THE CASE FOR SEQUENCING THE PACIFIC OYSTER GENOME

DENNIS HEDGECOCK,*1 PATRICK M. GAFFNEY,2 PHILIPPE GOULETQUER,3 XIMING GUO,4 KIMBERLY REECE5 AND GREGORY W. WARR6

1University of Southern California, Department of Biological Sciences, Los Angeles, California 90089-0371; 2College of Marine Studies, University of Delaware, Lewes, Delaware 19958; 3Institut Français de Recherche pour l’Exploitation de la Mer, Ronce les Bains, 17390 La Tremblade, France; 4Haskin Shellfish Research Laboratory, Institute of Marine and Coastal Sciences, Rutgers, the State University of New Jersey, Port Norris, New Jersey 08349; 5School of Marine Science, The College of William and Mary, Virginia Institute of Marine Science, Gloucester Point, Virginia 23062; 6Medical University of South Carolina, Hollings Marine Laboratory, Charleston, South Carolina 29412

ABSTRACT An international community of biologists presents the Pacific oyster Crassostrea gigas as a candidate for genome sequencing. This oyster has global distribution and for the past several years the highest annual production of any freshwater or marine organism (4.2 million metric tons, worth $3.5 billion US). Economic and cultural importance of oysters motivates a great deal of biologic research, which provides a compelling rationale for sequencing an oyster genome. Strong rationales for sequencing the oyster genome also come from contrasts to other genomes: membership in the Lophotrochozoa, an understudied branch of the Eukaryotes and high fecundity, with concomitantly high DNA sequence polymorphism and a population biology that is more like plants than any of the model animals whose genomes have been sequenced to date. Finally, oysters play an important, sentinel role in the estuarine and coastal marine habitats, where most humans live, environmental degradation is substantial, and oysters suffer intense fishing pressures and natural mortalities from disease and stress. Consumption of contaminated oysters can pose risks to human health from infectious diseases. The genome of the Pacific oyster, at 1C = 0.89 pg or –824 Mb, ranks in the bottom 12% of genome sizes for the Phylum Mollusca. The biologic and genomic resources available for the Pacific oyster are unparalleled by resources for any other bivalve mollusc or marine invertebrate. Inbred lines have been developed for experimental crosses and genetics research. Use of DNA from inbred lines is proposed as a strategy for reducing the high nucleotide polymorphism, which can interfere with shotgun sequencing approaches. We have moderately dense linkage maps and various genomic and expressed DNA libraries. The value of these existing resources for a broad range of evolutionary and environmental sciences will be greatly leveraged by having a draft genome sequence.

KEY WORDS: Pacific oyster, Crassostrea gigas genome sequence, Lophotrochozoa, nucleotide diversity, evolutionary and ecological genomics

INTRODUCTION

Complete genome sequences enable a more thorough understanding of biology, particularly of complex traits. Complete genome sequences are available for about 200 species, mostly bacteria and archaea (>150), because of their medical and environmental importance and small genome sizes. Only about half of the more than 40 eukaryotes that have been sequenced so far are metazoans—9 mammals (Bos taurus [cow], Canis familiaris [dog], Felis catus [cat], Homo sapiens [human], Mus musculus [mouse], Ovis aries [sheep], Pan troglodytes [chimpanzee], Rattus norvegicus [rat], Sus scrofa [pig]), 4 other vertebrates (Danio rerio [zebrafish], Gallus gallus [chicken], Takifugu rubripes [fugu], Tetraodon nigroviridis [pufferfish]), 4 insect genera (Anopheles gambiae [mosquito], Apis mellifera [honey bee], Bombyx mori [silk moth] and Drosophila spp. [fruit flies]), 2 sea squirts (Ciona intestinalis and C. savignyi) and 2 nematodes (Caenorhabditis elegans and C. briggsae).

The sea contains most of the higher-order biologic diversity on the planet, composed of invertebrates, most having complex life cycles with planktonic larval phases (Thorson 1950). Though a few marine species have been adopted as models for biologic research and genome sequencing (i.e., the purple sea urchin Stronglylocentrotus, the ascidian Ciona and the puffer fish Takifugu), these have been selected primarily for their advantages in addressing fundamental questions in genome evolution and development (“Evo-Devo”). Sequencing the genome of the Pacific oyster—a model marine invertebrate, with a complex life cycle, living in overexploited and heavily impacted coastal marine environments—would not only provide a mollusc for comparative genomics but also a model species for a broad spectrum of genome-level studies of shellfish biology.

The Joint Genome Institute (JGI) of the US Department or Energy (http://www.jgi.doe.gov/), one of four federally funded centers that participated in the international effort to sequence the human genome, established a Community Sequencing Program (CSP), in February 2004, to direct sequencing and informatics capacity towards issues of scientific and societal importance. This program is intended to fund proposals not targeting the human genome, human disease, or traditional model organisms, which could be funded by other genome programs. In response to this call for proposals, an international community of scientists, self-organized as the Oyster Genome Consortium (OGC), submitted a CSP proposal to generate raw sequencing reads for Pacific oyster DNA, to assemble these into a draft genome sequence and to house the data at JGI, in the near term, allowing community access according to the JGI data-sharing policy. Though it received favorable comments from the review panel and ranked fifth among large-genome proposals, the proposal was not accepted (only one large genome, that of the moss Physcomitrella, was approved for sequencing). The OGC has again submitted the proposal in February 2005. Whether this particular proposal is successful or not, declining costs of high throughput DNA sequencing and excess sequencing capacity at genome centers will eventually make possible the sequencing of a bivalve mollusc genome. Here, we, who served as a steering committee for the OGC, summarize the status...
SCIENTIFIC COMMUNITY

The community of biologists that would benefit from a bivalve mollusc genome sequence, is large and well organized. The National Shellfisheries Association (NSA), founded in 1908 by industry and academic scientists interested in the eastern oyster, now has more than 1,000 members from 20 nations. OGC participants organize genetics and disease sessions at most annual NSA meetings. A West Coast chapter of NSA meets annually with the Pacific Coast Shellfish Growers Association, whose members are involved in two USDA-funded breeding programs to increase yield of C. gigas on the West Coast: (1) the Molluscan Broodstock Program (http://hmsc.oregonstate.edu/projects/mbp), a commercial-scale selection program based at Oregon State University’s Hatfield Marine Science Center, where USDA ARS has located two shellfish biologists and (2) USDA Western Regional Aquaculture Center (WRAC) project, “Crossbreeding for High Yield” (http://www.hmsc.orst.edu/projects/wrac) based at a commercial hatchery, Taylor Resources, Inc. Quilcene, WA. In addition, NOAA National Sea Grant Oyster Disease Research Program has funded projects on stress responses. In 1995, shellfish geneticists established the first marine-theme USDA Western Region Coordinating Committee, WCC-99, “Broodstock management, genetics and breeding programs for mollusc shellfish.” Forty scientists are on Oregon State University’s list server for WCC-99, and 20–30 of these meet annually, usually in conjunction with the NSA annual meeting and with international participation. In 1997, 13 scientists, representing five countries, who were interested in oyster genomics, joined with scientists interested in other aquaculture species (shrimp, tilapia, salmon, catfish and more recently striped bass) to form USDA Northeast Regional Project, NE-186. This group also meets annually in conjunction with the International Plant & Animal Genome Conference (PAG). Last year, NE-186 fused with NRSP-8, the umbrella USDA national program for all animal genomics. Although oysters have received a good share of USDA competitive funds over the years, USDA resources fall far short of those needed for whole genome sequencing.

Internationally, OGC members interact at meetings such as PAG, the triennial meeting of the International Society for Genetics in Aquaculture, the biennial meeting of the International Society for Aquatic Genomics, the International Society for Developmental and Comparative Immunology (ISDCI), and Marine Biotechnology. OGC members have organized several United States-France bi-national conferences sponsored by the NSF (1994) and NOAA-IFREMER (2002 and 2004), attended by 30–50 scientists. It was at the PAG 2003 meeting that OGC formed to prepare the ultimately successful proposal to construct genomic DNA libraries for the Pacific and eastern oysters and, at the summer 2003 ISDCI meeting, that OGC members organized the printing of the first oyster microarray. The OGC is nonexclusive and welcomes all investigators, who have a theoretical or practical interest in oyster genetics and genomics. At present, there are nearly 70 participants in the OGC from 10 countries (Australia, Canada, China, France, Greece, Ireland, Japan, Spain, United Kingdom and the United States; list available from DH).

TECHNICAL INFORMATION ABOUT THE OYSTER GENOME

G + C Content and Genome Size

McLean & Whiteley (1973) reported G + C content for Crassostrea gigas (32.2%, from thermal denaturation, 33.6%, from buoyant density) and estimated haploid genome (1C) size from reassociation kinetics as 1.26 pg. This appears to be an overestimate based on more recent work by González-Tizón et al. (2000; 1C = 0.91 pg by Feulgen image analysis) and Guo (unpublished; 1C = 0.87 pg by flow cytometry). Consensus haploid genome size of ~0.89 pg is equivalent to 824 million nucleotide base pairs (Mb). Genome sizes of 174 molluscs (81 bivalves, 5 cephalopods, 81 gastropods and 7 chitons) range from 0.43 pg for the owl limpet Lottia gigantea to 5.88 pg for the Antarctic whelk Neobuccinum eatoni, with a mean of 1.80 ± 0.07 pg (Gregory 2003; Fig. 1). Only 23 molluscs have genome sizes smaller than that of the Pacific oyster, nine snails, eight limpets and of four bivalves, two clams and two oysters (the eastern oyster Crassostrea virginica [0.69 pg; Hinageardner 1974, Goldberg et al. 1975] and an unknown species). Genome duplication has occurred in the bivalves, but oysters represent the diploid lineage (Wang & Guo 2004).

Polymorphism

Protein polymorphism of the Pacific oyster was long ago found to be among the highest for animals; average heterozygosity is >20%, three to four times the mammalian average (Buroker et al. 1979, Fjugo 1982, Hedgecock & Sly 1990). High nucleotide polymorphism in bivalves is suggested by abundant nonamplifying PCR-null alleles for microsatellite DNA markers (McGoldrick et
al. 2000, Launey & Hedgecock 2001, Reece et al. 2002, Reece et al. 2004, Vadopalas & Bentzen 2000). In the Pacific oyster, null alleles are segregating at over half of ~100 loci tested in mapping families, even though families and microsatellite clones come from the same natural population (Li et al. 2003, Hedgecock et al. 2004a). Success in cross-specific PCR amplification decays rapidly with evolutionary distance, so that only 1 in 8 Pacific oyster markers amplifies from the eastern oyster, which diverged >5 million years ago (Mya). This decay in cross-specific amplification exceeds that observed across genera or even families of vertebrates (Schlötterer et al. 1991, Garza et al. 1995, Pépin et al. 1995, FitzSimmons et al. 1995, Rico et al. 1996) or species of Drosophila (Harr et al. 1998, Colson et al. 1999, Noor et al. 2001, Huttunen & Schlötterer 2002) and suggests rapid rates of sequence evolution in PCR primer binding sites. In their initial study of expressed DNA sequences, Curole & Hedgecock (2005) report a frequency of one SNP every 40 base pairs and one insertion or deletion (indel) every 33 base pairs. The causes of this high polymorphism—a main point of scientific interest in sequencing the oyster genome—and a strategy for reducing it to facilitate the assembly of shotgun DNA sequences are addressed later.

Repeat Structure

Reassociation kinetics of denatured DNA shows that 30% of the Pacific oyster genome is repetitive DNA (McLean & Whiteley 1973). Repeat structure has not been further characterized, except for description of a centromeric repeat (Clabby et al. 1996, Wang et al. 2001). Gaffney reported transposon-like repeat DNA sequences in eastern and Pacific oysters, many of which are associated with tetranucleotide repeat microsatellite loci (Gaffney 2002). One of these, Pearl, is characterized in C. virginica (Gaffney et al. 2003) and detected in C. gigas.

Available Resources

The community of oyster researchers has, collectively, developed diverse and substantial resources in areas ranging from the organismal to the molecular. These resources which are briefly summarized in this section, represent enormous potential for basic research into the fundamental biology of oysters, as well as for commercially relevant broodstock improvement.

Biological Resources

Aneuploid, triploid, tetraploid and gynogenetic Pacific oysters are routinely produced (Guo et al. 1993, Guo & Allen 1994a, Eudeline et al. 2000). Triploid oysters are farmed commercially because of their retarded gonadal development and superior growth. These stocks provide excellent biologic resources for gene-centromere mapping (Guo & Gaffney 1993, Guo & Allen 1996, Hedgecock et al. 2003) or investigations of the role of gene-dosage effects in heterosis (Birchler et al. 2003). Investigators have developed ~50 inbred lines for experimental crosses from the naturalized population of C. gigas in Dabob Bay, Washington (Hedgecock 1994), using self and brother-sister mating (Hedgecock et al. 1995). Factorial crosses among inbred lines produce F1 hybrids for comparisons of growth and survival in hatchery and field trials (progress reports at http://www.hmsc.orec.edu/projects/wrac). The 51 × 35 reciprocal cross, on which massively parallel signature sequencing (MPSS) expression profiling has been done (see later), is in commercial production. The third inbred generation (G3) of the 51 and 35 inbred lines was propagated in 2004 to produce the G4, which has an expected inbreeding coefficient of 0.59.

Linkage Maps

Framework linkage maps of >100 microsatellite DNA markers have been published for the Pacific oyster (Li et al. 2003, Hedgecock et al. 2003, Hubert & Hedgecock 2004). The consensus maps have 10 linkage groups (Fig. 2), in accord with the haploid chromosome number (see later), cover 70% to 80% of the Pacific oyster genome and have marker densities such that the expected distance of a new gene to the nearest marker on a map is 4–6 map units (cm). This microsatellite DNA scaffold can be fleshed out quickly with several hundred AFLP markers (Yu & Guo 2003, Li & Guo 2004). For example, Hedgecock et al. (2004b) have constructed an AFLP map of 341 markers for the 35 × 51 F2, which has an estimated coverage of up to 94%. USDA currently supports QTL mapping (Hedgecock NRICGP #2003–35205–12830) and development and mapping of 100 Type I SNP markers for the Pacific oyster (Gaffney NRICGP #DELR-2003–03620).

Cytological Maps

Cupped oysters have 10 pairs of metacentric or submetacentric chromosomes (Longwell et al. 1967, Ahmed & Sparks 1967, Leitão et al. 1999). The frequency of chiasmata is estimated as 1.0–1.3 per chromosome arm (Guo, unpbl.), implying a genetic map length of 500–650 cM. Chromosomal banding and FISH techniques with P1 clones, rRNA genes and repetitive sequences have recently been applied to chromosome identification and mapping (Wang 2001). A repeat that accounts to 1% to 4% of the genome has been isolated and mapped to centromeric regions of several chromosomes in the Pacific oyster (Clabby et al. 1996, Wang et al. 2001). Major ribosomal RNA genes have been mapped to 10 q in the Pacific oyster and 2 p in the eastern oyster (Xu et al. 2001, Wang et al. 2004, Fig. 3). Integration of linkage and cytogenetic maps is underway for Pacific and eastern oysters (Guo NICST #00–2004–0207–20). Clones from newly constructed BAC libraries (see below) are being mapped to cytogenetic maps by FISH and to linkage maps by polymorphism in BAC clone end sequences.

Expressed Sequence Tags from cDNA & MPSS Libraries

As of January 2005, the NCBI Entrez taxonomy browser retrieved 12,341 entries for “Ostreaeidae.” Of these, 12,059 entries are for the genus Crassostrea, composed mainly of 2,870 sequences for C. gigas and 9,102 entries for C. virginica. The complete mitochondrial genome is available for the Pacific oyster and has been submitted for the eastern oyster. Of the 2,835 nonmitochondrial Pacific oyster sequences in GenBank, 370 are microsatellite-containing clones. The combined number of expressed sequence tags (ESTs) for the eastern and Pacific oysters, 9,018, though useful, is woefully inadequate for characterizing the genome or identifying genes of interest to the community.

Pilot EST collection programs for the Pacific and eastern oysters used hemocyte and embryo cDNA libraries from C. virginica (Jenny et al. 2002, http://www.marinegenomics.org/; Tanguy et al. 2004, Peatman et al. in press; Fig. 4) and a hemocyte cDNA library from C. gigas (Gueguen et al. 2003, http://www.ifremer.fr/GigasBase). Although EST collections are small, the rate of discovery of genes of important function in the response of oysters to stress and immune challenge is excellent. Groups in Montpellier, France (Escoubas, Bachère), Auburn (Li), Baltimore (Vasta),
Rutgers (Guo) and Charleston (Warr, Chapman, Cunningham, Gross) have pooled resources to print the first cDNA-based microarray. Over 5,000 unigenes from *C. gigas* and *C. virginica* and 384 unigenes from the oyster parasite *Perkinsus marinus*, are assembled in Charleston for printing of the first microarray in spring 2005. Similarity of DNA sequences observed in initial comparisons of orthologous genes in *C. gigas* and *C. virginica* averages \(\sim 86\%\) (G. Warr, unpubl.) and suggests that the microarray should be useful for measuring transcriptomic responses in various oyster species.

In addition to the traditional oyster EST collections described earlier, a remarkable library of 4.6 million Pacific oyster ESTs is available from a genome-wide survey of gene expression in larval inbred and hybrid Pacific oysters carried out by Lynx Therapeutics (http://www.lynxgen.com/), using Megaclone and massively parallel signature sequencing or MPSS technologies (Brenner et al.).
Figure 4. A summary of 4,560 expressed sequence tags from \textit{Crassostrea virginica} classified into functional groups based on comparison with the NCBI database (excluding ribosomal RNAs). Approximately 65\% of the ESTs in the database (http://www.marinegenomics.org/) lacked significant homology to any entries in the NCBI nonredundant database based on BLASTX comparisons. The remaining 35\% with significant homology (e-value <10^{-7}) were clustered into functional groups based on broad molecular or biologic classification. Complete annotation based on Gene Ontology is viewable on the website above.

2000a, Brenner et al. 2000b). These ESTs are discussed in the section on growth heterosis.

Bacterial Artificial Chromosome Libraries

The Oyster Genome Consortium made a successful request, to the National Human Genome Research Institute (NHGRI) bacterial artificial chromosome (BAC) Library Construction Program in 2003 for developing genomic BAC libraries from \textit{C. gigas} and \textit{C. virginica}. The libraries were constructed by Clemson University Genomics Institute (Dr. Jeff Tomkins, Director) and are publicly available at (https://www.genome.clemson.edu/orders). These are deep coverage libraries (10x and 12x coverage for the two species, respectively), with average insert sizes of 134 kb and 150 kb, respectively. The libraries were constructed from sperm cells, which, in the case of \textit{C. gigas}, were taken from a 51×35 F_1 hybrid male. Thus, the \textit{C. gigas} BAC library will be of particular value in complementing a genomic sequence made from inbred line 51.

APPLICATIONS OF A PACIFIC OYSTER GENOME SEQUENCE

Applications of a genome sequence are numerous, but the most immediate identified by the OGC fall under 3 headings: (1) comparative genomics, in which oysters permit a phylogenetic contrast in studies of genome structure, function and diversity; (2) evolutionary biology, in which oysters may shed light on the evolution of sexuality or of reproductive isolating mechanisms in the sea and (3) adaptation, in which oysters are a model for understanding the genetic and physiologic bases of complex traits (e.g., growth and survival) that are strongly correlated with Darwinian fitness and population responses to environmental change and stresses, such as disease.

\textbf{Lophotrochozoa: A Major Branch of Life Unexplored}

Bilateral symmetry is found in 3 major clades of animals: the Deuterostomes (which includes vertebrates) and two clades of Protostomes, the Ecdysozoa (which includes arthropods and nematodes) and the Lophotrochozoa (which includes molluscs and annelids). The Deuterostomes are well represented by genomically enabled model species (e.g., Amphioxus, Oikopleura, Ciona spp., Takifugu, zebrafish, Xenopus, sea urchin, chicken, mouse and human), whereas the Ecdysozoans are represented by genome sequences for several species of Drosophila, the honeybee, the mosquito, the silkworm and two species of the nematode Caenorhabditis. An integrated picture of the diversity of animal life, however, requires detailed understanding of representative Lophotrochozoan species. Sequencing is underway at JGI for a gastropod, the limpet Lottia scutum and two annelids, the leech Helobdella robusta and the polychaete worm Capitella capitata and at the NHGRI for the fresh-water snail Biomphalaria glabrata. The phylum Mollusca has three other major classes besides the Gastropoda (snails): Cephalopoda (octopus, squid), Polyplacophora (chitons) and Bivalvia (oysters, mussels, clams). The Pacific oyster has one of the smallest bivalve genomes, at ~824 Mb of haploid DNA (Fig. 1). Lophotrochozoans are the focus of phylogenetic analyses of embryonic development and the diversity of body plans (Kourakis & Martindale 2001, Nederbragt et al. 2002, O’Brien & Degnan 2002, Tessmer-Raible & Arendt 2003) and the evolution of molecular mechanisms of immune recognition and response to stress, including heavy metals (Escoubas et al. 1999, Tanguy & Moraga 2001, Tanguy et al. 2001, Jenny et al. 2002, Gueguen et al. 2003, Huvet et al. 2004, Tanguy et al. 2004, Boutet et al. 2004, Jenny et al. 2004). Progress in these and other areas of comparative research would benefit enormously from the availability of the Pacific oyster genomic sequence.

\textbf{Genome Organization, Reproductive Isolation and Speciation}

There are two competing theories on reproductive isolation and speciation, one arguing for a central role of genic mutation and the other emphasizing chromosomal changes. Chromosomal mutations, particularly polyplody, have played a significant role in the evolution of plants (deWet 1980). Although polyplody is relatively rare in animals, there is increasing recognition of chromosomal changes as an important force in animal evolution (Furlong & Holland 2002, Spring 2002). Chromosomal rearrangements may play a major role in reproductive isolation and speciation, by creating barriers to recombination or reducing the fitness of hybrids (White 1978, King 1993). Many marine bivalves are sympatric broadcast spawners, for which mechanisms of reproductive isolation are interesting but largely unknown. In \textit{Crassostrea} oysters, differences in the rRNA-bearing chromosome clearly divide Asian-Pacific and Atlantic species, along a postzygotic hybridization barrier (Wang et al. 2004). Whether chromosomal rearrangement is responsible for the divergence can only be answered by genome-wide comparisons. Linkage mapping in the Pacific oyster reveals significant differences in recombination between genetic markers and in gene orders, suggesting chromosomal inversion polymorphism within this species (Hubert & Hedgecock 2004). A complete sequence of the Pacific oyster genome will provide a foundation for comparative genomics in marine bivalves, where macrogenomic events such as duplication (Wang & Guo 2004) and rearrangements (Wang et al. 2004) can be studied and their effects on speciation better understood.

The Asian clade of cupped oysters offers a spectrum of evolutionary differences and reproductive isolating mechanisms. Based on differences in allozymes and DNA sequences, the Kumamoto oyster \textit{Crassostrea sikamea} is believed to be the closest relative of the Pacific oyster (Buroker et al. 1979, Banks et al. 1994). The two species, which are sympatric in the Ariake Sea, Kyushu Island,
Japan, are reproductively isolated by ecologic and gametic isolating mechanisms (Banks et al. 1994). Nevertheless, sperm from allopatric populations of the Pacific oyster can fertilize the eggs of the Kumamoto oyster, producing viable progeny (Banks et al. 1994). With a genome sequence and subtractive cDNA hybridization methods, it should be possible to identify the gene or genes responsible for the gametic incompatibility and to trace the evolution of gametic isolation.

The eastern oyster *Crassostrea virginica* is distributed from the Maritime Provinces of Canada to the Yucatan peninsula. In the 1950s and 1960s, oyster biologists recognized significant geographic variation in reproductive and feeding physiology (Stauber 1950, Loosanoff & Nomejko 1951), which was later borne out by experiments (Haskin & Ford 1979, Ford & Haskin 1987, Vrijenhoek et al. 1990, Ford et al. 1990, Barber et al. 1991, Dittman et al. 1998). Disease resistance also varies among geographic populations (Bushek & Allen 1996). Physiologic races of eastern oyster have formed despite a dispersing larval stage that may spend several weeks drifting in the plankton. Analyses of protein (Buroker 1983) and DNA polymorphisms (Reeb & Avise 1992, Karl & Avise 1992, Cunningham & Collins 1994, McDonald et al. 1996) revealed strong genetic differentiation of populations from the Gulf and Atlantic Coasts. More recently, studies of DNA sequence polymorphisms in the mitochondrial large ribosomal subunit gene revealed distinct subpopulations even along the Atlantic coast (Wakefield & Gaffney 1996). Experimental dissection of the genetic basis of geographic variation in complex physiologic processes and local adaptation in the Eastern oyster would be greatly facilitated by genome sequences for the Pacific oyster.

**Sex Determination**

As a group, molluscs exhibit highly diverse modes of sexual reproduction, ranging from functional, simultaneous hermaphroditism to sequential hermaphroditism, to strict dioecy and genic determination (Coe 1943, Guo & Allen 1994b). Although dioecy seems to be the norm in molluscs, about 40% of the 5,600 genera are either simultaneous or sequential hermaphrodites (Heller 1993). Like all cupped oysters, the Pacific oyster is a protandric dioecious hermaphrodite, generally maturing first as a male, then as a female in most subsequent spawning seasons; nevertheless, there is a small but persistent fraction that presents simultaneous hermaphroditism (Coe 1943, Galtsoff 1964). Though modified by age and environment, sex determination seems to be controlled by a major gene (Guo et al. 1998) with the male being heterogametic; sex chromosomes have not been identified. Only preliminary attempts to isolate DNA markers for sex determination from a genomic BAC library of the oyster have been reported (Shimizu et al. 1998). A genomic sequence is likely to stimulate renewed and sustained interest in this topic.

**Biology of Highly Fecund Animals**

**High Mutational Load**

The enormous fecundity ($10^8$–$10^9$ eggs per female per season) and high larval mortality of most marine animals makes them fundamentally different from the more familiar and better-studied animal models ($10^3$–$10^4$ eggs per female, lifetime). A quarter century ago, Williams (1975) argued in his “Elm-Oyster Model” that sexual reproduction and genetic diversity are favored to a much greater degree in oysters and trees than in low fecundity species, such as human, fruit fly or mouse. Species with high fecundity likely generate more mutations than low fecundity species, owing to the large number of cell divisions required to produce millions or billions of gametes (cf. to the argument for male-driven evolution in humans, Li et al. 2002). Most mutations are likely deleterious and recessive and constitute a substantial genetic load on population fitness. As predicted by Williams, trees have larger mutational loads than most animals (cf. mean effective number of lethal mutations per zygote for conifers is 8.1 vs. 2.8 for mammals or *Drosophila*; Lynch & Walsh 1998, Tables 10.4–10.6). Williams’ specific prediction of high genetic load in oysters was recently confirmed by experiments revealing a minimum of 15–20 lethal mutations per oyster, about five times the genetic load of a human or fruit fly (Launey & Hedgecock 2001, Bucklin 2002).

High genetic load resolves two long-debated issues in bivalve genetics: (1) common distortions of Mendelian inheritance ratios in laboratory-reared progeny of wild parents (Wada 1975, Beaumont et al. 1983, Foltz 1986, Hu et al. 1993, McGoldrick & Hedgecock 1996, McGoldrick et al. 2000, Reece et al. 2004) and (2) correlation of heterozygosity with fitness (i.e., growth) in natural populations (see Zouros & Pogson 1994). These phenomena are uncommon or nonexistent in terrestrial animals (Houle 1989, Britten 1996). High mutational load further accounts for severe inbreeding depression and its converse, hybrid vigor (heterosis), in experimental crosses (Lannan 1980, Hedgecock et al. 1995, Hedgecock et al. 1996; Evans et al. 2003). Thus, deleterious recessive mutations in highly fecund marine animals are an important source of endogenous variation in the relative fitness of individuals, with the most severe consequences being expressed at the larval stage or at metamorphosis. The genome sequence, in the context of experimental crosses, would provide means for studying the genome-wide distribution of mutational load, dissecting the gene regulatory networks underlying larval development and survival and accelerating the development of SNPs needed for population genetic analyses of these phenomena.

**Heterosis for Growth, Survival and Other Fitness Traits**

The Pacific oyster provides an animal model for studying heterosis, a phenomenon more evident in plants and underlying the improvement of most crops (Gowen 1952, Fig. 5). Crossbreeding inbred lines of oysters to produce hybrids holds great promise for...
increasing the yields of farmed Pacific oysters. Hybrids are currently in commercial production on the United States West Coast (J. P. Davis, Taylor Resources, pers. comm.). Apart from the obvious economic importance, however, is the significance of heterosis for fundamental evolutionary genetics and physiology.

Because of the facility with which energy budgets can be constructed for animals and bivalves in particular, we have already achieved considerable insight into the physiologic basis of growth heterosis (Hedgecock et al. 1996, Bayne et al. 1999). These early results attracted collaboration with Lynx Therapeutics, Inc., which, as mentioned in the EST section, provided MPSS profiles of gene expression in inbred and hybrid larval oysters produced by a factorial cross of inbred lines 35 and 51. These profiles quantify genomic expression with great depth, to the equivalent of a few mRNA molecules per cell, for all expressed genes simultaneously (Jongeneel et al. 2003). Of the 4.6 million MPSS sequences of 52,828 unique signatures (Hedgecock et al. 2002), only ~9,100 signatures are present in all four families, a number that is intriguingly close to the estimate of ~8,500 genes expressed in the sea urchin embryo (Cameron et al. 2000) and the ~8,500 genes expressed in common by two human cell lines (Jongeneel et al. 2003).

From statistical contrasts among genotypes of MPSS expression data, we identify ~350 candidate heterosis genes for further genetic or functional analysis. Observed patterns of gene expression are more complex than predicted from the classic dominance and over-dominance explanations of heterosis (Gowen 1952, Crow 1998), in that hybrids show dominance for low expression and even under-expression. These expression patterns are consistent, on the other hand, with the metabolic efficiency hypothesis for growth heterosis (e.g., reduced rates of protein turnover in hybrid compared with inbred oysters; Hedgecock et al. 1996). Were a complete annotated genome sequence for C. gigas available, we could immediately identify 95% of these candidate genes (Brenner et al., 2000a; Jongeneel et al. 2003). A genomic sequence would thus greatly accelerate research into the physiologic mechanisms of heterosis, which are still not understood in crop plants, and speed the pace of discovery of genes and gene-regulatory elements affecting metabolic efficiency and growth, key components of fitness in the wild.

We are presently making QTL maps for yield heterosis in the Pacific oyster and preliminary results are promising (Hedgecock et al. 2004b). We aim not only to map heterosis QTL, but also to determine their mode of gene action, to test classic hypotheses about the genetic causes of heterosis (Crow 1998). We also seek to validate MPSS candidate genes, by mapping them to heterosis QTL, and are looking for cis- and trans-regulation of MPSS candidate gene expression, by following allele-specific expression using SNPs and quantitative PCR methods. However, understanding the physiologic mechanism and the functional basis of heterosis is a much more difficult task, requiring proof of function for every gene implicated by the MPSS approach. Genomic sequence would make this task easier by increasing the chances of identifying the MPSS candidates and their functions through searches of GenBank and other databases.

**Assessing Genomic Variability**

A genome sequence will provide means of sampling the oyster genome for nucleotide and haplotype polymorphism. An important scientific question is whether highly fecund animals, such as the purple sea urchin and Pacific oyster, share higher genomic variability than is typical of low fecundity animals.

**Ecological, Economic and Cultural Importance of the Oyster**

**Worldwide Production of Oysters**

The Pacific oyster has been introduced from Asia to all continents but Antarctica (Mann 1979) and for the past several years has had the highest annual production of any freshwater or marine organism (4.2 million metric tons, worth $3.5 billion US; FAO 2005). In the context of the ongoing programs for improving yield mentioned earlier, a genomic sequence will aid in gene discovery, permitting the quantitative and population genetic approaches presently used to reach their ultimate goal of understanding and utilizing genomic variation.

**Oysters As a Model for Marine Recruitment**

The oyster is a model organism for marine environmental science. It has the complex life history of most marine animals, the larvae of which develop for weeks in the plankton, are microscopic, weakly swimming and suffer high mortality, while being dispersed by ocean currents. Distance and direction of larval dispersal, factors dictating recruitment failure or success, and connectivity among adult populations are major concerns in marine fisheries management, ecology and conservation. Although the physical and biologic oceanography of recruitment are well studied, the endogenous genetic and physiologic components of larval fitness will receive attention in a project recently funded by Genomically Enabled Environmental Science (Gen-En) portion of the NSF Biocomplexity in the Environment program. This project extends experimental studies of oysters, which suggest that common assumptions about larval biology and recruitment need revision. Many oyster larvae may die, for example, not because they fail to encounter suitable environments or food during a critical period, but because they have genotypes that are not capable of developing and surviving in any environment (Launey & Hedgecock 2001). On the other hand, oyster larvae show genetically variable resilience to starvation (Moran & Manahan 2004, Yu & Manahan, unpublished), which is counter to classic ideas about a “critical period” for larval success. The Pacific oyster is poised for the application of genomic methods for identifying, quantifying and modeling mechanisms controlling the phenotypic variation at the heart of the marine recruitment problem. Critical experiments are now possible for these highly fecund animals because of the availability of inbred lines and F2 populations for quantitative genomic analysis and expression profiling using MPSS. Another advantage of oyster species is the ability to synthesize experimental data on endogenous sources of variation in larval fitness into a biologically explicit, individual-based model of larval population dynamics (Bochenek et al. 2001, Powell et al. 2002). Having a genome sequence for the Pacific oyster will help to infuse marine environmental science with a genomically enabled, Darwinian, perspective on individual differences, adaptation and evolution.

**Restoring Oysters and Oyster Habitats**

Nowhere is the ecologic, economic and cultural significance of oysters better illustrated than in the Chesapeake Bay, where the eastern oyster has declined to <1% of its original abundance, owing to a complex interaction of over-harvesting of oysters and shell reefs, anthropogenic impacts on the Bay, physical factors (low
rainfall causing increased salinity) and disease (parasitic infestation of oysters with *Perkinsus marinus* and *Haplosporidium nelsoni*). The collapse of the oyster industry in Maryland and Virginia has lead to contentious proposals to introduce a nonnative oyster as a solution to the environmental and economic crises (NRC 2004). In the meantime, loss of oysters means loss of their capacity to filter and to help control algal populations in coastal estuarine waters (Newell 1988, Wetz et al. 2002). Phenotypes of immediate interest in any attempt at restoration of the eastern oyster are disease and stress resistance, for which genomic information permits broader comparisons, benefiting from the rich literature and intense work on biomedical models.

The intense interest in oyster disease is evident in the focus of pilot EST collection programs on finding genes of importance in the response of oysters to stress and immune challenge (Jenny et al. 2002, Gueguen et al. 2003, Tanguy et al. 2004, Peatman et al. in press). The first cDNA-based microarray for oysters has genes from both species of oyster, as well as genes from the oyster parasite *Perkinsus marinus*. The availability of genomic sequence from the closely related *C. gigas* would greatly enhance ongoing research into the structure, diversity and function of genes encoding chaperones, such as Hsp70 (Clegg et al. 1998, Hamdoun et al. 2003), metallothionein genes (Jenny et al. 2004), genes that encode antimicrobial peptides and disease-resistance QTL (Tanguy et al. 2004). The comparison of Pacific and eastern oysters is especially interesting, because *C. gigas* tolerates diseases that kill *C. virginica* (Mann et al. 1991, Meyers et al. 1991, Barber & Mann 1994, Calvo et al. 1999).

The difference in infectivity and pathogenicity of *P. marinus* between the eastern and Asian oyster species is under scrutiny by several laboratories, aided by the development of *in vitro* culture methods for this parasite (Gauthier & Vasta 1993, Kleinschuster & Swink 1993, La Peyre et al. 1993). Challenge experiments demonstrate differences in *P. marinus*-host recognition or in the antimicrobial activity of hemocytes from different oyster species (Gauthier & Vasta 2002). Motility and respiratory burst activity of hemocytes are likewise affected in a species-specific manner by *P. marinus* (Garreis et al. 1999, Anderson 1999). Parasite expression of lower molecular weight proteases is inhibited by oyster homologues (Armstrong & Quigley 1992, Bender et al. 1992, Thogersen et al. 1992, Elsayed et al. 1999), protect against entry of protozoan parasites (Fuller & McDougald 1990, Polanski & Wilusz 1996). *C. virginica* has antiproteases with specific activity against *P. marinus* (Faisal et al. 1998, Oliver et al. 1999), levels of which positively correlate with differential survival across families (Oliver et al. 2000, Romestand et al. 2002), simultaneously demonstrating a heritable basis for disease resistance. *C. gigas* possesses protease inhibitors with significantly higher specific activities than those in *C. virginica* (Faisal et al. 1999). Study of the interaction between oyster serum and parasite proteases is providing insights into the role of proteases in pathogenesis (Muñoz et al. 2003).

Despite these advances, little is known about pathogenic mechanisms and host-parasite relationships at the molecular level. Unfortunately, continuous cultures of molluscan cell lines have not been established, despite attempts by several researchers, complicating study of host immune function and determination of the molecular basis for tolerance or resistance of different *Crassostrea* species and strains. Genomic information on the genus *Crassostrea* will facilitate identification and characterization of the genes involved in the immunologic response of the oyster hosts in the presence of these protozoan parasites. This information could be used to help predict relative disease tolerance of different oyster stocks and to facilitate optimization of selective breeding strategies.

Although the Pacific oyster tolerates or resists the parasites that kill eastern oysters, they do succumb to mass summer mortalities wherever this oyster is cultivated, including western North America (Cheney et al. 2000), Japan (Goulletquer et al. 1998) and France (Gauthier & Vasta 1998, Oliver et al. 1999), levels of which positively correlate with disease resistance (Mann et al. 1991, Meyers et al. 1991, Barber & Mann 1994). The coastal zone of the United States is home to a significant number of harmful algal species in the genera *Alexandrium*, *Dinophysys* and *Karenia* produce highly toxic environmental chemicals. Infectious agents and harmful algae are concentrated by filter-feeding shellfish, and the consumption of

Marine Environment and Human Health

The coastal zone of the United States is home to a significant and rising proportion of the population, and the pressures of residential, industrial and tourism-related development have resulted in degradation of the coastal marine environment. Numerous threats to human health arise from the marine environment, including infectious diseases and harmful algal blooms (NRC 1999). Infectious agents include viruses, such as hepatitis A and the caliciviruses, including the Norwalk virus, and bacteria, such as *Vibrio parahaemolyticus*, *V. vulnificus*, *Escherichia coli*, *Salmonella* spp. and *Shigella* spp. Harmful algae, such as some species in the genera *Alexandrium*, *Dinophysys* and *Karenia* produce highly toxic environmental chemicals. Infectious agents and harmful algae are concentrated by filter-feeding shellfish, and the consumption of
shellfish, including oysters, is thus an important mode of transmission of infectious disease and paralytic shellfish poisoning, amnesic shellfish poisoning, neurotoxic shellfish poisoning and diarrhetic shellfish poisoning (NRC 1999). Oysters (and other molluscs implicated in both infectious disease and shellfish poisoning) are not just passive vectors of the agents that they accumulate while feeding. Their relationship to these organisms is dynamic and intimately connected with the overall health of the marine coastal environment.

The relationship between oysters, their physical environment, human pathogens, oyster pathogens and the ecology of the coastal environment is complex; dissection of this relationship is being undertaken by many investigators, using not only traditional cellular and biochemical studies but also functional genomic approaches. It should be mentioned that oysters are also highly efficient bio-accumulators of toxic heavy metals such as lead and cadmium, which can lead to poisoning of humans who consume oysters from contaminated sites. Although this threat to human health has been substantially reduced by environmental standards in most developed nations, the genetics and biochemistry of oyster uptake and response to heavy metals exposure is an active area of research (e.g., Tanguy & Moraga 2001, Tanguy et al. 2001, Boutet et al. 2004). It should also be mentioned that oysters are prodigious preciptiators of carbonate and provide a metazoan model for studying bio-mineralization (Mount et al., 2004) and its impacts on carbon cycling.

Broader Impacts of Sequencing a Cultural Icon

An oyster is “something from which profit or advantage can be extracted [the world is his oyster]” (Webster’s New World Dictionary of the American Language). Much besides the scientific information summarized here could be gained by sequencing an oyster genome. Oysters hold a revered place in literature and gastronomy (Clark 1964, Fisher 1988). The oyster provides a natural vehicle for communicating to a broader lay audience the excitement of genomics and how genome sequences can lead to increased understanding, appreciation and wise use of biodiversity.

CONCLUSION AND RECOMMENDATIONS

The Oyster Genome Consortium, a nonexclusive entity open to any scientist with a bona fide interest in the application of genomics to understanding shellfish biology, proposes the Pacific oyster as a candidate for whole genome sequencing. The competition for access to sequencing facilities and resources is intense, so the shellfish community must support its best candidate for whole genome sequencing. The OGC believes that the best case can be made for the Pacific oyster. The genome of the Pacific oyster is among the smallest of bivalve genomes and is therefore a tractable genome to sequence. Several critical resources are available to aid in the genome sequencing effort, including inbred lines, whose reduced level of nucleotide polymorphism will help a shotgun sequencing effort, segregating F2 populations and moderately dense genomic linkage maps for QTL mapping, a BAC library and thousands of ESTs.

The OGC is well positioned to use a genome sequence for the Pacific oyster in 3 main areas of research, comparative genomics, evolutionary biology and understanding the genetic, physiologic and immunologic bases of adaptation. Advances in these areas are likely to have profound impacts on the worldwide culture of this species, which has had the highest production of any aquatic organism since 1998. A complete genome sequence for the oyster will add a bivalve to the small but growing list of sequenced Lophotrochozoan genomes. A genome sequence would also provide means for dissecting the genetic basis of reproductive isolation in these marine animals. Finally, the progress in understanding the genetic basis of variation in growth and survival that will result from having a genome sequence will also accelerate the pace of discovery of genes important in the adaptation and evolution of other bivalves and fucund marine species in general.

In addition to inviting individual participation in the OGC, we call on the National Shellfisheries Association and other scientific societies with interest in shellfish biology to become proactively involved in supporting and participating in genome sequencing efforts. Organized community support and effort are required to bring shellfish biology into the genomic and postgenomic era of biology.

LITERATURE CITED


in oyster populations and DNA markers. Available at: http://www.intlpag.org/pag/6/abstracts/shimizu.html.


