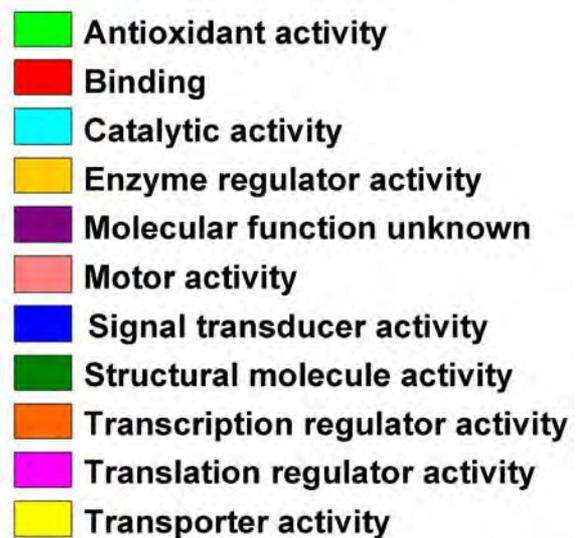
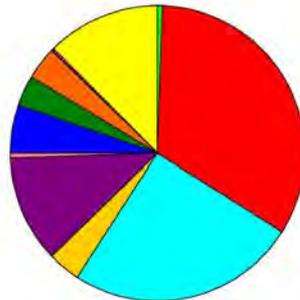
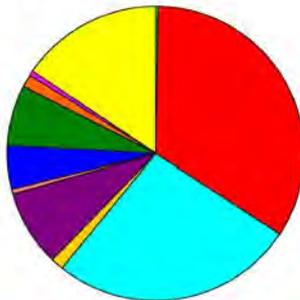
(50 µg/ml) and incubated overnight at 31°C. A Qbot (Genetix, New Milton, UK) was used to pick 4608 clones from each library. Cultures were incubated overnight in 96-well plates with LB media and carbenicillin (50 µg/ml). Expressed 5' sequence tags were generated by sequencing of plasmids using the standard sequencing primer for pTriplEx2 (Clontech Laboratories). The immune-challenged cDNA library was constructed by mixing equal amounts of total RNA from control and immune-stimulated hemocyte samples. The digestive gland cDNA library was constructed by combining total RNA from control, 4 and 24 h immune-stimulated, cadmium-treated, and hypoxia/hyperthermia-treated oysters.

cDNA Microarray Construction. Amplicons were generated via polymerase chain reaction (PCR) amplification from bacterial stocks or plasmid

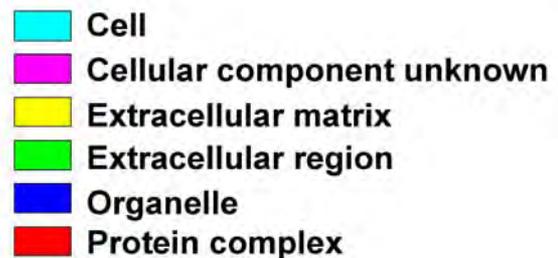
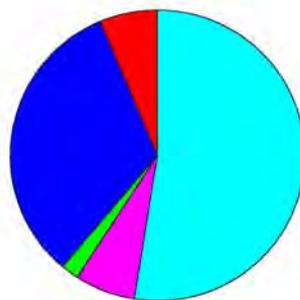
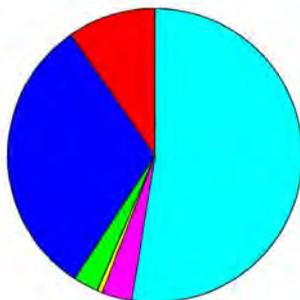
Biological Process (3,394; 11,211)



Molecular Function (322; 1,048)



Cellular Component (974; 2,504)



C. gigas

C. virginica

Figure 1. Gene ontology annotation of the *Crassostrea* microarray ESTs. Functional annotation of ESTs displaying homology to the GO database. Each pie chart represents the distribution of the GO annotations into the major functional categories for each division. The values in parentheses represent the number of functional annotations in each division for *C. gigas* (left) and *C. virginica* (right), respectively.

DNA preparations using SP6, T7, T3, M13F, M13R primers or pTriplEx2. PCR reactions were performed in 96-well plates and purified using the Qiagen® Turboprep 96 kits and the Qiagen® Biorobot 9600 (Qiagen). The success of the PCR amplification and purification was assessed by agarose gel electrophoresis. Purified amplicons were quantified using a SPECTRAMax PLUS spectrophotometer (Molecular Devices, Sunnyvale, CA), vacuum dried, and resuspended in 30% dimethyl sulfoxide to a final concentration of 250 µg/ml. The arrays were constructed using a Genetix QArrayMax microarray printer (Genetix) fitted with a 24-pin head, from 12×384-well source plates (Genetix) on GAPS II amino-silane coated slides (Corning, Corning, NY). A 575-feature per subarray pattern layout was used with amplicons printed as side-by-side duplicates. The cDNA microarray slide consists of two complete, side-by-side arrays, with each array containing a total of 24 subarrays (48 subarrays per slide). Slides were fixed by baking for 2 h at 80°C. Spot morphology was checked by hybridization with a SpotQC kit using a Cy3™ dye-labeled oligonucleotide probe (Integrated DNA Technologies, Coralville, IA).

Experimental Design. To determine tissue-specific gene expression profiles and occurrence of cross-hybridization between species, gill and digestive gland tissue were collected from each of three individual *C. virginica* and *C. gigas*; and digestive gland was also collected from three *C. ariakensis*. Tissues from the three individuals of each species were pooled together for a total of five separate tissue samples, three from digestive gland and two from gill. The tissues were homogenized in RLT buffer containing 0.05% (vol/vol) β-mercaptoethanol and stored at -20°C. To assess technical replicability of the microarray, and to compare gene expression profiles of tissues and species, three replicate labelings and hybridizations were performed for each RNA sample, for a total of 15 hybridizations.

Microarray Hybridization and Data Collection. Total RNA was isolated from oyster tissues using the RNeasy® Miniprep kits (Qiagen) with an on-column DNase treatment. The mRNA used as positive controls (*K. brevis* clones) was generated from plasmid templates by in vitro transcription using the MEGAscript® T7 kit (Ambion, Austin, TX) according to the manufacturer's protocol. In brief, double-stranded DNA template for in vitro transcription was generated by PCR from plasmid stocks [cloned transcripts include a poly(A) tail] using standard T7 and SP6 primers. The PCR products were gel purified using the NucleoTrap® Gel Extract Kit

(Clontech Laboratories) and 1 µg of DNA was used as template for the reaction. In vitro transcribed mRNA was purified using the RNeasy® Miniprep kits (Qiagen). Antisense RNA (aRNA) was synthesized from 1 µg of oyster total RNA or 1 µg of in vitro transcribed RNA for each of the positive controls using the Amino Allyl MessageAmp™ Kit (Ambion) according to the manufacturer's protocol. A mixture of "spiked in" positive control RNA's was created so that 10 µl of the mixture contained four separate control RNAs in a 10-fold serial dilution (100 ng, 10 ng, 1 ng, 0.1 ng; 111.1 ng total) to be added to the oyster aRNA. Amino allyl aRNA (mixture of 10 µg of oyster aRNA and 111.1 ng of control aRNA) was coupled with an amine-reactive Cy3 dye, column purified, and diluted (1:3) in hybridization buffer [50% formamide, 2.4% sodium dodecyl sulfate (SDS), 4× SSPE, 2.5× Denhardt's solution, plus 1 µl of blocking solution (1 µg Cot-1 DNA and 1 µg polydATP)]. The probe was then boiled for 1 min and incubated in the dark for 1 h at 50°C. Microarray slides were prewashed with 0.2% SDS for 2 min, boiled in deionized water for 2 min, rinsed in 70% ethanol for 2 min, and then dried. Slides were prehybridized in a

Table 2. Non-oyster genes used for spiked and negative controls

	NCBI accession no.
<i>Karenia brevis</i> (dinoflagellate, Dinophyceae)	
Chlorophyll A/B binding protein	CO059871
Flavodoxin	CO065421
Photosynthesisism core protein	CO062297
Photolyase	CO064781
<i>Anas platyrhynchos</i> : (duck, Chordata)	
IgY heavy chain	X65218
<i>Litopenaeus vannamei</i> : (shrimp, Arthropoda)	
Hemocyanin	CK572324
<i>Ictalurus punctatus</i> : (catfish, Chordata)	
BOB. 1	M. Lennard et al. (‘pers com’)
Oct-1	AJ000267
Oct-2α	Y12651
Oct-2β	Y12652
<i>Tursiops truncatus</i> : (dolphin, Chordata)	
Hemoglobin β chain	DT660258
IL-1α	DT660180
IL-8	DT660217
<i>White Spot Syndrome Virus</i> : (virus, Nimaviridae)	
Viral-coat protein 19	Robalino et al. 2005
Viral-coat protein 26	CK572418
Putative DNA polymerase	CK572871

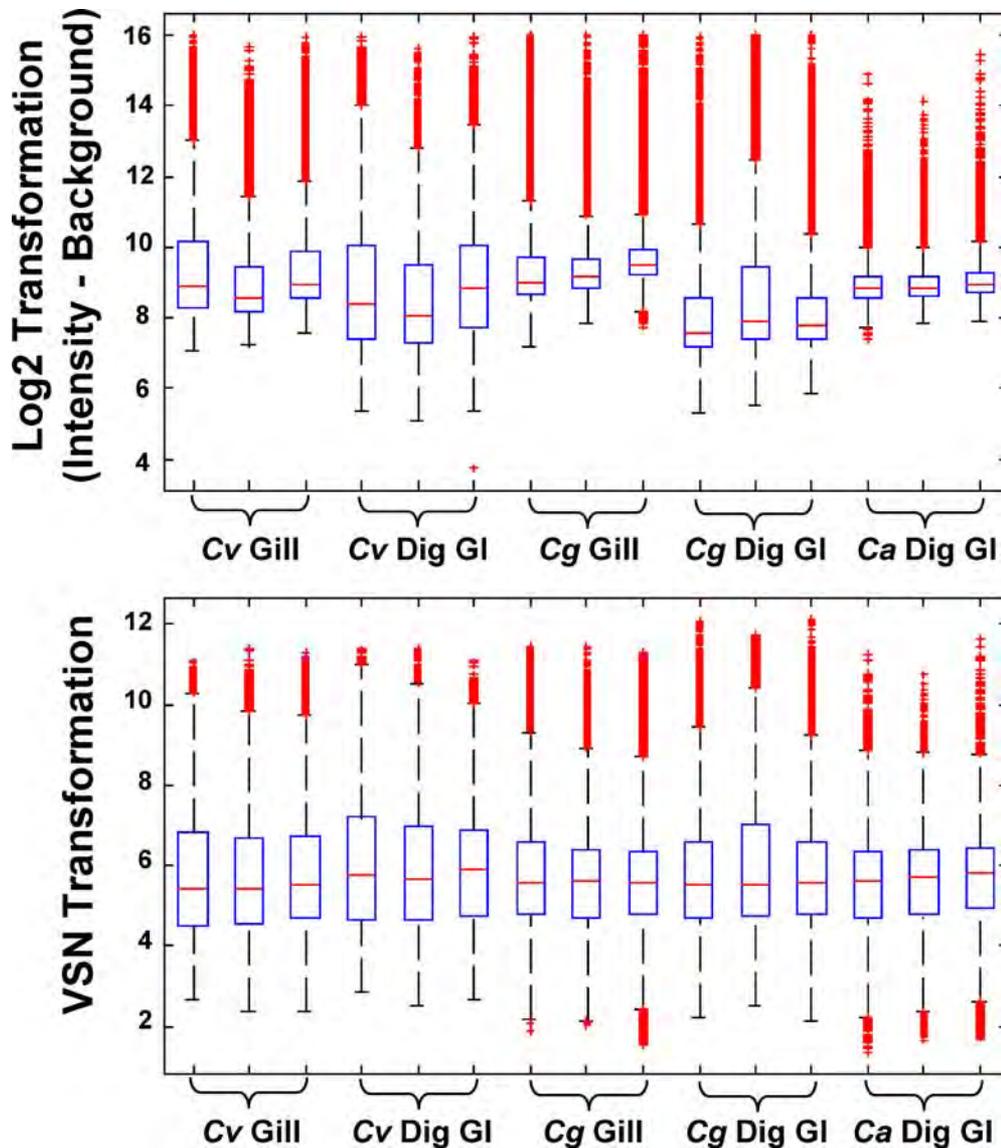


Figure 2. Boxplots of hybridization intensity before and after normalization. The top chart represents the array data that have been log transformed. The bottom chart represents the array data after normalization using the variance stabilized transformation (VSN). Each boxplot represents a single array (Dig GI = digestive gland). The red crosses represent extreme values exceeding the interquartile range by a factor of 1.5.

hybridization oven with a prehybridization buffer (33.3% formamide, 1.6% SDS, 2.6 SSPE, 1.6× Denhardt's solution, and 0.1 μ M salmon sperm DNA) in the dark for 1 h at 50°C. Slides were hybridized with Cy-3-aarRNA in the dark for 16 h at 50°C. After the hybridization, slides were rinsed in 2× saline sodium citrate (SSC), 0.1% SDS and soaked in 0.2× SSC, 0.1% SDS for 15 min in the dark at room temperature, followed by a rinse in 0.2× SSC, soaking in 0.2× SSC for 15 min, 0.1× SSC for 15 min, and finally deionized water for 5 min in the dark to remove residual SDS. The microarrays were then dried and immediately scanned.

Microarray Data Analysis. The microarray slides were scanned using a ScanArray Express microarray scanner at 75% photomultiplier tube (PMT) gain and 90% laser power in conjunction with the SpotArray software package (Perkin Elmer, Boston, MA). Hybridization data were obtained by determining spot intensity, background, and QA/QC features with the QuantArray Histogram method (Perkin Elmer). The microarray hybridization data and the MIAME protocols have been deposited at the GEO site (Series no. GSE5293; Accession nos. GSM119921 to GSM119935; www.ncbi.nlm.nih.gov/geo/) and at www.marinegenomics.org.

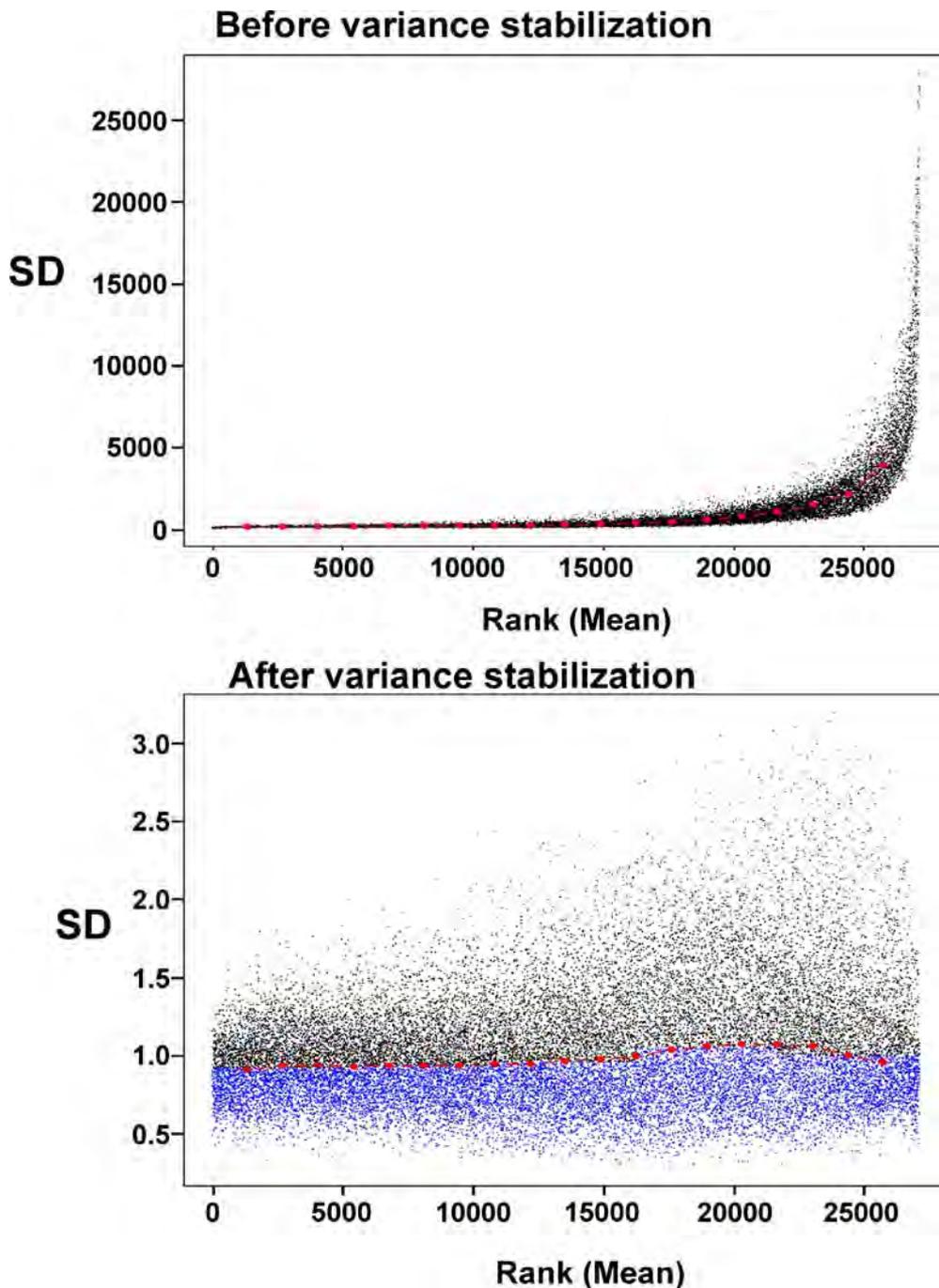


Figure 3. Variance stabilization after normalization. Variance-versus-mean dependence is apparent in the top chart, where the standard deviation (SD) of raw intensity values (after background subtraction) increases dramatically as the ranked mean intensity increases. The bottom chart illustrates the stabilization of the variance after VSN transformation. More details can be found in Huber et al. (2002).

A robust generalized log transformation, variance stabilized normalization (VSN) method (Huber et al. 2002), was used to normalize the data before our analyses to eliminate observed variance-versus-mean dependence in the intensity signals. A two-way ANOVA (Kleinbaum et al. 1998) was then used to assess differential gene expression between tis-

suues and among species for all of the oyster clones on the array. We estimated q values using the software Q -value (Storey 2002; Storey and Tibshirani 2003) to decide significant features on the arrays for multiple-hypothesis testing. This approach adjusting for multiple-hypothesis testing based on controlling the false discovery rate (Benjamini and

Table 3. Technical reproducibility of the *Crassostrea* cDNA microarray

Spearman correlation	Between replicate 1 and 2	Between replicate 1 and 3	Between replicate 2 and 3	Median correlation
Cv Gill	0.90	0.88	0.87	0.88
Cv Digestive gland	0.94	0.88	0.88	0.88
Cg Gill	0.74	0.69	0.75	0.74
Cg Digestive gland	0.83	0.86	0.85	0.85
Ca Digestive gland	0.73	0.69	0.61	0.69

Holchberg 1995) has become popular for genome-wide and high-throughput studies. The proportion of truly null features, nondifferential expressed genes among species and between tissues, was estimated, respectively. Differential expression between species for large numbers of genes was apparent during preliminary examination, and therefore, a more stringent false discovery rate method was chosen for evaluating the effect of species as compared to that of tissue. The Euclidean distance between normalized expression values was calculated and average linkage method was used for hierarchical cluster analysis (Statistical Toolbox of MATLAB™). Clustergrams (combination of expression heat map and dendrogram) were used to visualize the relation-

ships of differential gene expression between tissues and among species (Martinez and Martinez 2002). Pairwise Spearman rank correlation coefficients were calculated using mean expression values of the four replicate spots for all genes on the microarray to estimate the technical reproducibility of the oyster microarrays.

Results

The Oyster Microarray EST Collection. A combination of unmodified cDNA libraries, representing unbiased sampling of the transcriptome of diverse tissues and developmental stages, and SSH

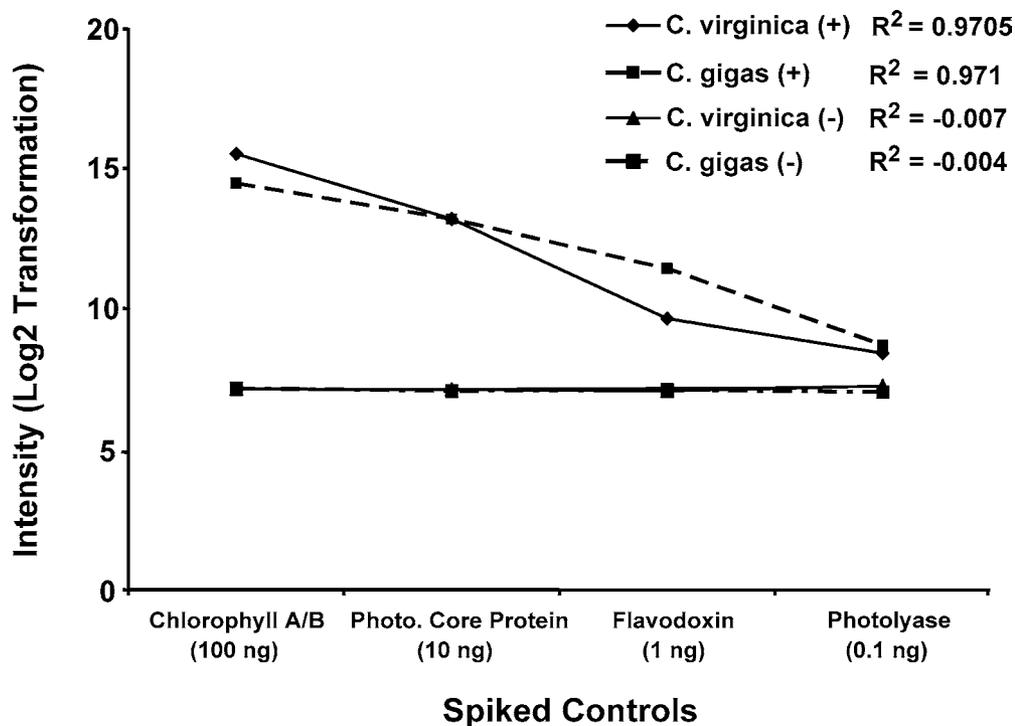


Figure 4. Validation of “spiked-in” RNAs from *K. brevis*. Chloroplast genes used as positive controls. In vitro transcribed RNA from four *K. brevis* genes involved in photosynthesis were added (+) to oyster RNA samples in a 10-fold dilution series before labeling and hybridization. The strong R^2 values are indicative of consistent labeling and hybridization. In contrast, those oyster RNA samples without the “spiked” RNA (-) do not display hybridization intensity values above normal background.

cDNA libraries for transcripts putatively induced or suppressed by a variety of challenges were used to create a collective pool of transcripts with which the oyster microarray was constructed (Table 1). A total of 5586 and 3961 sequences from *C. virginica* and *C. gigas*, respectively, were collected from these cDNA libraries and subjected to Gene Ontology (GO) analysis allowing putative functional annotation (molecular function, biological process, and cellular component) of the sequences (Ashburner et al. 2000) to determine the diversity within the EST collections used to construct the microarray. Approximately 67% of *C. virginica* and 60% of *C. gigas* sequences had no significant homology with annotations in the GO database, but of the sequences that could be putatively annotated, 11,211 and 3394 biological process, 1048 and 322 molecular function, and 2504 and 974 cellular component annotations were obtained for *C. virginica* and *C. gigas*, respectively (Figure 1). Each of the three major categories has also been divided into its major hierarchies (Figure 1) representing broadly specialized functions or components. Although *C. virginica* has a much greater number of annotations, the distribution of functional annotations is diverse and broadly equivalent for both oyster species. In an effort to determine the extent of mitochondrial sequences

within the collection of ESTs the 5586 and 3961 sequences from *C. virginica* and *C. gigas*, respectively, were compared to the complete mitochondrial genome sequences for both species (NC_005881, AY905542, NC_001276, AF177226). Of the *C. virginica* ESTs, 8.8% of the sequences were of mitochondrial origin, and of the *C. gigas* ESTs 10.4% of the clones were of mitochondrial origin. The EST collections from both species of *Crassostrea* were subjected to redundancy analysis and subsets of clones were chosen to be included on the microarray.

Design of the *Crassostrea* cDNA Microarray. The *Crassostrea* cDNA microarray slide consists of two complete, side-by-side arrays, with each array containing duplicate spots for 4460 *C. virginica* amplicons and 2320 *C. gigas* amplicons and 96 amplicons representing 16 different non-oyster controls (see later) for a total of 13,752 features per array (27,504 features per slide). Of the 6780 *Crassostrea* amplicons on the array, only 355 represent clones that have not been sequenced, and based on the redundancy analysis less than 8% of the total numbers of clones are estimated to represent sequences of mitochondrial origin. Each array contains a collection of non-oyster cDNAs to serve as negative controls for nonspecific background

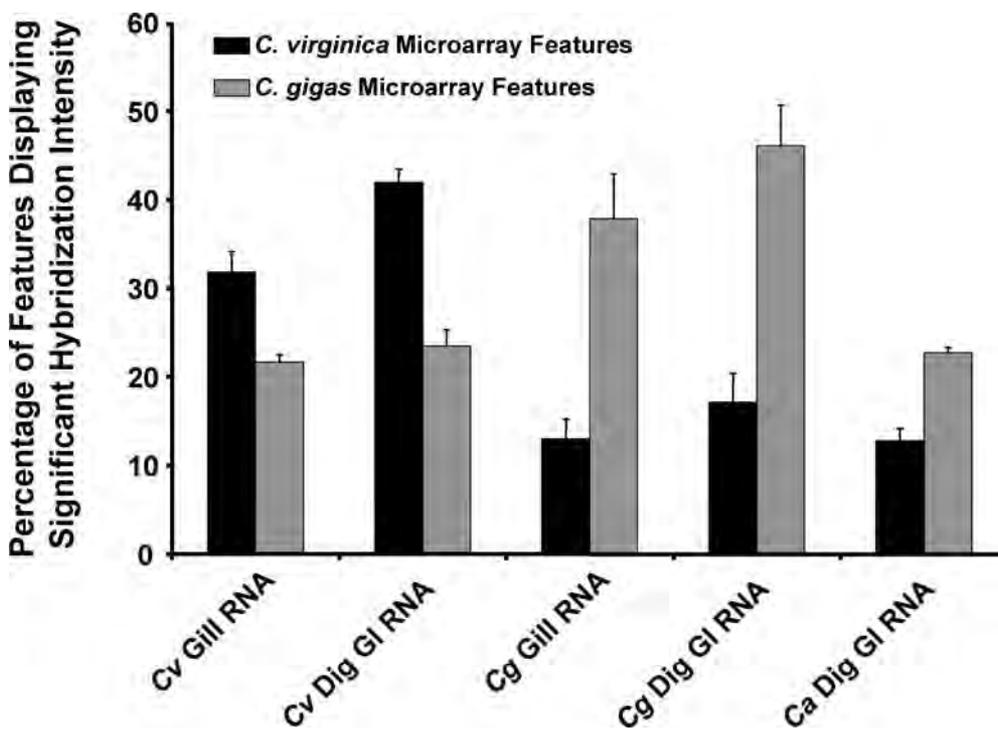


Figure 5. Species-specific hybridization to the oyster microarray. The bar graph highlights the percentage of features, separated by species, displaying significant intensity (at least 2-fold above background) after hybridization with samples from two different tissues collected from three different oyster species.

detection and as positive controls for “spiked-in” control RNA samples (Table 2). Included among the non-oyster transcripts spotted onto the array are four dinoflagellate (*Karenia brevis*) chloroplast transcripts involved in photosynthesis (Lidie et al. 2005), as well as copies of the duck (*Anas platyrhynchos*) IgY heavy chain and shrimp (*Litopenaeus vannamei*) hemocyanin transcripts, that are provided to serve as targets for “spiked-in” control RNA samples. Eight to twelve copies of all six of these transcripts were printed per slide. The remaining non-oyster transcripts (Table 2)

were printed at a higher frequency (16 to 88 copies per slide) and are intended to serve as negative controls for the detection of nonspecific hybridization.

Validation of the Microarray. To normalize the gene expression data sets, a robust generalized log transformation and variance stabilization and normalization (VSN) method was applied (Huber et al. 2002). The success of the VSN normalization in producing a smoother and more consistent intensity distribution among the various arrays compared to a

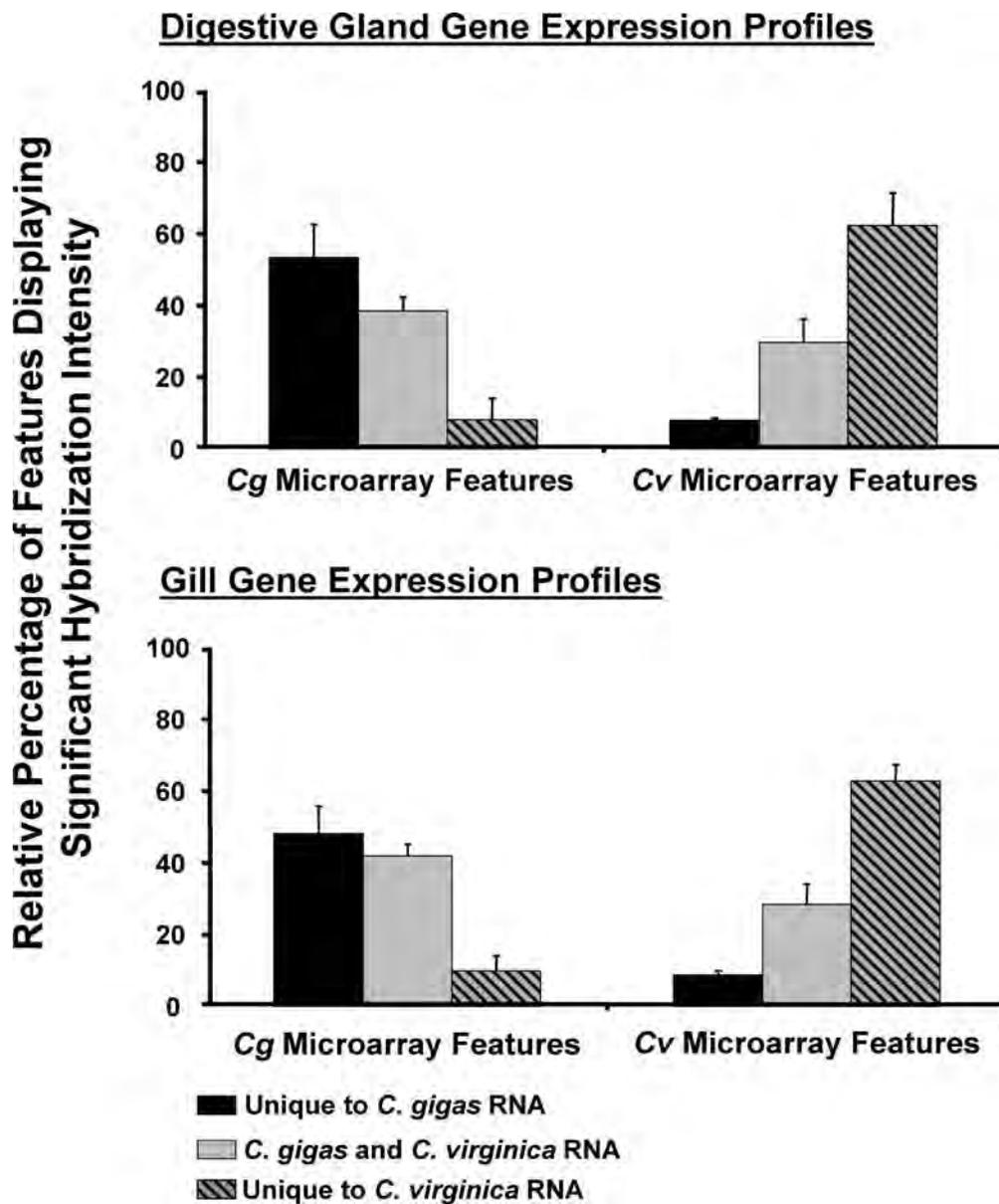


Figure 6. Homologous and heterologous hybridization with features on the oyster microarray. Pairwise comparisons of tissue-specific gene expression profiles between both *C. virginica* and *C. gigas* were performed to determine the specific percentages of clones displaying homologous and heterologous hybridization (determined as at least 2-fold above background).

standard log₂ transformation is illustrated in Figure 2. Moreover, the variance-versus-mean dependence was minimized after normalization before our analyses (Figure 3). The median Spearman correlation for all the replicated arrays was 0.85 and the correlation for each pairwise comparison of the technical replicates is shown in Table 3. Technical reproducibility of the oyster microarrays can also be observed in Figures 8 and 9.

The utility of the non-oyster genes added to the microarray, as both positive controls (with “spiked-in” labeled RNA) and as negative controls (to detect both background and any inappropriate hybridization of oyster sequences) was addressed. The addition of in vitro transcribed *K. brevis* RNAs to either *C. virginica* or *C. gigas* total RNA samples in a 10-fold serial dilution demonstrated that these transcripts would serve effectively as a positive control (Figure 4). The “spiked in” *K. brevis* RNA hybridized specifically to the features that contained its cognate sequences, and the *K. brevis* spots on the

array did not show hybridization with oyster transcripts when the “spiked-in” RNA was not added to the labeling reaction. Other non-oyster sequences printed on the microarray also serve as negative controls for the detection of nonspecific hybridization. None of the negative controls showed hybridization exceeding background intensity levels with RNA samples from either gill or digestive gland of the three oyster species tested (Figure 4).

Cross-Hybridization Detection Across Species. The utility of the oyster microarray as a multispecies tool can be enhanced by the degree of cross-hybridization it is capable of detecting. This potential was investigated by analysis of detectable hybridization when using the microarray to interrogate the transcriptomes not only of the two species represented on the microarray (*C. gigas* and *C. virginica*) but also that of a third species (*C. ariakensis*). For the purposes of classifying a feature on the microarray as positive for hybridization in

Genes Used for Comparison

- ◆ Cv Tubulin (96.75%)
(CD648525; CB617442)
- Cv Ribosomal Protein L5 (93.53%)
(CV132476; BQ426452)
- ◆ Cv B-actin (92.35%)
(CD646509; CB617448)
- ◆ Cv ARE Binding Prot (89.31%)
(CD649252; BG467427)
- Cv Fructose-bisPhosphate (87.97%)
(CD646914; BQ426705)
- Cv 40S Ribosomal (84%)
(CD648070; BQ426793)

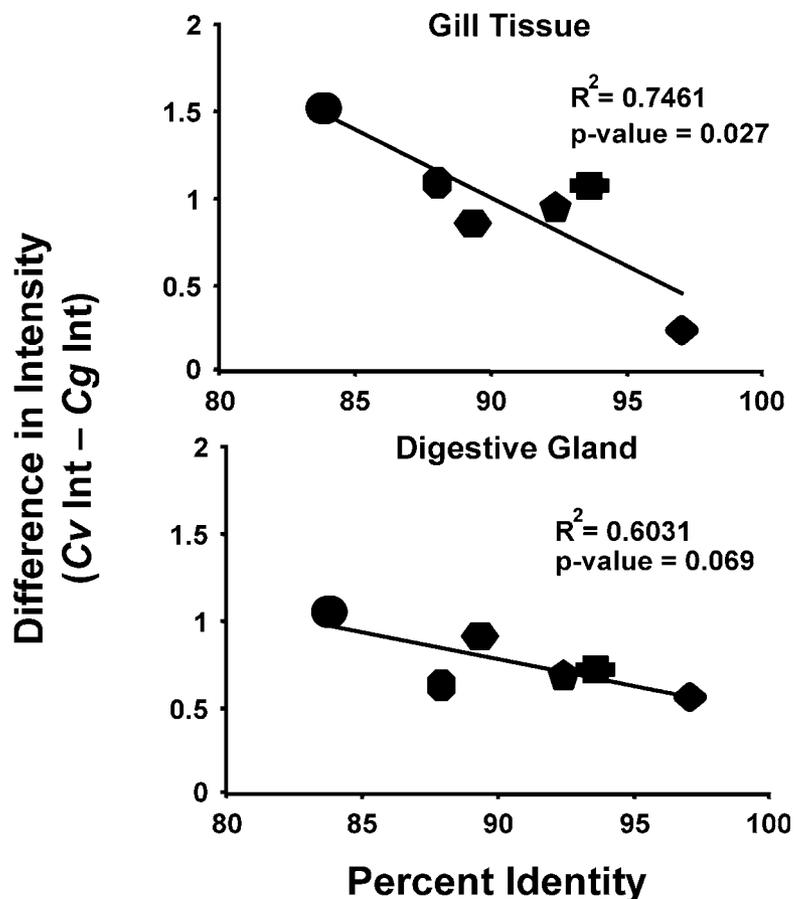


Figure 7. Relationship between cross-species hybridization and DNA sequence similarity. Six genes (percent sequence identity calculated from a minimum of 200 bp of overlapping contiguous sequence; values in parentheses) were compared for hybridization intensity with RNA from gill and digestive gland of *C. gigas* and *C. virginica*. The difference of VSN normalized intensity value from *C. gigas* RNA subtracted from the VSN values from *C. virginica* RNA was plotted against the % sequence identity. *C. virginica* and *C. gigas* accession numbers in parentheses, respectively.

this comparison, a conservative threshold of 2-fold above the average background was used (Gracey et al. 2001; de Lorgeril et al. 2005). The percentage of microarray features of *C. gigas* or *C. virginica* origin that were positive when hybridized with RNA from each of the three different oyster species is shown in Figure 5. As expected, RNA from *C. gigas* and *C. virginica* showed the highest levels of hybridization (between 32% and 45% of features) in a species-specific manner (Figure 5). However, significant cross-species hybridization was observed with all samples, including *C. ariakensis* digestive gland RNA, which hybridized with greater than 16% of the total number of oyster features on the microarray. Pairwise comparisons of tissue-specific *C. virginica* and *C. gigas* gene expression profiles demonstrate that approximately 30% to 40% of the clones (for both species) displaying significant hybridization were reciprocal for both species and 8% to 10% of the features displayed significant

hybridization solely with RNA samples from the opposite species (Figure 6).

To assess the effect of sequence difference on the observed cross-hybridization, a plot was made of changes in hybridization intensity against percent sequence identity for six genes (Figure 7). Despite the unknown impact upon the results of different transcript abundance between the two species of oyster, the observed strength of cross-hybridization intensity between the two species is clearly positively correlated with the degree of sequence identity (Figure 7).

Differential Expression of Genes Between Tissues and Species. In comparing the hybridization to the microarray observed with gill and digestive gland RNA from two and three species of oyster, respectively, it is reasonable to anticipate that significantly different levels of observed hybridization to given genes on the microarray will reflect both the

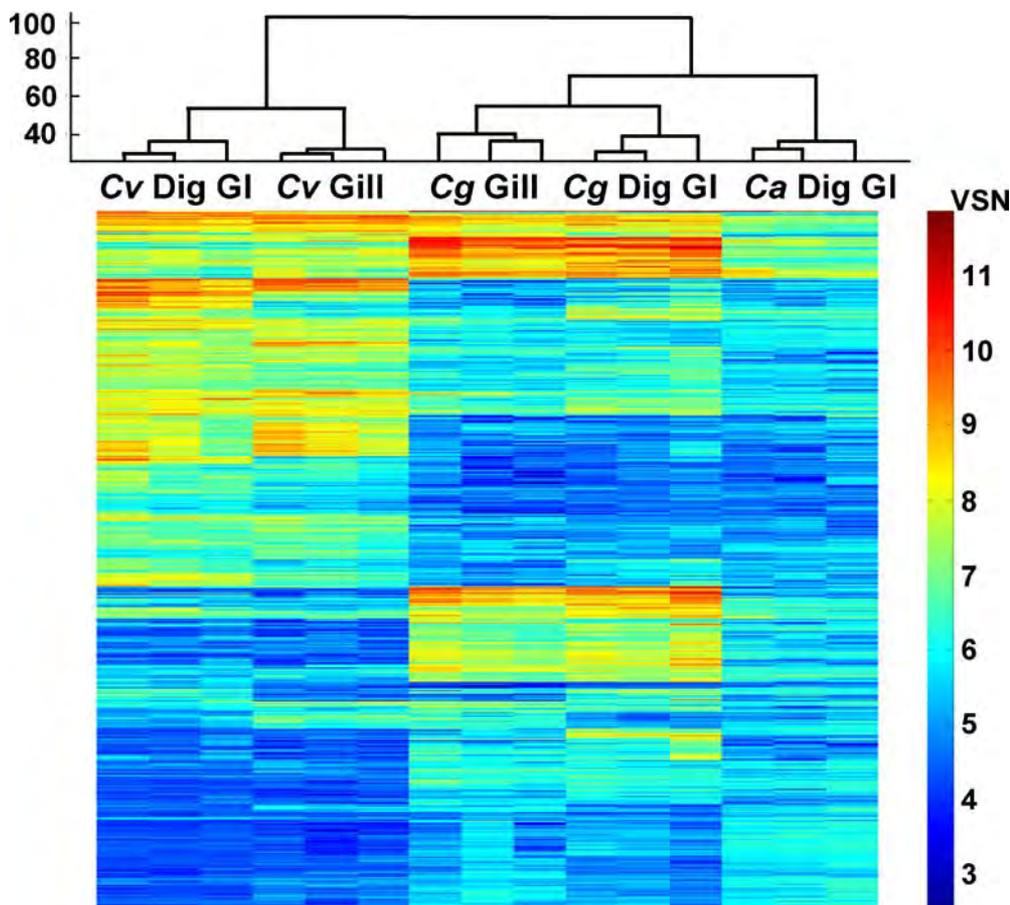


Figure 8. Clustergram of the 2717 genes differentially expressed between species and tissues. Each column represents a single array and the clustering of the arrays demonstrates the high reproducibility of the technical replicates. Gene expression intensity is represented by a color-coded scale based on the VSN transformed values. Tree branch length is a measure of Euclidean distance.

differential abundance of transcripts and the differential efficiencies of hybridization, as a result of nucleotide sequence divergence. In spite of these two potential effects, a global comparison identified 2541 genes with significantly different expression intensities between species (when controlling the FDR at 0.001) and 438 genes significantly different between tissues (when controlling the FDR at 0.01). A clustergram (Figure 8) of expression pattern of these genes showed a basal distinction at the species level, with a *C. virginica* branch and one that includes both *C. gigas* and *C. ariakensis*. The microarrays hybridized (in groups of 3) against single tissues from each species showed not only the closest relationships, but also very similar patterns of expression, confirming the technical reproducibility of the microarray (Figure 8). In a cluster analysis focused on differential expression between gill and digestive gland tissue of *C. virginica*, both the technical reproducibility and the ability of the microarray to distinguish tissue-specific gene expression (in this case of 438 differentially expressed genes between tissues when controlling the FDR at 0.01) could also be visualized (Figure 9).

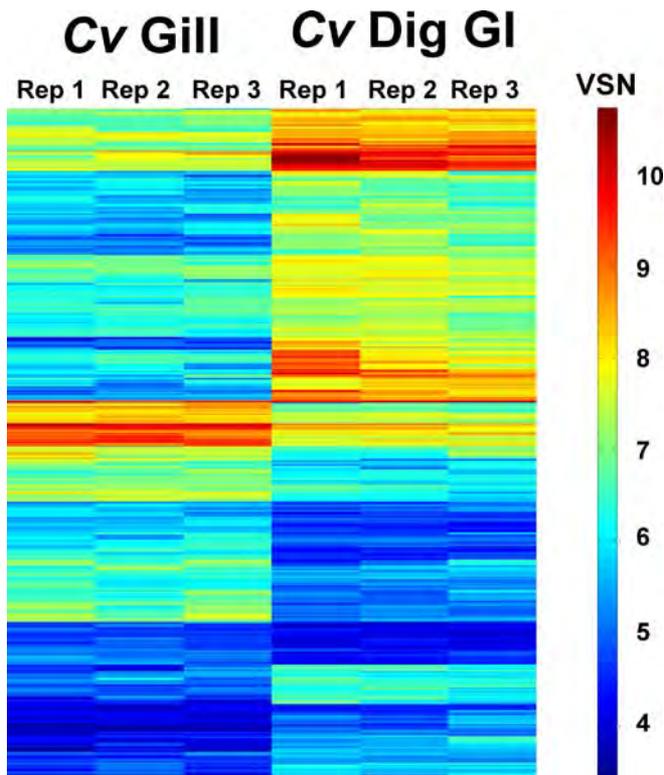


Figure 9. Tissue-specific gene expression profiles from *C. virginica*. A clustergram of the 438 genes differentially expressed between gill and digestive gland tissue. Each column represents a single array and gene expression intensity is represented by a color-coded scale based on the VSN transformed values.

Discussion

The oyster microarray described in this report contains sequences expressed in *C. gigas* and *C. virginica*. These sequences were derived from different tissues and developmental stages as well as oysters exposed to different physical, chemical, and biological challenges. The data presented document the reproducibility and utility of this microarray in probing gene expression profiles from different tissues and from different species of oyster (including *C. ariakensis*, which was not represented on the array).

The gill and digestive gland were tissues chosen for the initial characterization of the microarray because of their important physiological roles and relatively large size. While both tissues gave robust expression profiles, in the case of both *C. virginica* and *C. gigas*, the digestive gland gave the strongest profile in terms of the number of features displaying significant hybridization intensity. This is likely due to the cellular heterogeneity and functional diversity of this tissue, as well as to the strong representation of clones derived from digestive gland libraries on the array. However, the diversity of tissues used for EST collections, which included hemocytes, mantle, and gonad, suggests that the microarray should be a useful tool for examining gene expression in tissues other than gill and digestive gland.

During the initial planning for printing the oyster microarray it was decided that including genes from both *C. virginica* and *C. gigas* could enhance the effective size and experimental value of the array by heterologous hybridizations. Recent publications have highlighted the utility of this approach in other groups of related species (Adjaye et al. 2004; Renn et al. 2004; von Schalburg et al. 2005). Predicting the level of sequence identity necessary to ensure effective and specific cross-hybridization is not straightforward because factors beyond percent identity (such as GC content and the length of contiguous stretches of identity) play a role in the hybridization efficiency (Chen et al. 2006). However, in the case of heterologous hybridization an increase in phylogenetic distance correlates with a loss of hybridization efficiency (Renn et al. 2004). It is clear that in the present study, a substantial proportion of features on the oyster microarray are capable of providing useful gene expression information through cross-species hybridization. Although the columns representing *C. ariakensis* in Figure 8 may suggest that this species provided very limited hybridization to the array it is important to point out that greater than 16% of all the oyster features on the array (~13% of the *C. virginica* and 23% of the *C. gigas* features) displayed significant

hybridization with RNA from *C. ariakensis* (Figure 5). Because the clustergram (Figure 8) represents differences in expression, it is likely that the significant hybridization from *C. ariakensis* RNA is the result of greater shared homology to “housekeeping” genes. The unique observation that 8% to 10% of the features displaying significant hybridization resulted from heterologous hybridization (Figure 6) is most easily explained by differential gene expression between individual samples. Many of these results are similar to observations made with a salmonid gene array that assessed the levels of heterologous hybridization from several fish species against genes selected from the Atlantic salmon and the rainbow trout (von Schalburg et al. 2005). It seems clear that the potential utility of the oyster microarray has been greatly enhanced by the inclusion of features from both *C. virginica* and *C. gigas*.

In an effort to determine how much overlap in cross-species hybridization can be expected from the two oyster species, a comparison of differences in hybridization intensity of several putative “housekeeping” genes was used to highlight the relationship between sequence similarity and hybridization efficiency (Figure 7). Although the analysis is based on the assumption that the homologous genes were expressed at similar levels in both oyster species, the observed correlations are quite consistent with the expected relationship.

Even though the oyster microarray described here represents only a portion of the oyster transcriptome, the results presented, coupled with the deliberate bias of using genes regulated in response to stress and environmental challenge on the array, indicate that it should prove a useful tool for studies of the oyster transcriptional response to a variety of challenges, including those relevant to the marine estuarine habitat of the oyster.

Acknowledgments

The authors thank Dr. Fran van Dolah (NOAA) and Kristy B. Lidie for donation of the *K. brevis* chloroplast clones, Mara Lennard for the donation of the catfish Bob1 clone, and Dr. Javier Robalino for the donation of the white spot syndrome virus clones. This paper is Marine Genome Project Contribution no. 30 and no. 618 of the South Carolina Department of Natural Resources. Part of the research (graduate research support for M.J. Jenny) was conducted under an award from the Estuarine Reserves Division, Office of Ocean and Coastal Resource Management, National Ocean Service, National Oceanic and Atmospheric Administration and the Science to Achieve Results Grad-

uate Fellowship, U.S. Environmental Protection Agency. Construction and characterization of the *Crassostrea* cDNA microarray was also supported by National Oceanic and Atmospheric Administration Center of Excellence for Oceans and Human Health at the Hollings Marine Laboratory (National Ocean Service, National Centers for Coastal Ocean Science), as well as the Ernest F. Hollings Visiting Scholar Program (C. Cunningham). Additional funding came from the Oyster Disease Research Program, NOAA agreement no. NA16RG1039 and Sea Grant Project R/SAQ-08-NSI. The authors would also like to thank Drs. A.F. Holland and P.A. Sandifer for their encouragement and enthusiasm throughout all phases of the work. We thank Drs. Christopher Bayne, Philippe Roch, and Andrew Mount for careful reading of the manuscript and many helpful suggestions. The oyster cDNA microarray is publicly available to the research community on a cost-recovery basis (contact Paul Gross; grossp@mus.edu).

References

- Adjaye J, Herwig A, Herrmann D, Wruck W, BenKahla A, Brink TC, Nowak M, Carnwath JW, Hultschig C, Niemann H, Lehrach H (2004) Cross-species hybridisation of human and bovine orthologous genes on high density cDNA microarrays. *BMC Genomics* 5, 83
- Ashburner M, Ball AC, Blake JA, Botstein D, Butler H, Cherry M, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. *Nat Genet* 25, 25–29
- Benjamini Y, Holchberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B* 57, 289–300
- Boutet I, Tanguy A, Moraga D (2004) Response of the Pacific oyster *Crassostrea gigas* to hydrocarbon contamination under experimental conditions. *Gene* 329, 147–157
- Chen YA, McKillen DJ, Wu S, Jenny MJ, Chapman RW, Gross PS, Warr GW, Almeida JS (2004) Optimal cDNA microarray design using expressed sequence tags for organisms with limited genomic information. *BMC Bioinformatics* 5, 191
- Chen Y, Chou C-C, Lu X, Slate E, Peck K, Xu W, Voit E, Almeida J (2006) A multivariate prediction model for microarray cross-hybridization. *BMC Bioinformatics* 7, 101
- Cheney DP, MacDonald BF, Elston RA (2000) Summer mortality of Pacific oysters, *Crassostrea gigas* (Thunberg): initial finding on multiple environmental stressors in Puget Sound, Washington, 1998. *J Shellfish Res* 19, 353–359
- Cunningham C, Hikima J, Jenny MJ, Chapman RW, Fang G-C, Saski C, Lundqvist M, Wing RA, Cupit PM, Gross PS, Warr RW, Tomkins JP (2006) New resources for marine genomics: bacterial artificial chromosome libraries for the eastern and Pacific oysters (*Crassostrea virginica* and *C. gigas*). *Mar Biotechnol* 8, 521–533

- David E, Tanguy A, Pichavant K, Moraga D (2005) Response of the Pacific oyster *Crassostrea gigas* to hypoxia exposure under experimental conditions. *FEBS J* 272, 5635–5652
- de Lorgeril J, Saulnier D, Janech MG, Guegen Y, Bachere E (2005) Identification of genes that are differentially expressed in hemocytes of the Pacific blue shrimp (*Litopenaeus stylirostris*) surviving an infection with *Vibrio penaeicida*. *Physiol Genomics* 21, 174–183
- Gouletquer P, Soletchnik P, Le Moine O, Razet D, Geairon P, Faury N, Taillade S (1998) Summer mortality of the Pacific cupped oysters *Crassostrea gigas* in the Bay of Marennes-Oleron (France). *Cons Int Explor Mer CIEM* 14, 20
- Gracey AY, Troll JV, Somero GN (2001) Hypoxia-induced gene expression profiling in the euryoxic fish *Gillichthys mirabilis*. *Proc Natl Acad Sci USA* 98, 1993–1998
- Guegen Y, Cadoret JP, Flament D, Barreau-Roumiguere C, Girardot AL, Garnier J, Hoareau A, Bachere E, Escoubas JM (2003) Immune gene discovery by expressed sequence tags generated from hemocytes of the bacteria-challenged oyster, *Crassostrea gigas*. *Gene* 303, 139–145
- Hedgecock D, Gaffney PM, Gouletquer P, Guo X, Reece K, Warr GW (2005) The case for sequencing the Pacific oyster genome. *J Shellfish Res* 24, 429–441
- Huber W, von Heydebreck A, Sultmann H, Poustka A, Vingron M (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* 18(Suppl 1):S96–S104
- Huvet A, Herpin A, Degremont L, Labreuche Y, Samain J-F, Cunningham C (2004) The identification of genes from the oyster *Crassostrea gigas* that are differentially expressed in progeny exhibiting opposed susceptibility to summer mortality. *Gene* 343, 211–220
- Jenny MJ, Ringwood AH, Lacy ER, Lewitus AJ, Kempton JW, Gross PS, Warr GW, Chapman RW (2002) Potential indicators of stress response identified by expressed sequence tag analysis of hemocytes and embryos from the American oyster, *Crassostrea virginica*. *Mar Biotechnol* 4, 81–93
- Kleinbaum D, Kupper L, Muller K, Nizam A (1998) *Applied Regression Analysis and Other Multivariable Methods*, 3rd ed (Pacific Grove, CA: Duxbury)
- Koganezawa A (1974) Present status of studies on the mass mortality of cultured oysters in Japan and its prevention. USA–Japan Meeting on Aquaculture, Tokyo, Japan, pp 29–34
- Lidie KB, Ryan JC, Barbier M, Van Dolah FM (2005) Gene expression in Florida red tide dinoflagellate *Karenia brevis*: analysis of an expressed sequence tag library and development of DNA microarray. *Mar Biotechnol* 5, 481–493
- Martinez W, Martinez A (2002) *Computational Statistics Handbook with MATLAB*. (Boca Raton, FL: Chapman & Hall/CRC)
- McKillen DJ, Chen YA, Chen C, Jenny MJ, Trent HF, Robalino J, McLean DC, Gross PS, Chapman RW, Warr GW, Almeida JS (2005) Marine genomics: a clearing-house for genomic and transcriptomic data of marine organisms. *BMC Genomics* 6, 34
- Newell RIE (1988) Ecological changes in Chesapeake Bay: are they the result of over-harvesting the Eastern oyster (*Crassostrea virginica*)? In: *Understanding the Estuary: Advances in Chesapeake Bay Research*, Lynch MP, Krome EC, eds. (Solomons, MD: Chesapeake Research Consortium) pp 536–546
- Peatman EJ, Wei X, Feng J, Liu L, Kucuktas H, Li P, He C, Rouse D, Wallace R, Dunham R, Liu Z (2004) Development of expressed sequence tags from Eastern oyster (*Crassostrea virginica*): lessons learned from previous efforts. *Proc Mar Biotechnol* 6, S491–S496
- Renn SC, Aubin-Horth N, Hofmann HA (2004) Biologically meaningful expression profiling across species using heterologous hybridization to a cDNA microarray. *BMC Genomics* 5, 42
- Robalino J, Bartlett T, Shepard E, Prior S, Jaramillo G, Scura E, Chapman RW, Gross PS, Browdy CL, Warr GW (2005) Double-stranded RNA induces sequence-specific antiviral silencing in addition to nonspecific immunity in a marine shrimp: convergence of RNA interference and innate immunity in the invertebrate antiviral response? *J Virol* 79, 13561–13571
- Soletchnik P, Le Moine O, Faury N, Razet D, Geairon P, Gouletquer P (1999) Summer mortality of the oyster in the Bay Marennes-Oleron: spatial variability of environment and biology using a geographical information system (GIS). *Aquat Living Res* 12, 131–143
- Storey JD (2002) A direct approach to false discovery rate. *J R Statist Soc B* 64, 479–498
- Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 100, 9440–9445
- Tanguy A, Guo X, Ford, S (2004) Discovery of genes expressed in response to *Perkinsus marinus* challenge in eastern (*Crassostrea virginica*) and Pacific (*C gigas*) oysters. *Gene* 338, 121–131
- Tanguy A, Boutet I, Laroche J, Moraga D (2005) Molecular identification and expression study of differentially regulated genes in the Pacific oyster *Crassostrea gigas* in response to pesticide exposure. *FEBS J* 272, 390–403
- von Schalburg KR, Rise ML, Cooper GA, Brown GD, Gibbs AR, Nelson CC, Davidson WS, Koop BF (2005) Fish and chips: various methodologies demonstrate the utility of a 16,006-gene salmonid microarray. *BMC Genomics* 6, 126