Genomic perspectives in microbial oceanography

Edward F. DeLong[1] and David M. Karl[2]

The global ocean is an integrated living system where energy and matter transformations are governed by interdependent physical, chemical and biotic processes. Although the fundamentals of ocean physics and chemistry are well established, comprehensive approaches to describing and interpreting oceanic microbial diversity and processes are only now emerging. In particular, the application of genomics to problems in microbial oceanography is significantly expanding our understanding of marine microbial evolution, metabolism and ecology. Integration of these new genome-enabled insights into the broader framework of ocean science represents one of the great contemporary challenges for microbial oceanographers.

Marine ecosystems are complex and dynamic. A mechanistic understanding of the susceptibility of marine ecosystems to global environmental variability and climate change driven by greenhouse gases will require a comprehensive description of several factors. These include marine physical, chemical and biological interactions including thresholds, negative and positive feedback mechanisms and other nonlinear interactions. The fluxes of matter and energy, and the microbes that mediate them, are of central importance in the ocean, yet remain poorly understood. Detailed field studies over the past three decades have established the current 'microbial loop' hypothesis wherein microbes have a central position in the conversion of dissolved organic matter into higher trophic levels (Fig. 1). An explicit and comprehensive test of the microbial loop hypothesis, however, has not yet been achieved. In addition, the central role of microbial activities in maintaining the oxidative state of our planet, and biogeochemical cycles other than the carbon cycle, are not well captured in the current microbial loop hypothesis. Significant obstacles remain to be overcome in the measurement and modelling of ocean microbial processes.

For microbial oceanographers, a major challenge is to more accurately incorporate the details of diversity, physiology and ecology into oceanographic and biogeochemical models. For example, bacterioplankton community processes have often been modelled as being solely heterotrophic[1], understating the diverse physiologies and metabolic processes that operate *in situ*. Indeed, the importance of oxygenic photoautotrophic picoplankton, currently acknowledged to be dominant components of planktonic communities, was overlooked in early oceanographic models. In addition, planktonic microbial metabolism has sometimes been presumed to resemble that found in common laboratory strains, some of which have questionable ecological relevance. Recent advances, such as new cultivation approaches[2], cultivation-independent identification and enumeration strategies[3,4], and environmental genomics[5,6], are improving this situation. For example, we can now identify the genes and biochemical pathways that differentiate microbial species living in different habitats. It will soon be possible to map the metabolic and functional gene distributions of indigenous microbial species in space and time and within different oceanic provinces. The recent establishment of microbial observatories at selected sites will also

greatly aid in these efforts. The correlation between organism- and habitat-specific genomic features and other physical, chemical and biotic variables has the potential to refine our understanding of microbial and biogeochemical process in ocean systems.

All these advances — improved cultivation, environmental genomic approaches and *in situ* microbial observatories — promise to enhance our understanding of the living ocean system. Below, we provide a brief recent history of marine microbiology and outline some of the recent results from genomic technology and the challenges to be faced for integrating these new data into the larger fabric of ocean science.

A brief historical perspective

In the past 30 years, there has been a remarkable growth in understanding of marine microbiota (Fig. 2). During this time, researchers have recognized the crucial role that microbes play in ocean ecosystems. This stems in part from technical advances, such as improved epifluorescence microscope techniques and ATP-based biomass metrics[7], that have revealed bacterioplankton standing stocks that are several orders of magnitude greater than had been estimated by viable counting techniques[8]. The use of radiotracers[1,9] to estimate planktonic bacterial growth rates and turnover has also led to revised qualitative and quantitative models of microbial contributions to marine food webs[10,11]. Research using quantitative autofluorescent cell counts and flow cytometry eventually led to the discovery of abundant photoautotrophic picoplankton, including *Synechococcus*[12,13] and *Prochlorococcus*[14] species, that dominate photosynthetic activities in open ocean gyres. Around the same time, hydrothermal vents and their rich macrofauna and microflora were discovered[15], as well as the first bacterial isolates with an obligate growth requirement for elevated hydrostatic pressure[16]. The development of cultivation-independent phylogenetic surveys using ribosomal RNA (rRNA) sequencing[3,17] and fluorescence *in situ* hybridization[4] set the stage for contemporary environmental genomic studies. Soon after their development, rRNA-based phylogenetic survey techniques using the then-novel technique polymerase chain reaction (PCR) revealed the widespread distribution and abundance of several previously unrecognized marine microbial groups, including *Pelagibacter* (also known as SAR11)[18], abun-

[1]Department of Civil and Environmental Engineering & Division of Biological Engineering, 48-427 Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA; [2]School of Ocean and Earth Science and Technology, University of Hawaii, Honolulu, Hawaii 96822, USA.
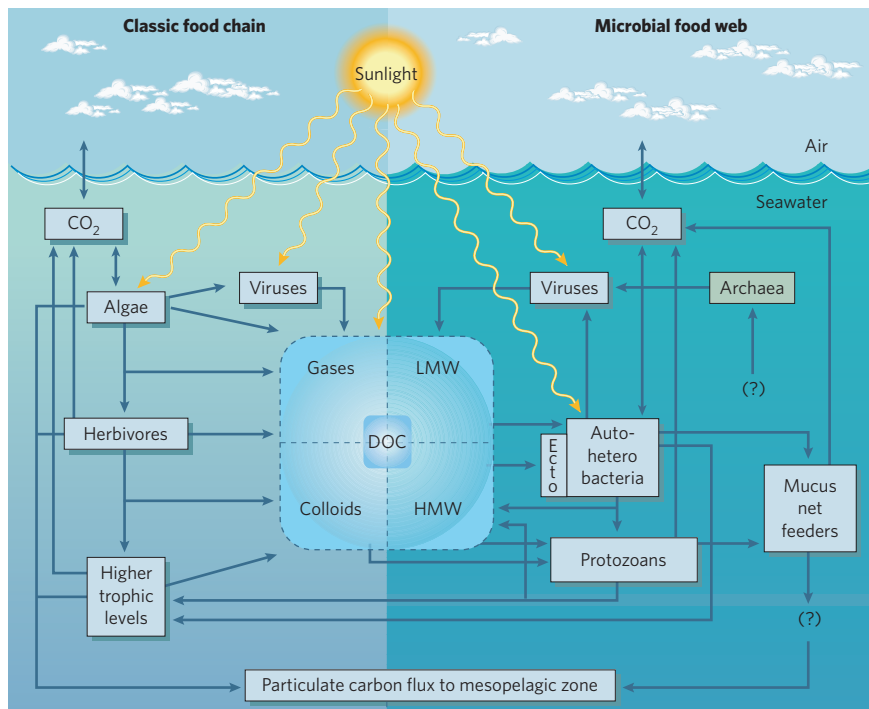
Figure 1 | Marine microbial interactions in the upper ocean. Schematic representation of the ocean food web showing on the left the classic pathway of carbon and energy flow through photosynthetic Eukarya, to herbivores and on to higher trophic levels. Depicted on the right is the microbial food web, which uses energy stored in the non-living, detrital carbon pool to produce microbial biomass that can re-enter the classic pathway of carbon and energy flow. Cell-associated ectoenzymes (Ecto) enable bacteria to use high-molecular-weight (HMW) DOC in addition to the more traditional low-molecular-weight (LMW) and gaseous carbon substances. Also shown in the microbial food web are viral particles and Archaea. At the present time, there is only rudimentary knowledge of the role of *Archaea* in the oceanic food web. Shown at the bottom of this diagram is the downward flux of particulate carbon (and energy), which is now thought to fuel most subeuphotic zone processes. The classic algae-herbivore grazer pathway (left side) is most important in this regard. Adapted from ref. 51.

dant new groups of planktonic marine Archaea[19-21] and novel eukaryotic picoplankton[22], to list a few examples.

More recently, the prevalence of bacteriochlorophyll-containing[23] and rhodopsin-containing[24] bacterioplankton was recognized, providing new perspectives on the nature of light-use strategies in ocean surface waters. Surprisingly, it has only very recently been realized that viral particles can exceed total microbial cell numbers by an order of magnitude in marine plankton[25] and that they represent potentially important vectors of bacterioplankton mortality and lateral gene transfer[26].

Many new marine microbial species continue to be described, and unexpected physiological and biochemical properties await further discovery and description. There remains much to learn about the distributions, variability and biogeochemical influence of naturally occurring microbes in the sea. To this end, new technologies are now providing life science and ocean science with remarkable new research opportunities. In particular, advances in genome science and technology are revealing the interdependences that link evolutionary, ecological and biogeochemical processes in natural marine microbial communities. For example, the co-evolution and genetic exchange between nuclear and chloroplast genomes record events that have profoundly altered the predominant modes and mechanisms of oxygenic photosynthesis on Earth[27].

Our understanding of the frequency and mechanisms by which genes and functional traits are acquired is being radically altered by genomic studies suggesting the pervasiveness of lateral gene transfer[28]. Similarly, the recent recognition of the tremendous sympatric species diversity in microbial plankton is adjusting current views of form, function and speciation in the sea[29,30]. As more samples of naturally occurring microbial genomic diversity become available, they will reveal more of the specific details of microbial evolutionary process and ecological dynamics, and how these relate to environmental processes. With the above observations and goals in mind, considerable effort is now being focused on sequencing whole marine microbial genomes as well as surveying entire marine microbial assemblages at the genomic level. The results of these efforts have the potential to improve descriptions of the structure and function of the living ocean system.

Genomics of cultivated marine microbial species

Adaptation to the unique characteristics of their environment defines the essence of marine microbial species. High salinity, low and variable nutrient concentrations and steep gradients of temperature, light, pressure and nutrients with increasing depth are some of the environmental variables that influence the evolution of form and function in microbial plankton. The adaptive instructions for coping with the marine environment are encoded in the genomic blueprints of ocean microbiota. Genome sequences are therefore useful tools for exploring the relationship between genotype, phenotype and environment in native marine microbial species (Table 1). A few examples provided below illustrate how recent sequence information has contributed to our understanding of the adaptations, physiology and ecology of marine microbes.

Genomic perspective on the forests of the sea

Diatoms, a type of algae, are one of the most conspicuous plant forms in the ocean. As a group they account for a large proportion of contemporary marine primary productivity, particularly in coastal regions. What can be learned about the remarkable evolutionary success of these protists by studying the structure of their chloroplast and nuclear chromosomes? How do genomic features of diatoms correlate with their phenotypic and ecological properties? Can genomic analyses provide new information about the metabolic and physiological properties of these important phytoplankters? Information gained from the recent sequencing of the 34-mega-base-pair (Mbp) nuclear genome of *Thalassiosira pseudonana* has answered some of these questions[31].

The 24 diploid chromosomes of *T. pseudonana* encode genes for proteins associated with the construction of the siliceous cell wall, including genes for silicic-acid uptake, polyamine synthesis and sillafin production. These newly identified genes can now be used in experiments to understand and model diatom cell-wall synthesis, and to model regulatory response mechanisms involved in growth limitation caused by silica starvation. *T. pseudonana* and other diatom species extrude chitin fibres from cell-wall pores, presumably to increase drag and thereby slow down their sinking rate. The genome of *T. pseudonana* encodes many genes for chitin synthesis and degradation, including 22 putative chitinases, which indicates a dynamic

and perhaps highly regulated chitin synthesis/degradation cycle to control buoyancy and nutrient cycling[31]. Diatoms are typically growth-limited by the lack of inorganic nutrients (nitrogen, phosphorus and silica) and need to respond to the variable nutrient flux in their surrounding environment. Features of the *T. pseudonana* genome that reflect its photosynthetic, but auxotrophic, lifestyle include a wide variety of nutrient transporters, nitrogen-acquisition genes that allow conversions from several different organic and inorganic nitrogen sources and multiple pathways for energy storage using both carbohydrate and lipid synthesis degradation pathways[31].

A genomic glimpse into coastal bacterial lifestyles

The physical, chemical and biological changes that occur along a transect from the coast to the open sea are profound. Do specific biological properties of coastal bacterioplankton differentiate them from their open-ocean relatives? Can the genomic and physiological properties of bacterioplankton explain in part their observed distribution? Can these different biological features tell us about potential regional differences in the microbial cycling of matter and energy? Although it is too early to generalize extensively, recent studies are starting to answer these questions. For example, some bacterial groups such as α-Proteobacteria in the *Roseobacter* clade seem more abundant in nearshore waters than the open ocean. The genome of *Silicibacter pomeroyi*, a member of the *Roseobacter* clade, has some adaptations that may attune this bacterium to life in coastal environments[32]. On the basis of its genome content, *S. pomeroyi* seems to consume not only dissolved organic carbon but also reduced inorganic compounds such as carbon monoxide and sulphur; it harbours the genes required to generate energy from these compounds. Preliminary experiments confirmed that carbon monoxide and reduced sulphur compounds are important substrates for *S. pomeroyi* metabolism[32]. In common with *T. pseudonana*, this coastal bacterium is equipped with a wide variety of nutrient transporters for uptake of different nitrogenous compounds such as amino acids, ammonium and urea.

The *S. pomeroyi* genome also encodes several transporters for uptake of osmolytes such as dimethylsulphoniopropionate (DMSP) and glycine betaine, presumably produced by its algal co-inhabitants, which are abundant in coastal environs[32]. Genomic and physiological analyses suggest a 'lithoheterotrophic' metabolic strategy for *S. pomeroyi*, allowing energy generation from organic and inorganic compounds alike. On the whole, these genomic and physiological features of *S. pomeroyi* seem to be specific adaptations to microniches typical of coastal ecosystems.

Figure 2 | Selected milestones in marine microbial oceanography. This timeline highlights a few of the advances and discoveries that have influenced marine microbiology over the past 30 years. Many important contributions could not be included simply owing to space limitation. WGS, whole-genome shotgun sequencing. PCR applications have led to more discoveries of novel microbial groups not shown here. QPCR, quantitative PCR.
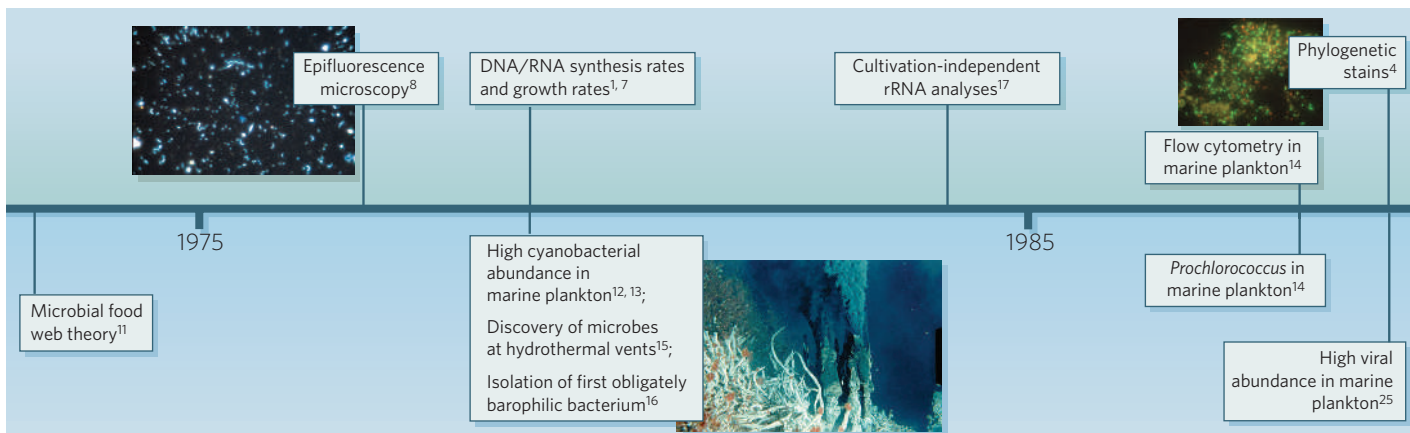
Variety is the spice of ocean life

Microbiologists find it convenient to organize microbial groups into functional guilds involved in specific processes, for instance sulphate-reducers, methanogens and oxygenic photoautotrophs. It is now well recognized that within any such group, or indeed even within a species, tremendous biological diversity exists. Closely related physiological and genetic variants that seem to be 'tuned' to specific ecological conditions have been referred to as 'ecotypes'[33]. What might genomic blueprints tell us about the ecophysiological differences among ecotypes? Could gene-linkage analyses of such ecotypes provide information about metabolic pathway interdependence? And at the genomic fine scale, how many differences must accumulate before one ecotype is ecologically distinguishable from another?

One of the best-documented examples of environmentally tuned ecotypes are photoautotrophic *Prochlorococcus* strains isolated from different depths. The genome sequences of these ecotypes reveal some of the unique adaptive strategies employed in different parts of the depth continuum. Several high- and low-light adapted *Prochlorococcus* strains were recently sequenced[33,34]. The low-light-adapted strain MIT9313 isolated from the bottom of the euphotic zone has significantly more genes on its 2.4 Mbp genome than its high-light-adapted relative MED4, with a genome size of 1.66 Mbp. Genome size, however, is not a strict indicator of light adaptation, because the genome of another low-light-adapted strain, *Prochlorococus* SS120, is 1.75 Mbp[33,34]. Genes found in the low-light strain but not in the high-light-adapted relative included those for nitrite transport and assimilation. This is consistent with the types of nutrient available at the base of the euphotic zone. Strain MED4 encoded many more genes for high-light-inducible proteins than its low-light-adapted cousin. By contrast, the low-light-adapted strain MIT9313 contained more genes associated with photosynthesis.

The genome sequence of *Synechococcus* WH8102 revealed different characteristics from those found in its *Prochlorococus* cousins. The *Synechococcus* genome suggests that it is more versatile generalist than *Prochlorococcus*, with a broad range of nutrient acquisition and catabolic capabilities[35]. All told, the genomic characteristics of these different marine oxygenic photoautotrophs are consistent with, and predictive of, their ecophysiology and environmental distributions.

Expanding databases

Whole-genome sequencing of marine microbes from different habitats and phylogenetic lineages promises to widen the scope of future comparative genomic analyses[36–38]. Other recently published marine microbial genome sequences include those of the psychrophilic (cold-loving) sulphate-reducer *Desulfotalea psychrophila*[38], a bioluminescent squid symbiont *Vibrio fisheri*[39], the marine planctomycete *Rhodopirellula baltica*[36], a heterotrophic bacterium from a hydrothermal vent site *Idiomarina loihiensis*[37] and the pressure-adapted deep-sea bacterium *Photobacterium profundum*[40]. Many other marine microbes are in the pipeline for sequencing. The Gordon and Betty Moore

Epifluorescence microscopy[8]

DNA/RNA synthesis rates and growth rates[1, 7]

Cultivation-independent rRNA analyses[17]

Phylogenetic stains[4]

Flow cytometry in marine plankton[14]

1975

1985

Microbial food web theory[11]

High cyanobacterial abundance in marine plankton[12, 13];

Discovery of microbes at hydrothermal vents[15];

Isolation of first obligately barophilic bacterium[16]

Prochlorococcus in marine plankton[14]

High viral abundance in marine plankton[25]

Foundation is currently funding a marine microbiology initiative that in part aims to generate draft genome sequences from many ecologically relevant marine microbial species (http://www.moore.org/program_areas/science/initiatives/marine_microbiology/initiative_marine_microbiology.asp).

The continued comparative analyses of marine microbial genomes promises to enhance both the databases and our current understanding. Nevertheless, the phenomenal 'microdiversity', disparate evolutionary histories and variable ecological strategies of marine microbial species necessitates careful sampling of more habitats and acquisition of even more genome sequences, from diverse as well as closely related species. New and improved methods for cultivating commonly occurring, but previously cultivation-resistant, microbial species are now being successfully developed[2], and these should greatly increase the relevance and use of existing genomic databases. An excellent example of this was the cultivation of one of the most abundant bacteria in the ocean *Pelagibacter ubique*[2,41]. The genome sequence of this ubiquitous marine microbe revealed its streamlined nature, as well as the presence of a proteorhodopsin-based photophysiology[24,41]. Continuing efforts will help to provide a comparative database for assessing the genomic repertoire of cultivated microbial species as well as the crucial context necessary for interpreting nascent environmental genomic surveys of marine microbial communities (see below).

Genomic surveys of natural microbial communities

Biological processes occurring in natural microbial communities have diverse, complex, interdependent intracellular and intercellular reactions. Much of this complexity is encoded in the structures, distribution and dynamics of interacting genomes in the environment. Recently, it has become possible to directly access the genomes of co-existing microbial species in natural communities en masse, without cultivation, using environmental genomic approaches[6,29]. Variously termed environmental genomics, metagenomics or ecogenomics, cultivation-independent genomic approaches can provide a new perspective on the naturally occurring microbial world.

Some of the first environmental genomic studies were focused on marine picoplankton[41,42]. Initially, large genomic fragments were cloned directly from marine microbial communities to survey genomic features of then-uncultivated phylogenetic groups, such as marine Archaea[5,24,42]. Similar genomic surveys recently led to the discovery of rhodopsins in bacteria, a domain of life not anticipated to contain these photoproteins[24]. Hypotheses about these novel bacterial rhodopsins could be tested *in vitro* and *in vivo* by heterologous expression in *Escherichia coli* and subsequent biophysical and biochemical characterization[24]. Subsequent field studies revealed that the bacterioplankton rhodopsins were diverse, widespread and abundant in the marine environment[43,44]. The progression from genome sequence to biochemical and biophysical experimentation to more refined ecological observations presents a promising model for future work aimed at obtaining new genome sequence data for physiological, ecological and
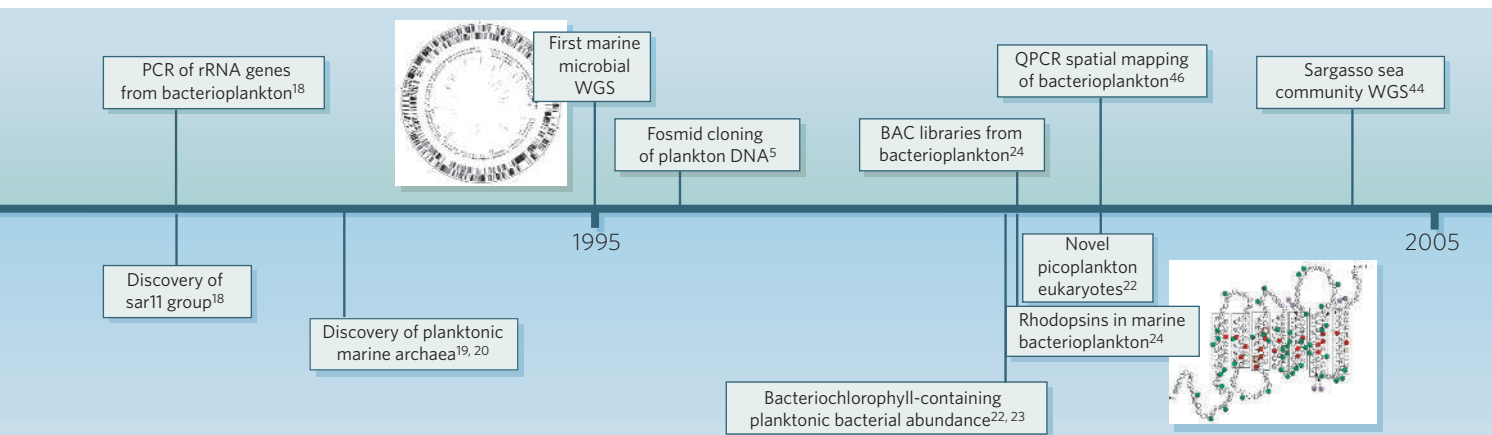
oceanographic research[29].

Another recent field application of marine genomics employed whole-genome shotgun (WGS) sequencing approaches. Spectacularly successful for sequencing whole microbial genomes, WGS involves the random sequencing of small (~3 kbp) DNA insert libraries and subsequent assembly of long consensus DNA sequences *in silico*. The high-throughput nature of this approach makes it extremely attractive, but environmental variation in microbial species richness, evenness and intraspecific genetic heterogeneity pose challenges to reassembling large contiguous DNA sequences (contigs) from smaller sequence fragments[29]. Nevertheless, large amounts of data can be recovered by WGS approaches. The most marked example was the recent genomic survey of the Sargasso Sea microbial assemblages, where over 1.6 billion bp of genome-sequence information and about 1.2 million genes were recovered from the collective microbial assemblage[44].

The raw power of a massive WGS DNA-sequencing survey was revealed in the Sargasso Sea study[29]. A remarkable number of new genes and gene families were recovered in this data set, and gene-mining efforts will continue to yield useful results. The reliable extraction of contiguous, long DNA-sequence assemblies, however, from short DNA sequences derived from complex, heterogeneous populations still pose a challenge to current assembly methods and algorithms. It seems evident from ongoing analyses that the availability of appropriate and relevant reference genomes, as well as the availability of large insert clone sequences as a reference, will play an important role in interpreting WGS sequencing surveys from complex microbial communities[29].

New approaches to the interpretation and analysis of microbial assemblage WGS data sets continue to be developed. Tringe *et al.*[45] recently compared, on a gene-by-gene basis, the similarities and differences between microbial communities from the Sargasso Sea, a whalebone, an acid-mine drainage assemblage and a farm silage assemblage using WGS data sets[45]. By taking a 'gene-centric', as opposed to an assembly-driven 'genome-centric', approach, it was possible to compare the patterns of occurrence of specific gene categories and assemble 'community profiles' of functional gene content. The over-represented specific categories of 'environmental gene tags' (EGTs) in different samples (for instance, a disproportionate representation of photosynthetic genes and rhodopsins in the Sargasso Sea sample) verified the usefulness of this approach for identifying metabolic features unique to specific microbial assemblages[45].

There are several ways in which microbial-community genomic data are likely to be applied in microbial oceanography and ecology in the future. First, a comprehensive microbial 'parts list' (for example, gene, taxon and biochemical pathway content) of the ocean's microbial inhabitants will provide a deeper perspective of their physiological potential. These data help formulate biochemical and physiological hypotheses that can be tested in the lab and in the field. Second, detailed sequence-based genomic data sets will also provide informatic tools and reagents, in the form of molecular probes and predicted mol-

PCR of rRNA genes from bacterioplankton[18]

First marine microbial WGS

QPCR spatial mapping of bacterioplankton[46]

Sargasso sea community WGS[44]

Fosmid cloning of plankton DNA[5]

BAC libraries from bacterioplankton[24]

1995

2005

Discovery of sar11 group[18]

Novel picoplankton eukaryotes[22]

Discovery of planktonic marine archaea[19, 20]

Rhodopsins in marine bacterioplankton[24]

Bacteriochlorophyll-containing planktonic bacterial abundance[22, 23]

ecular and organic geochemical targets, to better track and quantify specific microorganisms and microbial processes. For example, it is now becoming possible to quantitatively map multiple microbial groups in time and space using quantitative PCR[46], and in the future DNA-microarray techniques are likely to play a significant role as well. Improved gene and gene-expression mapping and inventories in the field, as well as more targeted and specific quantitative techniques, may lead to more meaningful biogeochemical measurements. This quantitative gene, taxon and metabolic mapping may, in turn, enable the development of models that more accurately reflect the *in situ* biological details. Third, comparative analyses of entire microbial assemblages have the potential to reveal larger-scale patterns of biochemical interactions and habitat-specific correlations that might otherwise be missed in autoecological studies of individual species. For example, comparative community genomics can reveal previously unrecognized biological interactions. For instance, rhodopsin-mediated light utilization in bacteria was entirely unknown until environmental genomics was used[24].

Ultimately, it may be possible to overlay microbial functional genes, metabolic pathways and taxonomic distributions upon other physical, chemical and biological oceanographic data to map microbial features onto oceanic provinces. Finally, the study of genome structure and dynamics *in situ* will provide the necessary information to interpret evolutionary processes — such as genetic drift, recombination and lateral gene transfer — that drive microbial adaptation and divergence. These data in turn provide predictive tools to model the variable complexity of form, function and process that underpin the ocean's physical, chemical and biotic interactions[47].

Contemporary challenges in microbial oceanography

One common criticism of current massive data-collection efforts is that muchinformation, but little knowledge, is accumulating. Although it is true that the informatics challenges are significant, they are not insurmountable. The immediate tasks ahead are to organize, standardize and coordinate annotation, access and integration of environmentally derived genomic data sets. The research community needs to recognize that environmental genomic data sets are fundamentally different from other existing databases because they are not derived from an individual strain, species, taxon or a single genetic locus. Rather, each discrete data set represents the collective genes recovered randomly from one single microbial assemblage. These data sets derive from dynamic microbial populations and environments, where species richness, evenness and composition (and so gene richness, evenness and composition) are variable in space and time. This is one major reason why any attempts at assembly need to be qualified and specified, because individual sequences in any given assembly could be derived from different samples, times and origins[44]. The original sequence reads, and their metadata, need to be linked explicitly with assemblies because they are of potentially of heterogeneous origin.

Many other informatic challenges need to be considered as the field of environmental genomics develops. Management, annotation and archiving of environmental data sets is an ongoing process and many, but not all, of the mechanisms to handle the large datasets now emerging are available. One issue of concern is establishment of standards and requirements for environmental genomic metadata. Additionally, tools and algorithms for the annotation of environmental genomic data sets need to be better developed. Hidden Markov model methods, effective for gene-calling in whole-genome sequences, are not as

Table 1 | Phenotypic and genomic characteristics of some recently sequenced marine microbes and communities

Isolate or environmental gene survey	Habitat or origin/habitat	Environmentally relevant phenotypic and genomic features	References
Thalassiosira pseudonana (K)	Surface water, plankton	Silicic acid metabolism, chitin buoyancy mechanisms, diverse nitrogen acquisition and metabolism, several energy storage and utilization strategies	31
Silicibacter pomeroyi DSS-3 (B)	Coastal water, plankton	Lithotrophic energy metabolism (carbon monoxide, reduced sulphur compounds), DMSP metabolism, many amino and carboxlic acid transporters	32
Prochlorococcus marinus MED4, SS120 (B)	Oligotrophic water, upper photic zone	High-light adapted, greater number of high-light-inducible proteins, photolyase genes, narrower range of nitrogen sources (nitrite reductase and transporter absent), fewer chlorophyll-binding antenna protein genes	33,34
Prochlorococcus marinus MIT9313 (B)	Oligotrophic waters, lower photic zone	Low-light adapted, smaller number of high-light-inducible proteins, no photolyase, wider range of nitrogen sources (nitrite reductase and transporter present), greater number of chlorophyll-binding antenna protein genes	33
Synechococcus sp. WH8102 (B)	Open ocean surface water	Wide range of nitrogen sources (organic, nitrate, nitrite), amino-acid transporters, swimming motility, more transporters than *Prochlorococcus*	35
Photobacterium profundum SS9 (B)	Deep-sea amphipod	Growth pressure optima > 100 atm, polyunsaturated fatty acids in membrane lipids, differential gene expression at high and low pressure growth, trimethylamine oxide respiration, possible anaerobic amino acid fermentation	40
Vibrio fischeri ES114 (B)	*Euprymna scolopes* light organ	Bioluminescence, specific association with squid *Euprymna scolopes*, many adhesions and toxin-like genes shared in common with other vibrio pathogens	39
Rhodopirellula baltica (B)	Kiel fjord seawater	Derived absence of peptidoglycan, formaldehyde oxidation pathway, abundant sulphatases, high proportion of genes most similar to Eukayra	36
Desulfotalea psychrophila Lsv54(B)	Permanently cold Arctic marine sediment	Growth below 0 °C, sulphate-reducer, potential oxygen-reducing cytochromes	38
Idiomarina loihiensis (B)	Mixed hydrothermal vent/cold seawater (1,296 m)	Amino-acid fermentation, incomplete amino acid biosynthetic pathways, exopolysaccharide and biofilm formation, reduced carbohydrate metabolism	37
Pelagibacter ubique	Oligotrophic coastal and open ocean	Represents a group that makes up 25% of planktonic marine bacteria, smallest known genome of free-living organism, contains proteorhodopsin	41
Sargasso Sea environmental gene survey	Oligotrophic surface water	High species diversity, high representation of photosynthetic and rhodopsin genes, significant intraspecific microheterogeneity, poor assembly	44
Eel River Basin Archaeal methane oxidizer environmental gene	Deep-sea methane seep (550 m)	High representation of methanogen-like genes, high representation of two major Archaeal methane oxidizer phylotypes, absence of one 'universal' methanogenic gene survey	50
Whale fall microbial environmental gene survey	Deep-sea whale bone	Low diversity, many glycine-betaine transporters and type IV secretion system	45

This is an illustrative but not exhaustive list of marine microbial strains whose genomes have been fully sequenced or partial shotgun environmental gene survey efforts.
*Domain affiliation in parentheses. K, Eukarya; B, Bacteria; A, Archaea.
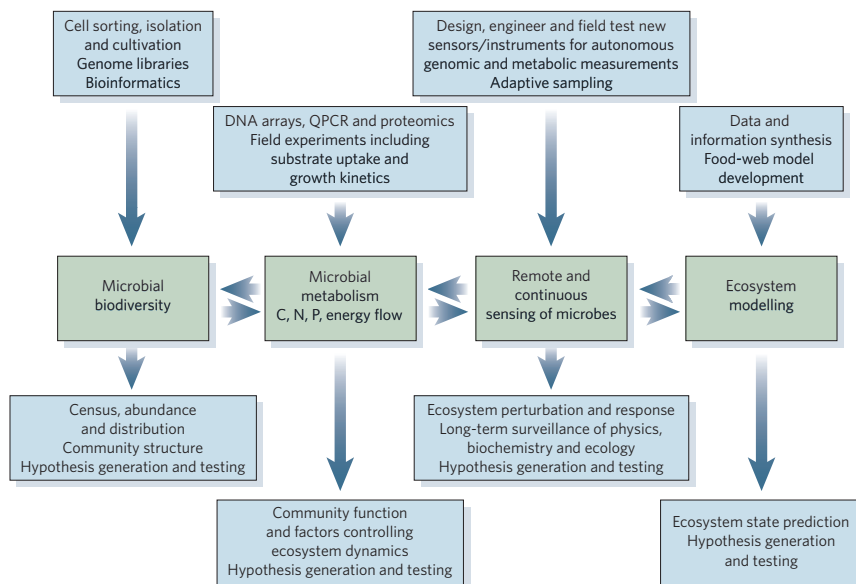
Figure 3 | Microbial systems analysis in oceanography. This flow chart outlines the research prospectus for a new international Center for Microbial Oceanography: Research and Education (CMORE), which the authors will direct. The flow path illustrates some of the linkages that will need to be fostered between and among subdisciplines in microbiology, oceanography, ecology and engineering to successfully complete the prospectus.

applicable for annotation of shorter sequences typical of metagenomic data sets. New gene-calling algorithms are another important area for development. In tandem with database issues, there is a strong need for further development of computational tools to analyse and compare microbial-community genomic data sets. Finally, interoperability between genome and metagenome sequence databases and environmental and ecological databases will be a crucial requirement in the future.

According to a recent report by the American Academy of Microbiology and by our own independent assessments, the fields of microbial ecology and genomics are currently poised at a 'crossroads of opportunity'[48]. It is important to understand ocean processes and to place the emergent genomic-based knowledge of marine microorganisms into the ecological context of carbon, nitrogen, phosphorus and related biogeochemical cycles. The scientific and societal benefits at stake are profound. This research prospectus must begin with the complete integration of the disciplines of microbiology, systems biology, ecosystem dynamics and oceanography, with the goal of developing comprehensive, mechanistic-based numerical models that are essential for scaling the interactions of specific microorganisms and processes across entire ocean basins[47]. Some of the tools, approaches and information flow required in this new synthesis are shown in Fig. 3. As an example of the magnitude of the challenge, consider that future climate predictions indicate a warmer and more stratified global ocean with more acidic near-surface waters, all resulting in large part from the increasing burden of carbon dioxide in the atmosphere. How will the present day oceanic 'genotype' respond to these new state variables, and are we prepared to observe and interpret the ocean's expressed 'phenotypic response'? New and integrative approaches to link microbial and global geochemical processes are required.

The establishment of a global network of ocean time-series stations for repeated observations of microbial diversity and function represents one strategy of ecosystem surveillance that would increase our understanding of these complex processes. Despite their recognized importance in Earth and ocean sciences, systematic long-term microbial observatories in key oceanic habitats are rare. Repeat observations of the North Pacific subtropical gyre over the past four decades have documented fundamental changes in microbial community structure and major bioelemental cycling. It has been hypothesized that climate variability in the North Pacific basin has led to an enhancement of microbial-based nitrogen fixation and a fundamental shift from nitrogen limitation to phosphorus limitation with several ecological consequences, including alterations in food web structure and changes in carbon sequestration[49]. These changes occurred because of subtle

changes in near-surface ocean stratification and decreases in the flux of inorganic nutrients from below. These habitat changes are controlled, in part, by the frequency and duration of the basin-scale El Niño and Pacific Decadal Oscillation climate systems. The nitrogen-fixation potential of the global ocean may ultimately be controlled by the atmospheric deposition of iron, which itself is a time-variable, climate-sensitive process. Total atmospheric dust transport is affected by humankind, including population demographics, global economies and land-use patterns. These complex natural and anthropogenic forces, each with multiple potential feedback loops and possible threshold effects, have significant global impacts on microbes in the sea that are only now being appreciated. The influence of climate variability, especially that which takes place on decadal to centennial time scales, on marine microbial biogeochemistry is not well understood.

Conclusion

Marine microbes are present at billions of cells per litre in seawater. They double every few days and are consumed at about the same rate by viral parasites and protistan predators. These activities capture and process energy and drive major elemental cycles. Hidden within these dynamic assemblages and diverse genomic structures are fundamental but, at present, incomplete lessons about environmental sensing, response and adaptation, gene regulation, species and community interactions and genomic plasticity and evolution. The genomic diversity, evolutionary dynamics and ecological processes contained in these populations have global effects on the rates and flux of energy and matter in the sea, biogeochemical cycling, the Earth's atmospheric composition and global climate trends. Over the 3.8 billion years of life on Earth, microbes have been the stewards of geochemical balance and as biotic recorders of evolutionary history their 'biological memories' extend backwards further than other lifeforms. Additionally, they are Nature's biosensors *par excellence*, but we still need to decipher their code to interpret their outputs and understand the underlying strategies and mechanisms of their survival in nature. The current convergence of microbiology, ecology, genomics and ocean science has the potential to be focused in unprecedented ways through the lens of microbial oceanography. By working together, molecular biologists, microbiologists and oceanographers have new opportunities to advance observation, method and theory, which together will better describe the living ocean system. ∎

1. Fuhrman, J. A. & Azam, F. Thymidine incorporation as a measure of heterotrophic bacterioplankton production in marine surface waters. *Mar. Biol.* **66,** 109–120 (1982).
2. Rappé, M. S., Connon, S. A., Vergin, K. L. & Giovannoni, S. J. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418,** 630–633 (2002).

3. Olsen, G. J., Lane, D. J., Giovannoni, S. J., Pace, N. R. & Stahl, D. A. Microbial ecology and evolution: a ribosomal RNA approach. *Annu. Rev. Microbiol.* **40,** 337–365 (1986).
4. DeLong, E. F., Wickham, G. S. & Pace, N. R. Phylogenetic stains: ribosomal RNA-based probes for the identification of single cells. *Science* **243,** 1360–1363 (1989).
5. Stein, J. L., Marsh, T. L., Wu, K. Y., Shizuya, H. & DeLong, E. F. Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *J. Bacteriol.* **178,** 591–599 (1996).
6. Handelsman, J. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* **68,** 669–685 (2004).
7. Karl, D. M. Cellular nucleotide measurements and applications in microbial ecology. *Microbiol. Rev.* **44,** 739–796 (1980).
8. Hobbie, J. E., Daley, R. J. & Jasper, S. Use of Nuclepore™ filters for counting bacteria by fluorescence microscopy. *Appl. Environ. Microbiol.* **33,** 1225–1228 (1977).
9. Karl, D. M. Measurement of microbial activity and growth in the ocean by rates of stable ribonucleic acid synthesis. *Appl. Environ. Microbiol.* **38,** 850–860 (1979).
10. Azam, F. *et al.* The ecological role of water-column microbes in the sea. *Mar. Ecol. Prog. Ser.* **10,** 257–263 (1983).
11. Pomeroy, L. R. The ocean's food web, a changing paradigm. *BioScience* **24,** 499–504 (1974).
12. Waterbury, J. B., Watson, S. W., Guillard, R. R. L. & Brand, L. E. Widespread ocurrence of a unicellular marine, planktonic, cyanobacterium. *Nature* **277,** 293–294 (1979).
13. Johnson, P. W. & Sieburth, J. M. Chroococcoid cyanobacteria in the sea: a ubiquitous and diverse phototrophic biomass. *Limnol. Oceanogr.* **24,** 928–935 (1979).
14. Chisholm, S. W. *et al.* A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* **334,** 340–343 (1988).
15. Corliss, J. B. *et al.* Submarine thermal springs on the Galapagos Rift. *Science* **203,** 1073–1083 (1979).
16. Yayanos, A. A., Dietz, A. S. & Van Boxtel, R. Obligately barophilic bacterium from the Mariana Trench. *Proc. Natl Acad. Sci. USA* **78,** 5212–5215 (1981).
17. Stahl, D. A., Lane, D. J., Olsen, G. J. & Pace, N. R. Analysis of hydrothermal vent-associated symbionts by ribosomal RNA sequences. *Science* **224,** 409–411 (1984).
18. Giovannoni, S. J., Britschgi, T. B., Moyer, C. L. & Field, K. G. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345,** 60–63 (1990).
19. DeLong, E. F. Archaea in coastal marine environments. *Proc. Natl Acad. Sci. USA* **89,** 5685–5689 (1992).
20. Karner, M. B., DeLong, E. F. & Karl, D. M. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409,** 507–510 (2001).
21. Fuhrman, J. A., McCallum, K. & Davis, A. A. Novel major archaebacterial group from marine plankton. *Nature* **356,** 148–149 (1992).
22. Moon-van der Staay, S. Y., De Wachter, R. & Vaulot, D. Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* **409,** 607–610 (2001).
23. Kolber, Z. S., Van Dover, C. L., Niederman, R. A. & Falkowski, P. G. Bacterial photosynthesis in surface waters of the open ocean. *Nature* **407,** 177–179 (2000).
24. Béjà, O. *et al.* Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289,** 1902–1906 (2000).
25. Bergh, O., Borsheim, K. Y., Bratbak, G. & Heldal, M. High abundance of viruses found in aquatic environments. *Nature* **340,** 467–468 (1989).
26. Paul, J. H., Sullivan, M. B., Segall, A. M. & Rohwer, F. Marine phage genomics. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **133,** 463–476 (2002).
27. Falkowski, P. G. *et al.* The evolution of modern eukaryotic phytoplankton. *Science* **305,** 354–360 (2004).
28. Lerat, E., Daubin, V., Ochman, H. & Moran, N. A. Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol.* **3,** e130 (2005).
29. DeLong, E. F. Microbial community genomics in the ocean. *Nature Rev. Microbiol.* **3,** 459–469 (2005).
30. Thompson, J. R. *et al.* Genotypic diversity within a natural coastal bacterioplankton population. *Science* **307,** 1311–1313 (2005).
31. Armbrust, E. V. *et al.* The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306,** 79–86 (2004).
32. Moran, M. A. *et al.* Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. *Nature* **432,** 910–913 (2004).
33. Rocap, G. *et al.* Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424,** 1042–1047 (2003).
34. Dufresne, A. *et al.* Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. *Proc. Natl Acad. Sci. USA* **100,** 10020–10025 (2003).
35. Palenik, B. *et al.* The genome of a motile marine *Synechococcus*. *Nature* **424,** 1037–1042 (2003).
36. Glockner, F. O. *et al.* Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. *Proc. Natl Acad. Sci. USA* **100,** 8298–8303 (2003).
37. Hou, S. *et al.* Genome sequence of the deep-sea gammaproteobacterium *Idiomarina loihiensis* reveals amino acid fermentation as a source of carbon and energy. *Proc. Natl Acad. Sci. USA* **101,** 18036–18041 (2004).
38. Rabus, R. *et al.* The genome of *Desulfotalea psychrophila*, a sulfate-reducing bacterium from permanently cold Arctic sediments. *Environ. Microbiol.* **6,** 887–902 (2004).
39. Ruby, E. G. *et al.* Complete genome sequence of *Vibrio fischeri*: a symbiotic bacterium with pathogenic congeners. *Proc. Natl Acad. Sci. USA* **102,** 3004–3009 (2005).
40. Vezzi, A. *et al.* Life at depth: *Photobacterium profundum* genome sequence and expression analysis. *Science* **307,** 1459–1461 (2005).
41. Giovannoni, S. J. *et al.* Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309,** 1242–1245 (2005).
42. Schmidt, T. M., DeLong, E. F. & Pace, N. R. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J. Bacteriol.* **173,** 4371–4378 (1991).
43. Béjà, O., Spudich, E. N., Spudich, J. L., Leclerc, M. & DeLong, E. F. Proteorhodopsin phototrophy in the ocean. *Nature* **411,** 786–789 (2001).
44. Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304,** 66–74 (2004).
45. Tringe, S. G. *et al.* Comparative metagenomics of microbial communities. *Science* **308,** 554–557 (2005).
46. Suzuki, M. T., Preston, C. M., Chavez, F. & DeLong, E. F. Quantitative mapping of bacterioplankton populations in seawater: field tests across an upwelling plume in the Monterey Bay. *Aquat. Microb. Ecol.* **24,** 117–127 (2001).
47. Doney, S. C., Abbott, M. R., Cullen, J. J., Karl, D. M. & Rothstein, L. From genes to ecosystems: the ocean's new frontier. *Front. Ecol. Environ.* **2,** 457–466 (2004).
48. Stahl, D. A. & Tiedje, J. M. *Microbial Ecology and Genomics: A Crossroads of Opportunity* (American Society for Microbiology, Washington D.C., 2002).
49. Karl, D. M., Bidigare, R. R. & Letelier, R. M. Long-term changes in plankton community structure and productivity in the North Pacific Subtropical Gyre: The domain shift hypothesis. *Deep-Sea Res. II* **48,** 1449–1470 (2001).
50. Hallam, S. J. *et al.* Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science* **305,** 1457–1462 (2004).
51. Karl, D. M. Accurate estimation of microbial loop processes and rates. *Microbiol. Ecol.* **28,** 147–150 (1994).

CORRIGENDUM

# Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks

Michael D. Shapiro, Melissa E. Marks, Catherine L. Peichel, Benjamin K. Blackman, Kirsten S. Nereng, Bjarni Jónsson, Dolph Schluter & David M. Kingsley

In this Article, we showed that *Pitx1* maps to the major locus controlling pelvic reduction in sticklebacks and shows altered expression, but not altered amino-acid sequence, in a pelvic-reduced population. Owing to a technical error, Figs 3 and 4 of the Article incorrectly show results from a *Pitx1* sense probe, rather than from an antisense probe. Similar staining is not observed for sense probes from other genes, raising the possibility that sticklebacks may express some endogenous antisense transcripts from the *Pitx1* region.

A new antisense probe (see Fig. 1, below) shows strong expression in the mouth, lower jaw and developing pelvis of marine larvae. *Pitx1* expression appears normal in the mouth and lower jaw of the Paxton benthic population, but is completely missing in the pelvic region. No expression was detected in neuromasts, thymus, olfactory pits or tail in either population, unlike the pattern seen with the original probe. Despite these differences, our major conclusion, that pelvic reduction results from *cis*-acting regulatory changes in the *Pitx1* locus, remains unchanged. This is because the endogenous sense transcript loses its expression along with the presumptive antisense transcript in the pelvic region of the larval progeny of fish missing the pelvis.
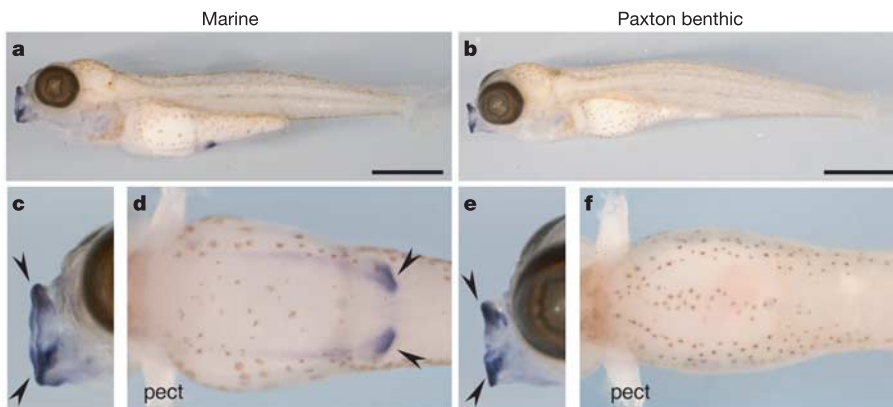
Figure 1 | *Pitx1* **is expressed in the prospective pelvic region of marine but not Paxton Lake benthic sticklebacks. a,** Whole-mount *in situ* hybridization shows *Pitx1* expression in the mouth, jaw and pelvic buds of stage-29 marine larvae. (Details are available from the corresponding author, D.M.K.) **c, d,** Enlarged views of the lateral head (**c**) and ventral pelvis (**d**) of stage-29 marine larvae; arrowheads indicate sites of expression. **b, e, f,** In Paxton benthic larvae, *Pitx1* expression is absent from the prospective pelvic region (**f**), but is present in the mouth and jaw (**e**, arrowheads). Scale bars, 1 mm; pect, pectoral fin.

CORRIGENDUM

Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data

Anders Moberg†, Dmitry M. Sonechkin, Karin Holmgren, Nina M. Datsenko, Wibjörn Karlén & Stein-Erik Lauritzen[1]

[1]Department of Earth Science, University of Bergen, 5007 Bergen, Norway. †Present address: Department of Physical Geography and Quaternary Geology, Stockholm University, 10691 Stockholm, Sweden

The authorship of this Letter is amended to include Stein-Erik Lauritzen. Details of the Søylegrotta Cave record (series 8), which should have been accredited to S.-E.L., were not supplied in the paper but are available from the corresponding author (A.M., anders.moberg@natgeo.su.se) on request.

In addition, the tree-ring-width data from the Indigirka river region (series G) were inadvertently used without the proper permissions: although the series has been discussed in the literature[1], they are unpublished data that have not been made publicly available; they may, however, be obtained through A.M.

1. Sidorova, O. V., Naurzbaev, M. M. Response of *Larix cajanderi* to climatic changes at the Upper Timberline and in the Indigirka River Valley [in Russian]. *Lesovedenie* **2**, 73–75 (2002).

CORRIGENDUM

Genomic perspectives in microbial oceanography

Edward F. DeLong & David M. Karl

It has been drawn to our attention (by J. A. Fuhrman) that Fig. 2 contains a citation error. Specifically, the citation associated with "Discovery of planktonic marine archaea" in Fig. 2 that was incorrectly given as ref. 20 should instead be ref. 21 (Fuhrman, J. A., McCallum, K. & Davis, A. A. *Nature* **356**, 148–149; 1992).