



Genomics-enabled research in marine ecology: challenges, risks and pay-offs

Gretchen E. Hofmann*, Sean P. Place

Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, California 93106-9610, USA

ABSTRACT: Genomics-enabled applications are becoming increasingly common in conjunction with research in marine ecology. In this Theme Section, we review the success of cases where techniques used to profile gene expression have been used to gain new insight into 3 areas of research: symbioses in marine invertebrates, physiological responses to environmental conditions, and examining the determinants of species-range boundaries in marine ecosystems. In addition, we briefly discuss the challenges facing new practitioners of these techniques, including an overview of essential equipment to conduct research in ecological genomics.

KEY WORDS: DNA microarrays · Gene expression · Genomics · Marine ecology · Transcriptomics

Resale or republication not permitted without written consent of the publisher

INTRODUCTION

Numerous recent reviews have highlighted the increased use of genomics-based techniques to address central questions in ecology and evolution (Gibson 2002, Jackson et al. 2002, Feder & Mitchell-Olds 2003, Klaper & Thomas 2004, Ranz & Machado 2006, van Straalen & Roelofs 2006). Increasingly, these approaches are being applied in marine systems with outstanding results (Hofmann et al. 2005, Wilson et al. 2005). The goal of the present study is to highlight the costs and benefits of genomics-based approaches in marine ecology, concentrating on the practical aspects and the hurdles that are faced by a new practitioner of these molecular tools. In addition, we have strategically chosen examples and particular techniques that study the transcriptome (i.e. a collection of all mRNAs in a cell) as a means to focus the article, illustrate the power of this approach, and to highlight the advances made using techniques that profile patterns of gene expression (Schena et al. 1995, Gracey & Cossins 2003, Allison et al. 2006). It should be noted that there are numerous other examples within marine ecological genomics (Venter et al. 2004), and other important analytical tools (Marsh & Fielman 2005). Broadly focused discussions of genomics specific to ecological

interests can be found in van Straalen et al. (2006) and recent reviews (Gibson 2002, Thomas & Klaper 2004).

IT'S WORTH THE EFFORT: INSIGHTS INTO BASIC AND APPLIED PROCESSES

For those considering entry into the genomics fray, a frequently asked question is: Are the costs of these technologies really worth it in proportion to what is being learned? In short, we believe the answer is an emphatic 'Yes'. The best way to illustrate this is by example. Thus, we have chosen studies of marine organisms that have used various methods (e.g. quantitative realtime PCR and cDNA [complementary DNA] macro- and microarrays) to profile gene expression in an ecological context (Appendix 1). Each of the examples share 2 characteristics: they each use a non-model marine organism, and they have each made leaps ahead in their respective fields owing to the insight generated by examining mRNA expression in their respective experimental systems. These examples include: (1) studying the mechanisms involved in important species interactions; in this case, in cnidarian-dinoflagellate symbiosis; (2) assessing the physiology of individuals such as the stress response; and

*Email: hofmann@lifesci.ucsb.edu

(3) examining expression across environmental gradients and linking physiological responses to large-scale ecological processes such as the determinants of biogeographic ranges in marine organisms.

Species interactions: exploring the cnidarian-algal symbioses

Some of the more compelling stories resulting from the use of DNA microarrays in marine ecology are the gene expression studies that identify genes involved in the mutualistic relationship between cnidarians and their intracellular algal symbionts. The outcome of this research illustrates the power of the 'discovery' process in genomics: results contribute to what is already known, and can often provide novel insight into important mechanisms. A recent study on the sea anemone *Anthopleura elegantissima* is exemplary of the utility of a genomics approach: Weis and colleagues demonstrated that the mechanisms involved in the maintenance and regulation of the relationship is perhaps more complex than might be expected (Rodríguez-Lanetty et al. 2006). In a comparison of symbiotic and aposymbiotic anemones, investigators found that genes from numerous metabolic processes displayed variation (Rodríguez-Lanetty et al. 2006). Using DNA microarray-based transcriptome analysis, 28 host genes were shown to vary in the symbiotic state; of these 28, functional-group analysis indicated that the results were underscoring that symbiosis had a more global effect on the host metabolism, rather than revealing a suite of genes unique to the symbiotic state (Rodríguez-Lanetty et al. 2006). In the supporting category, genes involved in lipid metabolism changed in a predictive fashion (i.e. some synthetic enzymes were down-regulated, and degradative ones were up-regulated). In the novel category, the study provided unprecedented insight into how apoptosis and cell-cycle genes may be related to maintaining the symbiosis by controlling the life of the host cell, something that investigators had observed in other symbioses, but was very new evidence in the cnidarian system.

Similarly powerful tools to assess the interaction of the host invertebrate and the algal symbiont are being built through ongoing efforts with coral genomics. Given the recent observations regarding changes in the strain of *Symbiodinium* that correlated with environmental conditions (Baker 2003, Rowan 2004), there is increasing evidence that the flexibility of the host-symbiont combination may be subject to environmental regulation. In order to characterize the nature of the symbiosis, Medina and colleagues have been constructing cDNA libraries for different stages of 2 important Caribbean coral species, *Acropora palmata*

and *Monastrea faveolata*, and eventually will use these to assess gene expression that is linked to the symbiosis in stages ranging from eggs to adults in colonies (Schwarz et al. 2006). Similarly, a recent annotated cDNA library for the squid-*Vibrio* symbioses will facilitate research on this invertebrate-bacterial system (Chun et al. 2006). These efforts will certainly pay off enormously for investigators, providing new foundational data that can be used to form hypotheses about topics ranging from how the symbiosis is established to what modulators may be regulating the presence of the intracellular symbionts.

Organismal-level studies: individual performance and stress responses

Another emerging application for microarray-based transcript profiling is being found in studies that address organismal physiology. Here, suites of differentially regulated genes provide 'physiological fingerprints' of an organism's response to changes in abiotic conditions, especially with respect to stress. Successfully demonstrated in model organisms such as yeast (Gasch et al. 2000) and *Arabidopsis* (Seki et al. 2002, Rizhsky et al. 2004), this approach has been applied to marine organisms to assess response to short-duration changes in temperature (Gracey et al. 2001, Podrabsky & Somero 2004, Buckley et al. 2006), to disease (Dhar et al. 2003), and in studies that use organisms as biosensors in response to toxins (Klaper & Thomas 2004, Almeida et al. 2005, Dondero et al. 2006). Although the cost of transcriptomics has been called into question for the assessment of stress in some cases (Feder & Walser 2005), these techniques are appropriate for those seeking a deeper understanding of mechanisms because such studies can provide the foundations for future hypothesis testing. Especially for investigators interested in physiological or cellular responses, gene-expression profiling can be very informative and impart information about single gene families (Jenny et al. 2004), or provide insight into patterns of expression in specific biochemical pathways (Gracey et al. 2004, Buckley et al. 2006).

In addition to stress responses, microarray applications are being applied to basic questions such as how a particular genotype leads to the phenotype—i.e. inter-individual variation in natural populations (Oleksiak et al. 2002)—and, interestingly, how gene regulation contributes to this process (Ranz & Machado 2006). For example, researchers working on killifish have shown distinct differences in gene expression in individuals with varying performance abilities (Oleksiak et al. 2005). These data are intriguing in that they suggest that physiological performance is multifactor-

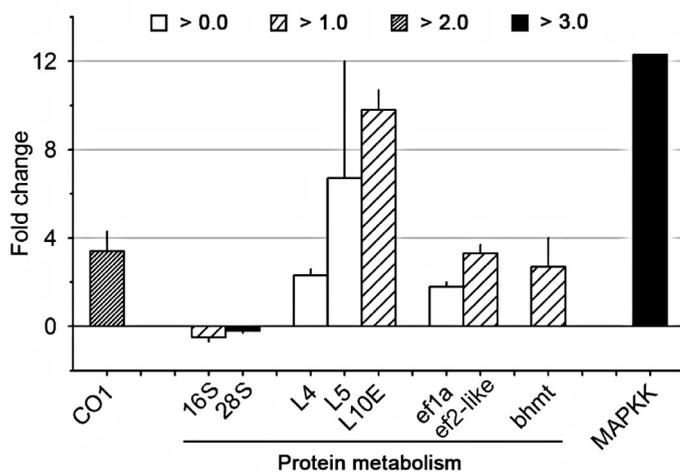


Fig. 1. Differences in gene expression detected by macroarray analysis of *Strongylocentrotus purpuratus* early, 4-arm echinoplutei following a 15 to 25°C transition (K. T. Fielman & G. E. Hofmann unpubl. results). Control (15°C) and treatment (25°C) larvae were held at their exposure temperatures for 24 h prior to RNA extraction. Changes in 3072 gene transcripts were evaluated using macroarrays and software provided by the Sea Urchin Genome Project. Fold change in gene expression as a function of temperature increase is shown on the y-axis for 10 genes: cytochrome oxidase 1 (CO1), mitochondrial ribosomal RNA (16S), cytosolic ribosomal RNA (28S), ribosomal proteins (L4, L5, L10E), translation elongation factors (ef1a and ef2-like), betaine-homocysteine methyltransferase (bhmt), and a mitogen activated protein kinase kinase (MAPKK). Bars indicate the magnitude of change based on locally normalized data. This macroarray-based transcriptional profiling illustrates the response among larval gene groups important in aerobic respiration (CO1), protein metabolism, and cell signaling (MAPKK).

ial, the sum of perhaps subtle changes in numerous metabolic pathways. Furthermore, for the genomics practitioner, such differences among individual specimens must be considered in experimental designs.

Exploring environmental gradients: biogeographic range patterns in marine populations

A central aim in ecology is to determine the processes and mechanisms that set species biogeographic range boundaries (Gaston 2003), and genomics-enabled techniques are contributing to answers for these types of questions. Increasingly, marine ecologists are interested in understanding the physiological state of an organism across its range (Partensky et al. 1999, Somero 2005, Sorte & Hofmann 2005, Sagarin & Somero 2006, Stillman et al. 2006, Osovitz & Hofmann 2007). Genomic approaches, particularly the use of gene-expression profiling, have a great potential to help guide this discussion by providing insight into organismal performance across a variety of spatial scales.

Thus, perhaps one of the greatest utilities of genomic techniques is the illumination of the response to, and thus role of, temperature in effecting organismal distribution, an often complex response that is difficult to comprehensively quantify. Here, genomics-enabled techniques have been used successfully to examine 2 ecologically significant processes that contribute to species distribution: thermotolerance and dispersal. First, as a functional example, a study profiling gene expression in larvae of the purple sea urchin *Strongylocentrotus purpuratus* found that the expression of genes involved in protein metabolism and cell signaling was strongly affected by high temperature stress (Fig. 1). Results such as these contribute insight into dispersal recruitment processes in marine invertebrates, especially the presumptive role of temperature (Gaylord & Gaines 2000). Also in the dispersal category is the routine application of genomics approaches in microbial ecology (Zhou 2003). In the marine environment, using quantitative real time PCR (qPCR), oceanographers have identified variation in the distribution of *Prochlorococcus* spp. ecotypes that strongly correlated with temperature as a function of depth in the Atlantic Ocean (Johnson et al. 2006). Importantly, both of these examples highlight significant advances made by the application of genomic-scale tools to the assessment of overall patterns of gene expression.

LIMITATIONS AND CHALLENGES

Although divergent in focus, the examples of success given above all share a similar subtext: each project was faced with a series of challenges related to working on a non-model organism with emerging technology in a discipline to which genomic technology was largely foreign. These challenges are worth discussing in detail because the bar is moving in terms of how great these hurdles may be, and to what extent they can be ameliorated within marine ecology.

There are several challenges for the marine ecologist making the leap into functional genomics, each having a unique timeframe for resolution. Obviously, funding is a major hurdle. These techniques are expensive, requiring access to costly equipment, and involve the work of experienced researchers for whom salaries are required. The real-world consequence of this situation is that collaborations with colleagues that have expertise and resources in molecular biology are essential. One of the biggest initial decisions is how to partition the work: how much can your own group do, and what proportion of the work should be performed 'out of house'? In many cases, someone with minimal training in molecular biology can handle the basics of preparing cDNA libraries and DNA sequencing; in contrast,

printing and scanning arrays may require collaboration with the neuroscientist down the hall, in another department of even a separate academic or commercial institution.

Another major hurdle, and perhaps one of the more intractable ones, is access to equipment. Appendix 2 describes some of the basic equipment, reagents, and personnel associated with the methods of transcriptome analysis. No matter which method is employed, any laboratory undertaking a functional genomics project needs a fundamental set of equipment (e.g. thermal cycler, electrophoresis equipment, spectrophotometers) to perform basic molecular work in-house. While much of the smaller equipment purchases inevitably fall within an individual laboratory's budget, many departments have been successful in securing funds for core facilities that house and/or operate many of the specialty equipment items listed in Appendix 2. In addition, many processes that require the most expensive tools (sequencing, array construction, DNA library construction, etc.) can often be contracted out to genomic facilities for less than it would cost in-house when personnel, reagents, service contacts for major equipment, and time to successful completion are tallied. Ultimately, these techniques may prove too expensive for widespread use, i.e. the development of a comprehensive cDNA or EST library for all species of interest, but alternative techniques such as qPCR could still be applied (Osovitz & Hofmann 2005). Although restricted to a specific set of *a priori* selected genes, qPCR has a significant advantage over the highly complex methods mentioned previously. Since the construction of a comprehensive cDNA library for the study organism is not a prerequisite, equipment and reagent costs are comparable with routine molecular applications. However, research on non-model organisms remains significantly hindered by the lack of readily available sequence information.

Perhaps one of the most daunting hurdles for new investigators in the field of functional genomics is related to bioinformatics: namely, how to mine the voluminous raw data once it is in hand, and how to best analyze large data sets. Fortunately, the expense of this hurdle is not much more than the usual software bundle, and programs are often freely distributed via the Internet. Still, the learning curve associated with present-day bioinformatics analysis software can be steep, even with regard to the current trend of moving from platform- and programming language-specific command-line execution to platform-independent, user-friendly interfaces. Hundreds of analysis programs are available for array data alone, all of which will allow some degree of normalization, clustering, and hierarchical analysis of the raw data. For ecologists familiar with modeling algorithms, this may prove to

be no hurdle at all because they may quickly adjust to the new, yet familiar parameters and analysis approaches.

IS THERE ANY GOOD NEWS?

In summary, and in our opinion, there is indeed good news in this arena. As described herein, exciting science is being conducted at the interface of genomics and marine ecology. Owing to the redundancy of the basic molecular biology techniques, many young investigators versed in molecular ecology have the tools with which to conduct these experiments. In addition, from a resource perspective, the printing of microarrays is becoming more accessible as more central facilities acquire the printers and microarray scanners. Individual principal investigators are also finding that prices for these major pieces of equipment are reducing and are affordable for a single laboratory group.

Cross-species hybridizations are increasingly commonplace and seem to be yielding reliable results (Ji et al. 2004, Renn et al. 2004). Although initially received with much skepticism, cross-species hybridizations have begun to gain acceptance within the comparative community. The feasibility of obtaining biologically meaningful data has started to be systematically assessed with favorable results. For example, using a microarray derived from the cDNA library of a cichlid fish, Renn et al. (2004) conducted heterologous hybridizations with several divergent fish species, showing that consistent expression profiles can be achieved for species that diverged as long as 65 million years ago. Combined with the efforts of genome projects currently underway for species central to ecological and evolutionary studies, these results show great promise for the application of molecular-based approaches to the elucidation of complex phenotypes.

Overall, these findings indicate that continued cooperation among colleagues within these fields will facilitate the use of genomic approaches. Increased cooperativity is already apparent within the marine and aquatic biology communities. Consortia such as the Marine Genomics group at the University of South Carolina (McKillen et al. 2005) and the consortium for Genomics Research on All Salmon (GRASP) exemplify the teamwork, cooperativity, and resource sharing that will ensure continued success in marine ecological and environmental genomics. Also, groups interested in genomics in specific biogeographic regions (Clark et al. 2004, Schwarz et al. 2006) are moving towards important shared resources for the marine community. Finally, governmental support for marine genomics is rapidly emerging. For example, in the USA, the Joint

Genome Institute (JGI) has assisted with sequencing large numbers of genes from non-model marine organisms (e.g. coral, intertidal mussels, and crabs). Yet, strong funding support for young, interdisciplinary investigators who will train the next generation of researchers in marine genomics is warranted.

Acknowledgements. We thank Dr. Kevin Fielman and Dr. Anne Todgham for helpful conversations and editorial comments that improved the manuscript. We acknowledge the US National Science Foundation for financial support during the course of writing (NSF grants OCE-0425107 and ANT-0440799 to GEH). This is Contribution No. 225 from PISCO (the Partnership for Interdisciplinary Studies of Coastal Oceans) funded primarily by the Gordon and Betty Moore Foundation and David and Lucile Packard Foundation.

LITERATURE CITED

- Allison DB, Cui X, Page GP, Sabripour M (2006) Microarray data analysis: from disarray to consolidation and consensus. *Nature Rev Genet* 7:55
- Almeida JS, McKillen DJ, Chen YA, Gross PS, Chapman RW, Warr G (2005) Design and calibration of microarrays as universal transcriptomic environmental biosensors. *Comp Funct Genom* 6:132–137
- Baker AC (2003) Flexibility and specificity in coral-algal symbiosis: diversity, ecology, and biogeography of *Symbiodinium*. *Annu Rev Ecol Syst* 34:661–689
- Buckley BA, Gracey AY, Somero GN (2006) The cellular response to heat stress in the goby *Gillichthys mirabilis*: a cDNA microarray and protein-level analysis. *J Exp Biol* 209:2660–2677
- Chun CK, Scheetz TE, Bonaldo MF, Brown B and 20 others (2006) An annotated cDNA library of juvenile *Euprymna scolopes* with and without colonization by the symbiont *Vibrio fischeri*. *BioMed Central (BMC) Genom* 7:154
- Clark MS, Clarke A, Cockell CS (2004) Antarctic genomics. *Comp Funct Genom* 5:230–238
- Dhar AK, Dettori A, Roux MM, Klimpel KR, Read B (2003) Identification of differentially expressed genes in shrimp (*Penaeus stylirostris*) infected with white spot syndrome virus by cDNA microarrays. *Arch Virol* 148:2381–2396
- Dondero F, Piacentini L, Marsano F, Rebelo M, Vergani L, Venier P, Viarengo A (2006) Gene transcription profiling in pollutant exposed mussels (*Mytilus* spp.) using a new low-density oligonucleotide microarray. *Gene* 376:24
- Feder ME, Mitchell-Olds T (2003) Evolutionary and ecological functional genomics. *Nature Rev Genet* 4:649–655
- Feder ME, Walsler JC (2005) The biological limitations of transcriptomics in elucidating stress and stress responses. *J Evol Biol* 18:901–910
- Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* 11:4241–4257
- Gaston KJ (2003) The structure and dynamics of geographic ranges. Oxford University Press, Oxford
- Gaylord B, Gaines SD (2000) Temperature or transport? Range limits in marine species mediated solely by flow. *Am Nat* 155:769–789
- Gibson G (2002) Microarrays in ecology and evolution: a preview. *Mol Ecol* 11:17–24
- Gracey AY, Cossins AR (2003) Application of microarray technology in environmental and comparative physiology. *Annu Rev Physiol* 65:231–259
- Gracey AY, Troll JV, Somero GN (2001) Hypoxia-induced gene expression profiling in the euryoxic fish *Gillichthys mirabilis*. *Proc Natl Acad Sci USA* 98:1993–1998
- Gracey AY, Fraser EJ, Li W, Fang Y, Taylor RR, Rogers J, Brass A, Cossins AR (2004) Coping with cold: an integrative, multitissue analysis of the transcriptome of a poikilothermic vertebrate. *Proc Natl Acad Sci USA* 101:16970–16975
- Hofmann GE, Burnaford JL, Fielman KT (2005) Genomics-fueled approaches to current challenges in marine ecology. *Trends Ecol Evol* 20:305–311
- Jackson RB, Linder CR, Lynch M, Purugganan M, Somerville S, Thayer SS (2002) Linking molecular insight and ecological research. *Trends Ecol Evol* 17:409–414
- 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
- Ji W, Zhou W, Gregg K, Yu N, Davis S, Davis S (2004) A method for cross-species gene expression analysis with high-density oligonucleotide arrays. *Nucleic Acids Res* 32:e93, doi:10.1093/nar/gnh084
- Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward EMS, Chisholm SW (2006) Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* 311:1737–1740
- Klaper R, Thomas MA (2004) At the crossroads of genomics and ecology: the promise of a canary on a chip. *BioScience* 54:403–412
- Marsh AG, Fielman KT (2005) Transcriptome profiling of individual larvae of two different developmental modes in the poecilognous polychaete *Streblospio benedicti* (Spionidae). *J Exp Zool* 304B:238–249
- McKillen D, Chen Y, Chen C, Jenny M and 7 others (2005) Marine genomics: a clearing-house for genomic and transcriptomic data of marine organisms. *BioMed Central (BMC) Genom* 6:34
- Oleksiak MF, Churchill GA, Crawford DL (2002) Variation in gene expression within and among natural populations. *Nat Genet* 32:261
- Oleksiak MF, Roach JL, Crawford DL (2005) Natural variation in cardiac metabolism and gene expression in *Fundulus heteroclitus*. *Nat Genet* 37:67
- Osovitz CJ, Hofmann GE (2005) Thermal history-dependent expression of the *hsp70* gene in purple sea urchins: biogeographic patterns and the effect of temperature acclimation. *J Exp Mar Biol Ecol* 327:134
- Osovitz CJ, Hofmann GE (2007) Marine macrophysiology: studying physiological variation across large spatial scales in marine systems. *Comp Biochem Physiol*
- Partensky F, Hess WR, Vault D (1999) *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* 63:106–127
- Podrabsky JE, Somero GN (2004) Changes in gene expression associated with acclimation to constant temperatures and fluctuating daily temperatures in an annual killifish *Austrofundulus limnaeus*. *J Exp Biol* 207:2237–2254
- Ranz JM, Machado CA (2006) Uncovering evolutionary patterns of gene expression using microarrays. *Trends Ecol Evol* 21:29–37
- Renn S, Aubin-Horth N, Hofmann H (2004) Biologically meaningful expression profiling across species using heterologous hybridization to a cDNA microarray. *BioMed Central (BMC) Genom* 5:42
- Rizhsky L, Liang H, Shuman J, Shulaev V, Davletova S, Mittler R (2004) When defense pathways collide. The

- response of *Arabidopsis* to a combination of drought and heat stress. *Plant Physiol* 134:1683–1696
- Rodriguez-Lanetty M, Phillips W, Weis V (2006) Transcriptome analysis of a cnidarian—dinoflagellate mutualism reveals complex modulation of host gene expression. *BioMed Central (BMC) Genom* 7:23
- Rowan R (2004) Coral bleaching: thermal adaptation in reef coral symbionts. *Nature* 430:742
- Sagarin RD, Somero GN (2006) Complex patterns of expression of heat-shock protein 70 across the southern biogeographical ranges of the intertidal mussel *Mytilus californianus* and snail *Nucella ostrina*. *J Biogeogr* 33:622–630
- Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467–470
- Schwarz J, Brokstein P, Manohar C, Coffroth MA, Szmant A, Medina M (2006) Coral reef genomics: developing tools for functional genomics of coral symbiosis. In: *Proc 10th Int Coral Reef Symp*. Japanese Coral Reef Society, Okinawa, p 274–281
- Seki M, Narusaka M, Ishida J, Nanjo T and 14 others (2002) Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J* 31:279–292
- Somero G (2005) Linking biogeography to physiology: evolutionary and acclimatory adjustments of thermal limits. *Front Zool* 2:1
- Sorte CJB, Hofmann GE (2005) Thermotolerance and heat-shock protein expression in Northeastern Pacific *Nucella* species with different biogeographical ranges. *Mar Biol* 146:985–993
- Stillman JH, Teranishi KS, Tagmount A, Lindquist EA, Brokstein PB (2006) Construction and characterization of EST libraries from the porcelain crab, *Petrolisthes cinctipes*. *Integr Comp Biol* 46:919–930
- Thomas MA, Klaper R (2004) Genomics for the ecological toolbox. *Trends Ecol Evol* 19:439–445
- van Straalen NM, Roelofs D (2006) Introduction to ecological genomics. Oxford University Press, Oxford
- Venter JC, Remington K, Heidelberg JF, Halpern AL and 19 others (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304:66–74
- Wilson K, Thorndyke M, Nilsen F, Rogers A, Martinez P (2005) Marine systems: moving into the genomics era. *PSZN I: Mar Ecol* 26:3–16
- Zhou J (2003) Microarrays for bacterial detection and microbial community analysis. *Curr Opin Microbiol* 6: 288

Appendix 1. Techniques used in analysis of the transcriptome

With the continuous refining of various methods used to profile gene expression, these techniques are becoming applied to non-model systems with ever greater numbers. Below we describe the more commonly used methods for analyzing the transcriptome

Quantitative real-time PCR (qPCR): a modification of the PCR in which cDNA is quantified after each round of amplification (real-time) as opposed to the end-point analysis of standard PCR reactions. Via reverse transcription, qPCR is used to quantify low abundance messenger RNA (mRNA), enabling a researcher to quantify relative gene expression at a particular time, or in a particular cell or tissue type. Two common methods of quantification involve the use of fluorescent dyes, which intercalate with double-stranded DNA, or modified DNA oligonucleotide probes that fluoresce when hybridized with complementary DNA. The amount of fluorescence emitted by these dyes is directly proportional to the number of amplicons produced in the reaction, and is measured by an optical module within a thermal cycler. Through the use of multiple dye combinations, researchers can ‘multiplex’ (monitor multiple genes in a single reaction); however, this technique requires a much greater understanding of the chemistry involved as well as more advanced real-time systems

Macroarray: a collection of DNA sequences, tens to thousands of bp in length, spotted onto a reusable membrane

(generally nylon). With a spot size of 300 μm , these arrays typically hold hundreds to thousands of features on a single membrane. Hybridization of the array with samples labeled with a fluorescent or radiolabeled reporter offers the simultaneous analysis of relative sequence abundance for thousands of genes in a sample. While offering the advantage of being reusable several times, analysis of the filters requires the use of a fluorescent scanner (e.g. Storm Phosphorimager, Molecular Dynamics) or the use of relabeled probes. In addition, owing to the size of the features spotted on the array, several membranes may be required to analyze the entire genome of an organism, entailing that the researcher generate more sample for each analysis

Microarray: similar to macroarrays, microarrays are a collection of single DNA sequences immobilized on a solid surface, in this case glass. Utilizing features of $\leq 200 \mu\text{m}$ in diameter, microarrays are printed in much greater densities, often containing tens to hundreds of thousands of genes on a single chip. By hybridizing microarrays with 2 alternatively labeled samples, researchers can measure the relative abundance of the entire transcriptome in a given sample. While allowing the entire genome to be printed on a single chip, these arrays are not reusable and require costly equipment such as robotic arrayers for printing and dual laser scanners for fluorescence capturing

Appendix 2. Basic equipment used in genomics research

Thermal cycler: thermal cyclers are used for PCR amplification of a specific DNA template. They use temperature-controlled blocks to cycle between programmed periods of DNA denaturation, primer annealing, and sequence elongation. Each 'cycle' of amplification results in exponential increases in the pool of the DNA sequence of interest. This particular piece of equipment is of use to any molecular application and is essential for the amplification of cDNA in preparation for microarray/macroarray printing as well as downstream sequencing applications

Real-time system: this system combines thermal cycling, fluorescence detection, and application-specific software to provide an integrated platform for the detection and quantification of nucleic acid sequences. These systems can become quite complicated with automated components for high-throughput applications; however, a standard 96-well compatible system will suffice for the most advanced ecological genomics laboratory. Reagents for real-time PCR can be a significant cost, requiring the regular purchase of fluorescent dyes, reverse transcriptase enzyme, *Taq* polymerase enzyme, and high purity oligonucleotide primers. The thermal cycler contained within these systems can often suffice for general thermal cycling applications as well

UV-spectrophotometer: used for the quantification of DNA/RNA concentration as well as the purity of a sample based on 260:280 nm ratios. These instruments range in price largely based on the number of samples (single samples up to 384 well plates) and the volume of sample used (1 µl up to 5 ml)

Horizontal gel electrophoresis: used for separation of DNA/RNA sequences, via agarose gels, based on fragment size for visualization, non-quantitative determination of concentration, molecular integrity, and even purification of fragments

cDNA library: a cDNA library refers to a complete, or nearly complete, set of all mRNAs contained within a cell or organism. Researchers use an enzyme called reverse transcriptase, which produces a DNA copy (cDNA) of each mRNA strand. These cDNAs are collectively known as the 'library'. Production of the library is a lengthy and often technically challenging endeavor. If not collaborating with a laboratory familiar with this process, all is not lost. Recently, several independent companies have begun to

create custom libraries, and a few even specialize in non-model organisms. The price for the custom production of a library has come down in recent years and one can expect to pay between US \$5000 and \$10 000, depending on the company and if you have the companies pick individual clone sets and print the clones on membranes in addition to producing the library

Dual laser microarray scanner: a confocal laser scanning device used to detect and quantify hybridization signals (532 and 635 nm), and specifically designed to scan DNA microarrays fabricated on glass slides using Cy3 and Cy5 fluorescent labels

Microarrayer: an environmentally controlled robotic printer capable of printing biological samples on standard-sized glass slides. Printing is from microtiter plates with 96 or 384 wells. Features are printed as 100 to 200 µm spots, size depending on tip type used; these robots can array hundreds of thousands of features per slide

DNA sequencer: automated processor used for determining the exact order of the bases A, T, C and G in a piece of DNA. The most commonly used method of sequencing DNA—the dideoxy or chain termination method is achieved by including in each reaction a nucleotide analogue that cannot be extended and thus acts as a chain terminator. In essence, DNA is used as a template to generate a set of fragments that differ in length from each other by a single base. Fragments are then separated by size, and the bases at the end are identified, recreating the original sequence of the DNA. The primers or nucleotides included in the reactions contain different fluorescent labels, allowing DNA strands terminating at each of the 4 bases to be identified. Reaction products are separated by gel electrophoresis. As the DNA strands pass a specific point, the fluorescent signal is detected and the base identified. Many outside sequencing facilities are available to researchers, and are often more cost effective when high-throughput (i.e library annotation) sequencing is not necessary; however, time can be lost due to sample delivery to these facilities

Bioinformatics: basic computing capabilities are necessary for all the methods described in the present study. Most data from these processes are produced in digital formats and therefore require software packs compatible with the output from each specific approach taken

Editorial responsibility: Howard Browman (Associate Editor-in-Chief), Storebø, Norway

*Submitted: August 21, 2006; Accepted: September 21, 2006
Proofs received from author(s): February 8, 2007*