Original Article

MARINE BIOTECHNOLOGY

New Resources for Marine Genomics: Bacterial Artificial Chromosome Libraries for the Eastern and Pacific Oysters (*Crassostrea virginica* and *C. gigas*)

Charles Cunningham,[1,5,★] Jun-ichi Hikima,[1] Matthew J. Jenny,[1,6] Robert W. Chapman,[2] Guang-Chen Fang,[3] Chris Saski,[3] Mats L. Lundqvist,[1] Rod A. Wing,[4] Pauline M. Cupit,[1,5] Paul S. Gross,[1] Greg W. Warr,[1] Jeff P. Tomkins[3]

[1]Marine Biomedicine and Environmental Sciences Center and Department of Biochemistry and Molecular Biology, Medical University of South Carolina, Hollings Marine Laboratory, Charleston, SC 29412, USA
[2]Marine Resources Research Institute, South Carolina Department of Natural Resources, Charleston, SC 29412, USA
[3]Clemson University Genomics Institute, Clemson, SC 29634, USA
[4]Arizona Genomics Institute, 303 Forbes Hall, Tucson, AZ 85721, USA
[5]Department of Biology, University of New Mexico, 216 Castetter Hall, Albuquerque, NM 87131, USA
[6]Department of Biology, Woods Hole Oceanographic Institution, Woods Hole, MA 02543-1050, USA

Received: 1 February 2006 / Accepted: 16 March 2006 / Published online: 11 August 2006

Abstract

Large-insert genomic bacterial artificial chromosome (BAC) libraries of two culturally and economically important oyster species, *Crassostrea virginica* and *C. gigas*, have been developed as part of an international effort to develop tools and reagents that will advance our ability to conduct genetic and genomic research. A total of 73,728 *C. gigas* clones with an average insert size of 152 kb were picked and arrayed representing an 11.8-fold genome coverage. A total of 55,296 clones with an average insert size of 150 kb were picked and arrayed for *C. virginica*, also representing an 11.8-fold genome coverage. The *C. gigas* and *C. virginica* libraries were screened with probes derived from selected oyster genes using high-density BAC colony filter arrays. The probes identified 4 to 25 clones per gene for *C. virginica* and 5 to 50 clones per gene for *C. gigas*. We conducted a preliminary analysis of genetic polymorphism represented in the *C. gigas* library. The results suggest that the degree of divergence among similar sequences is highly variable and concentrated in intronic regions. Evidence supporting allelic polymorphism is reported for two genes and allelic and/or locus specific polymorphism for several others. Classical inheritance studies are needed to confirm the nature of these polymorphisms. The oyster BAC libraries are publicly available to the research community on a cost-recovery basis at www.genome.clemson.edu.

Keywords: BAC library — *crassostrea gigas* — *crassostrea virginica* — oysters

Introduction

While bivalve molluscs, such as the oyster, are not widely used model organisms for basic biomedical research, they are of substantial economic and cultural significance and of great importance in the ecology of the coastal marine environment (Hedgecock et al., 2005). The coastal ocean zone is typically under pressure from overexploitation of natural resources, from pollution resulting from agricultural and industrial activities, and from coastal development. The importance of oysters in maintaining the health of a coastal ecosystem is illustrated dramatically by the condition of Chesapeake Bay, where the decline of the oyster population due to disease, climatic conditions, and anthropogenic influences is a major factor in the currently poor and probably irreversible condition of this ecosystem (Newell, 1988).

The lophotrochozoa, which include molluscs and annelids, represent one of the two great divisions of invertebrate life, the second branch being the ecdysozoa, including arthropods and nematodes. Clearly, since *Drosophila melanogaster* and *Caenorhabditis elegans* have been heavily exploited as

★Present address: Department of Biology, University of New Mexico, NM 87131, USA
Correspondence to: Charles Cunningham; E-mail: ccunnin@unm.edu

model organisms, the fundamental biology of ecdysozoans is much better understood than that of the lophotrochozoans. The oyster offers opportunities for basic research into the biology of a lophotrochozoan species. A number of recent articles have improved our understanding of the life cycle, development, immune function, and response to environmental contaminants of bivalve mollusks (Tanguy and Moraga, 2001; Jenny et al., 2002, 2004; Gueguen et al., 2003; Tanguy et al., 2004; Herpin et al., 2005a,b). There is also scope, however, for a myriad of highly practical investigations into the role oysters play in the health of marine ecosystems and their potential impact on human populations that not only consume them but also share the same marine environment for commercial and recreational activities.

An international group of investigators interested in the ecology, genetics, physiology, and molecular biology of oysters have come together to form the Oyster Genome Consortium (OGC). The purpose of this group is to promote and facilitate the development of critical tools and reagents that will permit genetic and genomic research on the oyster to prosper (Hedgecock et al., 2005). For example, Milbury and Gaffney (2005) have published a complete mitochondrial DNA sequence for *Crassostrea virginica* while Wang et al. (2005) have used fluorescent *in situ* hybridization techniques to begin identifying chromosomes in *C. virginica*. In addition, Hubert and Hedgecock (2004) have published the first linkage map for a bivalve mollusc (*Crassostrea gigas*) and Zhang et al. (2005) have identified polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) and microsatellite markers that should assist in the enhancement of *C. ariakensis* stocks. In 2004, the OGC was successful in its submission of a formal proposal to the U.S. Human Genome Research Institute requesting that bacterial artificial chromosome (BAC) library resources be developed for oysters. As a result, efforts were directed towards construction of BAC libraries from two important species of oyster; *C. gigas* (the Pacific oyster) and *C. virginica* (the Eastern oyster). To facilitate the exploitation of these publicly available genomic resources by the marine research community, we report the construction, characterization, and utility of these new BAC libraries. These new genomic tools provide a key resource advancing research in oyster genetics and genomics.

Materials and Methods

Oysters and Preparation of High Molecular Weight DNA. High molecular weight (HMW) DNA was prepared from two male hybrid (35×51) Pacific oysters (Taylor Shellfish Farms, Shelton, WA, USA; Hedgecock et al., 2004) and two male Eastern oysters taken from Wadmalaw Sound, SC, USA. Sperm was collected by aspiration from gonadal tissue using a 200-µl pipette and resuspended in 3.5% NaCl containing 5 mM EDTA. Only free-swimming sperm were used for HMW DNA preparation. An equal volume of each sperm suspension was mixed and the sperm were washed three times in saline by centrifugation at 1200 rpm for 3 min at 10°C. To ensure a clean preparation, only free-swimming sperm were collected after each wash. The sperm were finally resuspended in saline to a concentration of approximately 1×10^9 cells/ml, warmed to 45°C in a water bath for 7 min, and an equal volume of 1.5% low melting agarose (dissolved in 3.5% NaCl containing 5 mM EDTA) gently mixed in. The solution was pipetted into 75-µl gel plug molds (Bio-Rad). 10 Plugs were incubated in EPS buffer (0.5 M EDTA, 1 mg/ml Proteinase K, 1% sodium sarcosyl, pH 9.2) at 50°C with moderate rotation in a hybridization oven for 48 h with one buffer change after the first 24 h. Plugs were then washed twice at room temperature in TE (10 mM Tris, 1 mM EDTA) containing 1 mM phenylmethylsulfonyl fluoride (PMSF), the first wash for 4 h and the second overnight. Finally, plugs were washed three times in a 10-fold excess volume of TE for 2 h each and stored in 70% ethanol at −20°C until use.

BAC Library Construction and Clone Characterization. The BAC vector pCUGI-1 was utilized and prepared as described by Luo et al. (2001). Partial digests of HMW DNA (using *Hin*dIII for *C. virginica* and *Bsty*1 for *C. gigas*), size selections, and ligations were performed as described in detail by Peterson et al. (2000). Recombinant colonies were picked using a Q-bot (Genetix) and arrayed individually in 384-well microtiter plates (Genetix). Three copies of each library were made and stored in separate −80°C freezers.

To estimate insert size and determine the distribution of clone size, a total of 200 standard BAC mini-preparations in a 96-well format were performed for each library from 216 clones selected at random (Ammiraju et al., 2006). BAC DNA was digested with 7.5 U of *Not*I (10 h at 37°C) and analyzed by pulsed field electrophoresis through 1% agarose gels (6 V/cm, 5 to 15 s switch time, 15 h run time, 14°C). A total of 196 and 195 successful mini-preparations, restriction digest reactions, and gel reads were obtained for *C. gigas* and *C. virginica*, respectively.

BAC Library Screening. High-density BAC colony filter arrays prepared via a Q-bot (Genetix) were used for hybridization-based screening of the libraries. Clones were gridded in duplicate using a 4 × 4 array on 22.5-cm Hybond N+ filters (Amersham). This gridding pattern allows 18,432 clones to be represented per filter in duplicate.

Of the 14 genes chosen for screening the *C. virginica* BAC library (Table 1), 13 clones were supplied by the Marine Genomics Group, Charleston, SC, USA [NCBI accession nos. BG624486, BG624524 (Jenny et al., 2002); AY331697 (Jenny et al., 2004); CV087794, CV087870, CV087791, CV088804, CV089622, CV088821, CV088862, CV088694, CV088576, and CV088625 http://www.marinegenomics.org)] and the last clone was generously provided by Z. Lui, Auburn University, AL, USA (NCBI accession no. CD647860). Of the 13 genes chosen for screening the *C. gigas* BAC library (Table 2), 7 clones were generously provided from the IFREMER EST collection at the University of Montpellier, France [NCBI accession nos. BQ427137, BQ426631, BQ427181, BQ427355, BQ426799, BQ426657, BQ427361 (Gueguen et al., 2003)] and 6 clones were provided by CC [NCBI

Table 1. cDNA Templates and Oligonucleotide Primers Used to Generate Probes to Screen the *C. virginica* Library

cDNA template[a]	*GenBank accession nos.*	*Primers*	*Primer sequences (5′-3′)*
Metallothionein I	AY331697	F1	CCATGTAACTGTATTGAGACT
		R1	GCCTCTCATTGGTCGAGC
		R28	CGCGTAATACGACTCACTATAGGGAGAGCCTCTCATTGGTCG
Metallothionein III	BG624486	F2	CCAGCTGCACTTGTGCC
		R2	GCACTCGGCTGAGCAG
		R29	CGCGTAATACGACTCACTATAGGGAGAGCACTCGGCTGAGC
Metallothionein IV	CV087794	F3	GGAAAATGTGTGTGCTGCG
		R3	CAGCAACAGTTCTCGTCG
		R30	CGCGTAATACGACTCACTATAGGGAGACAACAGTTCTCGTCGC
NFκB	CV087870	F4	CGTGGTGTCCCAGTGTATTC
		R4	CTGCATCTTGACCTCTCG
		R31	CGCGTAATACGACTCACTATAGGGAGACATCTTGACCTCTCG
IκKinase γ	CV087791	F5	GGAGCAGCAACAGCAAC
		R5	GGACTGCCTTAACTCCAC
		R32	CGCGTAATACGACTCACTATAGGGAGAGGACTGCCTTACTCC
CD23/CD209	CV088804	F6	CGTCGTGAACTAGAGGAC
		R6	CATGTGACTGCACTGGAAG
		R33	CGCGTAATACGACTCACTATAGGGAGAGTGACTGCACTGGAAG
Crumbs	CV089622	F7	CCTCGAGAACTCTGTGAAC
		R7	CTTTATCACAGTCTTGTCCAG
		R34	CGCGTAATACGACTCACTATAGGGAGACACAGTCTTGTCCAG
Catalase	CV088821	F8	GGACGCTCTAGCTGGAC
		R8	TCCAAACGCTCCTGCTCC
		R35	CGCGTAATACGACTCACTATAGGGAGACCAAACGCTCCTGA
LPS BP	CV088862	F9	CAGAAATACTCCCACCAATC
		R9	CATTCTCCTCTGTGCTCCC
		R36	CGCGTAATACGACTCACTATAGGGAGACTCCTCTGTGCTCCC
GRAAL	CV088694	F10	GGAATCACGAACAAGACC
		R10	GAATGGTCTCATCCCACAC
		R37	CGCGTAATACGACTCACTATAGGGAGAGGTCTCATCCCACAC
RAL A BP1	CV088576	R11	CCAGCTGAAGCCACCAAAC
		F11	GCTCCTCTAGGAAAACAGG
		R38	CGCGTAATACGACTCACTATAGGGAGAGCTCCTCTAGGAAAAC
Defensin	BG624524	F12	CGGACGCTGTTGTAACATAG
		R12	CCGAGTTCCTTGTTCGTG
		R39	CGCGTAATACGACTCACTATAGGGAGACGAGTTCCTTGTTCGTG
Ferritin	CV088625	F13	GAGCACCTTCGGAACAGG
		R13	CTGGGTGATGTGGTCAGC
		R40	CGCGTAATACGACTCACTATAGGGAGAGGTGATGTGGTCAGC
Cactus	CD647860	F14	GTGGGACTCGGTCTCTTG
		R14	GACAGGCAGCTATCTCAAC
		R41	CGCGTAATACGACTCACTATAGGGAGAGACAGGCAGCTATCTC

[a]For ease of identification, each cDNA template was named for that protein with which it was most closely homologous in BLASTX analysis. This annotation does not necessarily impute function.

accession nos. AJ565473, AJ309316 (Herpin et al., 2002), AJ544074 (Herpin et al., 2005a), AJ577293 (Herpin et al., 2005b), AJ427420 (Herpin et al., 2005b), and AJ543432]. Hybridization-based detection was performed using digoxigenin-11-dUTP (DIG-11-UTP)-labeled probes. Initially, a specific DNA template from which to generate each probe was amplified by PCR using gene-specific primers (F1 to F27 and R1 to R27; Tables 1 and 2). The products of these reactions were electrophoresed through a 1% agarose-TAE gel and bands of the expected size excised and the DNA purified. This DNA was used as the template in a second PCR together with gene specific primers F1 to F27 and R28 to 54. Primers R28 to 54 contained a sequence corresponding to the reverse complement of the T7 polymerase binding site at their 5′-end. The amplified products were electrophoresed through a 1% agarose-TAE gel and bands of the expected size excised and the DNA purified. These products were generally 200 to 600 bp in length. To label each probe approximately 200 ng of template DNA, 2 μl of DIG-11-UTP (Roche), 2 μl of 10× transcription buffer (Promega), 2 μl of T7 RNA polymerase, and DNase-free water to a final volume of 20 μl were incubated at 37°C for 2 h. RNA probes were purified using an RNeasy minElute clean-up kit according to the manufacturer's instructions (Qiagen). Each probe was inspected by electrophoresis through a 0.8% agarose-

Table 2. cDNA Templates and Oligonucleotide Primers Used to Generate Probes to Screen the *C. gigas* Library

cDNA template[a]	*GenBank accession nos.*	*Primers*	*Primer sequences (5′-3′)*
Metallothionein I	AJ565473	F15	GTAACTGCACTGAGACTGGA
		R16	ACATCCACATCCGGAATCGT
		R42	CGCGTAATACGACTCACTATAGGGAGAACATCCACATCCGGA
Metallothionein IV	BQ427137	F16	ATGTCCGATCACTGTACCTG
		R16	CTTCTTACAGCAGCAGTCGG
		R43	CGCGTAATACGACTCACTATAGGGAGACTTCTTACAGCAGCA
Type 1 BMPR	AJ577293	F17	ATGGCGGTTCGAGTTTGGCG
		R17	CCGACATGAACGAATCACTC
		R44	CGCGTAATACGACTCACTATAGGGAGACCGACATGAACGAAT
MyD88	BQ426631	F18	ACGACGCTTTTGTGATTTAC
		R18	ATGGGCAACTTTAAGTTGAA
		R45	CGCGTAATACGACTCACTATAGGGAGAATGGGCAACTTTAAG
IκB	BQ427181	F19	ATGGAGACTCGCAACTACAC
		R19	CAAGTGCAGACAAGTCTGAC
		R46	CGCGTAATACGACTCACTATAGGGAGACAAGTGCAGACAAGT
Type I TGFβR	AJ544074	F20	CCCTATTATCTTTTAGCCTG
		R20	GGCTGCATGAAGGGCGGGTC
		R47	CGCGTAATACGACTCACTATAGGGAGAGGCTGCATGAAGGGC
Type II TGFβR	AJ427420	F21	TTCCAACCTATCATCGGCT
		R21	GGCACAAGAAACTGGTGGTC
		R48	CGCGTAATACGACTCACTATAGGGAGAGGCACAAGAAACTGG
Type I ActivinR	AJ309316	F22	GGTGGATGCCCATTTTCACA
		R22	TGGGGTAAGAGTTCCTCTCT
		R49	CGCGTAATACGACTCACTATAGGGAGATGGGGTAAGAGTTCC
Tolloid	AJ543432	F23	AGAAGAATATGCAGCCTGGA
		R23	CCACGACCAATGCGTTTACG
		R50	CGCGTAATACGACTCACTATAGGGAGACCACGACCAATGVGT
VAV2	BQ427355	F24	ATTGCGTTGACTGGCTCGTT
		R24	CGGTGGCCGATGAGAGGTCT
		R51	CGCGTAATACGACTCACTATAGGGAGACGGTGGCCGATGAGA
TIMP	BQ427361	F25	GGAACCCCGAGTACCTGACC
		R25	CCTCCCAACATCCGGTTTTG
		R52	CGCGTAATACGACTCACTATAGGGAGACCTCCCAACATCCGG
LMPX	BQ426799	F26	GAGGGTCATCTACACATTTG
		R26	AGTAAATATGGGTTGATTTC
		R53	CGCGTAATACGACTCACTATAGGGAGAAGTAAATATGGGTTG
Coagulation factor V	BQ426657	F27	ACAGACGGACAATAGTGATG
		R27	CACCAACCCAGAGTTGGTGT
		R54	CGCGTAATACGACTCACTATAGGGAGACACCAACCCAGAGTT

[a]For ease of identification each cDNA template was named for that protein with which it was most closely homologous in BLASTX analysis. This annotation does not necessarily impute function.

Table 3. Schematic of Procedure Used to Probe the Filters

Probe	*Filter 1*	*Filter 2*	*Filter 3*	*Filter 4*
1	0	0	0	1
2	0	0	1	0
3	0	0	1	1
4	0	1	0	0
5	0	1	0	1
6	0	1	1	0
7	0	1	1	1
8	1	0	0	0
9	1	0	0	1
10	1	0	1	0
11	1	0	1	1
12	1	1	0	0
13	1	1	0	1
14	1	1	1	0
15	1	1	1	1

0 indicates absence of probe from a filter; 1 indicates presence of probe in filter hybridization mix.

TAE gel and was denatured by incubation at 75°C for 3 min followed by cooling on ice before it was added to hybridization buffer.

Before use, membranes were soaked in 2× saline sodium citrate (SSC) and then incubated with hybridization buffer [6× SSC, 50 % deionized formamide, 0.5% sodium dodecyl sulfate (SDS), and 1% blocking solution (Roche)] at 65°C for 2 h, after which time the denatured probes were added to the hybridization buffer and incubated at 65°C with agitation. After overnight hybridization, the membranes were washed twice in 2× SSC containing 0.1% SDS for 5 min at room temperature with gentle rocking followed by two washes in 0.2× SSC containing 0.1% SDS for 15 min at 65°C with gentle rocking. The membranes were then blocked, treated with anti-DIG-alkaline phosphatase-labeled antibody, and washed according to the manufacturer's instructions (Roche). Signal was visualized by chemiluminescence after incubation of the membrane with disodium 3-(4-meth-oxyspiro {1,2-dioxetane-3,2′-(5′-chloro) tricyclo [3.3.1.1] decan}-4-yl) phenyl phosphate (CSPD, Roche) followed by exposure to BioMax ML film (Kodak, Rochester, NY, USA) for 60 min.

Initial screening of a library was carried out with four hybridizations, each containing a unique mixture of seven or eight probes. The approach assigned a unique binary code for each of the probes (rows in Table 3), which were applied to filters depending upon the column integer (0, absence; 1, presence; Table 3). This procedure permitted an economical use of 4 filter sets to screen for the selected genes, as opposed to the 13 to 14 filter sets that would have been required if each probe were applied individually. The positive (1) or negative (0) hybridization results for each colony on each filter provided a unique binary coding indicating the probe designation. The clones thus identified were picked, regrown, and rearrayed into 96-well microplates and onto nylon filters. The filters were hybridized with single gene-specific probes to confirm the original gene assignment.

The most probable gene copy number for each probe was obtained by the likelihood function

$$L = C(p)^r(1-p)^{(n-r)}$$

where n is the total number of clones in a library, r is the number of positive clones for a given probe,

Table 4. Primer Sets Used for PCR Amplification and Sequencing of *C. gigas* BAC Clones

Gene name	*Primers*	*Primer sequence (5′-3′)*	*PCR conditions*[a]	*Product length (bp)*
Type 1 BMPR	F28	TGTTGTATAGCTGACCTGGG	1	964
	R55	GCCATGTAACGTTTGGTTCC	1	
	F29	CATTGAAAAAAATGTACCGGTC	1	
Type 1 TGFβR	F30	CTCCGACAACTCGTTTCCCA	1	837
	R56	TCCCAAAGGTCAGACGCTGC	1	
Type I ActivinR	F31	TTGTACTGTCAAAATTCAGC	1	743
	R57	CAACTCTTCTTCCACTGTCA	1	
VAV2	F24	ATTGCGTTGACTGGCTCGTT	2	~4000
	R24	CGGTGGCCGATGAGAGGTCT	2	
TIMP	F32	CAGAACGGTTTCTGCTCAGC	1	866
	R58	TCGTCCCCAAACTTCAGTTC	1	
	F33	AACAATATACCGGTATTCACCA	1	
IκB	F19	ATGGAGACTCGCAACTACAC	2	467 or 484
	R19	CAAGTGCAGACAAGTCTGAC	2	
LMPX	F26	GAGGGTCATCTACACATTTG	1	~1500
	R26	AGTAAATATGGGTTGATTTC	1	
MyD88	F18	ACGACGCTTTTGTGATTTAC	1	453 or 349
	R18	ATGGGCAACTTTAAGTTGAA	1	

[a]1: 95°C/4 min ==> 30 cycles (95°C/30 s —> 55°C/30 s —> 72°C/45 s) ==> 72°C/7 min ==> 4°C/infinity.
2: 95°C/4 min ==> 30 cycles (95°C/30 s —> 58°C/30 s —> 72°C/2 min) ==> 72°C/7 min ==> 4°C/infinity.

p is the probability of selecting a given gene (assumed to be the number of clones divided by the genome coverage), and C is the factorial constant ($n!/(r!(n-r)!$).

PCR Amplicons for Cloning and Sequencing. Eight genes of *C. gigas* were selected for detailed study of their sequence variability. For these eight genes (Table 4), between five and eight of the positive BAC clones identified in the library screening were subjected to PCR amplification and sequencing of the resulting amplicons. The selected regions of each gene were amplified by PCR with Ex-Taq polymerase (Takara) using the primers shown in Table 4 and the BAC clones identified in the library screening as the templates. Two different sets of PCR cycles (depending on the primers) were used, as indicated in the footnote to Table 4. The amplified DNA fragments were cloned into pCR4-TOPO using the TOPO TA cloning kit for sequencing (Invitrogen). Cycle sequencing of the plasmid DNA was performed on the CEQ system (Beckman Coulter) according to the manufacturer's protocol (GenomeLab™ Dye Terminator Cycle Sequencing with Quick Start Kit, Beckman Coulter). Each plasmid DNA was sequenced at least twice. Sequence analysis and alignment was carried out using the Lasergene (DNAStar, Madison, WI, USA) suite of programs.

Results

BAC Library Construction and Characterization. We have constructed BAC libraries for the oyster species *C. gigas* and *C. virginica* that are suitable for physical mapping, map-based cloning, and high-throughput sequencing of selected genomic regions. The *C. gigas* BAC library was cloned using *Bsty*1 and consists of 73,728 clones stored in 192 384-well microtiter plates. Approximately 5%

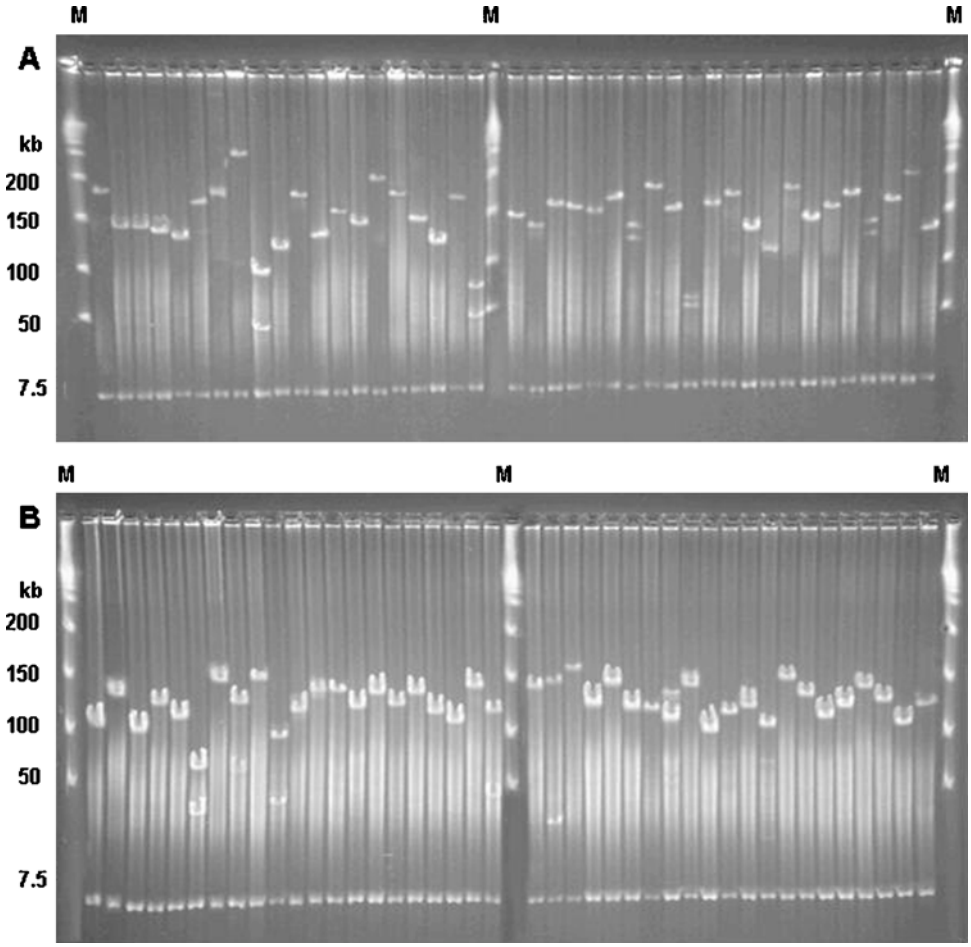


Fig. 1. Analysis of 42 randomly selected oyster BAC clones for *C. gigas* (**A**) and *C. virginica* (**B**). Ethidium bromide stained CHEF gels (5 to 15 s switch time, 15 h) showing insert DNA above the common 7.5-kb BAC vector band. The molecular weight marker (M) in outside lanes and center lanes is a lambda ladder.

(10/196 samples) of the *C. gigas* clones do not contain inserts as judged by random analysis of BACs sampled from the library. A random sampling of 196 BACs taken from the completed *C. gigas* library indicated an average insert size of 152 kb with a range of 30 to 370 kb. The *C. virginica* BAC library was cloned with *Hin*dIII and consists of 55,296 clones stored in 144 384-well microtiter plates. Approximately 0.5% (1/195 samples) of the *C. virginica* clones do not contain inserts as judged by random analysis of BACs sampled from the library. A random sampling *C. virginica* library indicated an average insert size of 150 kb with a range of 30 to 320 kb. Figure 1 shows 42 randomly selected clones from each oyster BAC library digested with *Not*I to release the insert and electrophoresed on pulsed field (CHEF) gels. The two *Not*I sites in pCUGI-1 flank the multicloning site. Because *Not*I recognizes an 8-bp GC sequence and the oyster genomes are not GC rich, digestion typically generates vector plus one to two insert bands per BAC clone.

To determine the size distribution of clones in the oyster libraries, the BACs analyzed with *Not*I digests were grouped by insert size and the insert size of each clone was plotted against the frequency of each group of clones represented in the library (Figure 2). Based on this analysis, more than 94% and 95% of the oyster clones in the libraries have an average insert size equal to or larger than 100 kb for *C. gigas* and *C. virginica*, respectively. Based on the average insert sizes, the percentage of empty clones and the haploid genome sizes of 900 Mb for *C. gigas* and 700 Mb for *C. virginica*, the libraries represent genome coverage of 11.8-fold for both species. In both libraries, this level of representation provides a greater than 99% probability of recovering any specific sequence of interest.

BAC Library Screening. The results of screening the libraries using the selected gene probes are shown in Table 5. Each of the 14 selected genes used to screen the *C. virginica* library, and the 13 genes used to screen the *C. gigas* library detected multiple clones, indicating that the construction of the libraries did not produce a gross bias in the sampling of the oyster genomes. The numbers of clones positive for any single gene probe ranged from 4 to 25 for *C. virginica*, and between 5 and 50 for *C. gigas*. The variations in the numbers of positive clones (Table 5) are interpreted as indicating that while some of the genes are probably single copy, others are present in multiple copies.

Analysis of Polymorphism in the Pacific Oyster Library. The *C. gigas* library was selected for further study because of its strong candidate position for whole genome sequencing (Hedgecock

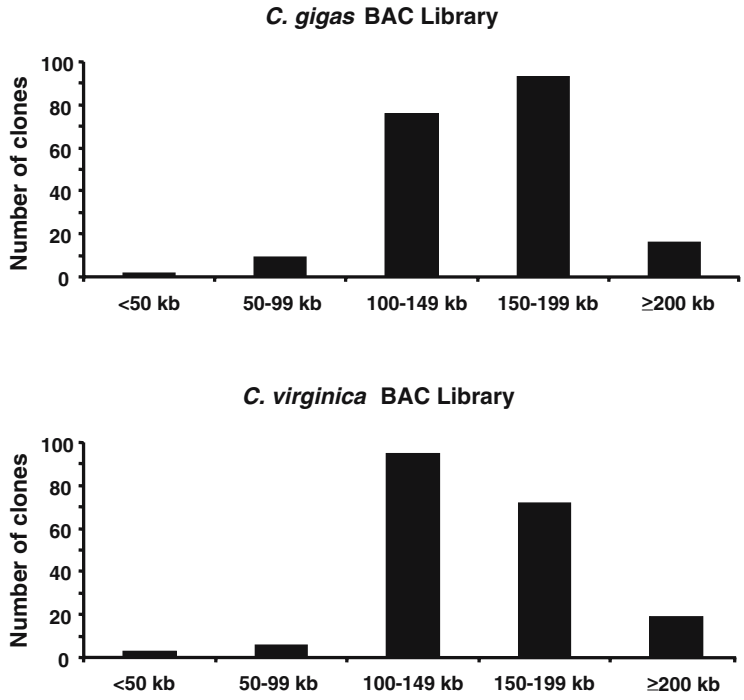


Fig. 2. Insert size distribution of BAC clones in the oyster BAC libraries. To estimate insert size ranges, BAC DNA was analyzed as shown in Figure 1.

Table 5. Frequency of Gene-Specific BAC Clones

Library	*Probe*	*Total number of +ve clones*	*Putative gene copy number* [a]	*Number of double +ve clones*
C. virginica	Metallothionein I	7	1	7 (Metallothionein IV)
C. virginica	Metallothionein III	9	1	
C. virginica	Metallothionein IV	9	1	7 (Metallothionein I)
C. virginica	NFκB	9	1	
C. virginica	IκKinase γ	8	1	
C. virginica	CD23/CD209	25	2	
C. virginica	Notch/Crumbs	16	1 or 2	
C. virginica	Catalase	7	1	
C. virginica	LPS BP	18	2	
C. virginica	GRAAL	8	1	
C. virginica	RAL A BP1	9	1	
C. virginica	Defensin	7	1	
C. virginica	Ferritin	12	1	
C. virginica	Cactus	4	1	
C. gigas	Metallothionein I	22	1 or ***2***	19 (Metallothionein IV)
C. gigas	Metallothionein IV	19	1 or ***2***	19 (Metallothionein I)
C. gigas	Type I BMPR	27	2 or 3	5 (Type I TGFβ Rec)
C. gigas	MyD88	22	2	
C. gigas	IκB	11	1	
C. gigas	Type I TGFβR	5	1	5 (Type I BMPR)
C. gigas	Type II TGFβR	49	4 or 5	28 (Type I ActivinR)
C. gigas	Type I ActivinR	28	2 or 3	21 (Type II TGFβR)
C. gigas	Tolloid	50	4 or ***5***	
C. gigas	VAV 2	28	2	
C. gigas	TIMP	12	1	
C. gigas	LMPX	15	1 or ***2***	
C. gigas	Coag. Factor V	16	1 or 2	

The putative copy number per genome was obtained for the likelihood function (see text). In cases where two values are given, the most likely is shown in bold.
[a]Based on the number of hybridizing BAC clones versus genomic coverage of the libraries. This is a minimal estimate; the members of some gene families are believed to be closely linked.

et al., 2005). The potential value of the *C. gigas* BAC library for defining polymorphisms in the genomes was explored by two methods. In the first, regions of selected BAC clones (five to eight clones for each selected gene) were PCR-amplified, cloned into plasmid vectors, and sequenced to detect allelic variations. Second, in selected cases, allele-specific primer sets were designed and used to determine, by PCR of all BAC clones containing the particular gene, the frequency of the alleles in the library. An

Table 6. Analysis of Sequence Variants in 8 Selected Genes from the *C. gigas* Library

Gene	*Presumptive number of loci*	*Sequences examined*	*Sequence variants*[a]	*Allele A*[c]	*Allele B*[c]	*Allele A*[d]	*Allele B*[d]
Type 1 BMPR	2 or 3	8	Two[b]	7	1	15	7
MyD88	2	8	Two[b]	7	1	–	–
IκB	1	5	Two[b]	4	1	–	–
Type I TGFβR	1	5	One	–	–	–	–
Type I ActivinR	2 or 3	7	One	–	–	–	–
VAV 2	2 or 3	7	One	–	–	–	–
TIMP	1	8	Two[b]	7	1	8	3
LMPX	1 or 2	7	Two[b]	5	2	–	–

[a]GeneBank accession nos. DQ336097–DQ336110.
[b]Alignments are shown in Figure 1.
[c]Values from sequencing of cloned PCR products.
[d]Values from allele-specific PCR of BAC clones (Figure 2).

analysis of sequences amplified by PCR from BAC clones representing eight *C. gigas* genes (Table 6) indicated that two sequence variants could be detected for five of the eight genes, with three genes (type I TGFβ receptor, type 1 activin receptor, and *VAV2*) yielding only a single sequence. Alignments of the sequences (Figure 3) permitted the calculation of the frequency of point mutations, insertions, and deletions in both exons and introns. Several conclusions can be drawn from the sequence alignment: first, the sequence variation is not uniformly distributed through the genes, being concentrated

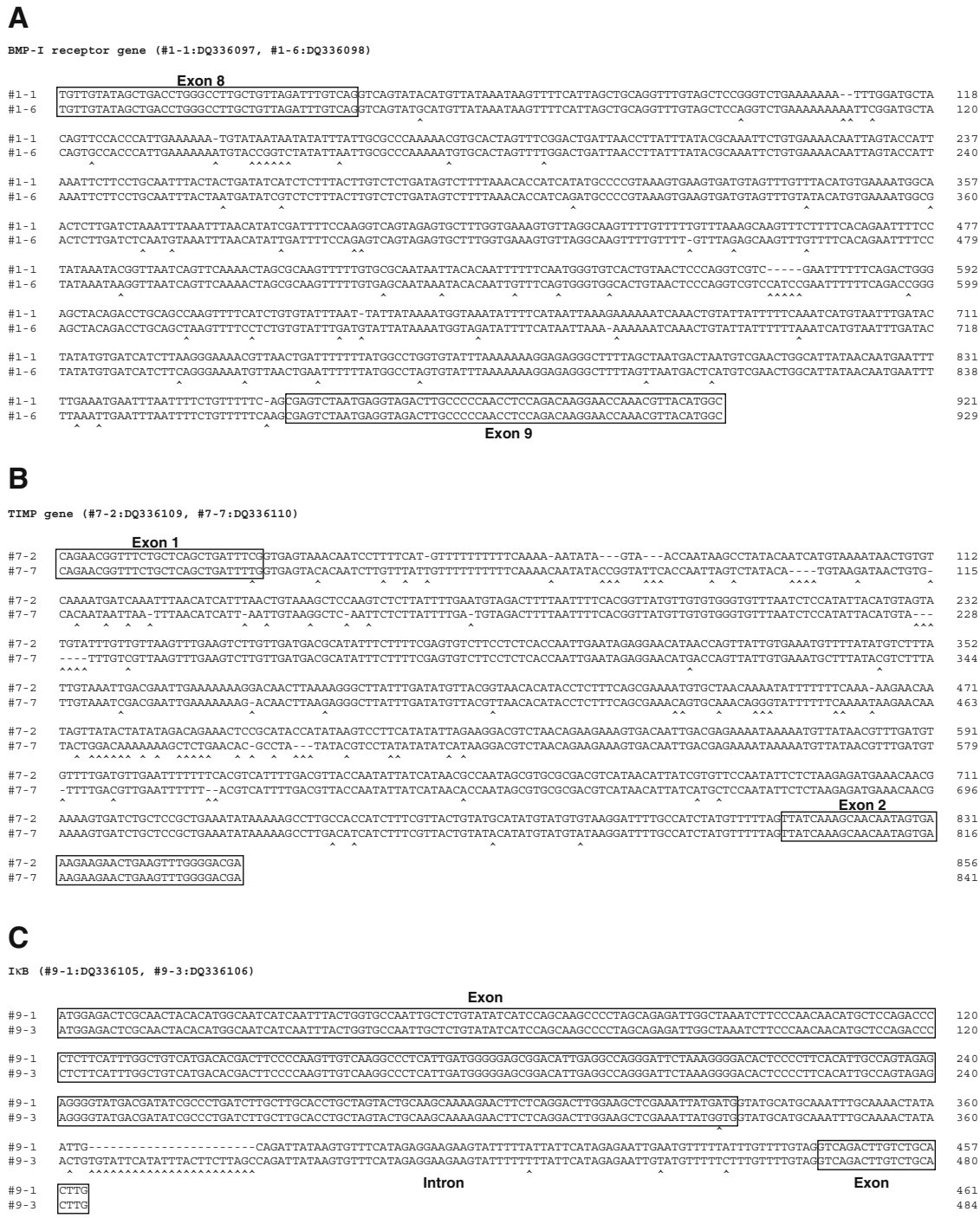
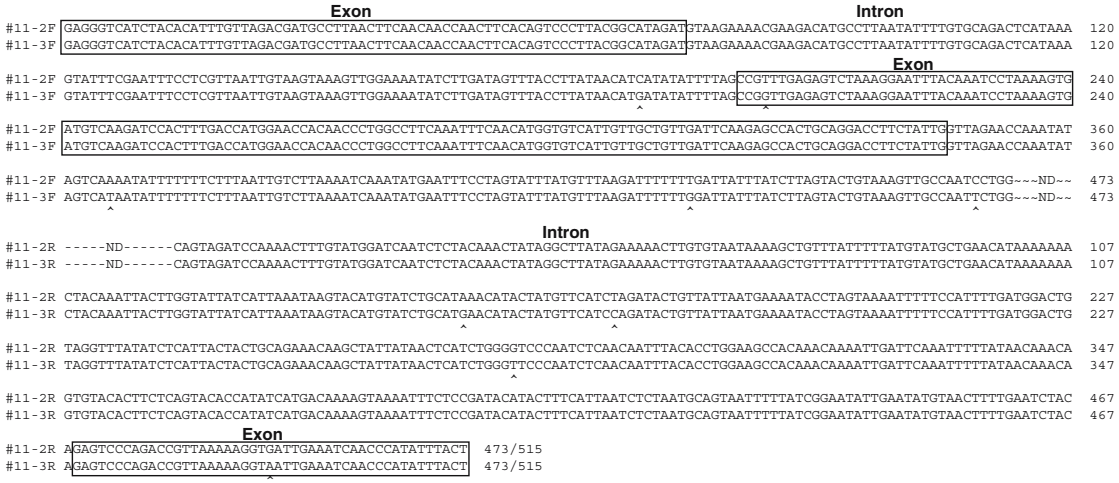


Fig. 3. Aligments are shown for the two sequence variants obtained for BMP type I receptor (**A**), *TIMP* (**B**), *IκB* (**C**), *LMPX* (**D**), and *MyD88* (**E**). Exon regions are indicated by boxes, "-" indicates gaps in the alignments, and "^" indicates variant positions. In (**F**) are shown the summary data for variations in sequence (mutation, insertion, or deletion) in the alignments.

Continues.

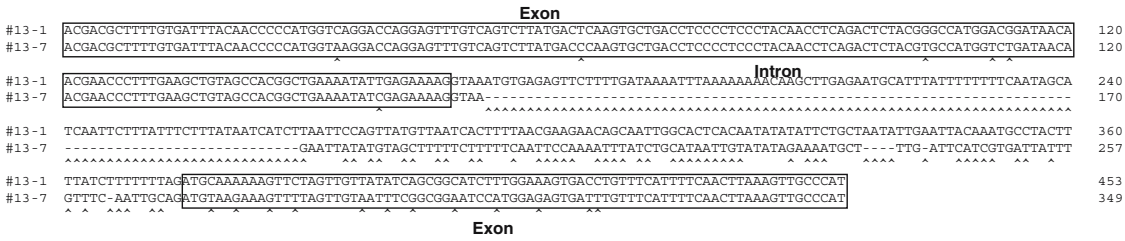
D

LMPX (#11-2F:DQ336101, #11-3F:DQ336102, #11-2R:DQ336103, #11-3R:DQ336104)



E

MyD88 (#13-1:DQ336107, #13-7:DQ336108)



F

Percentages of the allelic mutation, insertion, or deletion found in intronic regions

Gene	Length (bp)	Point substitution		Mutation >2bp		Insertion or deletion		Total	
		(nucleotide)	(%)	(nucleotide)	(%)	(nucleotide)	(%)	(nucleotide)	(%)
BMP-I receptor	921 (929)*	37	4.02 (3.98)	8	0.87 (0.86)	12	1.29 (1.30)	57	6.19 (6.14)
TIMP	856 (841)*	35	4.09 (4.16)	20	2.34 (2.38)	33	3.86 (3.92)	88	10.28 (10.46)
IκB	467 (484)*	4	0.86 (0.83)	0	0	17	3.64 (3.51)	21	4.50 (4.34)
LMPX	988**	6	0.61	0	0	0	0	6	0.61
MyD88	453 (349)*	4	0.89 (1.15)	46	10.15 (13.18)	104	22.96 (29.80)	154	34.00 (44.13)

* The length of the segment sequenced for each gene is shown, with the length of the corresponding allelic variant in parentheses.

** Sequencing of the LMPX gene was conducted from the 5' (473bp) and 3' (515bp) ends of the amplified segment, for a total of 988bp.

Fig. 3. Continued.

as predicted in the intronic regions. Second, the degree of sequence difference between alleles is variable (Figure 3F), ranging from 0.3% in the sequenced intronic regions of *LMPX* to 10.3% in the sequenced introns of *TIMP*. Third, the degree of variation seen for the *C. gigas MyD88* gene sequences is very high: 38% overall, and 34% in the intronic regions. This is sufficient to suggest that homologous sequences, rather than allelic sequences, have been compared. This would be consistent with the conclusion (Table 5) that the *MyD88* gene is probably present in two copies in the *C. gigas* genome. Fourth, no more than two sequences were observed for any one gene, and the number of individual sequences examined for each gene was not always high enough to provide confidence that all variation present in the library had been captured (Table 6); thus, for two genes (type 1 BMP receptor and *TIMP*) it was decided to examine all of the BACs in the library carrying

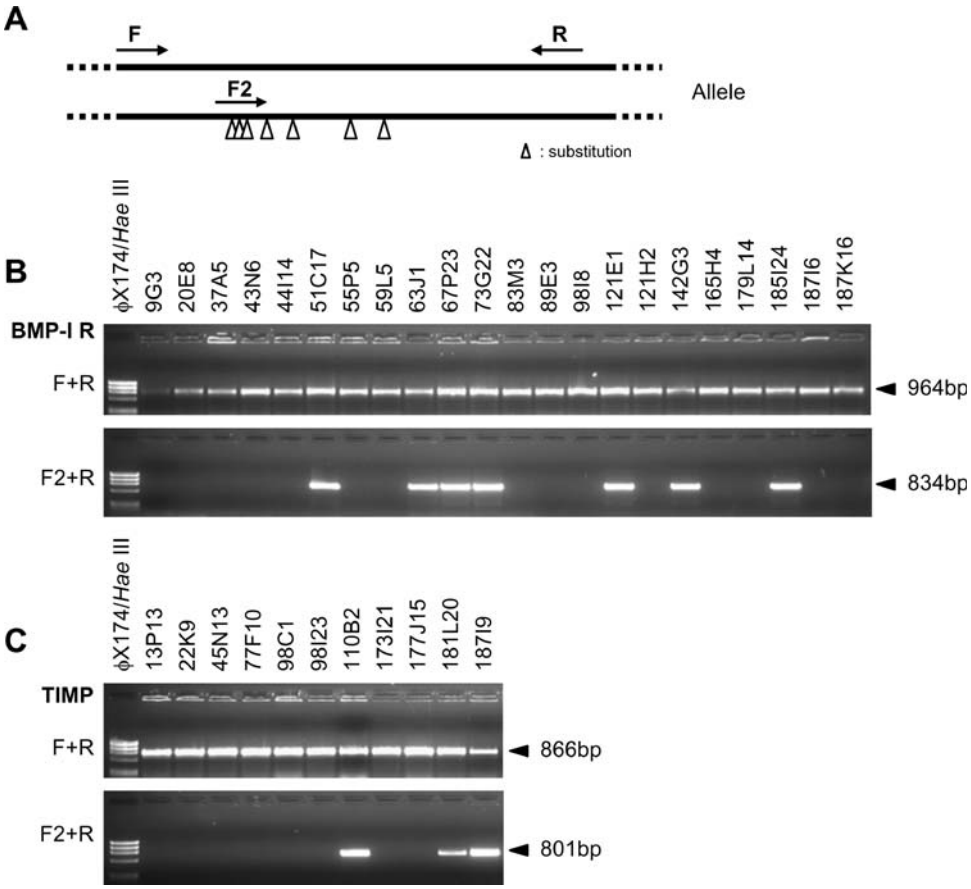


Fig. 4. Detection of allelic variants by PCR amplification from individual BAC clones. (**A**) Schematic showing the general method for the PCR amplifications of BMP type 1 receptor and *TIMP* genes. The triangles indicate the variant positions in the alignment of the BMP type 1 receptor sequences as shown in Figure 3A. (**B**) Amplified DNA fragments of BMP type I receptor gene on a 1.0% agarose gel. Primer F+R and F2+R denote primers F28+R55 and F29+R55 respectively, as listed in Table 4. (**C**) Amplified DNA fragments of the *TIMP* gene analyzed on 1.0% agarose gel. Primer F+R and F2+R denote primers F32+R58 and F33+R58, respectively, as listed in Table 4. The identities of the BAC clones used as PCR templates are shown above the pictures. The size (bp) of the amplicons is indicated by arrows to the right.

these genes. This analysis was carried out via PCR, using a primer set that would specifically anneal to (1) both sequences and (2) primer sets that were sequence specific. The results shown in Figure 4 indicate that the primer sets could detect both alleles, as predicted. In the case of the type 1 BMP receptor the nonspecific primer sets gave a product from 22 of the BAC clones, and of these, 7 were positive using the sequence-specific primers. As the data in Table 5 indicated that this gene is likely to be the product of two or more loci, we cannot determine if these observations are the result of allelic variation or differences between loci. Interestingly, the five clones that were positive for both type 1 BMP receptor and the type 1 TGFβ receptor (Table 5) did not yield a PCR product with either primer set. In the case of the *TIMP* clones, the sequence specific primer set yielded products from 3 of 11 BAC clones, with the twelfth clone yielding no product (data not shown).

Discussion

Collectively, oysters represent a key group of marine taxa for which relatively little genomic information is available, and for which genomic resources are correspondingly scarce. The availability of high-quality BAC libraries for *C. gigas* and *C. virginica*, two of the more economically important and scientifically investigated oyster species, represents a significant step forward in marine genomics research. We have demonstrated that these new resources can be effectively utilized for a variety of genomic investigations in conjunction with the oyster EST collection databased at http://www.marinegenomics.org and NCBI.

Both of these libraries have been shown to have good representation of the oyster genome as supported by the hybridization-based results obtained from screening the libraries using high-density BAC colony filter arrays. The oyster genes used as probes, 14 in the case of *C. virginica* and 13 in the case of *C. gigas*, all positively identified BAC clones in each of the libraries. While the number of BAC clones hybridizing with a specific probe can be used to evaluate library coverage based on gene copy number, such estimates must still be treated with some caution because libraries made with restriction enzymes will reflect a certain level of cloning bias. However, both *Hin*dIII and *Bsty*1 have proven to be excellent enzymes for BAC cloning owing to their capacity to produce good sequence-based genome representation combined with a lack of methylation sensitivity. In the cases in which a large number of positive signals were obtained during library screening with probes that were assumed to be single to low copy, the results were attributed to the targeting of gene families (e.g., metallothioneins; Tanguy and Moraga, 2001; Jenny et al., 2004).

Aspects of copy number estimates combined with the frequency of different probes targeting the same BACs on the oyster BAC filters (Table 5) may be informative concerning the close linkage of certain genes. For example, in *C. gigas*, metallothioneins (MT) I and IV appear to be tightly linked with all 19 of the MT IV clones being positive also for MT I. The close linkage between the MT I and MT IV genes seen in *C. gigas* was also observed for *C. virginica* along with apparent linkage via double-positive clones (Table 5). These results also suggest that there may have been a complex history of chromosomal rearrangement and gene duplication during the evolution of *C. gigas*. Clearly, this has not been a wholesale duplication of the genome, as both *C. gigas* and *C. virginica* have 10 chromosomes and the DNA content of *C. gigas* is only 20% larger (Hedgecock et al., 2004). The interpretation of putative gene copy numbers and their linkage associations in *C. gigas* requires further study.

The *C. gigas* library was examined, in the case of eight specific genes, to determine the degree of polymorphism that had been captured in the library. This is of interest because the library was constructed from two hybrid individuals, each derived from a cross between the 35 and 51 lines, which are partially inbred. In the case of five of the genes, two sequences were represented, and in three genes, only a single sequence was recovered. In the case of one gene (*MyD88*), the two variant sequences recovered were sufficiently distinct to imply homologues derived from a gene-duplication event. This hypothesis is reinforced by the number (22) of *MyD88* clones identified in the library, strongly suggesting that there are two copies of the *MyD88* gene. Quite different degrees of sequence polymorphism were apparent between the individual *C. gigas* genes examined (compare *LMPX* with 0.3% and *TIMP* with 10.3% variation in the intronic regions). While the higher values are consistent with those reported for *C. gigas* by Curole and Hedgecock (2005), the reasons for the gene-to-gene variation in the degree of polymorphism is unexplained. Allelic variation is the most likely explanation for *IκB* and *TIMP* sequence variation as these loci appear to be present in single copies.

Collectively, the characterization of these libraries suggests that they are not only representative of the genomes of these two species of oyster but, as clearly shown in the case of *C. gigas*, they also present the opportunity to investigate genetic polymorphism and linkage associations in the species.

Both of the oyster BAC libraries described in this report are publicly available to the research community on a cost-recovery basis as individual clones, whole libraries, and high-density colony filter arrays. The oyster BAC library resources may be ordered via a secure Web site at the Clemson University Genomic Institute (CUGI): http://www.genome.clemson.edu. Custom library pooling is also available at CUGI for the oyster BACs and quotes may be obtained by sending an email request to: bacrc-info@genome.clemson.edu.

Acknowledgments

We thank members of the Oyster Genome Consortium whose efforts resulted in the funding of the White Paper (http://www.genome.gov/10001852) advocating the construction of the oyster BAC libraries. Construction of the libraries was supported by the NIH Healthy People 2010 BAC library construction program of the National Human Genome Research Institute. The characterization of the libraries was supported as part of an Oceans and Human Health Initiative award from NOAA and by the Ernest F. Hollings Visiting Scholar Program at the Hollings Marine Laboratory, Charleston, SC. We thank Barry Ledford for access to his dock on Wadmalaw Sound for the collection of Eastern oysters.

References

Ammiraju JS, Luo M, Goicoechea JL, Wang W, Kudrna D, Mueller C, Talag J, Kim H, Sisneros NB, Blackmon B,

Fang E, Tomkins JB, Brar D, MacKill D, McCouch S, Kurata N, Lambert GH, Galbraith DW, Arumuganathan K, Rao K, Walling JG, Gill N, Yu Y, San Miguel P, Soderlund C, Jackson S, Wing RA (2006) The *Oryza* bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus *Oryza*. Genome Res 16, 140–147

Curole JP, Hedgecock D (2005) High frequency of SNPs in the Pacific oyster genome. Available at: http://intl-pag.org/13/abstracts/PAG13_W026.html

Gueguen Y, Cadoret JP, Flament D, Barreau-Roumiguiee C, Girardot A, 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, Li G, Hubert SK, Buckli A, Ribes V (2004) Widespread null alleles and poor cross-species amplification of microsatellite DNA loci cloned from the Pacific oyster, *Crassostrea gigas*. J Shellfish Res 23, 379–385

Hedgecock D, Gaffney PM, Goulletquer P, Gou X, Reece K, Warr GW (2005) The case for sequencing the Pacific oyster genome. J Shellfish Res 24, 429–441

Herpin A, Favrel P, Cunningham C (2002) Gene structure and expression of cg-ALR1, a type I activin-like receptor from the bivalve mollusc *Crassostrea gigas*. Gene 301, 21–30

Herpin A, Lelong C, Becker T, Rosa FM, Favrel P, Cunningham C (2005a) Structural and functional evidences for a type 1 TGF-beta *sensu stricto* receptor in the lophotrochozoan *Crassostrea gigas* suggest conserved molecular mechanisms controlling mesodermal patterning across bilateria. Mech Dev 122, 695–705

Herpin A, Lelong C, Becker T, Rosa F, Favrel P, Cunningham C (2005b) Structural and functional evidence for a singular repertoire of BMP receptor signal transducing proteins in the lophotrochozoan *Crassostrea gigas* suggests a shared ancestral BMP/activin pathway. FEBS J 272, 3424–3440

Hubert S, Hedgecock D (2004) Linkage maps of microsatellite DNA markers for the Pacific oyster *Crassostrea gigas*. Genetics 168, 351–362

Jenny MJ, Ringwood AH, Lacy ER, Lewitus AJ, Kempton JW, Gross PS, Warr GW, Chapman RW (2002) 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

Jenny MJ, Ringwood AH, Schey K, Warr GW, Chapman, RW (2004) Diversity of metallothioneins in the American oyster, *Crassostrea virginica*, revealed by transcriptomic and proteomic approaches. Eur J Biochem 271, 1702–1712

Luo M, Wang Y, Frisch D, Joobeur T, Wing R, Dean R (2001) Melon BAC library construction using improved methods and identification of clones linked to the locus conferring resistance to melon Fusarium Wilt (*Fom*-2). Genome 44, 154–162

Milbury CA, Gaffney PM (2005) Complete mitochondrial DNA sequence of the eastern oyster *Crassostrea virginica*. Mar Biotechnol 7, 697–712

Newell RIE (1988) Ecological changes in Chesapeake Bay: are they the result of overharvesting the eastern oyster (*Crassostrea virginica*)? In: *Understanding the Estuary: Advances in Chesapeake Bay Research*, Lynch MP, Krome EC, eds. Chesapeake Research Consortium Publication 129 (CBP/TRS 24/88), pp. 536–546

Peterson DG, Tomkins JP, Frisch DA, Wing RA, Paterson AH (2000) Construction of plant bacterial artificial chromosome (BAC) libraries: an illustrated guide. J Agric Genomics Vol 5. (http://www.ncgr.org/research/jag/index.html)

Tanguy A, Moraga D (2001) Cloning and characterization of a gene coding for a novel metallothionein in the Pacific oyster *Crassostrea gigas* (CgMT2): a case of adaptive response to metal-induced stress? Gene 273, 123–130

Tanguy A, Guo X, Ford SE (2004) Discovery of genes expressed in response to *Perkinsus marinus* challenge in eastern (*Crassostrea virginica*) and Pacific (*C. gigas*) oysters. Gene 38, 121–131

Wang Y, Xu Z, Pierce JC, Guo X (2005) Characterization of eastern oyster (*Crassostrea virginica* Gmelin) chromosomes by fluorescence in situ hybridization with bacteriophage P1 clones. Mar Biotechnol 7, 207–214

Zhang Q, Allen SK Jr, Reece KS (2005) Genetic variation in wild and hatchery stocks of Suminoe Oyster (*Crassostrea ariakensis*) assessed by PCR-RFLP and microsatellite markers. Mar Biotechnol 7, 588–599