



Marine ecological genomics: when genomics meets marine ecology

Samuel Dupont*, Karen Wilson, Mathias Obst, Helen Sköld, Hiroaki Nakano,
Michael C. Thorndyke

Kristineberg Marine Station, 566 Kristineberg, 45034 Fiskebäckskil, Sweden

ABSTRACT: Genomics, proteomics and metabolomics (the 'omic' technologies) have revolutionized the way we work and are able to think about working, and have opened up hitherto unimagined opportunities in all research fields. In marine ecology, while 'standard' molecular and genetic approaches are well known, the newer technologies are taking longer to make an impact. In this review we explore the potential and promise offered by genomics, genome technologies, expressed sequence tag (EST) collections, microarrays, proteomics and bar coding for modern marine ecology. Methods are succinctly presented with both benefits and limitations discussed. Through examples from the literature, we show how these tools can be used to answer fundamental ecological questions, e.g. 'what is the relationship between community structure and ecological function in ecosystems?'; 'how can a species and the phylogenetic relationship between taxa be identified?'; 'what are the factors responsible for the limits of the ecological niche?'; or 'what explains the variations in life-history patterns among species?' The impact of ecological ideas and concepts on genomic science is also discussed.

KEY WORDS: Sequencing · ESTs · Microarrays · Proteomics · Barcoding

Resale or republication not permitted without written consent of the publisher

INTRODUCTION

Genome-based technologies are revolutionizing our understanding of biology at all levels, from genes to ecosystems. Genomics is the study of the genomes of various organisms in their entirety, while genetics tends to study genes individually or in linked groups, relating DNA sequences to proteins and ultimately to heritable traits (Van Straalen & Roelofs 2006). The term 'genomics' appeared in the 1980s as the name of a new journal (McKusick & Ruddle 1987), but the genomics revolution really began in 1990 with the Human Genome Project and since then, thanks to rapid developments in molecular biology technologies, genomics-based discovery has grown exponentially. For example, the new sequencing system developed by Margulies et al. (2005) will be capable of sequencing 25 million bases in a 4 h-period — about 100 times faster than current state-of-the-art systems — with the same reliability and accuracy. The genomes of more than 300 organ-

isms have been sequenced and analyzed since the publication of the first complete genome in 1995, and today a new organism is sequenced nearly every week (Rogers & Venter 2005, Van Straalen & Roelofs 2006). The current challenge is no longer to collect sequence information but rather to analyze the data. Genomic approaches combine molecular biology with computing sciences, statistics and management. The intellectual infrastructure in genomics must be extended into bioinformatics (data storage and data query), computational biology (more complex, often hypothesis-driven analyses that may require the development of new algorithms and tools), and information technologies to share software and data.

Molecular ecology is a relatively new field in which techniques such as Polymerase Chain Reaction (PCR) and genetic engineering (recombinant DNA technology) has had an increasing role in the integration of genetic data with historical or field observations (White 1996). Through the study of single or small sub-

*Email: samuel.dupont@kmf.gu.se

sets of genes or small genomic regions (e.g. microsatellites), molecular ecology has been used to address classic questions in the areas of diversity, populations, and taxonomy. In contrast, the emerging field of ecological genomics is trying to answer larger ecological questions in areas such as nutrient cycling, population structure, life-history variations, trophic interactions, stress responses and ecological niches. Ecological genomics can be defined as 'the scientific discipline that studies the structure and functioning of a genome with the aim of understanding the relationship between the organism and its biotic and abiotic environments' (Van Straalen & Roelofs 2006). This new field crosses and interacts extensively with other disciplines such as microbiology, physiology, genetics and evolutionary biology. Ecological genomics investigates different levels of integration from the lower (functional mechanisms: physiology, biochemistry, cell biology, neuroscience, developmental biology etc.) to higher (ecology, evolution). The inclusion of 'function' is critical because the goal is to understand what genes/genomes and their variants do at higher levels of integration.

Marine ecological genomics is, then, the application of genomic sciences to attempt to understand the structure and function of marine ecosystems. Genomics provides biological information that is unobtainable by any other means, for example the biological capacities of marine organisms that underlie the ecology of oceanic ecosystems (see 'Genome sequencing: applications'). Approaches can include (1) whole genome sequencing of key organisms (e.g. genome comparison for phylogeny), or (2) genomic analysis of natural communities to understand how biodiversity supports ecosystem function (e.g. genomic analysis of microbial communities *in situ* with the concept of 'genome ecology', the collective genome in a given environment, also conceived as 'metagenomics'). For example, these approaches can be used to investigate life-history patterns (population ecology) and stress responses (physiological ecology).

Marine ecological genomics is a good example of a 21st century science that requires the mixing of scientific disciplines, hitherto historically and traditionally separated. Forging the link between marine ecologists, molecular biologists and genomics/bioinformatics scientists, and finding a common language, is a huge social challenge. The need for marine ecological genomics to be interdisciplinary is brought about by a number of factors including (1) the requirement for the use of highly specialized technologies, (2) the necessity for the development of new tools in key areas such as statistics and computational sciences, and (3) the lack of adequate funding for large-scale genome science research, especially in individual laboratories. As a

consequence, this field is not always fully amenable to the individual or individual research group, and it is often essential and more strategically viable to develop coordinated networks of collaborative interdisciplinary laboratories. Nevertheless, some techniques are more affordable than others. For example, expressed sequence tag (EST) libraries and microarrays limited to genes associated with a specific function, tissue or response pathway can be manufactured at relatively low costs for small research groups (Held et al. 2004).

With the exception of microbial ecology, genomic studies have until recently only been performed on a rather limited number of classic model species such as *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Mus musculus* or *Arabidopsis thaliana*. However, this is now changing and the number of new genomes is increasing (see www.ncbi.nlm.nih.gov, www.hgsc.bcm.tmc.edu/projects, <http://genome.jgi-psf.org>). The choice of the ideal model species for genomics is based on many practical (established reputation, genome size, possibility of genetic manipulation etc.) and scientific criteria (medical, biotechnological, agricultural or ecological significance, evolutionary position, comparative purpose, laboratory expertise etc.; see Feder & Mitchell-Olds 2003). This approach is not the tradition in ecology and there is a discrepancy between the available genomic models and ecologically interesting species. For example, *D. melanogaster* or *A. thaliana* are not sufficiently widespread in the environment and not very suitable for ecological studies. Moreover, no model is able to answer all questions. Consequently, the genomics revolution is the perfect time to move away from our fascination with model species, and the sequencing of the genomes of species such as *Ciona intestinalis* (<http://ghost.zool.kyoto-u.ac.jp/indexr1.html>, <http://genome.jgi-psf.org/Cioin2/Cioin2.home.html>) or the sea urchin *Strongylocentrotus purpuratus* (Sea Urchin Genome Consortium 2006, see also <http://sugp.caltech.edu>) as well as the amphioxus *Branchiostoma floridae* and the anemone *Nematostella* (www.jgi.doe.gov/sequencing/DOEprojseqplans.html) is the first step in this direction.

The marine ecology community must be prepared for the genomic era. The aim of this review is thus to explain general principles of the main genomic technologies and their applications to marine ecology with examples from the literature (for a more exhaustive presentation of genomic methods, see Van Straalen & Roelofs 2006). Genomic methods are succinctly presented with their strengths and limitations, and linked to marine ecological questions (Fig. 1). Marine ecological genomics is a new discipline merging genomics and marine ecology leading to new questions independent of both fields. Genomics is more than a toolbox

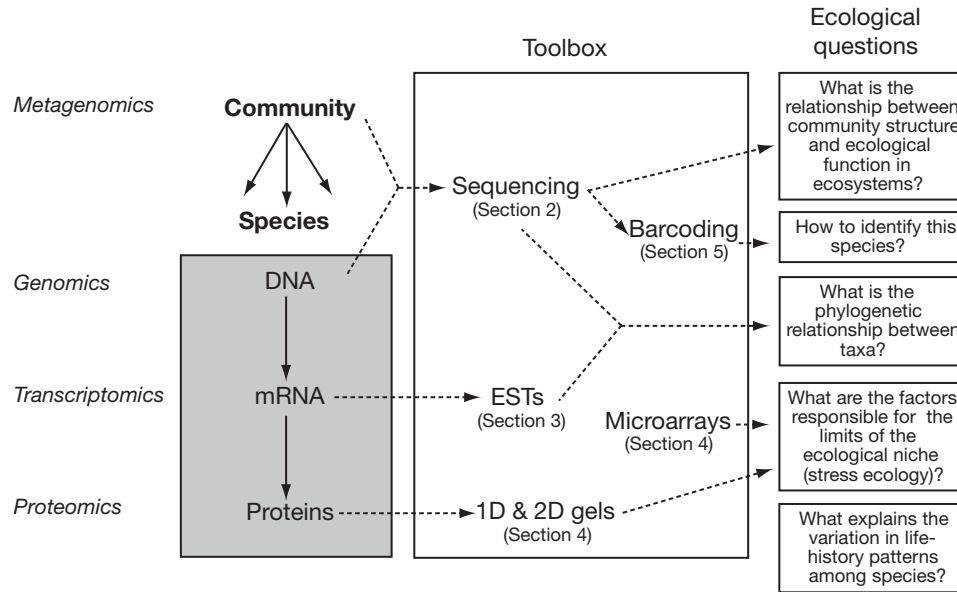


Fig. 1. Links between genomic tools and ecological questions

added to marine ecology; in the conclusion, therefore, some examples of feedback from ecology to genomics will be presented.

GENOME SEQUENCING

Method

Large-scale sequencing and first annotation is usually automated and based on the method initially developed by Sanger et al. (1977) and Smith et al. (1986). Nowadays, whole-genome sequencing is usually contracted out to commercial sequencing centres or is organised in collaborative networks comprising many different laboratories, often funded by national or international consortia (Van Straalen & Roelofs 2006). A complete description of methods and approaches is outside the scope of this article. For a list of websites and sequencing initiatives see 'Discussion'.

Applications

Environmental genome—microbial ecology

One of the most fundamental questions in community ecology is: what is the relationship between ecosystem processes and biodiversity? In other words, 'what do species do in ecosystems?' (Lawton 1994). In order to understand how biodiversity supports ecosystem function, it is necessary to estimate species diversity (richness, biomass, dominance structure, feeding

groups) and functions (production, respiration, degradation of organic matter, nitrification etc.). This is particularly difficult in marine microbial communities where it is not always clear what constitutes a microbial species and it is only possible to characterize species that can be cultured. Here, genomics provides a solution to the problem by reconstructing diversity and functions from the environmental genome (partial or whole sequence from 'environmental samples', i.e. DNA extracted from a seawater sample). The DNA of all species in a microbial environment can be assembled and functions characterized without attempting to put them into culture or separate them according to species. For example, the genome of the anammox bacterium *Kuenenia stuttgartiensis* was recently deduced from the DNA sequenced from a whole microbial community (Strous et al. 2006). This will enable insight into the metabolism and evolution of this bacterium, which is responsible for removing up to 50% of fixed nitrogen from the ocean. Using similar approaches, it is also possible to compare two communities, detect functional genes indicative of key steps in cycles (nitrogen, sulphur etc.) or reconstruct functions without the need for culture (Van Straalen & Roelofs 2006).

With an estimated 2 million species of bacteria in pelagic zones, a density of billions of cells per litre and a richness of 163 species per millilitre of ocean water (Curtis et al. 2002, DeLong & Karl 2005), microbes are major players in the structure and dynamics of marine ecosystems. It is of crucial importance to understand microbial roles in oceanic primary production, global carbon cycling and functioning of the biosphere.

Unfortunately, in the oceans, most microbes (>99%) resist efforts to grow them in pure culture. In consequence, very little is known about their physiology and their role in the environment. These organisms can be categorized into phylotypes using rRNA genes amplified from environmental DNA extracts; however, this does not reveal the physiology, biochemistry or ecological function of uncultured microbes (Giovannoni & Stingl 2005). Ecological genomics appears to be a new culture-independent tool with which to analyze microbial community structure and function in natural and engineered environments. Microbial communities can be explored by isolating large fragments of DNA directly from the environment, sequencing the fragments and assigning function to the genes based on their similarity to known genes or on functional studies. This process is referred to as community genomics or metagenomics. The recent genomic survey of the Sargasso Sea microbial assemblage is a perfect example. This led to 1.6 billion bp of genome sequence information and about 1.2 million genes identified from the collective microbial assemblage (Tringe et al. 2005). Such data frequently leads to the discovery of new genes (e.g. photorhodopsin, Béjà et al. 2000), gene functions, novel metabolic pathways, and other previously unknown properties of micro-organisms. These data can also shed light on physiological properties and ecological functions without consideration of species. Using such an approach, it is also possible to identify the genes and biochemical pathways that differentiate species living in different environments.

Microbial genomes are relatively small and allow rapid and relatively inexpensive sequence determination (Béjà 2004, Steele & Streit 2005). Cyanobacteria are a good example. They are amongst the most widespread and relevant organisms in marine habitats, and the genus *Prochlorococcus* has a key role in terms of global primary production (Hess 2004). The observation of the absence of the nitrate reductase gene from the *Prochlorococcus* genome changed the way we think about the ecological role of this organism in pelagic systems (García-Fernández et al. 2004). Genome sequencing of several biodegradation-relevant micro-organisms has provided the first whole-genome insights into the genetic background of the metabolic capability and biodegradative versatility of these organisms (Pieper et al. 2004).

Comprehensive approaches to describe and interpret oceanic microbial diversity and processes are only now emerging. Genomics applied to microbial ecology is significantly expanding our understanding of marine microbial evolution, metabolism and ecology. This new technology is revealing the links between evolutionary, ecological and biogeochemical processes in natural marine microbial communities (DeLong & Karl

2005). Genomics applied to microbial ecology is a striking example of true and successful marine ecological genomics that enhances our understanding of the living marine system, and that will lead to a new generation of more realistic oceanographic simulations, including improved climate change projections (Doney et al. 2004).

Comparing genomes—phylogenomics

Genomics has changed the way we define the term 'species'. Whole genome comparisons (size, G/C content, number of genes, gene distribution, sequence etc.) allow identification of core similarities and differences at each level of complexity. Whole genome comparisons for different strains suggest that polymorphism is common and in some cases reflects adaptability to different habitats. The genomic era is now providing the opportunity for phylogenetics to resolve a number of outstanding evolutionary questions through an increase of resolving power (Delsuc et al. 2005).

Despite extensive research, high-level phylogenetic relationships amongst animals remain contentious. Studies have been based upon several developmental, morphological and, more recently, molecular tools. Two main hypotheses are proposed (Fig. 2): (1) the Acoelomata-Pseudocoelomata-Coelomata (A-P-C) hypothesis, supported by morphological and whole-genome studies, divides animals according to the presence/absence of a coelom, lined (or not) by mesoderm; (2) the Lophotrochozoa-Ecdysozoa-Deuterostomia (L-E-D) hypothesis, supported by genetic studies, divides animals into Protostomia-Deuterostomia based

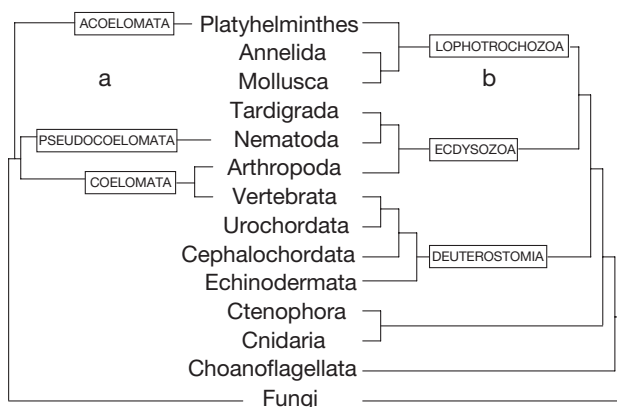


Fig. 2. Two main hypotheses of the relationship between animal phyla: the (a) Acoelomata-Pseudocoelomata-Coelomata phylogeny (A-P-C hypothesis) and (b) Lophotrochozoa-Ecdysozoa-Deuterostomia phylogeny (L-E-D hypothesis). From Jones & Blaxter (2005)

on origin of the mouth during development and the Lophotrochozoa-Ecdysozoa division of Protostomia based on moulting and presence of a lophophore. The use of molecular tools has resulted in some radical rearrangements of animal phyla. For example, phylogenetic analysis of 18S ribosomal DNA sequences supported the idea of Ecdysozoa (Winnepennincks et al. 1995, Aguinaldo et al. 1997, Adoutte et al. 1999, Peterson & Eernisse 2001). Nevertheless, analyses of whole genome sequences from a few species support older views (e.g. on human, fly, nematode and yeast by Mushegian et al. 1998, Blair et al. 2002, Wolf et al. 2004). These multigene analyses covered rather few taxa, and it is well known that the number of species represented in a phylogenetic study can induce systematic artefacts on tree reconstruction. For example, genome-scale analyses are especially sensitive to long-branch attraction (Felsenstein 1978). For these studies, the usual outgroup is yeast, very distantly related to animals and species such as the nematode *Caenorhabditis elegans*. Here, nematodes move to the base of the tree, generating support for the A-P-C hypothesis (Mushegian et al. 1998, Blair et al. 2002, Wolf et al. 2004). When analyses are set up to avoid long branch attraction, they do not support A-P-C hypothesis but rather L-E-D; for example, analyses of rare insertions and deletions of genomic features in some animal genomes (Roy & Gilbert 2005) or analysis of data from ESTs (see 'Expressed Sequence Tags' below) in addition to complete genome sequences (Philippe et al. 2005). This last study demonstrated that if only yeast is used as an outgroup, nematodes emerge at the base of the tree. However, by using outgroups closer to animals, nematodes cluster close to arthropods as predicted by the L-E-D hypothesis. This clustering is also improved when biased genes (those with greatest evolutionary rate in some species) are removed. Only 12 of the 35 animal phyla are currently represented in genomic studies and the use of genomics in phylogeny is still at its infancy. More genomes need to be sequenced and new analytical tools (e.g. algorithms and software) should be developed.

EXPRESSED SEQUENCE TAGS (EST)

Method

Development of an EST library is often the first step when starting a genomic project on a novel organism. Complete genome sequencing provides information about genome organization and promoter regions etc. It is, however, a major investment and not likely to be applied to the majority of organisms that are subjects for scientific investigation. In contrast, genes of an

increasing number of species are being investigated through generation of ESTs (www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html). ESTs are cost-effective and provide a rapid strategy with which to identify genes of the investigated organism. The sequence information contributes to the understanding of the dynamics of genome expression patterns and thereby to the understanding of the biology of the organism. ESTs can be used in expression profiling, evolutionary and taxonomy studies, systematics etc. It is important to remember that these are expressed sequences (RNA) and will vary according to the temporal and spatial (tissue/organ) origin of the cDNA.

ESTs are usually obtained by sequencing clones from a cDNA library and can be assembled into an EST database containing the fragments of the sequenced cDNAs. The cDNA library can be made by the individual researcher or commercially by various companies that offer such services. In most commercial kits for library construction, it is possible to make several libraries using the same kit. Commonly, about 1 g of tissue or 1 mg of total RNA is used for a standard-sized library, a factor important to bear in mind if working with limited amounts of material. A subtraction library can be made by removing identical genes present in 2 libraries from different conditions, and here you produce a library containing genes differentially expressed according to the 2 conditions chosen (for example 2 temperatures, pH etc.). To produce an ordinary but enriched library, it is possible to use organisms from particular environmental conditions in order to enrich transcripts induced by that particular treatment. For example, Kore-eda et al. (2004) analysed the profile of differentially expressed genes of well-watered and salinity-stressed specimens from the common ice plant *Mesembryanthemum crystallinum*. The same number (2782) of ESTs from each library (total = 8346 ESTs) were randomly selected and analysed. Their result showed differential expression of known genes related to stress responses, and also of novel and/or functionally unknown genes that may have a novel role in the salinity stress response. A similar approach using a subtractive hybridisation library has been used successfully to analyse hierarchical behaviour in rainbow trout (Sneddon et al. 2005). A fascinating use of this comparative approach is evident from the work of Kuo et al. (2004), who constructed 2 complementary DNA (cDNA) libraries from RNA isolated from symbiotic and aposymbiotic *Aiptasia pulchella* in order to understand algal-cnidarian interactions. Their systematic analysis of these ESTs provides a useful database containing numerous putative candidate genes for further investigations.

Functional annotation of ESTs from ordinary cDNA libraries by basic local alignment search tool (BLAST)

comparisons commonly identify unique sequences that share significant similarities to nucleotide or amino acid sequences of genes with known as well as unknown functions. In addition, relatively large numbers of ESTs often do not significantly match any genes in public databases. These may represent previously unidentified genes. Typically, a subsequent clustering analysis will further reveal higher expression of ribosomal genes and genes coding for metabolic pathway proteins, structural proteins, cell cycle proteins and proteins involved in cellular defence and stress responses (Ogasawara et al. 2002, Hackett et al. 2005, Watanabe et al. 2005, Simon et al. 2006). Genes involved in such processes or other highly expressed genes are likely to appear after sequencing about 1000 clones from a non-normalized cDNA library. Generally, high levels of expression indicate an important function in the organism. It may require more sequencing to obtain low expression genes or genes expressed only in a critical period, for example transcription factors. However, the highly expressed genes involved in energy metabolism, cellular defence and stress responses are important for homeostasis, and are thus potential candidates for sublethal markers against environmental stress and xenobiotics.

Applications

EST data contributes to the understanding of functional genes and gene networks, and has also been used for identification of non-protein coding mRNA with putative functions (Hirsch et al. 2006). Publicly available ESTs have also been used for subsequent novel phylogenetic analyses for species and groups (see subsection 'Genome sequencing: comparing genomes—phylogenomics'). Analysis of ESTs can also reveal the presence of microsatellite-containing genes, single nucleotide polymorphisms (SNPs), and other populational markers (Chen et al. 2006). Overall, the cDNA clones and EST sequence information (www.ncbi.nlm.nih.gov) are very useful for post-genomic functional analyses of the biology of the organism and for investigating links between evolution, ecology, physiology, genes and proteins.

Sequence information from an EST database can subsequently be used to quantify mRNA expression in a more focused experiment by gene-specific RT-PCR or other methods such as Northern blotting. *In situ* hybridization and antibody labelling can also reveal where in the organism the particular genes and proteins are expressed in both tissue sections and in whole animals. Identified genes coding for enzymes can, for example, be tested as putative novel biomarkers useful for simple enzyme activity based assays at the protein level.

The sequences can also be used for subsequent high throughput micro- or macroarray approaches (see subsection 'Microarrays/proteomics'), where clones or synthesized DNA oligos are arrayed for high throughput hybridization. Highly expressed housekeeping genes or structural genes that are likely to be obtained in an EST collection, such as actin or 18s, can be used as controls in expression experiments. A good example of a study combining ESTs with subsequent expression studies at a smaller scale is that carried out by Gueguen et al. (2003), who first sequenced 1142 cDNA clones made from an enriched library of hemocytes from bacteria-challenged oysters. After annotating their sequences, they identified 20 genes with putative immune function. Subsequent expression studies of 4 of these genes then revealed that 3 of them were indeed induced by bacteria. In an ecotoxicological study, Nakayama et al. (2006) constructed a DNA oligo array of 1061 sequences using sequence information from an EST collection from the common cormorant. They hybridized this array with cDNA from livers obtained from wild cormorants and could correlate levels of certain environmental contaminants found in the animals with altered expression of P450 and antioxidant enzymes in the liver.

The blue mussel has been widely investigated in bio-monitoring programs and recognized as a potential candidate species for marine genomic approaches in ecotoxicology (Wilson et al. 2005). An early EST project using multiple tissues from unstressed blue mussels revealed an expression profile and sequence data of known and unknown genes (Venier et al. 2003), and this information was recently used to design a low-density DNA oligo array for stress response detection (Dondero et al. 2006). Such small arrays may advance the use of genomics in marine biomonitoring.

In conclusion, ESTs provide sequence information useful for phylogenetic studies, population genetics, ecotoxicology, array projects and downstream gene- or protein-specific studies, all very useful for the understanding of the organism in its relationship with the environment. They can also be the starting point for more ambitious genome projects.

MICROARRAYS/PROTEOMICS

Methods

Microarrays

Transcription profiling using microarrays is expected to be the major activity of ecological genomics in the near future. This method allows analyses of the kinds and amounts of mRNA produced by a cell or tissue,

and therefore the facility to understand which genes are expressed. This in turn provides insights into how the cell/tissue responds when it grows or multiplies, changes function, or when it is subject to new or unnatural environmental conditions. Gene expression is a highly complex and tightly regulated process that allows a cell to respond dynamically both to environmental stimuli and to its own changing needs. This mechanism acts as both an 'on/off' switch to control which genes are expressed in a cell, and as a 'volume control' that increases or decreases the level of expression of particular genes as necessary. DNA microarray technology, in correlation with genome projects as well as phylogenetic and comparative genomic approaches, may also facilitate the identification and classification of DNA sequence information and the assignment of functions to newly identified genes (Wilson et al. 2006).

Common to all microarray approaches is the basic principle of complementary base pairing. A microarray operates by exploiting the ability of a given mRNA molecule to bind specifically and non-covalently to, or hybridize to, the DNA template from which it originated. By using a microarray, chip or slide, which consists of respective gene sequences or ESTs that are coated on a solid layer at high density, it is possible to determine, in a single experiment, the expression levels of hundreds or thousands of genes by measuring the amount of mRNA bound to each site on the array. The subsequent use of a computer driven microarray reader enables precise measurement of the amount of mRNA hybridized to the spots on the microarray. This generates a profile of gene expression for a cell or a cell population/tissue that can be used to build a molecular fingerprint. A judgement on the respective genes with regard to expression level is possible for distinct time points or response states. Moreover, besides qualitative assessment, the data also can be evaluated quantitatively, which may be highly relevant to both the ecological or ecotoxicological response of a species and its environmental management. Gene expression profiles thus provide a molecular fingerprint of the transcriptome. To date, ecologists have not used the global-gene expression response pattern per se as a 'signature response pattern' to changing conditions. Nevertheless, transcriptome pattern signatures, as a response to changing physiology, are increasingly used in medicine (in particular for diagnostic purposes), and it is only a matter of time before the approach crosses over to ecology (Chen et al. 2005, Jones et al. 2005, Selman et al. 2006).

To fabricate expression microarrays, EST complementary DNA (cDNA), or gene-specific sequences that are synthesized *in situ*, are spotted at defined positions on a surface (e.g. glass slide, nylon membrane). The mRNAs of interest (samples) and a control mRNA (reference) are then transformed into cDNAs, and each

sample and the reference are labelled by different fluorochromes and co-hybridized (Fig. 3). The detection of the hybridization signals requires a specific microarray scanner connected to a database, which is essen-

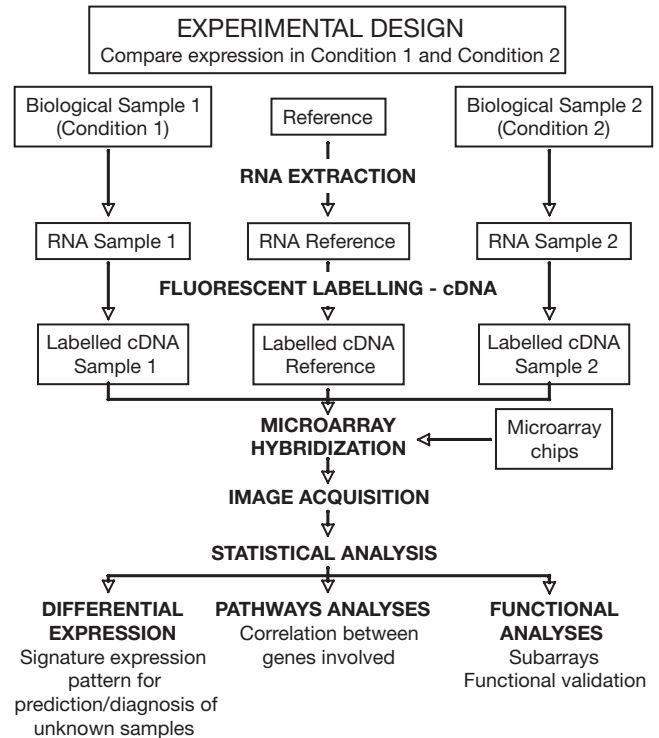


Fig. 3. Microarray experiment flowchart. cDNA microarrays are used in gene expression analysis. In this technique, RNA is isolated from 2 (or more) different samples derived from study subjects under different conditions, as well as from a common reference (which serves as a calibrator) (RNA EXTRACTION). The samples and common reference RNAs are transformed into cDNA and labelled by 2 fluorochromes respectively: 1 fluorochrome for the reference and 1 fluorochrome for the samples (generally the green cyanine 3 and the red cyanine 5 [Cy3, Cy5]) (FLUORESCENT LABELLING-cDNA). Each sample/common reference is then co-hybridised to a microarray replicate, which consists of large numbers of cDNAs/oligonucleotides orderly arranged/spotted onto a glass microscope slide (array spots) (MICROARRAY HYBRIDIZATION). After co-hybridization of each sample and common reference under stringent conditions on the microarray, a scanner records the emission of the 2 fluorochromes (for each sample-reference pair) for each spot on the microarray after excitation at given wavelengths (IMAGE ACQUISITION). The intensity of the fluorescence emission signals on each spot is proportional to transcript levels in the biological samples. For calibration purposes, the ratio of sample to reference emission for each microarray spot is used to compare the 2 (or more) study samples. Microarray data are then analysed using specific software that enables clustering of genes with similar expression patterns, which can be used to establish a differential expression signature for the samples compared (DIFFERENTIAL EXPRESSION). Responses of transcripts in various samples can also be clustered according to their affiliation with a particular intracellular signaling pathway (PATHWAY ANALYSES) or according to their common function (FUNCTIONAL ANALYSES)

tial for the analysis of the large amount of data. In addition, various algorithms must be applied to optimize the evaluation of the data. The objective is then to distinguish between random and significant patterns of gene expression among samples. After quality control of the sample and the hybridization comes image processing and the first analytical step to produce a large number of quantified gene expression values. These values represent absolute fluorescence signal intensities as a direct result of hybridization events on the array surface. The data are then normalized to compare the appropriate measured gene expression levels. The expression levels can then be used for several purposes. For example, it is possible to classify genes based on their expression levels in the different responses to the environment (e.g. environmental changes induce a number of genes to increase or decrease their expression) or to classify them in a functional way (e.g. all genes involved in cell membrane transport). Alternatively, if genes that change their expression belong to a biological network or pathway, they may be classified as such (e.g. genes involved in aerobic/anaerobic cell respiration that change their behavior during anoxic conditions).

Proteomics

As microarrays can be used to assess changes in the transcriptome, proteomics can be used to study the proteome, that is all the proteins that are synthesized by a particular cell at a particular time. The proteome is the protein complement of the genome and the study of proteomics is important because proteins are responsible for both the structure and the functions of all living things, whereas genes are simply the instructions for making proteins. Moreover, the proteome more accurately reflects the response because post-translational and post-transcriptional modifications as well as phosphorylations etc. can substantially change the nature of the expressed protein product. The set of proteins within a cell varies both from one differentiated cell type to another (e.g. in development) and over time, depending on the activities of the cell (e.g. division during algal cell blooms; repairing damage to DNA when pollutants occur; responding to a newly available nutrient or stress factor when the environment changes; responding to the arrival of a hormone during mating season etc.). In this way, the proteome is a genuine measure of the cell phenotype.

In proteomics, protein mixtures are extracted from cells or tissues that have been exposed to an environmental condition or that represent a temporal or spatial condition (sample). At the same time other 'normal' cells or tissues are used as controls (control). Each type

(sample and control) of protein mixture is subsequently subject to 2D gel electrophoresis, which separates the proteins in one dimension by their electrical charge and in the second dimension by their size. The gel is then stained to visualize various protein spots, and spots of sample and control gels are compared to identify differentially expressed proteins. Interesting (i.e. differentially expressed) spots are punched out of the gel, and analyzed. The analysis generally starts with treatment by a protease to digest the protein into a mix of peptides that can be run through a mass spectrometer to separate the peptides into sharply defined peaks. The result is mined against a database of all known proteins (which have been digested with the same enzyme) to see if a match can be found. If no match is found for the digested protein, a mass spectrometer can be used first to randomly break the peptide into a mix of fragments containing 1, 2 etc. amino acids and then to measure the mass of each fragment. The resulting data can be searched against a database that matches the mass data with known pairs, triplets etc. of amino acids. Subsequently overlapping fragments are assembled to reveal the entire sequence of the peptide. This can be searched against a genetic database to find the gene that encodes this particular peptide. In turn, translation of the matching gene reveals the entire sequence of the protein.

Another method frequently used to deliver valuable results for proteomics research is 2D nano-liquid chromatography-mass spectrometry (LC/MS). For instance, it has been used successfully in elucidating the proteome of several organisms (Washburn et al. 2001, Florens et al. 2002, Nägele et al. 2004).

Applications

DNA microarrays and proteomics have great potential to reveal community dynamics at different levels from individual genes to communities. This will become essential in population genetics and the analysis of biodiversity. These techniques are already being applied to marine ecology.

Large and medium environmental effects

The comprehensive description of transcriptomic responses provides useful information for conservation efforts, because it provides additional tools for early diagnostics. For example, a number of proteomic and genomic studies are underway to develop early markers for toxic algal bloom prediction (Chan et al. 2004, Lidie et al. 2005). Biomarkers for pollution in mussels are also being unveiled by proteomics, such as 2D gel

electrophoresis for peroxisome proliferation (Mi et al. 2005) or protein chip technology (Knigge et al. 2004). For restoration and bioregeneration efforts, genomics can help decipher the metabolic pathways involved in greenhouse gas balance in the ocean, such as those employed by the coccolithophore *Emiliania huxleyi*, which mediates oceanic and atmospheric carbon cycling (Nguyen et al. 2005, Dyhrman et al. 2006), or those used by methane-consuming bacteria (Hallam et al. 2004).

Ecotoxicology

By enabling the analysis of chemical effects at the molecular, tissue, and whole organism level, emerging technologies in the areas of genomics, proteomics, and metabolomics are important for the development of streamlined, cost-effective, and comprehensive testing approaches for evaluating environmental hazards. The genomic tools for ecotoxicogenomics have been reviewed by Wilson et al. (2005), and also recently by Miracle & Ankley (2005) with a particular emphasis on fish testing. Increasingly, more studies are emerging in this field, such as that of the effects of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) exposure on zebrafish caudal fin regeneration (Andreasen et al. 2006). Proteomics have also been used to follow protein profile alterations after exposure to cadmium in the marine alga *Nannochloropsis oculata* (Kim et al. 2005).

Adaptation and colonization of new habitats

Symbiotic associations are fundamental to the survival of many organisms, their diversity and even colonization of previously inhospitable habitats. Cnidarian-dinoflagellate intracellular symbioses are common mutualisms in the marine environment. They form the trophic and structural foundation of coral reef ecosystems and have played a key role in the radiation and biodiversity of cnidarian species. Proteomic studies to look at the interaction between host and symbiont have already begun (Barneah et al. 2006), as have systematic analyses of EST and cDNA microarray studies (Kuo et al. 2004, Rodriguez-Lanetty et al. 2006). This should ultimately lead to the discovery and characterization of symbiosis gene markers, which will enable early diagnosis of coral bleaching, a phenomenon that can ultimately lead to coral reef ecosystem breakdown owing to the loss of dinoflagellate symbionts from cnidarian hosts. One marker has currently been developed for the sea anemone *Anthopleura elegantissima* (Mitchelmore et al. 2002), but larger screenings will probably identify more.

Molecular responses permitting tolerance to extreme environments are also important to our understanding of how organisms have diversified and adapted. Certain halophilic archaea, for example, can develop anaerobic capabilities when high salt concentrations, elevated temperatures, and high cell densities promoted by aerobic growth and flotation reduce the availability of molecular oxygen. An operon with proteins responsible for and/or induced during anaerobic respiration was found in *Halobacterium* sp. by using transcriptome analysis as a complement to other methods such as phenotype analysis (Muller & DasSarma 2005).

The recent completion of several algal genome sequences and EST collections has facilitated functional genomic approaches for algal model systems. Ecological questions such as acquisition of increased metabolic versatility can be answered using these techniques. For example, the thermo-acidophilic unicellular red alga *Galdieria sulphuraria* can adopt heterotrophic and mixotrophic growth modes on more than 50 different carbon sources, and tolerate hot acidic environments as well as high concentrations of toxic metal ions, suggesting potential applications in bioremediation. To unravel the exceptional metabolic pathways of this organism, Weber et al. (2004) used a comparison between the *G. sulphuraria* transcriptome and the obligate photoautotrophic red alga *Cyanidioschyzon merolae*, which has a similar genome size. This study suggested that genes involved in the uptake of reduced carbon compounds and related enzymes were crucial to the metabolic flexibility of *G. sulphuraria* (Barbier et al. 2005). Proteomic approaches for dissecting molecular mechanisms of salinity tolerance in algae and higher plants are also in progress (Liska et al. 2004).

Ecophysiology and behavioural ecology

Atlantic salmon *Salmo salar* are known for spectacular marine migrations before homing to spawn in natal rivers. However, many males do not migrate before reproducing. Rather, these so-called 'sneaker' males mature early and reproduce at much smaller sizes than their migratory conspecifics without ever leaving freshwater. Early sexual maturity in salmon is the result of developmental plasticity, because the same genotype can express both types of reproduction tactics depending on the environment. Aubin-Horth et al. (2005) investigated the nature and extent of the coordinated molecular changes that accompany such a fundamental transformation by comparing brain transcriptional profiles of wild, mature sneaker males to age-matched, immature males (future large anadro-

mous males) and immature females. Of the ca. 3000 genes surveyed, 15% were differentially expressed in the brains of the 2 male types, and consistent patterns of gene expression were found for individuals of the same reproductive tactic. Notably, gene expression patterns in immature males differed from both immature females and sneakers, indicating that delayed maturation and sea migration by immature males, the 'default' life cycle, may actually result from an active inhibition of development into a sneaker. In this context, it is notable that a salmonid microarray containing cDNAs representing 16006 genes has been developed and assayed for intraspecific variation hybridization studies (von Schalburg et al. 2005).

A number of organisms change their activities and physiology during the circadian cycle: they emit chemical substances into the environment or bioluminesce, therefore influencing the ecosystem that they are part of. The first hints of temporal control within chloroplast proteins of *Arabidopsis thaliana* were identified by proteome analysis, and the technology has now been applied to the green alga *Chlamydomonas reinhardtii* (reviewed by Wagner et al. 2005), and chronobiological proteome assays have been performed for the dinoflagellate *Lingulodinium polyedra* (Akimoto et al. 2004).

Evolutionary ecological studies

The applications of functional genomics to evolutionary ecological studies have been explored by Feder & Mitchell-Olds (2003) and in the marine field were reviewed by Wilson et al. (2005), with a special focus on the plastic nature of the genome as seen by whole-genome comparisons. These applications also included assessment of influences on morphology and speciation brought about by variations in Quantitative Trait Loci (QTLs) and/or changes in non-coding regulatory sequences that control the expression of genes in time and space.

Limitations

Transcription profiling and proteomics are not ends in themselves and, for example, although very powerful and extremely useful, microarrays are simply exploratory instruments. This is only touching the surface of what ecological genomics has to offer. To be useful and worthwhile it needs to be grounded in physiological and biochemical knowledge, not to say understanding. It is rather more complicated than genomic analyses because, as noted earlier, a single gene can give rise to a number of different proteins through alternative

splicing of the pre-messenger RNAs, RNA editing of the pre-messenger RNAs, and/or post-translational processing such as attachment of carbohydrate residues to form glycoproteins and addition of phosphate groups to some of the amino acids in the protein (Black 2000, Schmucker et al. 2000). There is a disparity between mRNA and protein abundance and enzyme activity, supporting the contention that it is difficult to predict protein activity from genomic data such as microarrays or RT-PCR (Glanemann et al. 2003). Moreover, some evidence suggests that there is no direct correlation between mRNA and protein changes with phenotype and fitness (Jeong et al. 2001, Giaever et al. 2002, Carpenter & Sabatini 2004). These observations are not surprising and can be explained by variability in mRNA stability, translational control, post-translational modifications and regulation of enzyme activity. Moreover, genes physically adjacent in the genome often have similar expression profiles when comparing different environments. Genes present in these expression clusters proved to be no more similar in structure or function than could be expected by chance, and are not expressed because they play a particular role but because a neighbour is expressed (Spellman & Rubin 2002). In this regard, genomics, transcriptomics and proteomics go hand in hand and perhaps ideally should be used in parallel to study the same processes.

These techniques will certainly play a key role in ecology, but only in combination with other emerging tools used to try to unravel the complex questions surrounding the question of how genomes interact with their environment. A fully detailed picture of the state of any biological system requires knowledge of all its components (i.e. transcriptome, proteome, and metabolome).

BARCODING

Method

One of the beneficial side effects of the genomic revolution is that not only has it helped the discovery of sequences of interest for population genetics (microsatellites, SNPs, etc.), but also the identification of species using DNA barcoding. The concept of DNA barcoding has attracted much attention from a wide range of biological disciplines (Lipscomb et al. 2003, Seberg et al. 2003, Stoeckle 2003, Janzen 2004, Marshall 2005) and offers intriguing perspectives for applications in marine ecology (Schander & Willassen 2005). The method allows systematic screening of one or several reference genes for as many organisms as is feasible (Hebert et al. 2003). If assembled into a comprehensive database, these sequences can then be used as refer-

ence genes for the identification of species based on sequence comparisons. Large-scale DNA barcoding libraries are already under construction, for example the 'Barcode of Life Data Systems' (BOLD, www.barcodinglife.org) and the 'Consortium for the Barcode Of Life' (CBOL, <http://barcoding.si.edu>), and proponents of the method envisage that in the future the ability to determine species will no longer depend on the taxonomic expertise of a few specialists. Instead, by simply obtaining a DNA sequence from the organism in question, anybody should be able to determine species identification. The barcoding idea is partly built upon the already common practice of including molecular data in taxonomic studies. Electrophoresis (Thorpe & Ryland 1979) or sequencing of nuclear genes (Floyd et al. 2002) has earlier been used to discriminate between morphologically indistinguishable/identical species. Likewise, in ecological surveys, genetic or proteomic markers have become essential for species determination, as for example in commercially important marine species (López et al. 2002) or toxic strains of algae (Chan et al. 2004, Lidie et al. 2005). The novel idea with barcoding *sensu-stricto* (Hebert et al. 2003), however, is to find a single marker that is universally applicable to a large group of organisms such as animals or plants, and for which general primers can be used. One of the proposed barcoding genes for metazoans is the mitochondrial gene cytochrome *c* oxidase subunit 1, also referred to as 'COI' or '*cox1*' (Fig. 4).

Applications

The ecological applications of a universal molecular identification system resulting from a marine barcoding program are vast and would improve the quality of ecological surveys tremendously, in particular in those that contain species difficult to identify. Such species are found everywhere in the marine realm, especially when entering micro- or meio-faunal assemblages. The majority of organisms on earth are microscopic with body sizes <1 mm and, although these play a central role in marine ecosystem function (Blaxter et al. 2005), most them (e.g. nematodes) are as yet undescribed. Also, amongst larger animals, in particular those with few diagnostic features, identification is a complex exercise usually restricted to experts, e.g. platyhelminths, nemerteans, or nematodes. Schander & Willassen (2005) showed that major parts of the faunal composition in marine reports and inventories often remain undetermined. This greatly impairs the

comprehensibility of such studies and limits the conclusions that can be drawn.

There are many more potential applications arising from a marine barcoding program. Principally, it should be possible to determine species from all kinds of life-history stages, for example eggs or planktonic larvae. Stomach contents could also be utilised in order to resolve food webs in marine ecosystems. Also, faunal remains on the sea floor may be traced back to their living origins. Moreover, parasitic or other symbiotic relationships can be described (Tops & Okamura 2003) without the need to identify the symbiont visually. It seems there is great potential in characterizing faunal assemblages in such a detailed fashion. Further applications of the method lie in conservation and management efforts, for example in the monitoring of invasive species (DeSalle & Amato 2004). Dispersal in the marine environment is less hampered by geographical barriers compared with most terrestrial or limnic systems (Palumbi 1992), and invasive species are becoming an increasingly problematic side effect of globalization (Roman & Palumbi 2004). Here, barcodes could assist the tracing of invasive species, for example by scanning water samples and screening for the species in question.

Because of high variation at the species level, barcoding genes can also be used for other applications and vice versa. COI for example is a frequently applied marker in phylogeographic and phylogenetic studies. Hence, the genes used for species identification also have the potential to show the presence of cryptic species (Obst et al. 2005) and polymorphisms (Eriksson et al. 2006), describe the population structure within a species (Barber et al. 2002, Lessios et al. 2003), and test hypotheses of evolutionary relationships (Sorensen et

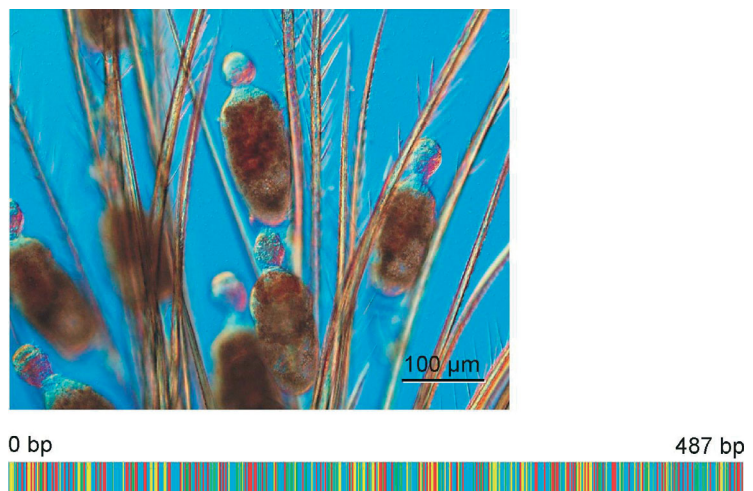


Fig. 4. A COI DNA-barcode (below) of the microscopic Cyclophora (above), an enigmatic protostome, so far only described from the mouth appendages of 3 nephropid lobsters (from Obst et al. 2005)

al. 2006). Thus, there is great potential for application of a marine barcoding program, and there are many examples of how genomic methods stimulate traditional fields such as taxonomy and ecology as well as training for a new generation of marine biologists with expertise in integrative approaches (Will et al. 2005).

Limitations

Recently, a number of examples have shown the ability of this technique to assign previously unidentified individuals to the right species using COI barcodes of (Hebert et al. 2003, 2004a, Hogg & Hebert 2004, Barrett & Hebert 2005) as well as the potential to discover new species (Hebert et al. 2004b, Obst et al. 2005). However, some principal issues regarding the method remain. For example, a recent study by Meyer & Paulay (2005) showed that the technique is successful only in those cases where the studied taxa are well known beforehand. This study revealed high error rates in the determination of species when the group was little studied. This is certainly the case for most of the biological diversity in the marine realm (May 1988). Furthermore, barcodes not only vary among but also within each species. This means that in order to reliably assign a sequence to the correct species, any reference library must take into account the entire intra-specific variation of the marker. Practically, this dictates the assortment of many barcodes necessary for each species, e.g. more than 10 sequences from the entire distribution area (Hajibabaei et al. 2005). Such problems will lead barcoding programs beyond the assembly of sequence information for a large number of organisms. Automatically, any barcoding project will necessitate parallel revisions of the groups under study, e.g. all sequence records need to be linked to voucher specimens that represent a species already described in considerable detail with regard to its taxonomy, morphology, and ecology (Hajibabaei et al. 2005).

ECOLOGY FEEDBACK TO GENOMICS

The genomics revolution provides some striking new insights for ecological studies. However, genomics is more than a toolbox added to marine ecology. Marine ecological genomics is new discipline merging genomics with marine ecology and leads to new questions independent of both fields. Despite such potential, genomic techniques present some limitations (e.g. see 'Genome sequencing: comparing genomes—phylogenomics' and 'Barcoding: limitations') that highlight the parallel importance of traditional taxonomy and ecological approaches.

One interesting example comes from the worm-like marine animal *Xenoturbella* spp. Even though they are neither parasitic nor microscopic, they lack a through-gut, gonads, coelomic cavities, and a brain. Owing to a simple body plan, their phylogenetic position has long remained puzzling. Based on morphology, they have been suggested to be a primitive flatworm (Westblad 1949), unique representatives of a plesiomorphic metazoan group (Jagersten 1959), an enteropneust, holothurian, or unique representatives of a deuterostome group (Reisinger 1960), a hemichordate (Pedersen & Pedersen 1986), an acoel flatworm (Franzen & Afzelius 1987, Lundin 1998, 2000, 2001), a primitive metazoan (Ehlers & Sopott-Ehlers 1997, Raikova et al. 2000), a bivalve (Israelsson 1997, 1999), or a bryozoan (Zrzavy 1998). This example shows that, even when using rather powerful techniques such as scanning or transmission electron microscopy, morphology alone cannot resolve the phylogenetic position of an animal.

In 1997, a gene sequence analysis of the mitochondrial *cox1* gene showed that *Xenoturbella* spp. are bivalves (Norèn & Jondelius 1997), solving the problem at last. However, in 2003, Bourlat et al. reported a different sequence of the same *cox1* gene from *Xenoturbella* spp., and suggested that they were deuterostomes. Are *Xenoturbella* spp. bivalves, or are they deuterostomes? This is a good example of where gene sequence analysis, despite wide acceptance today, is not all-powerful, and cannot alone determine the phylogenetic position of a simple animal.

Bourlat et al. (2003) showed that if you extract DNA from the epidermis alone, you obtain mostly deuterostome sequence. This suggests that the molluscan DNA was in the gut and was that of the prey of *Xenoturbella* spp.. This hypothesis clearly needs to be tested, and it is here that ecological data becomes useful, if not essential. The reported bivalve sequence showed 97.2% homology to *Nucula tenuis* at the nucleotide level. Through ecological projects such as the national Swedish monitoring program (Agrenius 2003), it is known that *N. tenuis* are present in the fjord where *Xenoturbella* spp. are found. Furthermore, the reported deuterostome sequence shows no match with other deuterostome animals in that area. These ecological and molecular data together support the contention that the bivalve DNA is that of *N. tenuis* in the area where *Xenoturbella* spp. feed, and that the deuterostome sequence is the genuine *Xenoturbella* spp. sequence. An immunohistochemical study has also supported the deuterostome status of *Xenoturbella* spp. (Stach et al. 2005). Thus, because the identity question has now gained morphological, gene sequence, and immunohistochemical evidential support together with ecological data, the answer to the long-lasting question has

finally been resolved: *Xenoturbella* spp. are deuterostomes.

Even when not investigating such a complicated example as *Xenoturbella* spp., it is vital to avoid contamination when applying genomic methods. This includes contamination from prey, parasites, animals attached to your intended animal, or contamination during experimental procedures in the lab. The first 3 factors are especially important when working with non-model organisms collected from nature, and highlights the importance of understanding ecology. Furthermore, there are always possibilities of artefacts, such as the formation of a concatemer of 2 separate genes during gene cloning (Hibino et al. 2004) or incorrect PCR priming (Quist & Chapela 2001, Metz & Futterer 2002, Kaplinsky et al. 2002), so thorough analyses, repeat sequencing and cross-checking are a necessity in genomic studies.

DISCUSSION

Previous ecological molecular studies have focused on limited numbers of genes and gene products. To understand complex life processes, a more integrative approach is necessary and 'approaches similar in spirit to systems biology should ultimately be adopted to enable genomics answers to ecological questions' (Van Straalen & Roelofs 2006). Recent advances in molecular techniques have made high throughput analyses of genomes, transcriptomes and proteomes possible and with this, ecology has entered a new era. Nevertheless, the incredibly powerful engine called genomics is still in its infancy and its inductive phase. One common criticism of current massive data-collection efforts is that much information, but little knowledge, is accumulating. This descriptive and not hypothesis-driven phase can be a source of impressive data sets and often unexpected information, and new hypotheses may be derived. The next step will be a more integrative approach and hypothesis-driven science. It will allow us to answer deep biological and evolutionary questions linking spatial and temporal considerations with the interaction between genome and the environment. As suggested by James Galagan, 'It's no longer enough to sequence a genome, catalogue the genes and come up with diagrams of signaling and so forth. We're expecting to get much more'.

What kind of evolution can we expect in the near future for marine ecological genomics? (1) At present, there are few sequenced genomes of ecologically relevant species in the marine environment. Technological advances in the near future will allow an increase in the number of these species and allow genome-wide analyses of ecological questions. (2) Metagenomics

approaches are particularly promising in ecology, and we can expect reconstruction of complete genomes from large-scale sequencing of the environment. (3) One limitation is that the great majority of genomics studies are conducted in the laboratory (perhaps with the exception of microbial ecology), so analysis performed directly in the field will allow us to answer new questions. (4) We can also expect some development of new methods for data analysis. (5) The study of epigenetic variants in natural populations has little influence in ecology now, but it will eventually have more impact thanks to 'omics' technologies (Van Straalen & Roelofs 2006).

This evolution is also a human challenge. Sequencing and analyzing a genome requires almost as many management skills as scientific ones. It often involves a large number of groups and therefore needs careful coordination (e.g. organization of conferences, workshops) between teams with different skills and goals. Moreover, it is an informatic challenge, and communication is crucial for the establishment of standards, tools and algorithms, for example for the annotation of environmental genomic data. The real challenge for marine ecological genomics is the creation of sufficiently large but effective collaborative networks around key model species. Some networks are already devoted to the development, utilization, and spreading of 'omics' approaches for the investigation of the biology and ecology of marine organisms. A marine genomics project is a functional genomics initiative developed in the USA to provide a pipeline for the curation of ESTs and gene expression microarray data for marine organisms (46000 ESTs from 19 species in the database; see www.marinegenomics.org). It has provided a clearing-house for marine-specific EST and microarray data available online (McKillen et al. 2005). In Europe, the Network of Excellence 'Marine Genomics Europe' (MGE; see www.marine-genomics-europe.org) is a major new enterprise funded by the European Community, comprising 44 laboratories and standing the crossroads between life sciences, ecology, environment, bioinformatics and high technologies within a multicultural European environment. MGE is devoted to the development, utilization, and spreading of high-throughput approaches for the investigation of the biology marine organisms. Uniquely, it has enabled the integration of a hitherto fragmented set of high-level expert groups to come together, share skills, state-of-the-art platforms and ambition. Benefits include large-scale sequencing projects, phylogenetic analyses, and the application of genomics technologies to functional, comparative, and environmental issues in marine biology. Thus, while marine ecological genomics is not completely beyond the reach of individuals, there are clear advantages to be gained from

formal, or informal, consortia brought together to solve common issues and shared ideals. The increasing need for multidisciplinary, combined with the costs of capital equipment and associated resources, makes such networks an important component for the future development of marine genomics, not least by providing opportunities for training the next generation of scientists and enabling the creation of sustainable collaborations.

Acknowledgements. This work is supported by VR (Swedish Research Council) EU RTNZ-2001-00029, Network of Excellence Marine Genomics Europe GOCE-04-505403 and the Royal Swedish Academy of Sciences.

LITERATURE CITED

- Adoutte A, Balavoine G, Lartillot N, de Rosa R (1999) Animal evolution—the end of the intermediate taxa? *Trends Genet* 5:104–108
- Agrenius S (2003) Övervakning av mjukbottenfaunan langs Sveriges vastkust: rapport fran verksamheten ar 2002. Naturvardsverket 1–7, available at: http://www.marecol.gu.se/digitalAssets/751833_Rapport_2003.pdf
- Aguinaldo AMA, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA (1997) Evidence for a clade of nematodes, arthropodes and other moulting animals. *Nature* 387:489–493
- Akimoto H, Wu C, Kinumi T, Ohmiya Y (2004) Biological rhythmicity in expressed proteins of the marine dinoflagellate *Lingulodinium polyedrum* demonstrated by chronological proteomics. *Biochem Biophys Res Commun* 315:306–12
- Andreasen EA, Mathew LK, Tanguay RL (2006) Regenerative growth is impacted by TCDD: gene expression analysis reveals extracellular matrix modulation. *Toxicol Sci* 92: 254–269
- Aubin-Horth N, Landry CR, Letcher BH, Hofmann HA (2005) Alternative life histories shape brain gene expression profiles in males of the same population. *Proc Biol Sci* 272: 1655–62
- Barber PH, Palumbi SR, Erdmann MV, Moosa MK (2002) Sharp genetic breaks among populations of *Haptosquilla pulchella* (Stomatopoda) indicate limits to larval transport: patterns, causes, and consequences. *Mol Ecol* 11:659–674
- Barbier G, Oesterhelt C, Larson MD, Halgren RG, Wilkerson C, Garavito RM, Benning C, Weber AP (2005) Comparative genomics of two closely related unicellular thermoacidophilic red algae, *Galdieria sulphuraria* and *Cyanidioschyzon merolae*, reveals the molecular basis of the metabolic flexibility of *Galdieria sulphuraria* and significant differences in carbohydrate metabolism of both algae. *Plant Physiol* 137:460–474
- Barneah O, Benayahu Y, Weis VM (2006) Comparative proteomics of symbiotic and aposymbiotic juvenile soft corals. *Mar Biotechnol* 8:11–16
- Barret RDH, Hebert PDN (2005) Identifying arachnids through DNA sequences. *Can J Zool* 83:481–491
- Béjà O (2004) To BAC or not to BAC: marine ecogenomics. *Curr Opin Biotech* 15:187–190
- Béjà O, Aravind L, Koonin EV, Suzuki MT and 7 others (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* 289:1902–1906
- Black DL (2000) Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology. *Cell* 103:367–370
- Blair JE, Ikeo K, Gojobori T, Hedges SB (2002) The evolutionary position of nematodes. *BMC Evol Biol* 2:7
- Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R, Abebe E (2005) Defining operational taxonomic units using DNA barcode data. *Phil Trans R Soc Lond B* 360: 1935–1943
- Bourlat SJ, Nielsen C, Lockyer AE, Littlewood DTJ, Telford MJ (2003) *Xenoturbella* is a deuterostome that eats molluscs. *Nature* 424:925–928
- Carpenter AE, Sabatini DM (2004) Systematic genome-wide screens of gene function. *Nature Rev Genet* 5:11–22
- Chan LL, Hodgkiss IJ, Wan JM, Lum JH, Mak AS, Sit WH, Lo SC (2004) Proteomic study of a model causative agent of harmful algal blooms, *Prorocentrum triestinum* II: the use of differentially expressed protein profiles under different growth phases and growth conditions for bloom prediction. *Proteomics* 4:3214–3226
- Chen ZM, Crone KG, Watson MA, Pfeifer JD, Wang HL (2005) Identification of a unique gene expression signature that differentiates hepatocellular adenoma from well-differentiated hepatocellular carcinoma. *Am J Surg Pathol* 29:1600–1608
- Chen C, Zhou P, Choi YA, Huang S, Gmitter FG (2006) Mining and characterizing microsatellites from citrus ESTs. *Theor Appl Genet* 11:1–10
- Curtis TP, Sloan WT, Scannell JW (2002) Estimating prokaryotic diversity and its limits. *Proc Natl Acad Sci USA* 99: 10494–10499
- DeLong EF, Karl DM (2005) Genomic perspectives in microbial oceanography. *Nature* 437:336–342
- Delsuc F, Brinkmann H, Philippe H (2005) Phylogenomics and the reconstruction of the tree of life. *Nature Rev Genet* 6: 361–375
- DeSalle R, Amato G (2004) The expansion of conservation genetics. *Nature Rev Genet* 5:702–712
- 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: 24–36
- Doney SC, Abbott MR, Cullen JJ, Karl DM, Rothstein L (2004) From genes to ecosystems: the ocean's new frontier. *Front Ecol Environ* 2:457–466
- Dyrhman ST, Haley ST, Birkeland SR, Wurch LL, Cipriano MJ, McArthur AG (2006) Long serial analysis of gene expression for gene discovery and transcriptome profiling in the widespread marine coccolithophore *Emiliania huxleyi*. *Appl Environ* 72:252–60
- Ehlers U, Sopott-Ehlers B (1997) Ultrastructure of the subepidermal musculature of *Xenoturbella bocki*, the adelphotaxon of the Bilateria. *Zoomorphology* 117:71–79
- Eriksson R, Nygren A, Sundberg P (2006) Genetic evidence of phenotypic polymorphism in the aeolid nudibranch *Flabellina verrucosa* (M. Sars, 1829) (Opisthobranchia: Nudibranchia). *Org Div Evol* 6:71–76
- Feder ME, Mitchell-Olds T (2003) Evolutionary and ecological functional genomics. *Nature Rev Genet* 4:649–655
- Felsenstein J (1978) Cases in which parsimony and compatibility methods will be positively misleading. *Syst Zool* 27: 401–410
- Floyd R, Abebe E, Papert A, Blaxter M (2002) Molecular barcodes for soil nematode identification. *Mol Ecol* 11: 839–850
- Franzen A, Afzelius BA (1987) The ciliated epidermis of

- Xenoturbella bocki* (Platyhelminthes Xenoturbellida) with some phylogenetic considerations. *Zool Scr* 16:9–17
- García-Fernández JM, Tandeau de Marsac N, Diez J (2004) Streamlined regulation and gene loss as adaptive mechanisms in *Prochlorococcus* for optimized nitrogen utilization in oligotrophic environments. *Microbiol Mol Biol Rev* 68:630–638
- Giaever G, Chu AM, Ni L, Connelly C and 69 others (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418:387–391
- Giovannoni SJ, Stingl U (2005) Molecular diversity and ecology of microbial plankton. *Nature* 437:343–348
- Glanemann C, Loos A, Gorret N, Willis LB, O'Brien XM, Lessard PA, Sinskey AJ (2003) Disparity between changes in mRNA abundance and enzyme activity in *Corynebacterium glutamicum*: implications for DNA microarray analysis. *Appl Microbiol Biotechnol* 61:61–68
- Gueguen Y, Cadoret JP, Flament D, Barreau-Roumiguere C and 5 others (2003) Immune gene discovery by expressed sequence tags generated from hemocytes of the bacteria-challenged oyster, *Crassostrea gigas*. *Gene* 303:139–145
- Hackett JD, Scheetz TE, Yoon HS, Soares MB, Bonaldo MF, Casavant TL, Bhattacharya D (2005) Insights into a dinoflagellate genome through expressed sequence tag analysis. *BMC Genomics* 6:80
- Hajibabaei M, DeWaard JR, Ivanova NV, Ratnasingham S, Dooh RT, Kirk SL, Mackie PM, Hebert PDN (2005) Critical factors for assembling a high volume of DNA barcodes. *Phil Trans R Soc Lond B* 360:1959–1967
- Hallam SJ, Putnam N, Preston CM, Detter JC, Rokhsar D, Richardson PM, DeLong EF (2004) Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science* 305:1457–1462
- Hebert PDN, Cywinka A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. *Proc R Soc Lond B* 270:313–321
- Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004a) Identification of birds through DNA barcodes. *PLoS Biol* 2: 1657–1663
- Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004b) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc Natl Acad Sci USA* 101:14812–14817
- Held M, Gase K, Baldwin IT (2004) Microarrays in ecological research: a case study of a cDNA microarray for plant-herbivore interactions. *BMC Ecol* 4:13
- Hess WR (2004) Genome analysis of marine photosynthetic microbes and their global role. *Curr Opin Biotech* 15:191–198
- Hibino T, Harada Y, Minokawa T, Nonaka S, Amemiya S (2004) Molecular heterotopy in the expression of *Brachyury orthologs* in order Clypeasteroidea (irregular sea urchins) and order Echinoidea (regular sea urchins). *Dev Genes Evol* 214:546–558
- Hirsch J, Lefort V, Vankersschaver M, Boualem A, Lucas A, Thermes C, d'Aubenton-Carafa Y, Crespi M (2006) Characterization of 43 non-protein coding mRNA genes in *Arabidopsis* including the MIR162a-derived transcripts. *Plant Physiol* 140:1192–1204
- Hogg ID, Hebert PDN (2004) Biological identification of springtails (Collembola: Hexapoda) from the Canadian Arctic, using mitochondrial DNA barcodes. *Can J Zool* 82: 749–754
- Israelsson O (1997) ...and molluscan embryogenesis. *Nature* 390:32
- Israelsson O (1999) New light on the enigmatic *Xenoturbella* (phylum uncertain): ontogeny and phylogeny. *Proc R Soc Lond B* 266:835–841
- Jagersten G (1959) Further remarks on the early phylogeny of Metazoa. *Zool Bidr Upps* 33:79–108
- Janzen DH (2004) Now is the time. *Phil Trans R Soc Lond B* 359:731–732
- Jeong H, Mason SP, Barabási AL, Oltvai ZN (2001) Lethality and centrality in protein networks. *Nature* 411:41–42
- Jones M, Blaxter M (2005) Evolutionary biology: animal roots and shoots. *Nature* 434:1076–1077
- Jones J, Otu H, Spentzos D, Kolia S and 8 others (2005) Gene signatures of progression and metastasis in renal cell cancer. *Clin Cancer Res* 11:5730–5739
- Kaplinsky N, Braun D, Lisch D, Hay A, Hake S, Freeling M (2002) Maize transgene results in Mexico are artefacts. *Nature* 416:601
- Kim YK, Yoo WI, Lee SH, Lee MY (2005) Proteomic analysis of cadmium-induced protein profile alterations from marine alga *Nannochloropsis oculata*. *Ecotoxicology* 14:589–596
- Knigge T, Monsinjon T, Andersen OK (2004) Surface-enhanced laser desorption/ionization-time of flight-mass spectrometry approach to biomarker discovery in blue mussels (*Mytilus edulis*) exposed to polyaromatic hydrocarbons and heavy metals under field conditions. *Proteomics* 4:2722–2727
- Kore-eda S, Cushman MA, Akselrod I, Bufford D, Fredrickson M, Clark E, Cushman JC (2004) Transcript profiling of salinity stress responses by large-scale expressed sequence tag analysis in *Mesembryanthemum crystallinum*. *Gene* 341:83–92
- Kuo J, Chen MC, Lin CH, Fang LS (2004) Comparative gene expression in the symbiotic and aposymbiotic *Aiptasia pulchella* by expressed sequence tag analysis. *Biochem Biophys Res Commun* 318:176–186
- Lawton JH (1994) What do species do in ecosystems? *Oikos* 71:367–374
- Lessios HA, Kane J, Robertson DR (2003) Phylogeography of the pantropical sea urchin *Tripneustes*: contrasting patterns of population structure between oceans. *Evol Int J Org Evol* 57:2026–2036
- Lidie KB, Ryan JC, Barbier M, Van Dolah FM (2005) Gene expression in Florida red tide dinoflagellate *Karenia brevis*: analysis of an expressed sequence tag library and development of DNA microarray. *Mar Biotechnol* 7: 481–493
- Lipscomb D, Platnick N, Wheeler Q (2003) The intellectual content of taxonomy: a comment on DNA taxonomy. *Trends Ecol Evol* 1865–66
- Liska AJ, Shevchenko A, Pick U, Katz A (2004) Enhanced photosynthesis and redox energy production contribute to salinity tolerance in *Dunaliella* as revealed by homology-based proteomics. *Plant Physiol* 136:2806–2817
- López JL, Marina A, Alvarez G, Vazquez J (2002) Application of proteomics for fast identification of species-specific peptides from marine species. *Proteomics* 2: 1658–1665
- Lundin K (1998) The epidermal ciliary rootlets of *Xenoturbella bocki* (Xenoturbellida) revisited: new support for a possible kinship with the Acoelomorpha (Platyhelminthes). *Zool Scr* 27:263–270
- Lundin K (2000) Phylogeny of the Nemertodermatida (Acoelomorpha, Platyhelminthes). A cladistic analysis. *Zool Scr* 29:17–27
- Lundin K (2001) Degenerating epidermal cells in *Xenoturbella bocki* (phylum uncertain), Nemertodermatida and Acoela (Platyhelminthes) *Belg J Zool* 131:153–157
- Margulies M, Egholm M, Altman WE, Attiya S and 52 others (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380

- Marshall E (2005) Taxonomy—will DNA bar codes breathe life into classification? *Science* 307:1037
- May RM (1988) How many species are there on earth? *Science* 241:1441–1449
- McKillen DJ, Chen YA, Chen C, Jenny MJ and 7 others (2005) Marine genomics: a clearing-house for genomics and transcriptomic data of marine organisms. *BMC Genomics* 6:34
- McKusick VA, Ruddle FH (1987) A new discipline, a new name, a new journal. *Genomics* 1:1–2
- Metz M, Fuetterer J (2002) Suspect evidence of transgenic contamination. *Nature* 416:600–601
- Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biol* 3:2229–2238
- Mi J, Orbea A, Syme N, Ahmed M, Cajaraville MP, Cristobal S (2005) Peroxisomal proteomics, a new tool for risk assessment of peroxisome proliferating pollutants in the marine environment. *Proteomics* 5:3954–3965
- Miracle AL, Ankley GT (2005) Ecotoxicogenomics: linkages between exposure and effects in assessing risks of aquatic contaminants to fish. *Reprod Toxicol* 19:321–326
- Mitchelmore CL, Schwarz JA, Weis VM (2002) Development of symbiosis-specific genes as biomarkers for the early detection of cnidarian-algal symbiosis breakdown. *Mar Environ Res* 54:345–349
- Muller JA, DasSarma S (2005) Genomic analysis of anaerobic respiration in the archaeon *Halobacterium* sp. strain NRC-1: dimethyl sulfoxide and trimethylamine N-oxide as terminal electron acceptors. *J Bacteriol* 187:1659–1667
- Mushegian AR, Garey JR, Martin J, Liu LX (1998) Large-scale taxonomic profiling of eukaryotic model organisms: a comparison of orthologous proteins encoded by the human, fly, nematode, and yeast genomes. *Genome Res* 8: 590–598
- Nägele E, Vollmer M, Hörth P (2004) Improved 2D Nano-LC/MS for proteomics applications: a comparative analysis using yeast proteome. *J Biomol Tech* 15:134–143
- Nakayama K, Iwata H, Kim EY, Tashiro K, Tanabe S (2006) Gene expression profiling in common cormorant liver with a oligo array: assessing the potential toxic effects of environmental contaminants. *Environ Sci Technol* 40: 1076–1083
- Nguyen B, Bowers RM, Wahlund TM, Read BA (2005) Suppressive subtractive hybridization of and differences in gene expression content of calcifying and noncalcifying cultures of *Emiliana huxleyi* strain 1516. *Appl Environ Microbiol* 71:2564–2575
- Norèn M, Jondelius U (1997) *Xenoturbella*'s molluscan relatives. *Nature* 390:31–32
- Obst M, Funch P, Giribet G (2005) Hidden diversity and host specificity in cycliophorans: a phylogeographic analysis along the North Atlantic and Mediterranean Sea. *Mol Ecol* 14:4427–4440
- Ogasawara M, Sasaki A, Metoci H, Shin-I T, Kohara Y, Satoh N, Satou Y (2002) Gene expression profiles in young adult *Ciona intestinalis*. *Dev Gen Evol* 212:173–185
- Palumbi SR (1992) Marine speciation on a small planet. *Trends Ecol Evol* 7:114–118
- Pedersen KJ, Pedersen LR (1986) Fine structural observations on the extracellular matrix (ECM) of *Xenoturbella bocki* Westblad 1949. *Acta Zool* 67:103–113
- Peterson KJ, Eernisse DJ (2001) Animal phylogeny and the ancestry of bilaterians: interferences from morphology and 18S rDNA gene sequences. *Evol Dev* 3:170–205
- Philippe H, Lartillot N, Brinkmann H (2005) Multigene analyses of bilaterian animals corroborate the monophyly of Edzysozoa, Lophotrochozoa, and Protostomia. *Mol Biol Evol* 22:1246–1253
- Pieper DH, Martins do Santos VAP, Golyshin PN (2004) Genomics and mechanistic insights into the biodegradation of organic pollutants. *Curr Opin Biotech* 15:215–224
- Quist D, Chapela IH (2001) Transgenic DNA introgressed into traditional maize landraces in Oaxaca, Mexico. *Nature* 414:541–543
- Raikova OI, Reuter M, Jondelius U, Gustafsson MKS (2000) An immunocytochemical and ultrastructural study of the nervous and muscular systems of *Xenoturbella westbladi* (Bilateria inc. sed.). *Zoomorphology* 120:107–118
- Reisinger E (1960) Was ist *Xenoturbella*? *Z Wiss Zool* 164: 188–198
- Rodriguez-Lanetty M, Phillips WS, Weis VM (2006) Transcriptome analysis of a cnidarian—dinoflagellate mutualism reveals complex modulation of host gene expression. *BMC Genomics* 7:23
- Rogers YH, Venter JC (2005) Genomics: massively parallel sequencing. *Nature* 437:326–327
- Roman J, Palumbi SR (2004) A global invader at home: population structure of the green crab, *Carcinus maenas*, in Europe. *Mol Ecol* 13:2891–2898
- Roy SW, Gilbert W (2005) Complex early genes. *Proc Natl Acad Sci USA* 102:4403–4408
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463–5467
- Schander C, Willassen E (2005) What can biological barcoding do for marine biology? *Mar Biol Res* 1:79–83
- Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, Dixon JE, Zipursky SL (2000) *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* 101:671–684
- Seberg O, Humphries CJ, Knapp S, Stevenson DW, Petersen G, Scharff N, Andersen NM (2003) Shortcuts in systematics? A commentary on DNA-based taxonomy. *Trends Ecol Evol* 18:63–65
- Selman M, Pardo A, Barrera L, Estrada A and 5 others (2006) Gene expression profiles distinguish idiopathic pulmonary fibrosis from hypersensitivity pneumonitis. *Am J Respir Crit Care Med* 173:188–198
- Simon A, Glockner G, Felder M, Melkonian M, Becker B (2006) EST analysis of the scaly green flagellate *Mesostigma viride* (Streptophyta): Implications for the evolution of green plants (Viridiplantae). *BMC Plant Biol* 6:2
- Smith LM, Sanders, JZ, Kaiser RJ, Hugues P and 5 others (1986) Fluorescence detection in automated DNA sequence analysis. *Nature* 321:674–679
- Sneddon LU, Margareto J, Cossins AR (2005) The use of transcriptomics to address questions in behaviour: production of a suppression subtractive hybridisation library from dominance hierarchies of rainbow trout. *Physiol Biochem Zool* 78:695–705
- Sorensen MV, Sterrer W, Giribet G (2006) Gnathostomulid phylogeny inferred from a combined approach of four molecular loci and morphology. *Cladistics* 22:32–58
- Spellman PT, Rubin GM (2002) Evidence for large domains of similarity expressed genes in the *Drosophila* genome. *J Biol* 1:5
- Stach T, Dupont S, Israelson O, Fauville G, Nakano H, Kanneyby T, Thorndyke M (2005) Nerve cells of *Xenoturbella bocki* (phylum uncertain) and *Harrimania kupfferi* (Enteropneusta) are positively immunoreactive to antibodies raised against echinoderm neuropeptides. *J Mar Biol Assoc UK* 85:1519–1524
- Steele HL, Streit WR (2005) Metagenomics: advances in ecology and biotechnology. *FEMS Microbiol Lett* 247:105–111

- Stoeckle M (2003) Taxonomy, DNA, and the bar code of life. *Bioscience* 53:796–797
- Strous M, Pelletier E, Mangenot S, Rattei T and 33 others (2006) Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* 440:790–794
- Thorpe JP, Ryland JS (1979) Cryptic speciation detected by biochemical genetics in 3 ecologically important intertidal bryozoans. *Estuar Coast Mar Sci* 8:395–398
- Tops S, Okamura B (2003) Infection of bryozoans by *Tetracapsuloides bryosalmonae* at sites endemic for salmonid proliferative kidney disease. *Dis Aquat Org* 57:221–226
- Tringe SG, von Mering C, Kobayashi A, Salamov AA and 9 others (2005) Comparative metagenomics of microbial communities. *Science* 308:554–557
- Van Straalen NM, Roelofs D (2006) An introduction to ecological genomics. Oxford University Press, Oxford
- Venier P, Pallavicini A, De Nardi B, Lanfranchi G (2003) Towards a catalogue of genes transcribed in multiple tissues of *Mytilus galloprovincialis*. *Gene* 314:29–40
- von Schalburg KR, Rise ML, Cooper GA, Brown GD, Gibbs AR, Nelson CC, Davidson WS, Koop BF (2005) Fish and chips: various methodologies demonstrate utility of a 16,006-gene salmonid microarray. *BMC Genomics* 6:126
- Wagner V, Gessner G, Mittag M (2005) Functional proteomics: a promising approach to find novel components of the circadian system. *Chronobiol Int* 22:403–415
- Washburn MP, Wolters D, Yates III JR (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnol* 19:242–247
- Watanabe H, Tatarazuko N, Oda S, Nishide H, Uchiyama I, Morita M, Iguchi T (2005) Analysis of expressed sequence tags of the water flea *Daphnia magna*. *Genome* 48:606–609
- Weber AP, Oesterhelt C, Gross W, Brautigam A and 13 others (2004) EST-analysis of the thermo-acidophilic red micro-alga *Galdieria sulphuraria* reveals potential for lipid A biosynthesis and unveils the pathway of carbon export from rhodoplasts. *Plant Mol Biol* 55:17–32
- Westblad E (1949) *Xenoturbella bocki* n.g., n.sp., a peculiar, primitive turbellarian type. *Arkiv Zool* 1:3–29
- White TJ (1996) The future of PCR technology: diversification of technologies and applications. *Tibtech* 14:478–483
- Will KW, Mishler BD, Wheeler QD (2005) The perils of DNA barcoding and the need for integrative taxonomy. *Syst Biol* 54:844–851
- Wilson K, Thorndyke M, Nilsen F, Rogers A, Martinez P (2005) Marine systems: moving into the genomics era. *Mar Ecol* 26:3–16
- Wilson AC, Dunbar HE, Davis GK, Hunter WB, Stern DL, Moran NA (2006) A dual-genome microarray for the pea aphid, *Acyrtosiphon pisum*, and its obligate bacterial symbiont, *Buchnera aphidicola*. *BMC Genomics* 7:50
- Winnepennincks B, Backeljau T, Mackey LY, Brooks JM, de Wachter R, Kumar S, Garey JR (1995) 18S rRNA data indicate that Aschelminthes are polyphyletic in origin and consists at least three distinct clades. *Mol Evol Biol* 12:1132–1137
- Wolf YI, Rogozin IB, Koonin EV (2004) Coelomata and not Ecdysozoa: evidence from genome-wide phylogenetic analysis. *Genome Res* 14:29–36
- Zrzavy J, Mihulka S, Kepka P, Bezdek A, Tietz D (1998) Phylogeny of the Metazoa based on morphological and 18S ribosomal DNA evidence. *Cladistics* 14:249–285

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

*Submitted: July 5, 2006; Accepted: October 13, 2006
Proofs received from author(s): February 19, 2007*