Systems Biology for Ecology: From Molecules to Ecosystems

KEVIN J. PURDY, PAUL J. HURD, JORDI MOYA-LARAÑO, MARK TRIMMER, BRIAN B. OAKLEY AND GUY WOODWARD

	Summary	88
I.	Introduction	88
	A. Towards a Systems Biology for Ecology?	88
	B. The Microbial Black Box	91
II.	A Brief History of Molecular Microbial Ecology	92
	A. Characterising Diversity, Abundance and Functional Traits	92
	B. Genomics and Post Genomics in Microbial Ecology	95
III.	Next-Generation Sequencing Technologies	96
	A. A New Paradigm	96
	B. 454 Pyrosequencing (Roche)	100
	C. Illumina Genome Analyzer	100
	D. HeliScope	106
	E. SOLiD (Life/Applied Biosystems)	108
	F. SMRT Technology (Pacific Biosciences)	108
	G. Improving Analysis of Multiple Samples and Increasing Read Length in NGS Runs	110
IV.	Genome Sequencing: Functional Diversity in Ecology	113
V.	Transcriptomics: Functional Expression	114
VI.	Application of Sequencing in Ecology	115
	A. Unveiling the Planet's Hidden Biodiversity	115
	B. Four Ecological Metagenomic Case Studies	116
	C. Moving Beyond Traditional Metagenomics	119
	D. Next-Generation Sequencing and Biodiversity	119
VII.	Linking Across Multiple Levels of Organisation: The Key to Understanding the System	122
	A. Scaling from Molecules to Ecosystems	122
	B. Linking Microbial Community Structure and Ecosystem Functioning	123
	C. New Questions in Ecology: Opening the Microbial Black Box	125
	D. Functional Redundancy: Do Species Matter?	127
	E. Beyond Bacteria: Opening the Eukaryotic Black Box	128
	F. Food Webs: Identifying the Missing Links	128
	G. NGS and the Link Between Ecology and Evolution	132
VIII.	Socioeconomic Applications: Ecosystem Goods and Services and Bioprospecting	133
IX.	Conclusions	134
	Acknowledgements	135
	References	135

ADVANCES IN ECOLOGICAL RESEARCH VOL. 43 0065-2504/10 $35.00

DOI: 10.1016/S0065-2504(10)43003-2

SUMMARY

Ecology stands on the edge of a true paradigm shift, fuelled by a recent technological revolution in our ability to measure both taxonomic and functional biodiversity via the application of metagenomics and transcriptomics. The advent of 'next generation sequencing' (NGS) in molecular biology is rapidly opening the black box of microbial ecology, providing us with some of the first glimpses of a previously hidden world. This is now enabling microbial ecology to become firmly embedded as a core subdiscipline within ecology, and to test general theories about biodiversity, biogeography and ecosystem functioning using a combination of molecular and more traditional techniques. In addition, NGS offers a means of not only measuring the abundance and diversity of the main drivers of many of the planet's key biogeochemical processes, but also of linking the microscopic and macroscopic worlds that have, until now, been largely studied in isolation. We provide a detailed review of the rise of NGS, as well as highlighting areas that offer special promise for addressing general ecological questions across a range of levels of organisation, from individuals to ecosystems: essentially, how a 'systems biology for ecology' might be developed. We consider the current limitations and future prospects for NGS, and also how it offers potential economic benefits, for instance via bioprospecting the environment for commercially valuable genes and their products within the metagenome of natural ecosystems.

I. INTRODUCTION

A. Towards a Systems Biology for Ecology?

All organisms interact with the physical and biological world that surrounds them, from the human gut microbiome to tropical forests and the open ocean. Any study aimed at understanding the development and existence of a multi-organism system falls within the broad remit of ecology, although most of us are more familiar with far a narrower definition. Despite this, ecology has, often out of necessity, been studied in relative isolation from many other fields of biology and has itself been calved into numerous discrete subdisciplines that deal with different levels of organisation (e.g. molecular, organismal, community and ecosystem ecology; Figure 1). In reality, of course, these are rather arbitrary and subjective delimitations, and there are numerous links and feedbacks across organisational levels (Woodward *et al.*, 2010b; Yvon-Durocher *et al.*, 2010b). In recent years ecology has become an increasingly holistic discipline that now draws on, for instance, molecular biology at one extreme and socioeconomics at the other. Here we

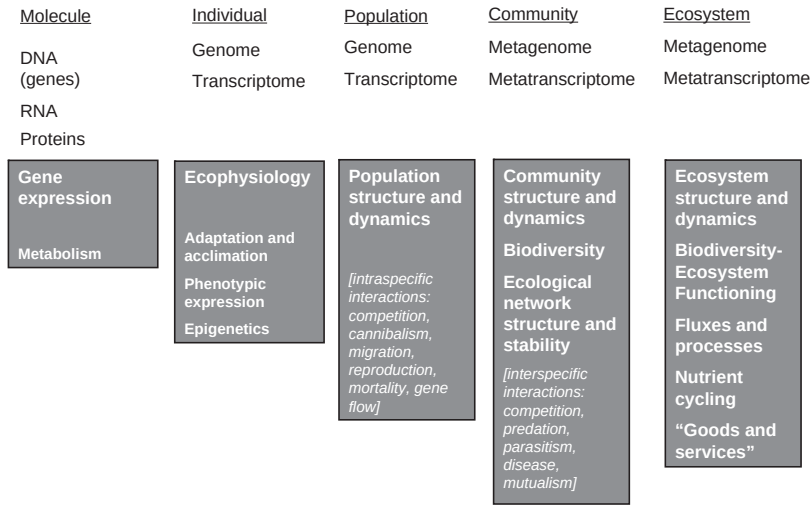


Figure 1 The hierarchy of biological organisation, in increasing order of complexity from molecules to ecosystems. Beneath each heading the relevant genetic and ecological (shaded boxes) measures and fields of study are highlighted.

consider the future of ecology in the context of understanding a system from the chemicals and molecules to the individual organisms that comprise the community and, ultimately, the processes they perform that contribute to the functioning of the ecosystem as a whole. We propose that an inclusive and integrative ecology should be able to encompass all these aspects: essentially, a 'Systems Biology for Ecology' (Figure 2). The opening section of the paper focuses on the dramatic advances that have been opened up in microbial ecology by the advent of new molecular techniques, as a case study, with subsequent consideration of the potential application of these technologies to more general ecology, particularly at the higher levels of organisation.

A major challenge facing ecology is the need to determine how whole ecosystems operate and, hence to be able to predict their responses to future environmental change. The ecological responses at these higher levels of organisation and larger spatial-temporal scales (Figure 1), however, cannot be predicted by simply scaling up from studies of single species in isolation: it is the diversity of species, their ecological roles, and the interactions between them that are key to understanding ecosystem functioning, and we need to understand these links between different levels (Woodward *et al.*, 2010a; Yvon-Durocher *et al.*, 2010b). Achieving this understanding of complex multi-species systems requires far more information than can be obtained by simply measuring bulk stocks and fluxes in the classic Lindeman (1942) tradition of ecosystem ecology: the community and the ecosystem are inextricably interlinked, as emphasised by the plethora of biodiversity–ecosystem functioning (B–EF)

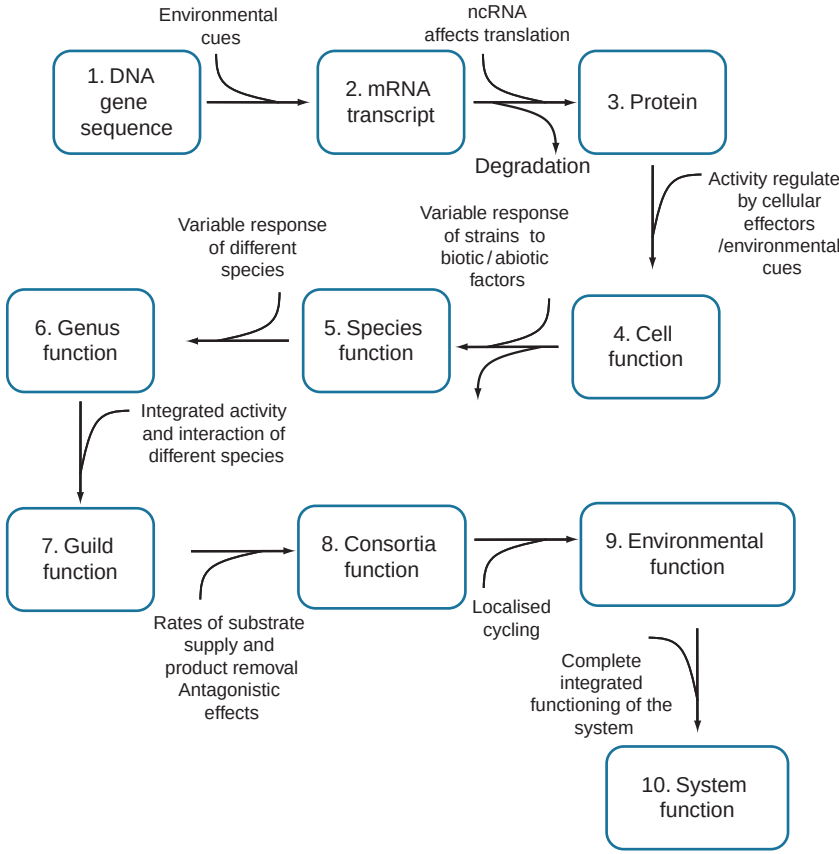


Figure 2 From molecules to ecosystems: a systems biology perspective highlighting a range of potential biotic and abiotic drivers and responses, and their interactions.

studies that have appeared in recent years (Woodward, 2009; Ptacnik *et al.*, 2010 and references therein). The resilience and stability of an ecosystem ultimately depends on the contributions made by its constituent individual organisms, from the smallest viruses and microbes to the largest plants and animals.

Unfortunately, the long-standing dichotomy between community and ecosystem ecology has hindered progress in linking structure to function, with most studies to date being compartmentalised into one or another approach. Much of community ecology, for instance, has focused on the primacy of species identity and diversity as the key variables at this level of organisation, leading some to suggest that it is in danger of being enslaved by the 'the curse of the Latin binomial' (Raffaelli, 2007; Woodward *et al.*, 2010c). Conversely, ecosystems ecology has tended to apply a more broad-brush approach to taxonomy by

focussing on stocks and fluxes of biomass, nutrients, or energy among often coarse aggregations of functional units, which often contain unrelated groups of taxa (e.g. cyanobacteria, diatoms and plants lumped as primary producers). The recent rise of B–EF research has, however, started to address structure–function links explicitly, but primarily from an experimental and macroscopic perspective, rather than in the natural environment (e.g. Perkins *et al.*, 2010). Also most B–EF studies have focused on higher plants or metazoans as the focal taxa of interest, whilst largely ignoring the critical role that microscopic organisms in general, and bacteria and archaea in particular, play in driving ecosystem processes in the real world (but see Reiss *et al.*, 2010a).

Much of the reason for the disparity between microbial community and ecosystem ecology can be ascribed to a methodological bottleneck: until relatively recently it has simply not been possible to characterise microbial diversity accurately, or to link it to specific ecosystem processes, in natural systems. For instance, taxonomic identification of the vast majority of bacterial and archaeal species by microscopy is impossible because of a lack of defining morphological characteristics, while microscopic identification of microbial eukaryotes is extremely time consuming, expensive, and also reliant on the assumption that, once detected, individual cells can be ascribed to species by visual inspection. Genetic techniques have assisted with the latter problem in particular and now underpin most of the taxonomy applied in microbial systematics and ecology. These molecular approaches have also been limited until very recently by the rate of data generation and also potential biases resulting from, for instance, the patchy availability of markers and the lack of quantitative resolution beyond simple presence–absence information on community composition. The use of molecular biology and whole community analysis approaches, such as metagenomics, offer exciting new ways to integrate structure and function, and ecology and evolution (DeLong, 2009; Falkowski *et al.*, 2008; He *et al.*, 2010b). Clearly, however, enormous challenges still remain when attempting to link the physical and biological parameters that characterise the functioning of ecosystems across multiple levels of organisation, but we are now at least able to move considerably faster in this direction than was possible even just a few years ago (Figure 2).

B. The Microbial Black Box

While understanding the functional role of plants and animals in ecosystems is itself a challenge, the microbial components of ecosystems, whether bacteria, archaea or microscopic eukaryotes, have for the most part been confined to a functional 'black box'. Material goes in and comes out altered, but with no comprehension of the processes or identity of the organisms that perform these functions (Vandenkoornhuyse *et al.*, 2010). Given that all biogeochemical cycles depend on the organisms within these black boxes (Falkowski *et al.*,

2008; Zak *et al.*, 2006), the fact that most studies have not explicitly identified the critical drivers of ecosystem functioning represents a fundamental lacuna in our current understanding (Woodward, 2009). In this context, it is widely perceived that the sum of an organism's functional traits defines their contribution to ecosystem functioning (Fitter, 2005). Which traits are important, or how they interact with one another or respond to environmental change, is still poorly defined even for plants and animals (Reiss *et al.*, 2009) and often entirely undefined for the microbial components of an ecosystem.

The primary reason for treating the microbes as a black box is simply that these organisms have not been amenable to study using traditional ecological methods: just 20 years ago, studying natural microbial communities in any real depth and consistency was almost impossible (Fenchel, 1992). The application of molecular tools has revolutionised microbial ecology, such that it is now possible to investigate these communities in a scientifically robust and ecologically meaningful way (Hugenholtz *et al.*, 1998). The immense taxonomic and functional diversity that is still being unearthed in microbial communities (Falkowski *et al.*, 2008; Hugenholtz *et al.*, 1998; Oren, 2004) implies that to truly understand ecosystem functioning it is necessary to view a system across the full range of levels of organisation that affect it. This starts at the molecules that an individual consumes and produces, the genes and proteins that facilitate that usage, and the species populations within the community that carry these capabilities through to the manifestation of processes at the ecosystem level (Figures 1 and 2). There is a need for an analysis of the systems that can integrate across all levels: a systems biology for ecology that works from molecules to ecosystems. The incorporation of molecular and microbial ecology into the more established fields of community and ecosystem ecology is a therefore critical prerequisite for developing such a conceptual framework. At this juncture it is necessary to evaluate briefly how microbial ecology, and the application of molecular techniques within this particular field, has developed over the last 30 years or so, and how these different disciplines have started to converge more recently, before returning to discuss the new research opportunities that are now emerging on the horizon.

II. A BRIEF HISTORY OF MOLECULAR MICROBIAL ECOLOGY

A. Characterising Diversity, Abundance and Functional Traits

Microbes play a key role in the functioning of all ecosystems, but the lack of clear defining characteristics meant microbial ecology was, for many years, perceived as being effectively impossible to study (Stanier and Smith, 1960;

Stanier and van Niel, 1962): even the three most basic questions in ecology (who is there? how many are there? and what do they do?) could not be addressed. However, the field was revolutionised in the late 1970s by the recognition that Zuckerkandl and Pauling's (1965) idea of using molecules as markers of evolutionary history could overcome the lack of defining morphological characteristics that hindered microbial taxonomy, and hence ecology. The application of molecular biology to microbial taxonomy, using the small-subunit ribosomal RNA gene (16S rRNA in bacteria and archaea and 18S rRNA in eukaryotes) revolutionised not just microbiology but altered our view of the phylogeny of all organisms with the proposal of the Three Domains Tree of Life (Fox *et al.*, 1980; Woese, 1983; Woese and Fox, 1977). This replaced Whittaker's Five Kingdoms tree (Whittaker, 1969), dismissed the eukaryotic/prokaryotic dichotomy (Chatton, 1938; Stanier and van Niel, 1962) as a natural division in taxonomy, and showed that the vast majority of biological diversity on the planet was microbial and not in the macroscopic animalia and planta. The so-called Tree of Life is still under constant refinement and revision, via, for instance, the application of phylogenomics and other advanced molecular techniques (Dunn *et al.*, 2008).

The initial sequencing of 16S rRNA sequences led to the development of short 16S rRNA-targeted oligonucleotide probes that could be used to detect microbes of specific phylogenetic lineages (DeLong *et al.*, 1994; Giovannoni *et al.*, 1988). These phylogenetic probes, whether radio- or fluorescently labelled (Amann and Fuchs, 2008), proved that previously unknown and uncultured bacteria and archaea dominated bacterioplankton communities in the global oceans (DeLong *et al.*, 1994; Giovannoni *et al.*, 1996; Morris *et al.*, 2002).

The advent of the polymerase chain reaction (PCR, Saiki *et al.*, 1986, 1988) made previously inaccessible, uncultured natural microbial communities available for analysis for the first time by the amplification of specific gene fragments from either whole communities or specific organisms (Britschgi and Giovannoni, 1991; Giovannoni *et al.*, 1990; Pace *et al.*, 1986). Since then the recognised diversity of the microbial world has increased enormously (Hugenholtz *et al.*, 1998), with many new, previously unrecognised clades of microbes being detected in a host of unexpected and often seemingly very hostile environments, from the rocks of the deep subsurface to high in the atmosphere (Hugenholtz *et al.*, 1998; Oren, 2004). In many cases, however, all that exists of most of these new 'organisms' are short sequences of phylogenetic marker genes, such as the 16S/18S rRNA gene, which give little or no insight into the physiology of the organism detected. PCR analysis of functional genes associated with specific respiratory processes, such as sulphate reduction (*dsrAB*; Joulian *et al.*, 2001; Wagner *et al.*, 1998) or denitrification (*nirS* and *nirK*; Braker *et al.*, 1998; Throbäck *et al.*, 2004) or metabolic processes, such as ammonia oxidation (*amoA*;

Sinigalliano *et al.*, 1995, Ward and O'Mullan, 2005) allowed the diversity of organisms involved in specific processes to be investigated: at last links could start to be forged between microbial community structure and specific ecosystem processes. Further links between community structure and functioning could be made by analysing rRNA or functional gene mRNA using reverse transcriptase PCR (RT-PCR) as the detection of either rRNA or mRNA from an organism is an indication that it is active *in situ* (Bernhard *et al.*, 2007; Chin *et al.*, 2008; Dar *et al.*, 2007; Kandeler *et al.*, 2006; Smith *et al.*, 2007). There is good evidence that the level of rRNA in a bacterial cell is correlated with the level of its activity (Kemp *et al.*, 1993; Rosset *et al.*, 1966), although such a link is less clear for the presence of mRNA and activity (Feder and Walser, 2005).

The more recent development of quantitative PCR has made it possible to address the issues of abundance more directly, by counting the number of specific genes or mRNA transcripts in a sample (Bernhard *et al.*, 2007; Smith *et al.*, 2007). Even so, qPCR counts the number of copies of specific target genes, not organisms themselves. As gene copy number per cell can vary, especially for 16S rRNA genes and even for protein-coding genes (VanGuilder *et al.*, 2008), such data needs to be interpreted with a modicum of caution.

The enormous growth in the number of sequences in the global databases has made it possible to target a large number of organisms using array and microarray technologies. Here short pieces of DNA (probes), either from a specific organism or derived from database analyses, are attached to a slide and then hybridised to DNA and/or RNA extracted from a sample, to produce a fluorescent signal proportional to the amount of binding of the applied DNA/RNA to the probes on the array. Microarray technologies have been used to study microbial communities, as these allow very large numbers of potential organisms or genes to be analysed in a single experiment. Both phylogenetic arrays (DeSantis *et al.*, 2007) and functional gene arrays (He *et al.*, 2007, 2010a) have been used with some success. The major limitations of microarrays are that only those organisms targeted by the probes put on the slide can be detected and arrays tend to have relatively poor dynamic detection ranges and poor signal-to-noise ratios (Bloom *et al.*, 2009; van Vliet, 2010).

Despite having access to markers for the presence and activity of microbes it is still difficult to make links between the structure of a community (who is present, and in what numbers) and the processes taking place *in situ* (what ecological roles are being performed): taking microbial ecology out of the laboratory and into the natural environment is still a challenging step. Specific analytical methods have been used to make such links, including the uptake of stable and radio isotopes of tracer molecules or by the experimental manipulation of communities to determine the effects of added

substrates (Boschker *et al.*, 1998; Lee *et al.*, 1999; Manefield *et al.*, 2002; Ouverney and Fuhrman, 1999; Purdy *et al.*, 2002, 2003b; Radajewski *et al.*, 2000) with some notable successes. However, these are still inevitably focused on specific processes and organisms and involve perturbations and manipulations of communities, rather than assessing structural–functional linkages in natural ecosystems.

B. Genomics and Post Genomics in Microbial Ecology

A further major step forward in molecular microbial ecology came with the advent of genome sequencing, initially of single strains of organisms (Fleischmann *et al.*, 1995) and then of genomic DNA fragments extracted directly from the environment (Handelsman, 2004). These new environmental analyses, initially called environmental, community or functional genomics, or ecogenomics, but now generally known as metagenomics (Handelsman, 2004; Tringe and Rubin, 2005), largely avoids the selective problems associated with PCR-based methods and could conceivably allow access to the whole gene pool within an environment. This has opened up a vast new range of applications and research vistas and has finally placed molecular and microbial biology firmly within the core of general ecology, much of which has been dominated for decades by community-level studies of macro-organisms. The potential to completely sequence all the genes in an environment suggested that it was also possible to define all the metabolic pathways that could occur within a system (Buckley, 2004). Metagenomics was rapidly followed by the application of post-genomics technologies, transcriptomics, proteomics and metabolomics (DeLong, 2009 and references therein) and the potential to define not just what the microbes in a system could potentially do but also what they were actually doing. Thus, it appeared that all of the barriers that had stood in the way of microbial ecology were finally falling, due to the application of large scale analytical methods that simply sequenced everything to expose the capacity and functions of all of the organisms within a sample (DeLong, 2009). Arguably, the challenge now is not so much the difficulty of extracting sufficient data, but rather the need for advanced bioinformatics approaches to handle the vast amounts of information that can now be generated rapidly from even a small number of samples.

Although they have provided important breakthroughs in their own right, these technical advances and sophisticated methods do not in themselves create the understanding and integration across different ecological levels that is needed to build a true systems-based approach to ecology (Vandenkoornhuyse *et al.*, 2010). This could be provided, however, by the range of next generation sequencing (NGS) technologies that have appeared

very recently and which seem destined to make a major contribution to ecology. Given their potential to generate significant paradigm shifts in all spheres of ecology, but particularly in integrating community and ecosystems approaches, we will consider NGS technologies in more detail in the context of what they are, what they offer, and the prospects for future advances.

III. NEXT-GENERATION SEQUENCING TECHNOLOGIES

A. A New Paradigm

Occasionally a new technology is developed that revolutionises biology. Between 1975 and 1977 Sanger and colleagues devised a method of DNA sequencing based on dideoxynucleotide chain-termination (Sanger and Coulson, 1975; Sanger *et al.*, 1977a,b). With subsequent technological developments (Hunkapiller *et al.*, 1991; Smith *et al.*, 1986), this method resulted in the successful sequencing of the ~3 billion base pairs of the human genome (Collins *et al.*, 2004; Lander *et al.*, 2001) after 13 years of work by hundreds of researchers and hundreds of DNA sequencing machines running in parallel, at an estimated cost of $3 billion. For comparison, in 2009, a human genome was sequenced for just $48,000 by three people using only four DNA sequencing machine runs (Pushkarev *et al.*, 2009); an indication of the exponential increase in sequencing power and declining costs over that period (Figure 3). This monumental shift in both cost and efficiency is based on new chemical techniques, miniaturization, parallelization, high-throughput and technological advances in computing and image detection, which in combination are commonly referred to as NGS. The paradigm shift associated with NGS is rapidly changing the field of biology at all levels of organization, from molecules to ecosystems, and promises to alter the way ecologists view the natural world in many unexpected ways.

There are currently four commercially available NGS platforms, and one near-commercial platform, which differ in both protocols and in the nature of the data generated. All have particular strengths and weaknesses, with some more suited to specific biological questions than others. Here we present a brief introduction and technical review, to highlight these new technologies, and discuss the relative merits and potential applications of NGS in an ecological context.

NGS requires only a few microgram quantities of DNA, and therefore the time-consuming cloning into bacterial hosts followed by multiple rounds of PCR amplification can be bypassed. Instead, the template is isolated directly from the source, fragmented, amplified in a limited way (or not at all), and

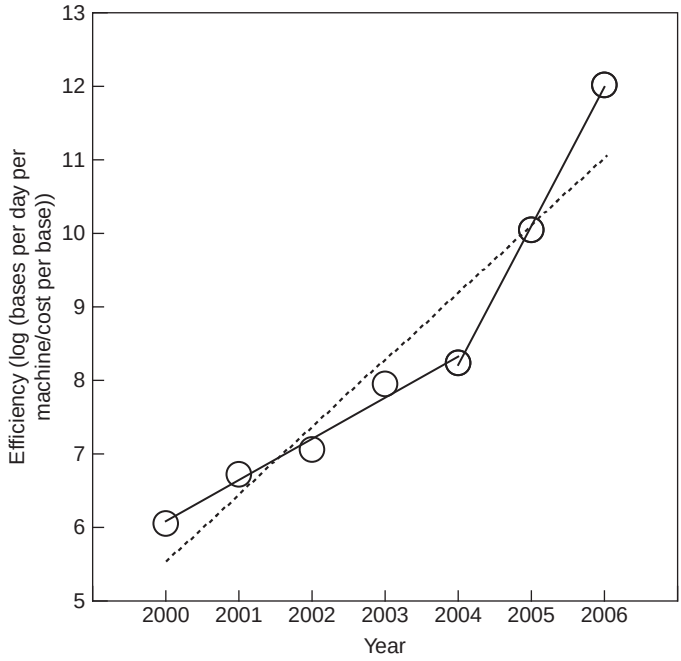


Figure 3 More bang for your buck: the rapid growth of sequencing power and declining costs over the last decade, expressed as an exponential increase in efficiency over time (data extracted from Hudson (2008)). On average there is close to an order of magnitude ($\beta = 0.9$) increase in efficiency per year over the time series (dashed line; $r^2 = 0.906$). However, the relationship is better described by two separate regressions, shown as the solid lines ($r^2 = 0.978$ and $r^2 = 0.999$, respectively), as a step change occurs from 2004 onwards as NGS technologies emerge and improvements in efficiency increase dramatically, with the scaling coefficient increasing from $\beta = 0.6$ to 2 orders of magnitude per year.

attached to a solid surface or support (see Figures 4–8). Either clonally amplified templates from single DNA molecules or single DNA-molecule (non-amplified) templates can be used here, with the former being derived from either emulsion PCR (emPCR) (Dressman *et al.*, 2003) or solid-phase amplification (Fedurco *et al.*, 2006). NGS platforms use three types of mechanism for sequencing the templates: (1) single-nucleotide addition (454 pyrosequencing); (2) cyclic reversible termination (CRT) (Illumina/Solexa and Helicos) and (3) real-time sequencing (Pacific Biosciences). Sequencing by ligation (ABI SOLiD) is also discussed. The key feature of all these methodologies is sequencing by synthesis: that is, the DNA templates are sequenced at the same time as DNA polymerization takes place. A description of these platforms is given below and a brief outline of some of their features and how they have been used is given in Table 1.

Table 1 Comparison of next-generation sequencing platforms and their advantages and disadvantages in the context of ecological research

Platform	454 Pyrosequencer GS FLX Titanium (Roche)	Solexa GA_{II} (Illumina)	HeliScope (Helicos BioSciences)	SOLiD 4 (Life/Applied Biosystems)	SMRT (Pacific Biosciences)
Library/template preparation	Fragment Mate-pair emPCR	Fragment Mate-pair Solid-phase	Fragment Mate-pair (Single molecule)	Fragment Mate-pair emPCR	Fragment (Single-molecule)
Chemistry	Single-nucleotide addition	Cyclic-reversible termination	Cyclic-reversible termination	Sequencing by ligation	Real-time
Read length (basepairs)	~400	≤150 (average 75)	≤55 (average 35)	≤50	964 (average)
Gbp per run	~0.4	Up to 18	Up to 35	Up to 100	N/A
Pros	• Fast run times • Longer reads (improve mapping)	• Most widely used platform • Fast run times	• Single-molecule sequencing, therefore non-biased representation of templates	• Two-base encoding provides inbuilt error correction	• Longest read-length • Single-molecule sequencing, therefore non-biased representation of templates
Cons	• High error rates in homopolymer repeats.	• Short reads	• High error rates • Short reads	• Long run times • Short reads	• High error rates
Applications	• *De novo* genome assembly • Metagenome sequencing • PCR amplicon sequencing	• Variant discovery by whole-genome resequencing • Seq-bases methods (RNA-Seq and ChIP-Seq)	• Seq-based methods (RNA-Seq and ChIP-Seq)	• Variant discovery by whole-genome resequencing	• *De novo* genome assemblies • Full-length transcriptome sequencing

Microbiome studies (within-organism ecosystems)	• Analysis of obesity associated gut microbiome (Turnbaugh *et al.*, 2006) • Oral microbiome (Zaura *et al.*, 2009)	• Human gut microbial gene catalogue (Qin *et al.*, 2010)	• None at present	• Used to confirm the role of a proposed causative agent of a disease in citrus trees (Tyler *et al.*, 2009)	• None at present
Environmental ecosystems	• Functional profiling of nine biomes (Dinsdale *et al.*, 2008) • Analysis of honeybee colony collapse disorder (Cox-Foster *et al.*, 2007)	• Genome sequencing from enrichment culture of novel microbe (Ettwig *et al.*, 2010)	• None at present	• None at present	• None at present

B. 454 Pyrosequencing (Roche)

The 454 system was the first commercially available NGS platform (Margulies *et al.*, 2005). Genomic DNA is randomly fragmented and universal DNA adaptors of known sequence are ligated to the free ends (Figure 4A). The resultant DNA fragments are denatured into single strands and captured on micro-beads coated with complementary adaptors under conditions that favour the addition of only one DNA molecule per bead (Figure 4B). Each DNA template is clonally amplified using emPCR. This reaction takes place within single aqueous droplets, each containing a single micro-bead with a bound DNA template (Figure 4C). Following amplification, each micro-bead, containing millions of identical copies of the DNA template, is enriched and deposited onto an array (Figure 4D). It is here that the DNA sequencing takes place via single-nucleotide addition, in a method known as pyrosequencing (Ronaghi *et al.*, 1998). Additional beads carrying the enzymes sulphurylase and luciferase are added to each well followed by the introduction of a single-nucleotide species, which flows across the wells (Figure 4E). Where this results in the incorporation of nucleotide, pyrophosphate is released which, in an enzymatic-catalyzed reaction, is converted into a burst of light (Figure 4F) whose intensity is proportional to the number of incorporated nucleotides and recorded as a series of peaks called a flowgram, from which a sequence can then be determined (Figure 4G). The sequencing cycle is repeated with each of the four nucleotides added sequentially until the desired read length is attained.

Using the GS FLX Titanium platform, up to 1 million beads, each coated with a clonally amplified DNA molecule, are pyrosequenced in parallel. With individual sequence read lengths of up to 400 bases, a single run can generate 400 Mb of sequence in 10 h. Using an improved sequencing chemistry, Roche currently report an average read length of 657 bases, which would boost coverage to over 650 Mb per run. The long reads (400+ bps) now being produced by 454 sequencing are sufficient to allow robust phylogenetic analysis of rRNA or functional genes. The capacity to acquire data that can be analysed robustly without the need to assemble contigs greatly reduces analysis time and increases the level of confidence with which such data can be interpreted. These are important advantages in ecological studies that may involve analyzing many samples, and are particularly well-suited to quantifying biodiversity in natural or experimental communities.

C. Illumina Genome Analyzer

Commonly referred to as ‘Solexa’, this platform currently dominates the market (Bentley *et al.*, 2008). DNA libraries are constructed from fragmented genomic DNA and clonal amplification is achieved via solid phase.

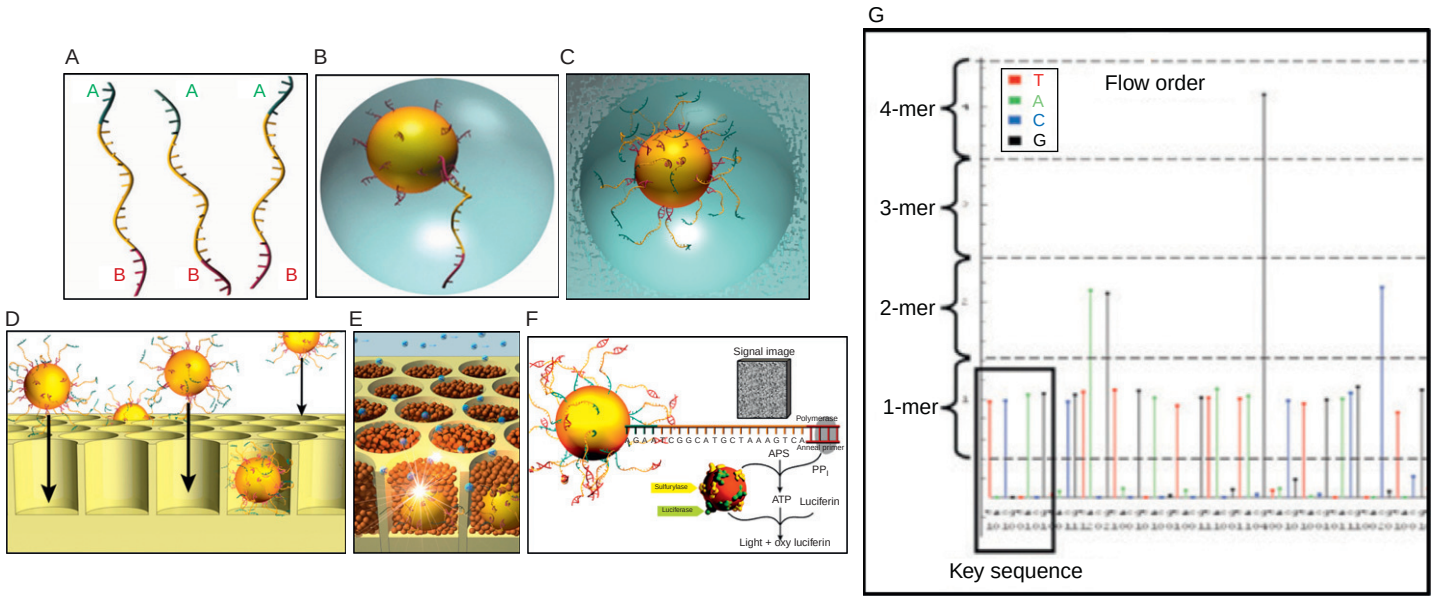


Figure 4 Outline of the GS FLX titanium 454 sequencing workflow. (A) Randomly fragment genomic DNA and ligate to 454-specific adapters (indicated as A and B). (B) The DNA is separated into single strands and captured onto beads through hybridisation of adapter sequences with primer sequences coating the bead. Conditions are used that favour one DNA molecule per bead. (C) An oil-aqueous emulsion is created to encapsulate single bead-DNA complexes within a single aqueous drop. emPCR is performed within these drops to create beads containing thousands of copies of the same DNA template. (D) Beads are then loaded into individual PicoTitre wells. (E) Additional beads coupled with sulphurylase and luciferase are added to each well, followed by a single type of nucleotide (in this case adenine). Incorporation of the nucleotide generates light, which is recorded by a CCD camera located underneath the PicoTitre plate. The cycle is then repeated with the other three nucleotides in a step-wise, pre-determined order until the desired read-length is achieved. (F) The pyrosequencing reaction. (G) The light generated by the enzymatic cascade is recorded as a series of peaks called a flowgram.

Single-stranded DNA fragments are ligated to universal adapters of known sequence (Figure 5, Step 1) and randomly arrayed on a glass slide containing adapters which also act as primers for the sequencing reaction (Figure 5, Step 2). Each DNA fragment is clonally amplified by bridge amplification to

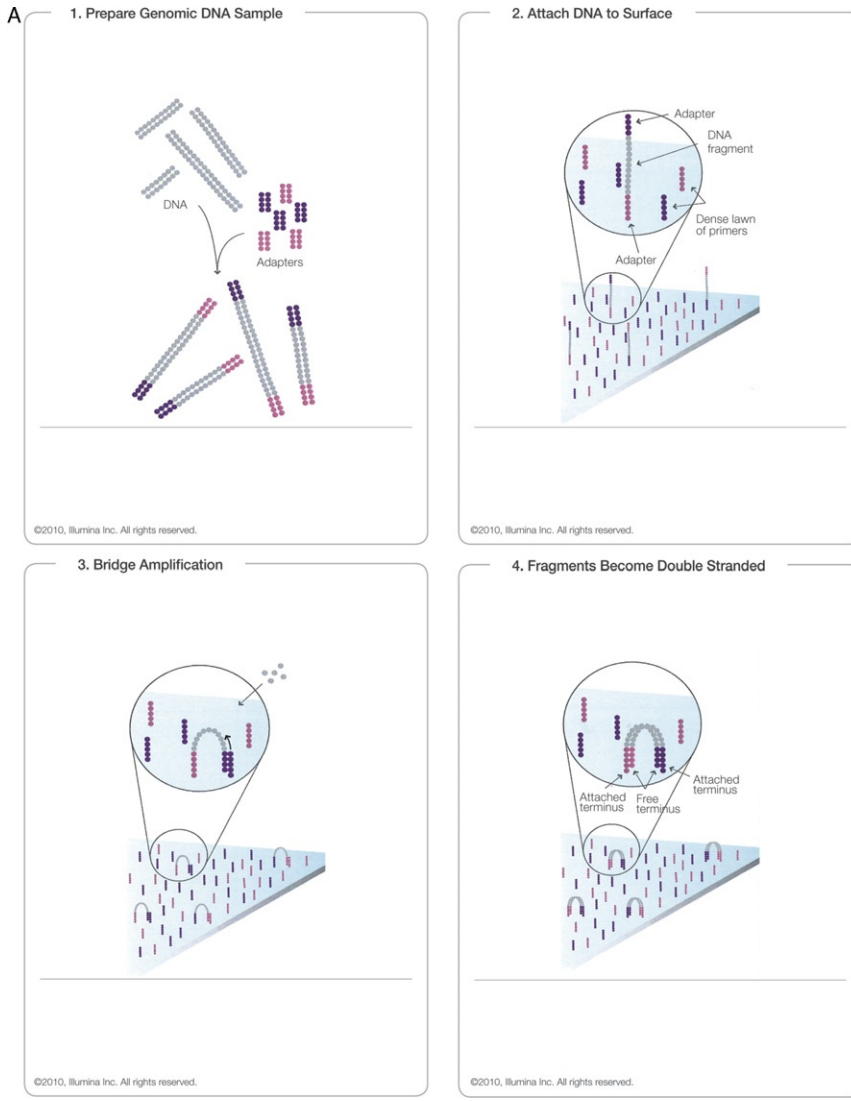


Figure 5 (Continued)

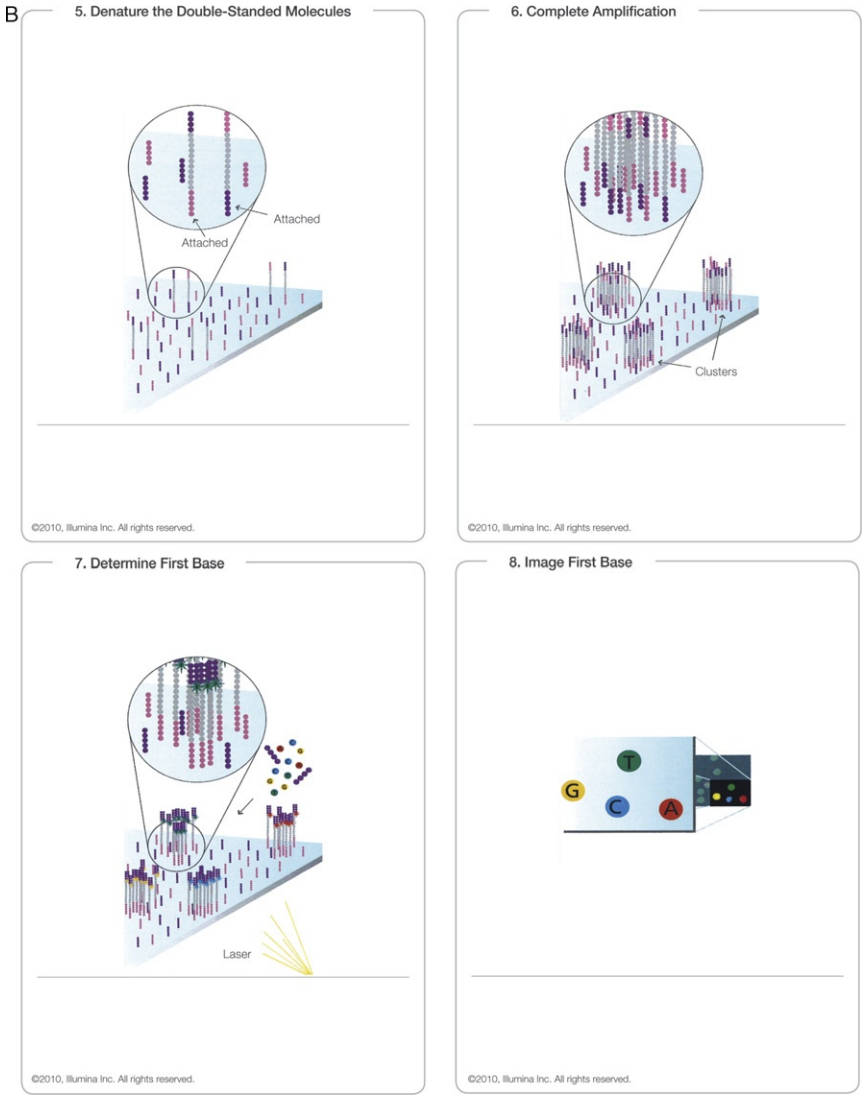


Figure 5 (Continued)

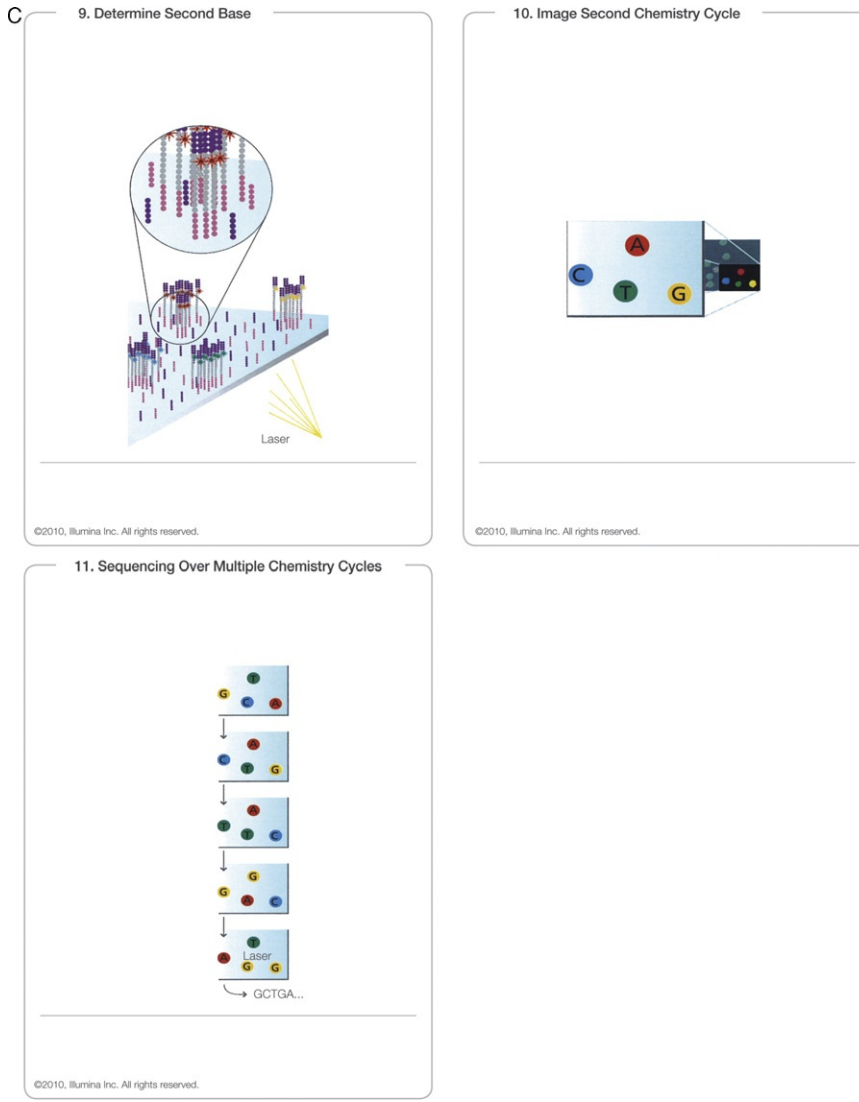


Figure 5 Outline of the Illumina Genome Analyzer workflow. (1) Randomly fragment genomic DNA and ligate adapters to both ends of the fragments. (2) Bind single-stranded fragments randomly to the inside surface of the flow cell channels. (3) Add unlabelled nucleotides and enzyme to initiate solid-phase bridge amplification. (4) The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate. (5) Denaturation leaves single-stranded templates anchored to the substrate. (6) Several million dense clusters of double-stranded DNA are

produce clusters (Figure 5, Steps 3–6). Eight independent lanes are contained on a single flow-cell, meaning eight independent libraries can be sequenced in parallel during the same instrument run, thus reducing time and reagent costs, whilst increasing throughput.

After cluster generation, the single-stranded amplicons are sequenced by CRT (Figure 5, Step 7). All four dye-labeled modified nucleotides containing reversible blocking groups are added in a cyclical process that includes nucleotide incorporation, fluorescence imaging and cleavage (Bentley *et al.*, 2008). Due to the presence of the blocking group on the modified nucleotide, DNA polymerase adds just one nucleotide, complementary to the template strand, before synthesis terminates. Unincorporated nucleotides are washed away and the incorporated nucleotide is detected by laser imaging (Figure 5, Step 8). A cleavage step then removes the blocking group from the incorporated nucleotide, allowing another modified nucleotide to be added and the cycle is repeated (Figure 5, Steps 9 and 10). In the Illumina/Solexa method, each of the four nucleotides carries a different fluorophore, allowing discrimination of the incorporated base. By repeated cycles of this method with imaging after cycle, the DNA sequence of a particular cluster of templates is determined by computer analysis of that cluster on each sequential image (Figure 5, Step 11).

Read lengths of up to 150 bases are already routinely possible but, to date, generally average about 75 bases, with the number predicted to rise rapidly. The recently introduced Genome Analyzer$_{IIx}$ can generate up to 320 million reads (640 million paired-end reads), with a total throughput of up to 6.5 Gb per day, the equivalent of two complete human genome a day. The massive sequencing capacity and reasonable read-length delivered by Solexa makes deep sequencing of many samples possible, as shown by Qin *et al.* (2010) in an analysis of the human gut microbiome of over 100 people. It is also extremely useful in analyzing small molecules such as non-coding RNA (ncRNA (Perkins *et al.*, 2009)).

generated in each channel of the flow cell. (7) The first sequencing cycle begins by adding four differently labelled reversible terminators, primers and DNA polymerase. (8) After laser excitation, the emitted fluorescence from each cluster is captured and the first base is identified. (9) The next cycle repeats the incorporation of four labelled reversible terminators, primers and DNA polymerase. (10) After laser excitation, the image is captured as before, and the identity of the second base is recorded. (11) The sequencing cycles are repeated to determine the sequence of bases in a fragment, one base at a time until the desired read-length is achieved. The data are subsequently aligned and compared to a reference, and sequencing differences are identified. Images copyright of Illumina, Inc. (2010). All rights reserved. Used with permission of Illumina, Inc.

D. HeliScope

A more recent advancement by Helicos Bioscience Corporation has seen the introduction of NGS based on single-molecule templates (Harris *et al.*, 2008). Here, no amplification is required and <1 μg of starting material can be analysed. The key advantage of single-molecule sequencing is that no PCR-

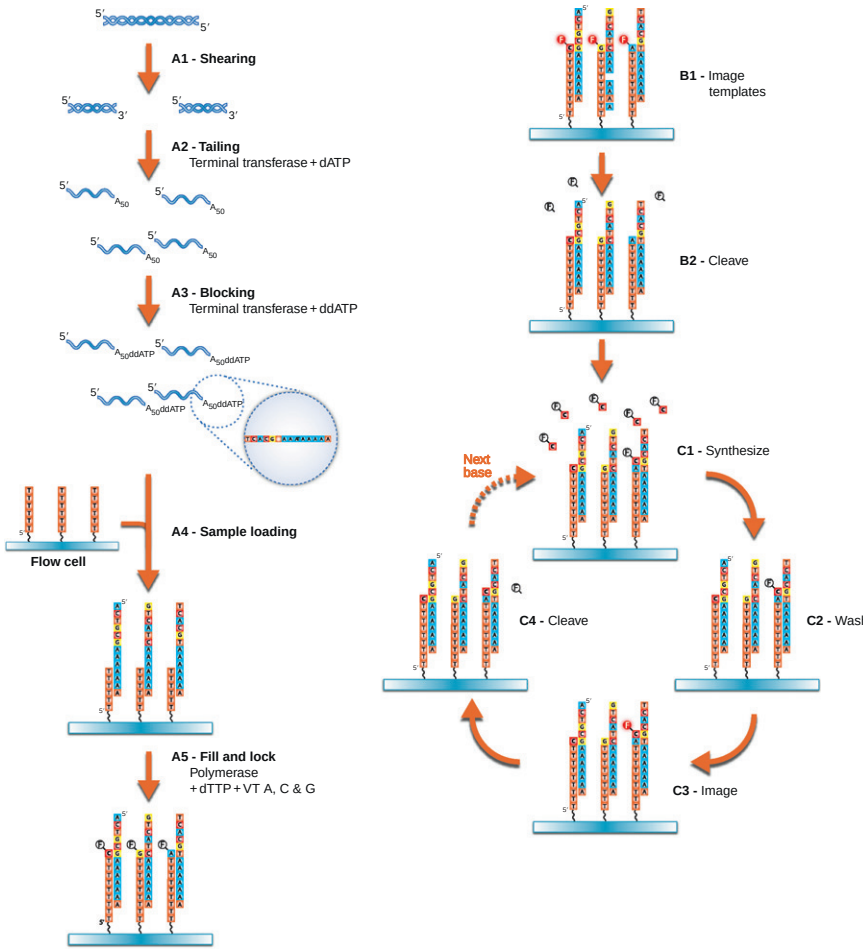


Figure 6 Outline of the HeliScope workflow. (A) 1. Genomic DNA is randomly fragmented. 2. DNA fragments are separated into single strands and a polyA universal priming sequence is added to the 3′-end using Terminal Transferase. 3. Each strand is then blocked by addition of a modified adenosine nucleotide. 4. These labelled DNA strands are then hybridised to the Helicos flow cell, which contains millions of oligo-dT

based errors or amplification biases are introduced into the sequencing template. Genomic DNA is randomly fragmented and single-stranded DNA molecules are generated, to which polyA tails are added in an enzyme-catalysed reaction (Figure 6, Steps A1–A3). DNA templates are then immobilised on a solid support attached to the flow cell, which contains spatially distributed primers (oligo-dT), complementary to the polyA tails (Figure 6, Step A4).

The CRT method of sequencing differs from that provided by Illumina/Solexa since each dye-labeled nucleotide is added sequentially. Following incorporation of the first dye-labeled nucleotide, a laser illuminates the surface of the flow cell, which determines the location of each fluorescently labeled single-molecule template. A camera then produces a map of the flow cell surface by taking sequential images across it. A cleavage step then removes the dye and the blocking groups to permit the addition of the next dye-labeled nucleotide in the following cycle (Figure 6, Steps B2–C4). This sequencing cycle is repeated with the other three dye-labeled nucleotides, with the number of cycles determining read length. The DNA sequence of a particular single molecule is determined by analysis of the images taken after each sequencing round. Read lengths of up to 55 bases (average read length 36 bases) have been reported with up to 35 Gb per run and as much as 1 Gb per hour.

This technology can produce 10 times as much data per run as the Solexa sequencer but the very short reads are best suited to analyses of small molecules and transcriptomes of previously sequenced organisms. While it is still difficult, at present, to see obvious broad ecological applications, it has been used to accurately quantify a yeast transcriptome over 4 orders of

capture sites that are immobilised on a solid-support. 5. The extra adenosine nucleotides on the template strands are filled in by addition of thymine nucleotides, and the template strand is locked in place at the first non-adenosine base by the addition of Virtual Terminator (VT) nucleotides. This also serves to label the templates. (B) 1. A laser then illuminates the surface of the flow cell showing the location of each fluorescently labelled template. A CCD camera then produces a map of the templates on the flow cell surface by taking a sequence of images across the flow cell in a step-wise pattern. 2. After the templates have been imaged the template label is cleaved and washed away. (C) 1. The sequencing reaction begins by the introduction of DNA polymerase and a fluorescently labelled nucleotide (in this case cytosine). The labelled nucleotide is incorporated in a template-dependent manner. 2. A wash step removes the DNA polymerase and any unincorporated labelled nucleotide. 3. The templates that have incorporated a cytosine are then visualised by illuminating and imaging the entire flow cell surface. 4. After imaging a cleavage step removes the fluorescent labels and the process (C1–C4) continues by addition of each of the other three labelled nucleotides in a step-wise, pre-determined manner and repeats until the desired read-length is achieved. Images copyright 2010. Helicos BioSciences Corporation. All rights reserved. Used with permission of Helicos BioSciences Corporation.

magnitude (Lipson *et al.*, 2009), so it could be especially useful for characterising the functioning and gene expression of certain key microbial taxa.

E. SOLiD (Life/Applied Biosystems)

In the Support Oligonucleotide Ligation Detection (SOLiD) system, sequencing is achieved by using a DNA ligase rather than a polymerase (Valouev *et al.*, 2008). DNA libraries are prepared and amplified on microbeads by emPCR (as in 454 pyrosequencing, see Figure 4A–C). Sequential rounds of ligation to a collection of fluorescently labeled probes (octamers) are performed, with imaging used to determine their identity after every round of probe hybridization (Figure 7A, Steps 1–4). The octamers are structured such that the identity of specific positions within the probes correlate with the identity of the fluorescent label (see Figure 7B). After ligation, the octamer is chemically cleaved to remove the fluorescent label, and the cycle is repeated nine more times to give 10 colour calls (Figure 7A, Steps 1–4). The number of ligation cycles determines the eventual read length. The primer is then removed and a second round of ligation is performed with an '$n-1$' primer, which shifts the interrogation bases and the corresponding colour calls one position to the left (Figure 7A, Steps 5 and 6). Nine more ligation cycles follow, followed by three more rounds of ligation cycles with 'base-shifted' primers ($n-2$, $n-3$ and $n-4$). Colour calls from the five ligation rounds are then arranged into a linear sequence (Figure 7A, Step 7). Using the SOLiD 4 system, read lengths of up to 50 bases are currently possible, with an expected throughput of up to 100 Gb per run.

As with the HelioScope the short reads of the SOLiD system make widespread ecological applications more difficult to envisage than with 454 pyrosequencing, but the technology has been used, along with Solexa sequencing, to analyse ncRNA in grapevines (Picardi *et al.*, 2010) and to identify the causative agent of a disease in citrus trees (Tyler *et al.*, 2009).

F. SMRT Technology (Pacific Biosciences)

The near-commercial SMRT technology from Pacific Biosciences offers a further revolution in DNA sequencing (Eid *et al.*, 2009). Like the HelioScope, single DNA molecules are sequenced but in real-time, such that nucleotides do not halt the DNA polymerase and synthesis. Incorporated dye-labeled nucleotides are imaged continuously during the latter process. In this approach, genomic DNA is randomly fragmented and hairpin structures

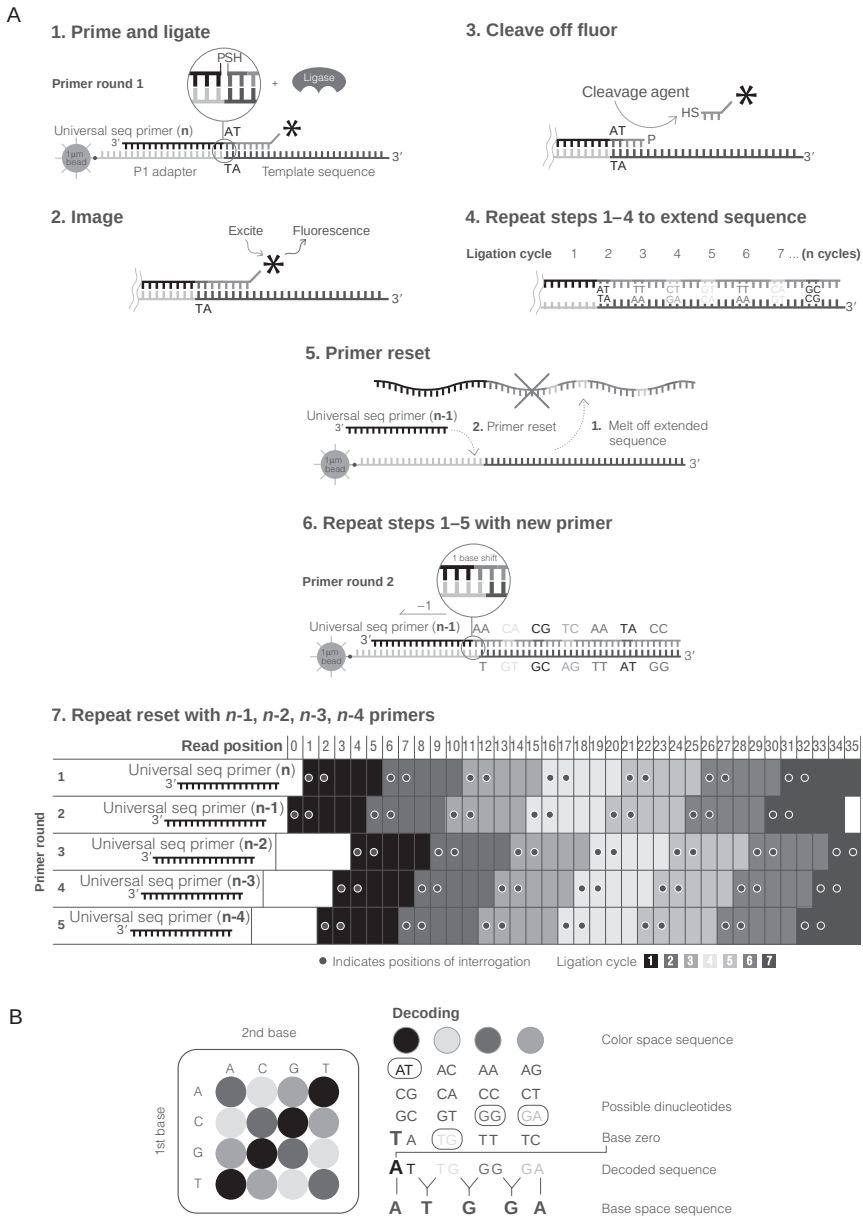


Figure 7 Outline of the SOLiD 4 workflow. DNA libraries are prepared and amplified on beads by emPCR (see Figure 4A–C). (A) Step 1: Universal sequencing primers hybridise to the P1 adapters. Libraries of fluorescently labelled 1,2-nucleotide probes compete for ligation to the template strand. The 1,2-nucleotide probes are designed to

are ligated to both ends of the DNA (Figure 8A). Single molecules of a strand-displacing bacteriophage DNA polymerase are attached to a solid support and template molecules are added to the immobilised polymerase (Figure 8B) within individual nanostructures (zero-mode waveguide [ZMW]; Levene *et al.*, 2003) with an observation volume of just 10^{-21} l. At this volume, the activity of a single molecule can be detected (Figure 8C) and the real-time incorporation of one of four differently dye-labeled nucleotides can be observed by monitoring the fluorescence pulse (Figure 8D).

With reported average read lengths of 964 bases (Metzker, 2010), and with each ZMW capable of producing up to 400 kb of sequence per day, an array of only 14,000 ZMWs would be required to produce the equivalent of one human diploid genome per day (Eid *et al.*, 2009). Currently SMRT DNA sequencing is performed on proprietary SMRT Cells, each having an array of approximately 75,000 ZMWs: that is, offering the potential detection of approximately 75,000 single-molecule sequencing reactions in parallel. As this technology has not yet been released onto the market there are currently no published reports of its use in any area of ecology, however, the very long reads could enable deep sequencing of samples that produce robust data of environmental samples. As with all new technologies, the real test is not in the claims of what it can do but what it actually does when faced with real samples.

G. Improving Analysis of Multiple Samples and Increasing Read Length in NGS Runs

Various strategies exist in order to utilise the sequencing capacity of NGS platforms fully. Although application-dependent, for many studies, including organisms with small genomes or for those involving the genome-wide

interrogate the first and second positions adjacent to the sequencing primer (shown here as TA), such that the 16 possible di-nucleotides are encoded by four dyes (see also Figure 7B). Step 2: Following ligation, fluorescence is detected by laser excitation during four-colour imaging. Step 3: The ligated 1,2-dinucleotide probes are chemically cleaved between the fifth and sixth nucleotides with silver ions to generate a free 5′-phosphate group. Step 4: The SOLiD cycle (Steps 1–3) is then repeated nine more times. Step 5 and 6: The extended primer is then stripped off and a second round of 10 ligations is then performed with a universal sequencing primer one base shorter than the previous one ('$n-1$'). Step 7: Three more rounds of ligation cycles are performed with $n-2$, $n-3$ and $n-4$ primers. Colour calls from the five ligation rounds are then arranged into a linear sequence. Through the primer reset process, virtually every base is interrogated in two independent ligation reactions by two different primers. For example, the base at read position 5 is assayed by primer number 2 in ligation cycle 2 and by primer number 3 in ligation cycle 1. (B) A two-base encoding scheme in which four di-nucleotide sequences are associated with one colour. The decoding matrix allows a sequence of di-nucleotides to be converted to a base sequence, as long as one of the two bases is known. Images copyright 2010. Life/Applied Biosystems. All rights reserved. Used with permission of Life/Applied Biosystems.

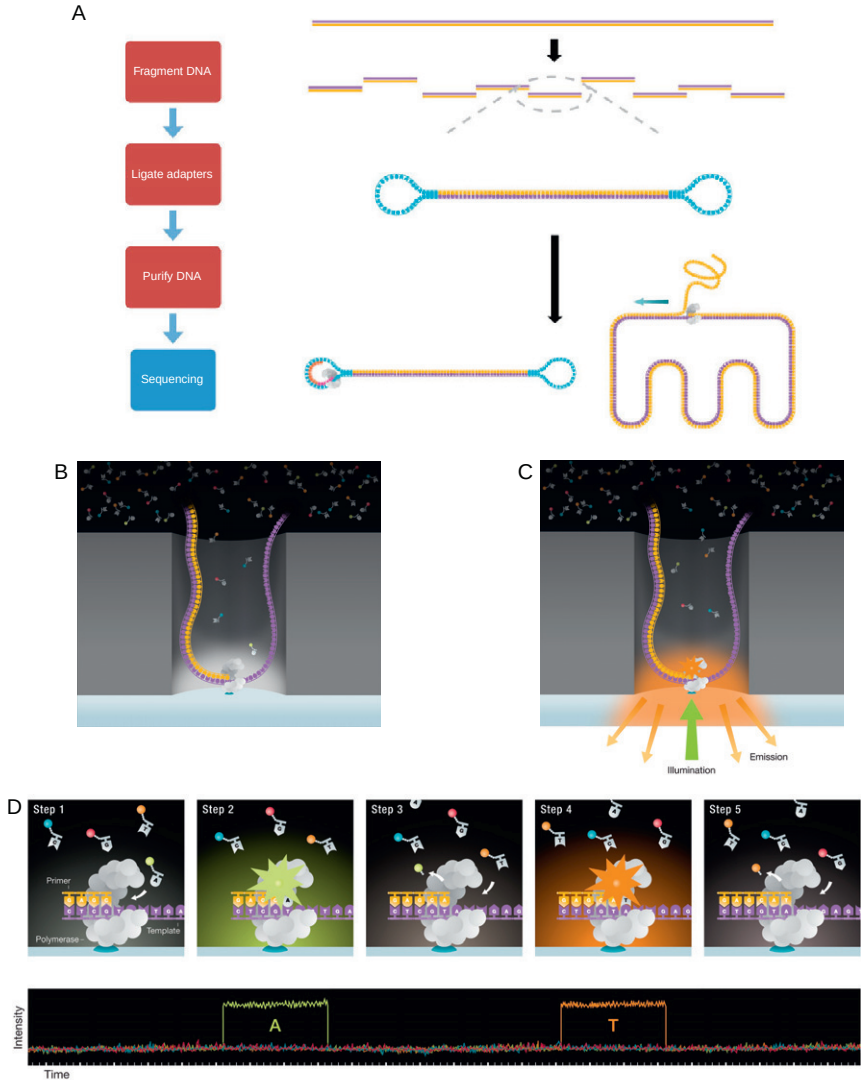


Figure 8 Outline of the SMRT workflow. (A) Genomic DNA is randomly fragmented and hairpin adapters are ligated to both ends. The templates can then go through the sequencing reaction. A strand displacing bacteriophage DNA polymerase opens up the DNA structure into a circular template and can generate independent reads, both forward and reverse of the same DNA molecule. (B) The bacteriophage DNA polymerase is immobilised at the bottom of each ZMW, nucleotides diffuse into the ZMW chamber and are processively incorporated in real-time to produce a long continuous template-directed DNA molecule. (C) Only the bottom 30 nm of the ZMW is illuminated, therefore only those nucleotides near the bottom fluoresce. Both excitation and detection occur without interruption through the transparent

mapping of DNA-associated factors where only a fraction of a genome is sequenced (Hurd and Nelson, 2009), the full capacity of NGS platforms is not required. Furthermore, there are also limitations to the number of different samples that can be processed in parallel at any one time. Two recently reported adaptations to NGS protocols that employ a multiplex (or bar-coding) strategy combined with either the 454 pyrosequencing platform (Meyer *et al.*, 2007, 2008) or the Solexa/Illumina platform (Lefrancois *et al.*, 2009) offer a solution to both these issues. Each uses a bar-coding system, whereby unique 'tags' are included in the oligonucleotide adapters ligated to DNA library fragments. After individual library generation, libraries can be pooled and sequenced in parallel and the tags used subsequently to identify the sample origin of individual reads. Bar-coding of multiple samples reduces the depth of sequencing (number of sequences reads per sample) for each sample but this can be extremely useful when using an NGS technology to analyse an amplified gene, where several thousand sequences will probably prove to be sufficient for a comprehensive analysis. Such an approach has been used recently in a temporal analysis of a lake protist community Nolte *et al.* (2010).

Mate-paired reads also offer increased sequencing reads per machine run. This strategy is particularly useful for the *de novo* assemblies of genomes and in the analysis of structural variants (SV) between related genomes (see Section V and references therein). Here, DNA libraries are prepared with fragments of a known size, with sequencing directed from both ends of the DNA fragments, rather than from just one, thus doubling the amount of sequence obtained. Since the sequenced ends are of a known distance apart, reads can be aligned to a reference genome and distances compared (as with SV analysis), or overlapping mate-pair libraries can be used to build larger contiguous sequences of DNA for *de novo* assembly.

These methods can be useful in an ecological context, where for instance we might wish to split a previously pooled sample to identify, for instance, replicates within an experimental treatment, dates within a seasonal study or quadrats positions within a spatial survey. The great advantage of this is that the efficiency of the yield-to-cost ratio can be tweaked to acquire the optimum output, to keep running costs down whilst also minimizing machine time, by running multiple samples simultaneously (e.g. replicates can be marked, pooled, sequenced, and then subsequently disentangled).

---

glass bottom of the SMRT Cell. (D) Step 1: Fluorescent phospholinked labeled nucleotides are introduced to the DNA polymerase. Step 2: The base being incorporated is held in the detection volume for tens of milliseconds, producing a bright flash of light, which is recorded as a pulse. Step 3: The phosphate chain is cleaved, releasing the attached dye molecule. Steps 4 and 5: The process repeats.

IV. GENOME SEQUENCING: FUNCTIONAL DIVERSITY IN ECOLOGY

Once DNA sequences have been generated, they can be aligned to a known reference sequence or assembled *de novo*. The alignment method is useful for identifying genetic variation, SV, or single-nucleotide variants (SNV) in highly related genomes. For example, genome variation has been analyzed in this way for various species of bacteria (Moran *et al.*, 2009; Nusbaum *et al.*, 2009; Srivatsan *et al.*, 2008) and eukaryotes (Hillier *et al.*, 2008; Ossowski *et al.*, 2008; Sarin *et al.*, 2008; Shen *et al.*, 2008), including humans (Ahn *et al.*, 2009; Bentley *et al.*, 2008; Kim *et al.*, 2009; Wang *et al.*, 2008). The SOLiD and HeliScope platforms have also been used to sequence the human genome (McKernan *et al.*, 2009; Pushkarev *et al.*, 2009). Within these new technologies there exists the capacity to analyse a sample to detect the real extent of diversity within specific organismal groups. This is particularly relevant for microbial populations where small differences between strains can be critical to their functional role in the environment (see references above).

Alignment of sequencing reads to complex genomes can be complicated by the presence of repetitive regions. Mate-pair reads are commonly employed to overcome this particular issue, where one read of the pair is unique to the genome, and in this way reads can be anchored to the reference genome. This approach also allows for the identification of SVs: larger regions of genomic variation such as substitutions, insertions, deletions, inversions, duplications and copy-number variations (Korbel *et al.*, 2007).

Existing NGS platforms generate relatively short reads, and therefore *de novo* assemblies have so far been limited to relatively small genome sizes, mostly among bacteria and archaea (Chaisson and Pevzner, 2008; Hernandez *et al.*, 2008; Margulies *et al.*, 2005). More recently Li *et al.* (2010b) published the first true *de novo* assembly of a large eukaryotic genome (giant panda) based entirely on short-reads generated by NGS, and the complete genomes of humans have also now been assembled *de novo* (Li *et al.*, 2010c).

Since relatively small amounts of starting material are required for NGS approaches, the ability to sequence genomes from ancient organisms has been revolutionised, enabling ecologists and evolutionary biologists to look back through deep time. Palaeogenomics has previously focused on sequencing the relatively abundant mitochondrial DNA that routinely requires many rounds of PCR amplification prior to traditional DNA sequencing. Owing to the low amount of genomic material from fossilised or preserved materials, however, the nuclear genome has thus far proved elusive. NGS allows not only an unbiased approach, but by aligning sequencing reads to phylogenetically related species, this allows contaminating DNA from microbial species to be removed from the analyses, or studied in parallel. These types of studies facilitate the

exploration of evolutionary pathways and phylogenetic relationships, which could, for instance, be mapped onto previous climates or fossil communities within long-extinct ecosystems: NGS has resulted in the partial sequencing of the nuclear genomes of the cave bear, *Ursus spelaeus* from 40,000-year-old bones (Noonan *et al.*, 2005), the woolly mammoth, *Mammuthus primigenius*, from 20,000-year-old hair shafts (Miller *et al.*, 2008), a 4000 year-old male Palaeo-Eskimo from permafrost-preserved hair (Rasmussen *et al.*, 2010) and from two 38,000-year-old fossilised Neanderthals (Green *et al.*, 2006; Noonan *et al.*, 2006).

This ability of NGS to deal with tiny sample sizes also holds great promise for its potential applications to contemporary as well as palaeo-ecology. For instance, it could be particularly useful for 'remote sensing' of the diets and gut flora of rare or protected species via faecal analysis in food web studies. Similarly suctorial feeders (e.g. aphid herbivores and many predatory beetles) have no visually recognizable gut contents and are therefore often omitted from traditional food web studies (e.g. Woodward *et al.*, 2010c), despite the fact that they can be very abundant and may perform key functional roles. Sufficient DNA could potentially be extracted from their guts and characterised using molecular techniques, rather than the less efficient and more time-consuming immunological responses (Carreon-Martinez and Heath, 2010). As previously noted, NGS can also be used to detect interactions between eukaryotes and potential pathogens within food webs, which have also been largely ignored in traditional ecology, for instance, via characterizing viromes in the faeces of vectors or intermediate hosts (e.g. Li *et al.*, 2010a).

V. TRANSCRIPTOMICS: FUNCTIONAL EXPRESSION

The transcriptome is the complete set of RNA transcripts in a cell, including mRNA, ncRNAs, and small RNAs. Such information allows the elucidation and determination of the transcriptional structure of genes, in particular enabling changes in expression levels and patterns of transcripts to be quantified in response to different cellular conditions/stimuli or changes in whole communities within environments (metatranscriptomics).

NGS allows the mapping and quantification of transcriptomes (RNA-Seq), via the isolation of total RNA (or enrichment of a particular RNA species) and enzymatically converting it to a library of complementary DNAs (cDNA) using reverse transcriptase. Universal adapters of known sequences are then ligated to the ends of the cDNAs, and NGS is then performed. Transcripts can be analysed from both 5′ and 3′ ends to determine both the start and the end of a particular transcript, and a mate-pair strategy (see Section IV.G) allows sequencing of both ends simultaneously. Transcriptome studies have already been carried out in a small number of eukaryote species (Cloonan *et al.*, 2008; Lister *et al.*, 2008; Morin *et al.*, 2008; Mortazavi *et al.*, 2008; Nagalakshmi *et al.*, 2008;

Wilhelm *et al.*, 2008). Using these methods it has been shown that up to 90% of the genome in yeast appears to be transcribed, with many transcripts being non-coding. Such results indicate the complexity of eukaryotic transcriptomes and that their analysis, even solely in terms of the data produced, is challenging even before the data is interpreted in a functional sense.

Transcriptome analyses in bacteria and archaea are further complicated by the fact that unlike eukaryotic mRNAs, most bacterial and archaeal mRNAs do not have a polyA-tail. Furthermore, most bacterial and archaeal RNA preparations contain up to 80% rRNA and tRNA. Thus far, transcriptome studies in bacteria have required the enrichment of mRNA, usually by the selective depletion of tRNA and/or rRNA (Liu *et al.*, 2009; Passalacqua *et al.*, 2009; Perkins *et al.*, 2009; Yoder-Himes *et al.*, 2009). In the study by Yoder-Himes *et al.* (2009) two strains of *Burkholderia cenocepacia* were compared, one from soil and another from a cystic fibrosis patient, at the transcriptome level under differing conditions and identified a subset of RNAs that may explain their different habitats and pathogenic potential.

Metatranscriptomics have been applied recently in a few studies that have attempted to characterise gene expression across multiple species, such as the recent study of microscopic eukaryotes (including protists and small metazoans) within soil communities (Bailly *et al.*, 2007). Recent reports suggest, however, that metatranscriptomes are still subject to very poor coverage and are not reliably reproducible (Gifford *et al.*, 2010; Stewart *et al.*, 2010). However, the development of single-molecule technologies could directly address these issues and make metatranscriptomics a more viable methodology in the near future.

VI. APPLICATION OF SEQUENCING IN ECOLOGY

A. Unveiling the Planet's Hidden Biodiversity

The advent of DNA sequencing has revolutionised molecular, microbial and general ecology, by enabling scientists to identify which and how many species are present in the system under study, and to consider both the microscopic and macroscopic world together. Most fundamentally the Tree of Life, now changed beyond recognition, and the discovery of lateral gene transfer as a factor in the evolution of all organisms has altered our understanding of both the ecological and evolutionary relationships between organisms. In the context of building a systems biology for ecology, probably the most dramatic recent change is the application of NGS to analyse communities, whether via metagenomics or by the analysis of amplified PCR products. We will highlight some metagenomics studies and critique the use of this, and other post-genomics approaches in ecology.

Most published metagenomics studies predate NGS, and were based on Sanger sequencing of cloned vectors, often plasmids. Such studies can be characterised as 'discovery' science, as there were generally no specific *a priori* hypotheses to be tested. Even now, many studies are primarily 'fishing expeditions', although a few have started to move into more sophisticated empirical end experimental cause-and-effect studies as the field has begun to mature as a scientific discipline (e.g. He *et al.*, 2010b). To illustrate the types of studies, and the data and insight they have tended to produce, four case studies are discussed briefly below.

B. Four Ecological Metagenomic Case Studies

There are many genomics and post-genomics studies in microbial ecology, but here we have selected four different case studies to highlight some general points. These are an analysis of an extreme, low diversity acid mine drainage (AMD) biofilm system by the Banfield group (Baker *et al.*, 2006; Lo *et al.*, 2007; Ram *et al.*, 2005; Tyson and Banfield, 2005; Tyson *et al.*, 2004); Venter *et al.*'s ambitious Global Ocean Sampling (GOS) expedition (Rusch *et al.*, 2007; Venter *et al.*, 2004); the application of Solexa sequencing to the human gut microbiome (Qin *et al.*, 2010); and the use of single species metagenomics by Rodriguez-Valera *et al.* (2009) to investigate the role of viral predation on biodiversity.

The Banfield group's work on the environmentally extreme (pH <1) and low diversity AMD biofilm system produced two almost complete genomes, from a Group II *Leptospirillum* and a type II *Ferroplasma*. Analysis of these genomes showed that, while both organisms had the potential? or did they measure rates? aerobically oxidised ferrous iron to ferric iron (Fe^{2+} to Fe^{3+}), the bacterial *Leptospirillum* fixed carbon which the archaeal *Ferroplasma* utilised. Interestingly, the *Leptospirillum* genome showed very little nucleotide polymorphism, suggesting either strong selection for this genotype or a very strong founder effect following the colonisation of the mine. In contrast the *Ferroplasma* genome appeared to be a mosaic of three ancestral genomes, with as many as 400 individual recombination events creating the contemporary genome. This suggests significant localised diversity in the *Ferroplasma* population, which could increase species resilience to perturbation. Several other partial *Leptospirillum* and *Ferroplasma* genomes were also constructed showing distinct divergence from the two complete genomes. In the almost complete genomes about 20% of the genes detected could not be assigned a cellular function (Tyson *et al.*, 2004). Within this metagenome of 76.2 Mbp just a single complete nitrogen-fixation operon was found and further work led to the isolation of the bacterium *Leptospirillum ferrodiazotrophum*, the putative sole nitrogen-fixing organism in the system (Tyson *et al.*, 2005).

Subsequent metaproteomics analysis showed high levels of expression of genes associated with oxidative stress and iron oxidation, matching the harsh conditions and predominant oxidative process in the biofilm (Ram *et al.*, 2005). This study still represents one of the most complete and successful of all genomics and post-genomics analyses of a defined ecosystem, and points the way for a host of future studies into real-world ecotoxicology and B–EF relationships in response to environmental stressors.

The GOS *Sorceror II* expedition was an extremely large and ambitious shotgun sequencing project, led by Craig Venter. Results from the survey within the Sargasso Sea (Venter *et al.*, 2004) were surprising in that, despite collecting over 1 Gbp of data, the study could not reconstruct even a single genome from these oligotrophic waters. An important finding in the study was the linking of a gene for aerobic ammonia oxidation to an archaeal scaffold which, along with a similar finding in soil (Treusch *et al.*, 2005) and the isolation of marine crenarchaeaon capable of aerobic ammonia oxidation (Könneke *et al.*, 2005), contributed to a shift in our understanding of this part of the global nitrogen cycle (Erguder *et al.*, 2009). The authors concluded that their samples contained at least 1800 distinct genomic species and identified a surprising diversity of genes encoding for bacteriorhodopsin (now called proteorhodopsin), a light-harvesting protein once thought to be limited to halophilic archaea but that had been previously identified in bacterioplankton (Beja *et al.*, 2000). The detection of such a natural diversity of proteorhodopsin genes indicated this gene was widespread in bacterioplankton, suggesting that it plays an important role in carbon cycling in the oceans. Of the 1.2 million genes identified in the study, $\sim$65% could not be assigned a cellular function nor aligned with any known organism. This study highlighted a major issue with metagenomics: that the diversity within even what was perceived as a relatively simple system simply overwhelmed the enormous sequencing effort applied to it.

NGS have been used to characterise an 'internal ecosystem' – the human gut microbiome in over 100 people. Qin *et al.* (2010) exploited the massively paralleled sequencing capacity of the Solexa technology to analyse this ecosystem: over 500 Gbp of data were collected, encoding 3.3 million genes from 1000 to 1150 microbial species. These data showed that each host had a bacterial population of $\sim$160 species. Analysis of a subset of the data showed that there were clear differences in the gene complements of healthy and diseased subjects, suggesting a role for gut microbiota in chronic conditions such as Crohn's disease and ulcerative colitis. The data has expanded the depth of knowledge of the potential microbial metabolism within the human gut by highlighting the potential for gut microbes to use carbohydrates, such as pectin and sorbitol, which are poorly absorbed by humans. Presumably these sugars are used by the microbes as an energy source that the human host does not require. However, even though nearly 200 gut microbes have

been genome-sequenced, only 12% of the sequenced genes in this study could be associated with existing human gut microbial genome sequences and >80% could not be assigned a cellular function at all.

Rodriguez-Valera *et al.* (2009) provide an excellent example for how metagenomics can be used to understand microbial diversity driven by phage predation. Using single-species metagenomics these authors identified that natural populations of bacteria and archaea are relatively rich in metagenomic islands (MGIs), regions of the bacterial genome that are under-represented in the population, but that include an over-representation of genes related to phage resistance. These genes are therefore among the most variable in the population. They then used computer simulations to propose that Constant Diversity (CD) dynamics, in which density-dependent phage predation prevents competitively superior strains from dominating the community, could be behind the maintenance of relatively high strain richness, with each strain being maintained with a sub-optimal and nearly equal fitness. They further proposed that the maintenance of such highly diverse populations can increase ecosystem efficiency (i.e. the decomposition of organic material), by each strain specializing in exploiting a different micro-resource instead of an over-abundant one taking most of them. These findings have direct parallels with classic ecology: first, they show how (biotic) heterogeneity can explain the maintenance of high diversity (the 'paradox of the plankton' – Hutchinson, 1961); and second, they show how predation may contribute to maintain strains within populations by suppressing inter-strain competition, in a manner similar to that shown in classic experiments with metazoans (Paine, 1966).

Over the 6 years since publication of the Tyson *et al.* (2004) paper the amount of data that can be collected via NGS has increased by nearly four orders of magnitude, yet proportionately fewer of the genes sequenced can be unambiguously assigned to a cellular function. The initial issues in metagenomics of poor depth of sampling has thus given way to a less conclusive analysis of more data, as well as concerns about systematic errors in metagenome libraries (Gomez-Alvarez *et al.*, 2009). Clearly, there are still major issues to be resolved in the use of NGS technologies, some of which may be remedied by more sophisticated bioinformatics approaches when analysing these vast datasets. It should be noted that the majority of metagenomic studies still tend to conclude with rather generic statements that the majority of the genes detected are associated with core metabolic processes, carbohydrate utilisation, protein metabolism and lipid biosynthesis. It is perhaps difficult to be excited by the unsurprising finding that cells use carbon and make/degrade proteins and lipids, and these rather bland conclusions highlight the embryonic state of the field at present; the potential for truly novel insights has barely been tapped and the field is still wide open for exploration and discovery.

C. Moving Beyond Traditional Metagenomics

There have been some important and potentially powerful developments in the field of metagenomics field, including its use in comparative studies (Gianoulis *et al.*, 2009; Tringe *et al.*, 2005) to determine real differences among systems from the complex mass of data. These studies have also highlighted the clear need for proper replication in order to draw statistically significant conclusions: in many instances statistical rigour has been largely ignored, undermining the validity of interpretations based on the data collected (Gilbert *et al.*, 2008, 2009). Other developments include using metagenomics to build genomes of important uncultured organisms, either from low diversity environments or from enrichment cultures (Chivian *et al.*, 2008; Erkel *et al.*, 2006; Strous *et al.*, 2006) and as an analysis tool for hypothesis-driven experimental studies, including those using stable isotope probing (Chen and Murrell, 2010 and references therein) and some simple, but elegant, microcosm or even field experiments (He *et al.*, 2010b; Mou *et al.*, 2008).

The problems and limitations that constrain metagenomics also affect metatranscriptomics, the analysis of all the transcripts from a system. At present, transcriptomes often suffer from poor rates of assignment for detected putative genes (Gilbert *et al.*, 2008), poor coverage (Gifford *et al.*, 2010) and technical replication such that it has been reported that only 17% of transcripts were found in common in duplicate runs of the same templates (Stewart *et al.*, 2010). These issues are exacerbated by the problems associated with studying microbial mRNA, including rapid degradation and poor separation from ncRNA species (Feder and Walser, 2005; van Vliet, 2010). There is clearly much that still needs to be done before metatranscriptomics becomes a readily available ecological tool for molecular analysis of communities or ecosystems. Even so, successful studies that have focused on clearly defined subsets of transcripts have been reported, with a large number of presumably regulatory small non-coding RNA (sRNA) molecules detected in ocean waters (Shi *et al.*, 2009), and a small scale metatranscriptomics analysis was used to characterise eukaryotic functional diversity within soil ecosystems (Bailly *et al.*, 2007). Given the rapid advances made using NGS within the space of just a few years, it seems likely that metatranscriptomics also has the potential to become an integral part of the next generation of molecular ecology techniques.

D. Next-Generation Sequencing and Biodiversity

NGS technologies have also been used in non-genomics studies, as the very large numbers of sequences reads make it possible to sample microbial communities completely. Using 454 sequencing ($\sim$400 bp reads) PCR-amplified functional gene fragments (*dsrB*) from a specific bacterial genus

(*Desulfobulbus*) was sampled completely sampled from 10 sites along an estuary in the United Kingdom (Figure 9A, unpublished data). As few as 1000 sequence reads were required to effectively analyse the complete diversity of this genus, with nearly 100 operational taxonomic units (OTUs, which are analogous to 'morphospecies') detected along the estuary. Furthermore, the data from the 454 sequences matched the differential distributions of four specific *Desulfobulbus dsrB* genotypes detected independently by molecular fingerprinting analysis and genotype-targeted qPCR (Figure 10, unpublished data). Therefore, the perennial problem in microbial ecology of

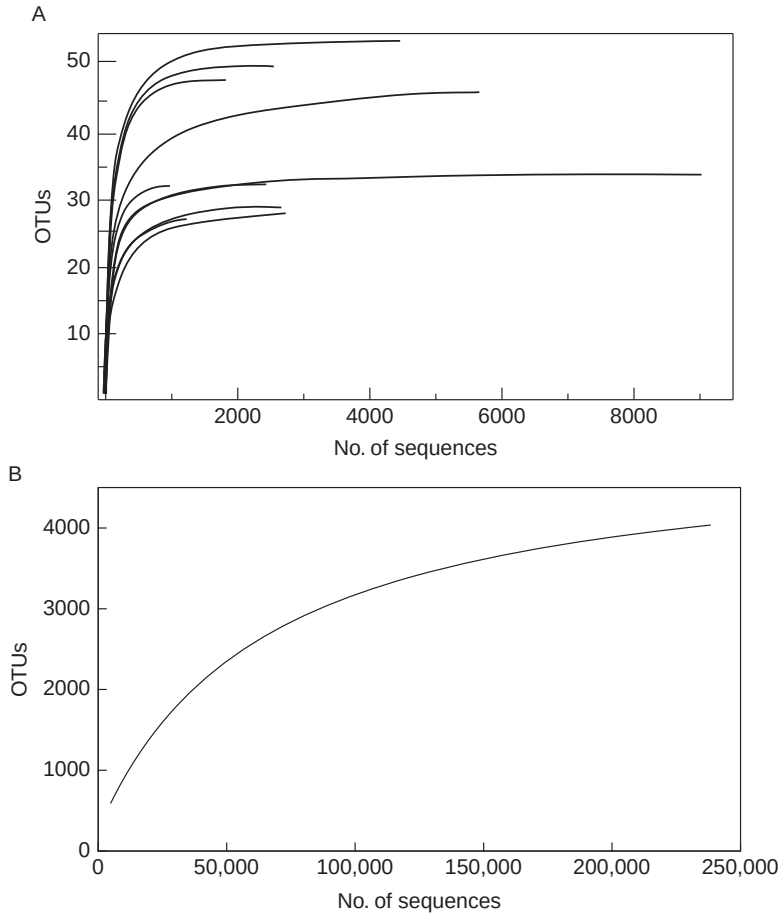


Figure 9 Rarefaction analysis of (A) 454 data (~400 bp) of amplified *Desulfobulbus dsrB* gene fragments from 10 sites along the Colne estuary, UK. All samples approach an asymptote; Coverage is >98% (Good, 1953). (B) Protist diversity in a lake ecosystem, as revealed from 454 sequencing, redrawn after Nolte *et al.* (2010).

undersampling may be solved using NGS (Figure 9). However, these methods are not necessarily the magic bullet that will solve all diversity problems in microbial ecology. The first major 454 diversity study, by Sogin *et al.* (2006), was unable to completely sample their targeted community, leading to the proposal of the idea of the rare biosphere, in which there exists a very long tail of low abundance microbial species. While this was initially seen as a major breakthrough in understanding microbial biodiversity, the validity of Sogin *et al.*'s analysis has been vigorously challenged (Kunin *et al.*, 2009; Quince *et al.*, 2009). Criticisms include the use of very short reads, which make accurate phylogenetic assignments very difficult, poor quality control, and a lack of real support for the general idea of the rare biosphere.

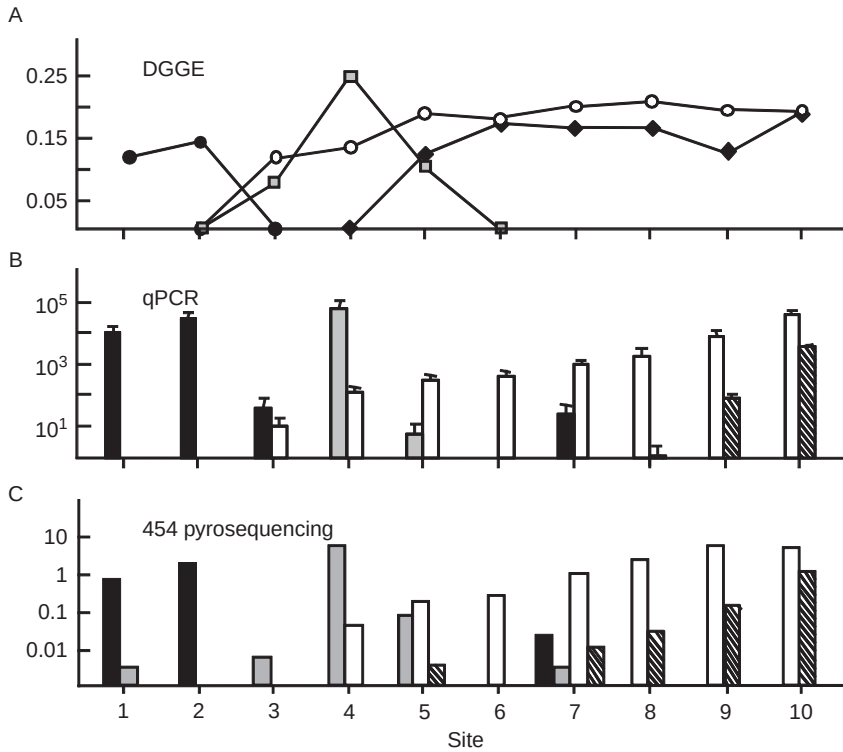


Figure 10 Analysis of the distribution of four genotypes of *Desulfobulbus* along the Colne estuary, UK, using amplified *dsrB* by (A) DGGE fingerprinting, (B) genotype-specific qPCR, (C) 454 pyrosequencing. The distribution pattern seen in the DGGE fingerprint is matched by the qPCR and 454 pyrosequencing data showing how *Desulfobulbus* is differentially distributed along an estuary.

VII. LINKING ACROSS MULTIPLE LEVELS OF ORGANISATION: THE KEY TO UNDERSTANDING THE SYSTEM

A. Scaling from Molecules to Ecosystems

A systems approach to ecology ultimately requires the ability to scale from molecules to ecosystems, but making these connections is an extremely challenging task (DeLong, 2009). Figure 2 shows a simple diagram defining the flow of information from a gene sequence to an ecosystem process. While this simplistic figure is useful it fails to highlight the full range of feedback interactions that can and do occur in the environment. All aspects of organismal functioning are related to the external environmental conditions and so, by extension, ecosystem functioning will also be dependent on those factors (Yvon-Durocher *et al.*, 2010c). For example, the extent to which sulphate reduction or methanogenesis dominates terminal oxidation of organic matter is dependent on many factors, including sulphate and organic carbon availability, temperature, and the presence of alternative respiratory compounds, such as nitrate or oxygen (Purdy *et al.*, 2003b; Raskin *et al.*, 1996). A systems approach to ecology will require the incorporation of the effects of such environmental feedback.

The importance of environmental feedback can be seen by considering at what level it might affect organisms. Microbes are very directly affected by their immediate surrounding and control their metabolism as a consequence of this. This can be represented as regulation during gene expression (Figure 2), by mRNA half-life and the effects of ncRNA, translational control and of course protein functional regulation, all of which are basic biological concepts (Nelson and Cox, 2009). These regulation controls will be dependent on the environment a cell finds itself in at that particular time and are further complicated by inter-cellular functions. Consequently there may be substantial disconnects between the genes present in a system and the activity of their products (Feder and Walser, 2005), and we therefore need more robust empirical data to gauge how strong (or weak) these feedbacks might be. Furthermore, ecologists commonly assume a substantial phenotypic coherence within defined taxonomic groups, but this is not a valid assumption for many microbes, which are much less clearly defined as species (Cohan, 2001): recent studies on microbial genomes suggest that bacteria and archaea have a flexible genome structure that includes a core genome, comprising the essential and defining genes of a particular clade, and also an accessory genome, which can vary substantially between individual strains. This has led to the definition of the pangenome, the total potential genome of a microbial species group (Medini *et al.*, 2005; Tettelin *et al.*, 2008). Therefore,

it would appear that no matter how well defined the species concept becomes for microbes, it will never define the exact phenotype of any particular strain within that species. Despite these potential caveats and limitations, which add complexity, but also reality, to a system, the study of microbial members of communities in terms of ecosystem functioning is becoming increasingly feasible in natural or experimental systems (Figure 11).

B. Linking Microbial Community Structure and Ecosystem Functioning

There are many reports that make coherent and well-supported links between microbial community structure and ecosystem functioning, including experimental manipulations of microbial communities that have linked specific organisms to the degradation of specific substrates (He *et al.*, 2010b; Purdy *et al.*, 1997, 2002, 2003a; Figure 11). Although these studies have made some very definitive links between organisms and substrates, they have also highlighted the complexity and variation of microbial responses to environmental stimuli. For example, in fatty acid amended sediment slurries from an

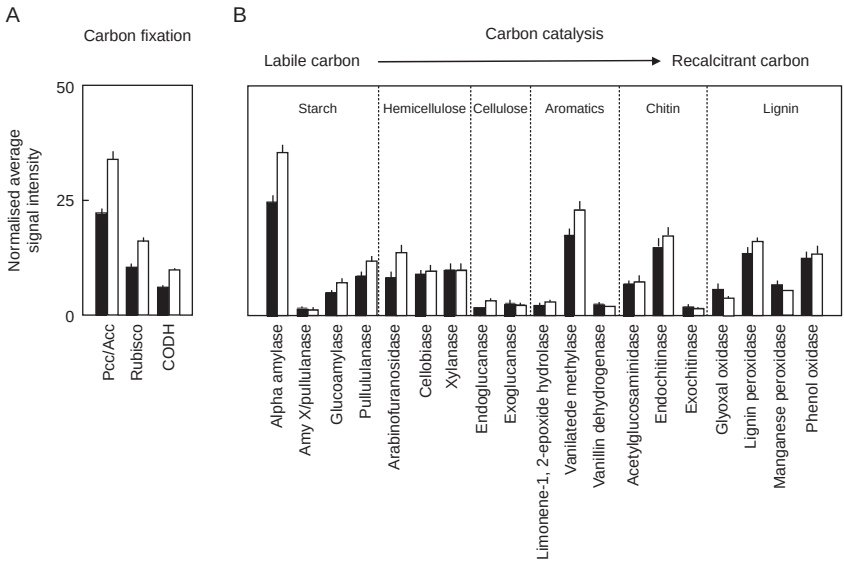
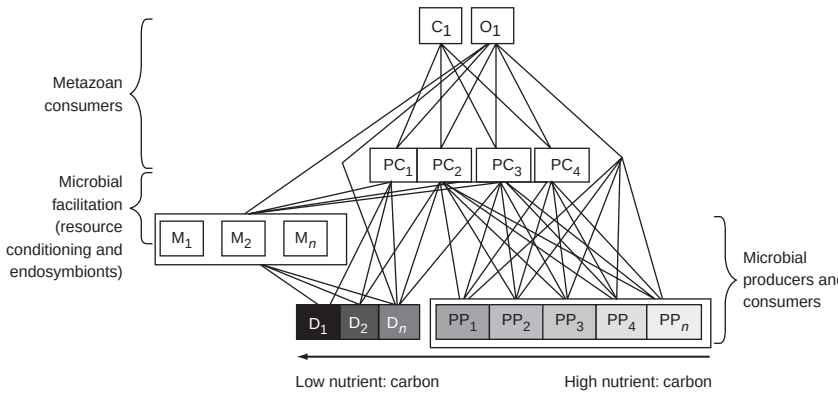


Figure 11 Linking microbial diversity to ecosystem functioning in a long-term field experiment. The metagenomic expression of metabolic processes associated with the carbon cycle differed markedly between soils that had been exposed to a decade of elevated CO_2 concentrations relative to ambient controls. Redrawn after He *et al.* (2010b).

estuary the addition of acetate, or its *in situ* production from lactate, significantly enriched the sulphate-reducing bacteria *Desulfobacter*, whereas the production of acetate from butyrate did not (Purdy *et al.*, 1997). This suggests that the immediate biological history of a system can affect the way in which it responds to a perturbation. The development of the use of stable isotope analysis has further advanced the range of methods that directly link organisms to activities (Boschker *et al.*, 1998; Chen and Murrell, 2010; Radajewski *et al.*, 2000). It is also important to recognise that the activity of many microbes is intimately linked to the activity of other, often taxonomically unrelated, organisms (Dolfing, 2001; Jackson and McInerney, 2002) making the assignment of a role to a specific microbial group often dependent upon at least some knowledge of the other organisms in the system. Madsen (2005) has detailed the very real issues associated with defining an ecological role to specific microbes.

The task of linking the different organisational levels in an ecosystem that incorporates microbial activity is clearly challenging (Figure 12). This is where systems approaches can be very powerful, as a discovery-based, wide-ranging approach to investigation is likely to throw up new and unexpected drivers and processes, such as light-harvesting by heterotrophs (via proteorhodopsin, Beja *et al.*, 2000). Such structural-functional interactions



Key:
C, carnivorous predator; O, omnivorous predator; PC, primary consumer; M, microbial decomposer; D, detritus (e.g. leaf-litter); PP, microbial/protist primary producers (e.g. diatoms, cyanobacteria).
Redrawn after Woodward (2009).

Figure 12 The microbial 'black box' within a schematic community food web. The vast majority of community and ecosystem studies have ignored the diversity and functional roles of the taxa at the base of the web, even though these are the ultimate drivers of bottom-up processes and the supply of energy and nutrients to the higher trophic levels.

will then need to be proven to have a real biological effect on a system (Fuhrman and Steele, 2007), which requires carefully designed and rigorous controlled experiments and manipulations. On the ecosystem side, manipulating environmental conditions in reasonably sized and replicated meso- or microcosms (e.g., Yvon-Durocher *et al.*, 2010a,c) will allow important questions related to the links between structure and functioning at the higher levels of organisation to be addressed (Ptacnik *et al.*, 2010; Reiss *et al.*, 2010a). Large-scale mesocosm experiments, which are needed to identify potential mechanistic relationships in the field, will facilitate the application of molecular microbial ecological studies within a wider framework and so should facilitate the development of links across different scales of organisation within an ecosystem.

C. New Questions in Ecology: Opening the Microbial Black Box

The aim of a systems approach is to define an ecosystem in terms of its components to predict how it will respond to change. If it is possible to open the microbial black box, and to understand what we find inside, then comparative analyses of different ecosystems will improve, as will our ability to predict responses to future change.

The strength of integrating NGS technologies within general ecology includes the ability to analyse a community without the need to amplify the DNA, to achieve complete coverage of a community and to handle multiple samples with limited sample manipulation. A 'complete' metagenome would, at least in theory, allow the mapping of all possible metabolisms, even those that are presently unknown (but which could be revisited repeatedly as bioinformatic databases mature), that were encoded within a system, and hence the full range of processes an ecosystem could potentially sustain. Such data could raise a vast new range of questions that could be asked of ecosystems, and importantly to test whether many of the existing theories of community ecology, which have developed primarily from macroorganismal studies, can also be applied to the microscopic world (Ptacnik *et al.*, 2010; Reiss *et al.*, 2010a,b). This leads to the challenge of marrying traditional ecology with microbial ecology.

There is an ever-increasing body of data that supports the idea that many aspects of the ecology of large organisms are not necessarily fundamentally different to that of microbes (Brown *et al.*, 2004; Ptacnik *et al.*, 2010). Nonetheless, there are in fact many obvious differences (e.g. reproduction via binary fission, reliance on nutrient uptake via diffusion, etc.) that could be critically important, depending on the question being addressed. The sheer diversity of microbes and the difficulty of defining an all-encompassing species concept

pose real challenges (Reiss *et al.*, 2010b). Delong *et al.* (2010), for example, have recently reported that allometric mass-metabolism scaling coefficients in bacteria and archaea differ from that in microscopic and large eukaryotes, suggesting that not all factors are independent of body size, although differences in metabolic scaling are far less than differences in body mass *per se* (Makarieva *et al.*, 2008). Notwithstanding some of these more obvious differences, there are also many areas of solid ground that link microbial and general ecology: for example, there is now strong evidence that microbes exhibit biogeographic patterns of distribution (Cho and Tiedje, 2000; Oakley *et al.*, 2010; Whitaker *et al.*, 2003) and that, similar to larger organisms, the number of species increases in a power-law relationship with increases in area or volume (Green and Bohannan, 2006; van der Gast *et al.*, 2005). Latitudinal diversity gradients (with higher species richness occurring at lower latitudes; Fuhrman *et al.*, 2008) and spatial heterogeneity explaining the structure of local microbial communities (Östman *et al.*, 2010) have also been reported recently.

NGS offers an excellent opportunity to test hypotheses for the processes involved in determining these diversity patterns, such as the suggestion that latitudinal diversity gradients are formed in part due to the enhancement that higher temperatures, and in terrestrial environments water availability, have on encounter rates among organisms and thus on the rate of biotic interactions (Moya-Laraño, 2010). One prediction of this hypothesis is that in aquatic ecosystems, and across individuals within a single species, functional diversity should increase with temperature (up to a certain threshold). NGS, metagenomics, metaproteomics and metabolomics can be used to test the hypothesis that once controlled for community size (cell number), species richness and phylogenetic distance, communities at higher temperatures (up to 30 °C) hold a higher number of functional units. In aquatic bacterial and archaeal cells, it has been hypothesised that temperature enhances cell-to-cell encounter rates and the diversity of interactions is even more likely because cell and metabolite mobility is higher in environments of low viscosity (Kummerli *et al.*, 2009) and viscosity decreases with increasing temperature (Woodward *et al.*, 2010a).

The advent of NGS may resolve the issue of undersampling of microbes but differences in sampling and analysis between the two fields will also need to be bridged. Perhaps the most relevant area where new questions can be raised is in the contribution microbial studies could make to ecology. Ecology has a strong history of theoretical developments, but rarely directly testing the validity of these theories (Raffaelli and Moller, 2000) due to the difficulties of performing controlled and replicated experiments at suitable spatiotemporal scales in field ecology (Woodward *et al.*, 2010a,b). Microbial communities, however, because of their large population sizes and rapid generation times, can be manipulated and experimented upon under clearly defined and controlled conditions (Petchey *et al.*, 1999). Hence, a microbial ecologist should be able to test the wide variety of mechanisms that have been put forward to

explain macroecological patterns among the larger taxa (McGeogh and Gaston, 2002), in addition to testing microbial ecology theories in the real world. For instance, the apparent stability in terms of the species composition and metabolic potential of viral and microbial assemblages observed in natural ecosystems appears to contradict theoretical models and viral–microbial dynamics observed in simple systems, where cycling of the dominant taxa is more typical (Rodriguez-Brito *et al.*, 2010).

D. Functional Redundancy: Do Species Matter?

A common assumption in general ecology has been that microbes can be treated as a black box because of massive functional redundancy (Vandenkoornhuyse *et al.*, 2010), yet there is little compelling evidence to support this view. In fact, data on reconstructed microbial communities suggests that more diverse microbial communities are also more productive (Bell *et al.*, 2005), in line with the positive B–EF relationships reported for many macro-organismal studies (Cardinale *et al.*, 2006). In contrast, rapid changes at the fine-grained level of viral genotypes and microbial strains overlain on the apparent community and ecosystem stability reported by Rodriguez-Brito *et al.* (2010) suggest that functionally redundant taxa might cycle at the level of viral genotypes and virus-sensitive microbial strains. In a different study, the cyanobacterium, *Prochlorococcus*, which is the most abundant oxygen fixing phototroph in the world's oceans, with an estimated global population of 10^{27} cells, underpins the pelagic food webs of the nutrient-poor tropical and subtropical gyres – among the largest ecosystems on Earth. Recent metagenomics studies have revealed that uncultured *Prochlorococcus* possess the genetic potential for nitrate assimilation, and metatranscriptomics data from the field have confirmed that these genes are expressed (Martiny *et al.*, 2009); yet earlier work on isolates found no evidence for this metabolic capacity (Rocap *et al.*, 2003). That this functional trait might not be ubiquitous even within a single species has implications for understanding not only the biogeography of this key primary producer and its role within the oceanic carbon and nitrogen cycles, but also the level of organization at which functional redundancy or uniqueness is expressed.

Whether microbial communities are functionally redundant is clearly important for understanding the role of microbes in ecosystem functioning, and this represents one of the great unanswered questions in ecology. A useful starting point would be to determine whether 'poorly-performing' or seemingly unstable ecosystems contain reduced diversity in specific critical ecosystem processes. If ecosystem process rates can be manipulated by increasing or reducing functional capacity, which could be validated via NGS techniques, in controlled small-scale experiments, then this would

suggest that redundancy is not as prevalent as is often imagined. This is a non-trivial issue and its resolution is critical to understanding the role of microbial biodiversity in an ecosystem context.

E. Beyond Bacteria: Opening the Eukaryotic Black Box

Although we have focused primarily on bacterial and archaeal examples, as this is where most of the recent advances have been made in this field, metagenomics and NGS techniques also offer promise in the future for studying eukaryotic micro- and macro-organisms, such as host–pathogen and host–symbiont ecology and also in the study of eukaryotic assemblages *per se*. Woyke *et al.* (2006) for instance revealed that bacteria can provide their eukaryote host with multiple sources of nutrition, and a recent study by Coolon *et al.* (2010) has characterised metagenomic shifts in the microbial assemblages in the environment and also in vertebrate guts across a pollution gradient. Widespread distributions of associations between bacteria and sponges have also been revealed in a range of marine ecosystems on a biogeographical scale (Lafi *et al.*, 2009). Similarly, in terrestrial systems, Jumpponen *et al.* (2010) have discovered strong seasonality in the diversity and dynamics of oak–ectomycorrhizae associations in urban versus rural environments.

Piganeau *et al.* (2008) have found unexpected picoeukaryotic sequences within the Sargasso Sea metagenome, and other recent studies have found evidence of surprisingly diverse eukaryotic assemblage in seemingly very harsh environments, including anoxic marine sediments (Stoeck *et al.*, 2010; Table 2). Clearly, as more sequences and datasets are assigned to functions in eukaryotic cells, both from uni- and multi-cellular organisms, we will be better able to use NGS to understand the functions present in an entire system as well as to assign each function to a particular species (or group of species) within communities, thus contributing to solving the structure of ecological networks accurately and to understanding the reasons behind their apparent robustness and stability.

F. Food Webs: Identifying the Missing Links

Microbes play many key roles in food webs, particularly at the lower trophic levels, where they are important conduits for the flux of energy and materials to the consumer assemblage. They can also be significant pathogens, but this has been almost completely ignored in food web research to date (Ings *et al.*, 2009). Much of food web ecology has suffered from patchy taxonomic resolution, which tends to be high at the top of the web but poor at the bottom, where the

Table 2 Molecules to ecosystems: selected examples of recent applications or reviews of genomics, metagenomics and NGS as tools for testing ecological theory

Ecological subdiscipline	Approach	Key findings
Organismal functional ecology		
Renaut *et al.* (2010)	454	Transcriptomics: identification of potential adaptive single-nucleotide polymorphisms in fishes
Wolf *et al.* (2010)	454	Transcriptomics: nucleotide divergence and gene expression differentiation in the carrion crow and hooded crow hybrid zone
Population structure and dynamics		
Van Bers *et al.* (2010)	Solexa	Baseline generation of >2bn basepairs of novel sequence information for an ecological model species, the great tit *Parus major*
Goetz *et al.* (2010)	454	Identifies genetic basis for phenotypic differences within trout populations
Microbiomes (*internal ecosystems*)		
Woyke *et al.* (2006)	Sanger	Symbiosis: bacteria provide eukaryote host with multiple sources of nutrition
Kloch *et al.* (2010)	454	Characterisation of gut parasite loads in bank voles
Community structure and dynamics		
Jumpponen *et al.* (2010)	454	Seasonality of diversity and composition in oak ectomycorrhiza across ecosystems
Schütte *et al.* (2010)	454	Shifts in bacterial diversity over glacial chronosequences
Stoeck *et al.* (2010)	454	Discovery of complex eukaryotic community in anoxic marine ecosystem
Piganeau *et al.* (2008)	Sanger	Picoeukaryotic sequences detected in the Sargasso Sea metagenome
Gilbert *et al.* (2009)	454	Seasonal dynamics of marine microbial communities
Rodriguez-Brito *et al.* (2010)	454	Viral and microbial community dynamics in four aquatic environments
Biogeography		
Dinsdale *et al.* (2008)	454; SEED	Potential for viromes to influence global evolutionary and metabolic processes
Breitbart and Rohwer (2005)	n/a (review)	Review: potential for viruses to form 'global' diaspora
Cuvelier *et al.* (2008)	Sanger; FISH	Discovery of widespread distribution of a unique marine protistan lineage
Biddle *et al.* (2008)	454	Peru Margin subseafloor biosphere revealed as a genetically distinct environment

(*continued*)

Table 2 (*continued*)

Ecological subdiscipline	Approach	Key findings
Lafi *et al.* (2009)	Sanger	Widespread distribution of associations between bacteria and sponges
Biodiversity & ecosystem functioning		
Pernthaler *et al.* (2008)	Magneto-FISH	Nitrogenase genes within metagenome and incorporation in biomass of methane-oxidizing consortia suggest role in new nitrogen inputs by syntrophic assemblages
Béjà *et al.* (2000)	Sanger	New form of phototrophy discovered in marine ecosystems
Food webs and trophic interactions		
Jones and Lennon (2009)	qPCR	Potential importance of indirect trophic interactions in planktonic food webs
Kant and Baldwin (2007)	n/a (review)	Review: Integration of molecular biology and ecology to characterise herbivory
Phillips *et al.* (2003)	n/a (review)	Review: Model of multitrophic molecular signals in the rhizosphere: effects of molecular 'control points' could propagate through other trophic levels
Proulx *et al.* (2005)	n/a (review)	Review: Identifies need to develop conceptual framework for integrating complex ecological, genetic and metabolic networks
Environmental stressors and biomonitoring		
Debroas *et al.* (2009)	Sanger	Overrepresentation of genes involved in xenobiotic degradation in a polluted lake
Dinsdale *et al.* (2008)	454, SEED	Increase in heterotrophic microbes and potential pathogens in human-impacted sites with widespread coral loss
Coolon *et al.* (2010)	454	High rates of microbial turnover in soils and consumer guts across a gradient of heavy metal pollution
He *et al.* (2010b)	454; GeoChip microarray	Long-term exposure of grassland to elevated CO_2 altered structure and functional potential of soil microbial communities
Ecosystem goods and services		
D'Costa *et al.* (2007) Kennedy *et al.* (2008, 2010) Hofmann and Gaines (2008) Marco (2010)	n/a (reviews)	Reviews: Potential for bioprospecting for antibiotics, pharmaceuticals and other bioproducts
Sharma *et al.* (2010)	Database	Bioinformatics database for bioprospecting for commercially useful enzymes

black box approach is still very much in evidence (Figure 12). Kant and Baldwin (2007) have recently highlighted the increasing convergence between ecology and molecular biology for studying plant–herbivore interactions, which are often extremely difficult to characterise using standard ecological techniques. Many consumers, such as suctorial herbivores or predators, have unrecognisable gut contents, and many resources have insufficient hard tissue to allow reliable identification based on morphological structures, and in such instances dissection of gut contents is largely pointless (Woodward *et al.*, 2010c). In such situations, stable isotope analysis can help to detect assimilated tissue, but it lacks sufficiently high-level taxonomic resolution to identify trophic interactions between species pairs, which may be addressed by immunological or, increasingly, molecular techniques (Figure 13). The use of metagenomics has been particularly useful for resolving otherwise difficult to characterise feeding links and species responses to exploitation, but proteomics and metabolomics

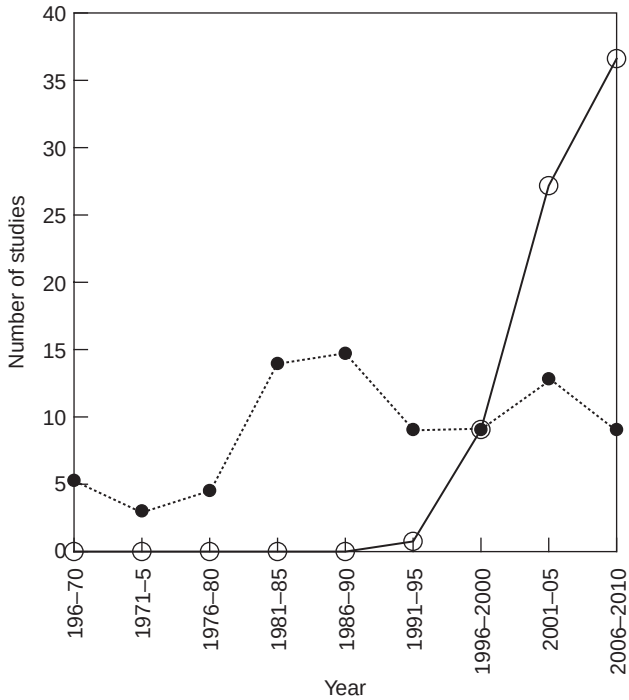


Figure 13 Potential for molecular applications in food web ecology. The dashed line shows the slow increase in studies using immunological techniques to characterise diets, the solid line shows the rapid rise of those based on molecular techniques. Redrawn after Carreon-Martinez and Heath (2010).

have also been used to describe the dramatic metabolic reconfigurations that occur in response to herbivory (Kant and Baldwin, 2007).

Foraging theory (FT), which has been recently used successfully to explain the structure of food webs (Petchey *et al.*, 2008; Woodward *et al.*, 2010c), explains animal foraging decisions, among other things, based on the profitability of food items. Since the microbial community within animal guts may largely determine the digestibility and profitability of the food ingested (Hooper *et al.*, 2002; Stevens and Hume, 1998), microbial-mediated foraging efficiency can contribute to determine predator–prey interactions and food web structure. In other words, the optimality of prey items may change depending on the microbial gut community. In well-resolved food webs, NGS could be used to test the hypothesis that metazoan food web structure is associated with the functional diversity of the microbial gut community. Microbial gut community NGS analysis could then be used to uncover missing links in other food webs.

G. NGS and the Link Between Ecology and Evolution

The link between ecology and evolution is experiencing a renaissance, particularly when it comes to understanding communities and ecosystems (Ellers, 2010; Johnson and Stinchcombe, 2007). Both at micro- and macro-scales, the ecological context may largely drive evolutionary outputs and, conversely, adaptive evolution and phylogenetic structure can determine ecological processes. The need for studying such reciprocal links is also obvious in microbial research, with authors requiring experimental approaches to determine the selective pressures to which microbes are exposed (Fraser *et al.*, 2007) or arguing that to understand adaptive microbe evolution we need to include the ecological context in which evolution is occurring (Kassen and Rainey, 2004). Importantly, adding the spatial context into this ongoing link, and explicitly adding the study of dispersal among communities (i.e. the metacommunity context) may greatly improve our understanding of how species and populations interact in real scenarios and the role of eco-evolutionary interplay in the stability of communities in the long term (Urban *et al.*, 2008).

In addition, the study of ecological networks, that is, networks in which many species interact with each other (e.g. food webs, parasite–host, competitive or pollination networks), is also contributing to the link of ecology and evolution, as both factors may affect network structure, robustness and stability (Bascompte, 2010; Bascompte and Jordano, 2007). The amount of genetic and phenotypic variation (and diversity) of the multifarious phenotypes that are present in populations have been proposed to affect food web structure in a manner that enhance its robustness and stability (Moya-Laraño, in press), for instance by increasing food web connectance (i.e. the proportion of potential links that are realised). However, finding that higher

phenotypic diversity correlates with higher connectance in networks would not be conclusive evidence in favour of this hypothesis, because networks may be maintained from phenotypic feedback, in which phenotypic diversity may affect connectance, which may, in turn, facilitate the maintenance of phenotypic diversity, and so on (J. Moya-Laraño, A. Rossberg and J.M. Montoya, manuscript in preparation). Furthermore, these networks of interactions may have complex eco-evolutionary dynamics in which genotypic variation can affect both evolutionary, as well as ecological processes, and these two processes may dynamically affect each other over ecological time scales (Fussmann *et al.*, 2007; Jones *et al.*, 2009; Pelletier *et al.*, 2009; Yoshida *et al.*, 2003). Lastly, it has been suggested that heritability at the level of communities and ecosystems may be substantial if we explicitly consider 'Interspecific Indirect Genetic Effects', by which the genetics of individuals of a species in a community may affect the expression of traits in individuals of a different species, thus allowing the community and its associated ecosystem functional attributes to evolve by natural selection (Shuster *et al.*, 2006).

The above hypotheses could potentially be tested by using NGS across space and time in both natural and experimental communities and by following the fate of sets of species, genes and/or molecules. In addition, by comparing local communities through time, one can potentially track among-community migration rate and include the metacommunity context to understand the ecology and evolution of complex interaction networks. NGS is thus a promising tool that may greatly accelerate our understanding for how ecology and evolution are interconnected via feedback loops.

VIII. SOCIOECONOMIC APPLICATIONS: ECOSYSTEM GOODS AND SERVICES AND BIOPROSPECTING

Ecosystem processes underpin socio-economic 'goods and services' of value to humans, for instance, the supply of clean drinking water, maintenance of a breathable atmosphere, and the provision of medicines. Many of these are supplied by microbes, yet as we have already discussed, their biodiversity and resultant functional roles are still largely unknown: hence their contributions to both relatively directly measurable goods and services (e.g. isolation of commercially exploitable antibiotics) and those that are more difficult to quantify (e.g. maintenance of the global carbon cycle) are also still poorly understood.

The huge untapped potential of biodiversity in natural systems offers great prospects for commercial applications in the future. For instance, bioprospecting for particular enzymes or genes that could offer new medical treatments is a rapidly growing field that is starting to trawl the vast

bioinformatics databases that have sprung up in the wake of NGS (D'Costa *et al.*, 2007; Hofmann and Gaines, 2008; Kennedy *et al.*, 2008, 2010; Sharma *et al.*, 2010). To date most of the literature in this field is in the form of reviews that highlight the potential commercial benefits of NGS, rather than concrete examples of metagenomics data being applied in a socioeconomic context *per se*, but this is now changing (Marco, 2010). No doubt these examples will appear at a rapidly increasing rate in the near future, given that the potential is vast: from purely medical applications to the use of bioengineering organisms to deal with pollutants and toxic spills, by identifying genes in the environment associated with metabolising xenobiotics (Marco, 2010). Some of the seemingly more outlandish suggestions that have been put forward include the use of certain bacteria as potential agents for terraforming other planets, based on their unusual metabolism and ability to survive in extreme environments (Ardelean *et al.*, 2008). On a more down-to-earth level, numerous biotechnological applications of metagenomics have been proposed in recent years, including bioremediation of damaged ecosystems and the development of new industrial bioproducts and novel medical treatments (Marco, 2010; Table 2).

IX. CONCLUSIONS

The revolution in ecology that began with the advent of molecular studies has developed increasingly rapidly via the exponential growth in our ability to extract data from a system, as exemplified by the recent emergence of NGS technologies. The application of these methods will undoubtedly lead to a new more inclusive ecology and allow us to begin to make the links required to understand how the different levels of organisation are interconnected, from molecules to whole ecosystems. However, we are not advocating these new technologies as a panacea, as many of the questions NGS data will raise will require testing using the rigorous experimental techniques that have underpinned traditional ecology for decades. It is in this area that a new and more integrative ecology will really alter our understanding: when ecological ideas developed from plants and animals are tested and validated using experimental microbial systems, then a real systems biology for ecology will emerge. When fully integrated within traditional ecology, NGS has the potential to ask many questions that previously could simply not be answered, and to not only forge links across multiple levels of organisation within ecology, but also to bind ecology and evolution more tightly together. We can therefore 'expect the unexpected' in the near future: whilst we have attempted to outline some of the prospects for the development of this field, many of the new and exciting advances that are undoubtedly around the corner will come as surprises to us all.

ACKNOWLEDGEMENTS

K. J. P. and B. B. O. were funded by the EU Commission (MicroComXT a Marie Curie Excellence Team MEXT-CT-2005-024112). The authors thank Andrew R. Leitch and A. Mira for helpful comments, and Murray Thompson for his assistance with the graphics.

REFERENCES

Ahn, S.M., Kim, T.H., Lee, S., Kim, D., Ghang, H., Kim, D.S., Kim, B.C., Kim, S.Y., Kim, W.Y., Kim, C., Park, D., Lee, Y.S., *et al.* (2009). The first Korean genome sequence and analysis: Full genome sequencing for a socio-ethnic group. *Genome Res.* **19**, 1622–1629.

Amann, R., and Fuchs, B.M. (2008). Single-cell identification in microbial communities by improved fluorescence *in situ* hybridization techniques. *Nat. Rev. Microbiol.* **6**, 339–348.

Ardelean, I.I., Moisescu, C., and Popoviciu, D.R. (2008). Magnetotactic bacteria and their potential for terraformation. In: *From Fossils to Astrobiology: Records of Life on Earth and the Search for Extraterrestrial Biosignatures* (Ed. by J. Seckbach and M. Walsh), pp. 335–350. Springer-Verlag, New York.

Bailly, J., Fraissinet-Tachet, L., Verner, M.C., Debaud, J.C., Lemaire, M., Wesolowski-Louvel, M., and Marmeisse, R. (2007). Soil eukaryotic functional diversity, a metatranscriptomic approach. *ISME J.* **1**, 632–642.

Baker, B.J., Tyson, G.W., Webb, R.I., Flanagan, J., Hugenholtz, P., Allen, E.E., and Banfield, J.F. (2006). Lineages of acidophilic archaea revealed by community genomic analysis. *Science* **314**, 1933–1935.

Bascompte, J. (2010). Structure and dynamics of ecological networks. *Science* **329**, 765–766.

Bascompte, J., and Jordano, P. (2007). Plant-animal mutualistic networks: The architecture of biodiversity. *Annu. Rev. Ecol. Evol. Syst.* **38**, 567–593.

Beja, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P., Jovanovich, S., Gates, C.M., Feldman, R.A., Spudich, J.L., Spudich, E.N., and DeLong, E.F. (2000). Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. *Science* **289**, 1902–1906.

Bell, T., Newman, J.A., Silverman, B.W., Turner, S.L., and Lilley, A.K. (2005). The contribution of species richness and composition to bacterial services. *Nature* **436**, 1157–1160.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., Boutell, J.M., Bryant, J., *et al.* (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59.

Bernhard, A.E., Tucker, J., Giblin, A.E., and Stahl, D.A. (2007). Functionally distinct communities of ammonia-oxidizing bacteria along an estuarine salinity gradient. *Environ. Microbiol.* **9**, 1439–1447.

Biddle, J.F., Fitz-Gibbon, S., Schuster, S.C., Brenchley, J.E., and House, C.H. (2008). Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment. *Proc. Natl. Acad. Sci. USA* **105**, 10583–10588.

Bloom, J.S., Khan, Z., Kruglyak, L., Singh, M., and Caudy, A.A. (2009). Measuring differential gene expression by short read sequencing: Quantitative comparison to 2-channel gene expression microarrays. *BMC Genomics* **10**, 221.

Boschker, H.T.S., Nold, S.C., Wellsbury, P., Bos, D., de Graaf, W., Pel, R., Parkes, R.J., and Cappenberg, T.E. (1998). Direct linking of microbial populations to specific biogeochemical processes by C^{13}-labelling of biomarkers. *Nature* **392**, 801–805.

Braker, G., Fesefeldt, A., and Witzel, K.P. (1998). Development of PCR primer systems for amplification of nitrite reductase genes (*nirK* and *nirS*) to detect denitrifying bacteria in environmental samples. *Appl. Environ. Microbiol.* **64**, 3769–3775.

Breitbart, M., and Rohwer, F. (2005). Here a virus, there a virus, everywhere the same virus? *Trends Microbiol.* **13**, 278–284.

Britschgi, T.B., and Giovannoni, S.J. (1991). Phylogenetic analysis of a natural marine bacterioplankton population by rRNA gene cloning and sequencing. *Appl. Environ. Microbiol.* **57**, 1707–1713.

Brown, J.H., Gillooly, J.F., Allen, A.P., Savage, V.M., and West, G.B. (2004). Toward a metabolic theory of ecology. *Ecology* **85**, 1771–1789.

Buckley, M.R. (2004). *The Global Genome Question: Microbes as the Key to Understanding Evolution and Ecology*. American Academy of Microbiology, Washington, DC.

Cardinale, B.J., Weis, J.J., Forbes, A.E., Tilmon, K.J., and Ives, A.R. (2006). Biodiversity as both a cause and consequence of resource availability: A study of reciprocal causality in a predator-prey system. *J. Anim. Ecol.* **75**, 497–505.

Carreon-Martinez, L., and Heath, D.D. (2010). Revolution in food web analysis and trophic ecology: Diet analysis by DNA and stable isotope analysis. *Mol. Ecol.* **19**, 25–27.

Chaisson, M.J., and Pevzner, P.A. (2008). Short read fragment assembly of bacterial genomes. *Genome Res.* **18**, 324–330.

Chatton, E. (1938). *Titres et Travaux Scientifiques [Titles and Scientific Works] (1906-1937)*. E. Sottano, Ste, France.

Chen, Y., and Murrell, J.C. (2010). When metagenomics meets stable-isotope probing: Progress and perspectives. *Trends Microbiol.* **18**, 157–163.

Chin, K.-J., Sharma, M., Russell, L., O'Neill, K., and Lovley, D. (2008). Quantifying expression of a dissimilatory (bi)sulfite reductase gene in petroleum-contaminated marine harbor sediments. *Microb. Ecol.* **55**, 489–499.

Chivian, D., Brodie, E.L., Alm, E.J., Culley, D.E., Dehal, P.S., DeSantis, T.Z., Gihring, T.M., Lapidus, A., Lin, L.-H., Lowry, S.R., Moser, D.P., Richardson, P.M., *et al.* (2008). Environmental genomics reveals a single-species ecosystem deep within Earth. *Science* **322**, 275–278.

Cho, J.C., and Tiedje, J.M. (2000). Biogeography and degree of endemicity of fluorescent *Pseudomonas* strains in soil. *Appl. Environ. Microbiol.* **66**, 5448–5456.

Cloonan, N., Forrest, A.R.R., Kolle, G., Gardiner, B.B.A., Faulkner, G.J., Brown, M.K., Taylor, D.F., Steptoe, A.L., Wani, S., Bethel, G., Robertson, A.J., Perkins, A.C., *et al.* (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods* **5**, 613–619.

Cohan, F.M. (2001). Bacterial species and speciation. *Syst. Biol.* **50**, 513–524.

Collins, F.S., Lander, E.S., Rogers, J., and Waterston, R.H.The International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945.

Coolon, J.D., Jones, K.L., Narayanan, S., and Wisely, S.M. (2010). Microbial ecological response of the intestinal flora of *Peromyscus maniculatus* and *P. leucopus* to heavy metal contamination. *Mol. Ecol.* **19**, 67–80.

Cox-Foster, D.L., Conlan, S., Holmes, E.C., Palacios, G., Evans, J.D., Moran, N.A., Quan, P.L., Briese, T., Hornig, M., Geiser, D.M., Martinson, V., vanEngelsdorp, D., *et al.* (2007). A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* **318**, 283–287.

Cuvelier, M.L., Ortiz, A., Kim, E., Moehlig, H., Richardson, D.E., Heidelberg, J.F., Archibald, J.M., and Worden, A.Z. (2008). Widespread distribution of a unique marine protistan lineage. *Environ. Microbiol.* **10**, 1621–1634.

Dar, S.A., Yao, L., van Dongen, U., Kuenen, J.G., and Muyzer, G. (2007). Analysis of diversity and activity of sulfate-reducing bacterial communities in sulfidogenic bioreactors using 16S rRNA and *dsrB* genes as molecular markers. *Appl. Environ. Microbiol.* **73**, 594–604.

D'Costa, V.M., Griffiths, E., and Wright, G.D. (2007). Expanding the soil antibiotic resistome: Exploring environmental diversity. *Curr. Opin. Microbiol.* **10**, 481–489.

Debroas, D., Humbert, J.F., Enault, F., Bronner, G., Faubladier, M., and Cornillot, E. (2009). Metagenomic approach studying the taxonomic and functional diversity of the bacterial community in a mesotrophic lake (Lac du Bourget – France). *Environ. Microbiol.* **11**, 2412–2424.

DeLong, E.F. (2009). The microbial ocean from genomes to biomes. *Nature* **459**, 200–206.

DeLong, E.F., Wu, K.Y., Prézelin, B.B., and Jovine, R.V.M. (1994). High abundance of Archaea in Antarctic marine picoplankton. *Nature* **371**, 695–697.

DeLong, J.P., Okie, J.G., Moses, M.E., Sibly, R.M., and Brown, J.H. (2010). Shifts in metabolic scaling, production, and efficiency across major evolutionary transitions of life. *Proc. Natl. Acad. Sci. USA* **107**, 12941–12945.

DeSantis, T.Z., Brodie, E.L., Moberg, J.P., Zubieta, I.X., Piceno, Y.M., and Andersen, G.L. (2007). High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. *Microb. Ecol.* **53**, 371–383.

Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M., Furlan, M., Desnues, C., Haynes, M., Li, L., McDaniel, L., Moran, M.A., *et al.* (2008). Functional metagenomic profiling of nine biomes. *Nature* **452**, 629–632.

Dolfing, J. (2001). The microbial logic behind the prevalence of incomplete oxidation of organic compounds by acetogenic bacteria in methanogenic environments. *Microb. Ecol.* **41**, 83–89.

Dressman, D., Yan, H., Traverso, G., Kinzler, K.W., and Vogelstein, B. (2003). Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc. Natl. Acad. Sci. USA* **100**, 8817–8822.

Dunn, C., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., *et al.* (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* **452**, 745–749.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., *et al.* (2009). Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133–138.

Ellers, J. (2010). Evolutionary processes in community ecology. In: *Community Ecology: Processes, Models and Applications* (Ed. by H.A. Verhoef and P.J. Morin), pp. 151–162. Oxford University Press, Oxford.

Erguder, T.H., Boon, N., Wittebolle, L., Marzorati, M., and Verstraete, W. (2009). Environmental factors shaping the ecological niches of ammonia-oxidizing archaea. *FEMS Microbiol. Rev.* **33**, 855–869.

Erkel, C., Kube, M., Reinhardt, R., and Liesack, W. (2006). Genome of Rice Cluster I Archaea – The key methane producers in the rice rhizosphere. *Science* **313**, 370–372.

Ettwig, K.F., Butler, M.K., Le Paslier, D., Pelletier, E., Mangenot, S., Kuypers, M.M.M., Schreiber, F., Dutilh, B.E., Zedelius, J., de Beer, D., Gloerich, J., Wessels, H.J.C.T., *et al.* (2010). Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature* **464**, 543–548.

Falkowski, P.G., Fenchel, T., and Delong, E.F. (2008). The microbial engines that drive Earth's biogeochemical cycles. *Science* **320**, 1034–1039.

Feder, M.E., and Walser, J.C. (2005). The biological limitations of transcriptomics in elucidating stress and stress responses. *J. Evol. Biol.* **18**, 901–910.

Fedurco, M., Romieu, A., Williams, S., Lawrence, I., and Turcatti, G. (2006). BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res.* **34**, e22.

Fenchel, T. (1992). What can ecologists learn from microbes: Life beneath a square centimetre of sediment surface. *Funct. Ecol.* **6**, 499–507.

Fitter, A.H. (2005). Darkness visible: Reflections on underground ecology. *J. Ecol.* **93**, 231–243.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., *et al.* (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512.

Fox, G.E., Stackebrandt, E., Hespell, R.B., Gibson, J., Maniloff, J., Dyer, T.A., Wolfe, R.S., Balch, W.E., Tanner, R., Magrun, L., Zabien, L.B., Blakemore, R., *et al.* (1980). The phylogeny of procaryotes. *Science* **209**, 457–463.

Fraser, C., Hanage, W.P., and Spratt, B.G. (2007). Recombination and the nature of bacterial speciation. *Science* **315**, 476–480.

Fuhrman, J.A., and Steele, J.A. (2007). Community structure of marine bacterioplankton: Patterns, networks, and relationships to function. *Aquat. Microb. Ecol.* **53**, 69–81.

Fuhrman, J.A., Steele, J.A., Hewson, I., Schwalbach, M.S., Brown, M.V., Green, J.L., and Brown, J.H. (2008). A latitudinal diversity gradient in planktonic marine bacteria. *Proc. Natl. Acad. Sci. USA* **105**, 7774–7778.

Fussmann, G.F., Loreau, M., and Abrams, P.A. (2007). Eco-evolutionary dynamics of communities and ecosystems. *Funct. Ecol.* **21**, 465–477.

Gianoulis, T.A., Raes, J., Patel, P.V., Bjornson, R., Korbel, J.O., Letunic, I., Yamada, T., Paccanaro, A., Jensen, L.J., Snyder, M., Bork, P., and Gerstein, M.B. (2009). Quantifying environmental adaptation of metabolic pathways in metagenomics. *Proc. Natl. Acad. Sci. USA* **106**, 1374–1379.

Gifford, S., Sharma, S., Rinta-Kanto, J., and Moran, M.A. (2010). Metatranscriptomic analysis of a summer bacterioplankton community in Southeastern U.S. coastal waters. *In13th International Symposium on Microbial Ecology*. ISME, Seattle.

Gilbert, J.A., Field, D., Huang, Y., Edwards, R., Li, W.Z., Gilna, P., and Joint, I. (2008). Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. *PLoS ONE* **3**, e3042.

Gilbert, J.A., Field, D., Swift, P., Newbold, L., Oliver, A., Smyth, T., Somerfield, P.J., Huse, S., and Joint, I. (2009). The seasonal structure of microbial communities in the Western English Channel. *Environ. Microbiol.* **11**, 3132–3139.

Giovannoni, S.J., DeLong, E.F., Olsen, G.J., and Pace, N.R. (1988). Phylogenetic group-specific oligodeoxynucleotide probes for identification of single microbial cells. *J. Bacteriol.* **170**, 720–726.

Giovannoni, S.J., Britschgi, T.B., Moyer, C.L., and Field, K.G. (1990). Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345**, 60–63.

Giovannoni, S.J., Rappé, M.S., Vergin, K.L., and Adair, N.L. (1996). 16S rRNA genes reveal stratified open ocean bacterioplankton populations related to the green non-sulfur bacteria. *Proc. Natl. Acad. Sci. USA* **93**, 7979–7984.

Goetz, F., Rosauer, D., Sitar, S., Goetz, G., Simchick, C., Roberts, S., Johnson, R., Murphy, C., Bronte, C.R., and Mackenzie, S. (2010). A genetic basis for the phenotypic differentiation between siscowet and lean lake trout (*Salvelinus namaycush*). *Mol. Ecol.* **19**, 176–196.

Gomez-Alvarez, V., Teal, T.K., and Schmidt, T.M. (2009). Systematic artifacts in metagenomes from complex microbial communities. *ISME J.* **3**, 1314–1317.

Good, I.J. (1953). The population frequency of species and the estimation of population parameters. *Biometrika* **40**, 237–262.

Green, J., and Bohannan, B.J.M. (2006). Spatial scaling of microbial biodiversity. *Trends Ecol. Evol.* **21**, 501–507.

Green, R.E., Krause, J., Ptak, S.E., Briggs, A.W., Ronan, M.T., Simons, J.F., Du, L., Egholm, M., Rothberg, J.M., Paunovic, M., and Paabo, S. (2006). Analysis of one million base pairs of Neanderthal DNA. *Nature* **444**, 330–336.

Handelsman, J. (2004). Metagenomics: Application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* **68**, 669–685.

Harris, T.D., Buzby, P.R., Babcock, H., Beer, E., Bowers, J., Braslavsky, I., Causey, M., Colonell, J., Dimeo, J., Efcavitch, J.W., Giladi, E., Gill, J., *et al.* (2008). Single-molecule DNA sequencing of a viral genome. *Science* **320**, 106–109.

He, Z., Gentry, T.J., Schadt, C.W., Wu, L., Liebich, J., Chong, S.C., Huang, Z., Wu, W., Gu, B., Jardine, P., Criddle, C., and Zhou, J. (2007). GeoChip: A comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J.* **1**, 67–77.

He, Z., Deng, Y., Van Nostrand, J.D., Tu, Q., Xu, M., Hemme, C.L., Li, X., Wu, L., Gentry, T.J., Yin, Y., Liebich, J., Hazen, T.C., *et al.* (2010a). GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. *ISME J.* **4**, 1167–1179.

He, Z., Xu, M., Deng, Y., Kang, S., Kellogg, L., Wu, L., Van Nostrand, J.D., Hobbie, S.E., Reich, P.B., and Zhou, J. (2010b). Metagenomic analysis reveals a marked divergence in the structure of belowground microbial communities at elevated CO_2. *Ecol. Lett.* **13**, 564–575.

Hernandez, D., Francois, P., Farinelli, L., Osteras, M., and Schrenzel, J. (2008). De novo bacterial genome sequencing: Millions of very short reads assembled on a desktop computer. *Genome Res.* **18**, 802–809.

Hillier, L.W., Marth, G.T., Quinlan, A.R., Dooling, D., Fewell, G., Barnett, D., Fox, P., Glasscock, J.I., Hickenbotham, M., Huang, W.C., Magrini, V.J., Richt, R. J., *et al.* (2008). Whole-genome sequencing and variant discovery in *C. elegans*. *Nat. Methods* **5**, 183–188.

Hofmann, G.E., and Gaines, S.D. (2008). New tools to meet new challenges: Emerging technologies for managing marine ecosystems for resilience. *Bioscience* **58**, 43–52.

Hooper, L.V., Midtvedt, T., and Gordon, J.I. (2002). How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annu. Rev. Nutr.* **22**, 283–307.

Hudson, M.E. (2008). Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol. Ecol. Resour.* **8**, 3–17.

Hugenholtz, P., Goebel, B.M., and Pace, N.R. (1998). Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* **180**, 4765–4774.

Hunkapiller, T., Kaiser, R.J., Koop, B.F., and Hood, L. (1991). Large-scale and automated DNA-sequence determination. *Science* **254**, 59–67.

Hurd, P.J., and Nelson, C.J. (2009). Advantages of next-generation sequencing versus the microarray in epigenetic research. *Brief. Funct. Genomic Proteomic* **8**, 174–183.

Hutchinson, G.E. (1961). The paradox of the plankton. *Am. Nat.* **95**, 137–145.

Ings, T.C., Montoya, J.M., Bascompte, J., Bluthgen, N., Brown, L., Dormann, C.F., Edwards, F., Figueroa, D., Jacob, U., Jones, J.I., Lauridsen, R.B., Ledger, M.E., *et al*. (2009). Ecological networks – Beyond food webs. *J. Anim. Ecol.* **78**, 253–269.

Jackson, B.E., and McInerney, M.J. (2002). Anaerobic microbial metabolism can proceed close to thermodynamic limits. *Nature* **415**, 454–456.

Johnson, M.T.J., and Stinchcombe, J.R. (2007). An emerging synthesis between community ecology and evolutionary biology. *Trends Ecol. Evol.* **22**, 250–257.

Jones, S.E., and Lennon, J.T. (2009). Evidence for limited microbial transfer of methane in a planktonic food web. *Aquat. Microb. Ecol.* **58**, 45–53.

Jones, L.E., Becks, L., Ellner, S.P., Hairston, N.G., Yoshida, T., and Fussmann, G.F. (2009). Rapid contemporary evolution and clonal food web dynamics. *Philos. Trans. R. Soc. B-Biol. Sci.* **364**, 1579–1591.

Joulian, C., Ramsing, N.B., and Ingvorsen, K. (2001). Congruent phylogenies of most common small-subunit rRNA and dissimilatory sulfite reductase gene sequences retrieved from estuarine sediments. *Appl. Environ. Microbiol.* **67**, 3314–3318.

Jumpponen, A., Jones, K.L., Mattox, D., and Yaege, C. (2010). Massively parallel 454-sequencing of fungal communities in *Quercus* spp. ectomycorrhizas indicates seasonal dynamics in urban and rural sites. *Mol. Ecol.* **19**, 41–53.

Kandeler, E., Deiglmayr, K., Tscherko, D., Bru, D., and Philippot, L. (2006). Abundance of *narG, nirS, nirK*, and *nosZ* genes of denitrifying bacteria during primary successions of a Glacier Foreland. *Appl. Environ. Microbiol.* **72**, 5957–5962.

Kant, M.R., and Baldwin, I.T. (2007). The ecogenetics and ecogenomics of plant-herbivore interactions: Rapid progress on a slippery road. *Curr. Opin. Genet. Dev.* **17**, 519–524.

Kassen, R., and Rainey, P.B. (2004). The ecology and genetics of microbial diversity. *Annu. Rev. Microbiol.* **58**, 207–231.

Kemp, P.F., Lee, S.-H., and LaRoche, J. (1993). Estimating the growth rate of slowly growing marine bacteria from RNA content. *Appl. Environ. Microbiol.* **59**, 2594–2601.

Kennedy, J., Marchesi, J.R., and Dobson, A.D.W. (2008). Marine metagenomics: Strategies for the discovery of novel enzymes with biotechnological applications from marine environments. *Microb. Cell Fact.* **7**, 27.

Kennedy, J., Flemer, B., Jackson, S.A., Lejon, D.P.H., Morrissey, J.P., O'Gara, F., and Dobson, A.D.W. (2010). Marine metagenomics: New tools for the study and exploitation of marine microbial metabolism. *Mar. Drugs* **8**, 608–628.

Kim, J.I., Ju, Y.S., Park, H., Kim, S., Lee, S., Yi, J.H., Mudge, J., Miller, N.A., Hong, D., Bell, C.J., Kim, H.S., Chung, I.S., *et al.* (2009). A highly annotated whole-genome sequence of a Korean individual. *Nature* **460**, 1011–1015.

Kloch, A., Babik, W., Bajer, A., Sinski, E., and Radwan, J. (2010). Effects of an MHC-DRB genotype and allele number on the load of gut parasites in the bank vole *Myodes glareolus*. *Mol. Ecol.* **19**, 255–265.

Könneke, M., Bernhard, A.E., de la Torre, J.R., Walker, C.B., Waterbury, J.B., and Stahl, D.A. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**, 543–546.

Korbel, J.O., Urban, A.E., Affourtit, J.P., Godwin, B., Grubert, F., Simons, J.F., Kim, P.M., Palejev, D., Carriero, N.J., Du, L., Taillon, B.E., Chen, Z.T., *et al.* (2007). Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318**, 420–426.

Kummerli, R., Griffin, A.S., West, S.A., Buckling, A., and Harrison, F. (2009). Viscous medium promotes cooperation in the pathogenic bacterium *Pseudomonas aeruginosa*. *Proc. R. Soc. B-Biol. Sci.* **276**, 3531–3538.

Kunin, V., Engelbrektson, A., Ochman, H., and Hugenholtz, P. (2009). Wrinkles in the rare biosphere: Pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ. Microbiol.* **12**, 118–123.

Lafi, F.F., Fuerst, J.A., Fieseler, L., Engels, C., Goh, W.W.L., and Hentschel, U. (2009). Widespread distribution of Poribacteria in Demospongiae. *Appl. Environ. Microbiol.* **75**, 5695–5699.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.

Lee, N., Nielsen, P.H., Andreasen, K.H., Juretschko, S., Nielsen, J.L., Schleifer, K.H., and Wagner, M. (1999). Combination of fluorescent in situ hybridization and microautoradiography – A new tool for structure-function analyses in microbial ecology. *Appl. Environ. Microbiol.* **65**, 1289–1297.

Lefrancois, P., Euskirchen, G.M., Auerbach, R.K., Rozowsky, J., Gibson, T., Yellman, C.M., Gerstein, M., and Snyder, M. (2009). Efficient yeast ChIP-Seq using multiplex short-read DNA sequencing. *BMC Genomics* **10**, e37.

Levene, M.J., Korlach, J., Turner, S.W., Foquet, M., Craighead, H.G., and Webb, W.W. (2003). Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299**, 682–686.

Li, L.L., Victoria, J.G., Wang, C.L., Jones, M., Fellers, G.M., Kunz, T.H., and Delwart, E. (2010). Bat Guano virome: Predominance of dietary viruses from insects and plants plus novel mammalian viruses. *J. Virol.* **84**, 6955–6965.

Li, R.Q., Fan, W., Tian, G., Zhu, H.M., He, L., Cai, J., Huang, Q.F., Cai, Q.L., Li, B., Bai, Y.Q., Zhang, Z.H., Zhang, Y.P., *et al.* (2010). The sequence and de novo assembly of the giant panda genome. *Nature* **463**, 311–317.

Li, R.Q., Zhu, H.M., Ruan, J., Qian, W.B., Fang, X.D., Shi, Z.B., Li, Y.R., Li, S.T., Shan, G., Kristiansen, K., Li, S.G., Yang, H.M., *et al.* (2010). De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* **20**, 265–272.

Lindeman, R.L. (1942). The trophic-dynamic aspect of ecology. *Ecology* **23**, 399–418.

Lipson, D., Raz, T., Kieu, A., Jones, D.R., Giladi, E., Thayer, E., Thompson, J.F., Letovsky, S., Milos, P., and Causey, M. (2009). Quantification of the yeast transcriptome by single-molecule sequencing. *Nat. Biotechnol.* **27**, 652–658.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**, 523–536.

Liu, J.M., Livny, J., Lawrence, M.S., Kimball, M.D., Waldor, M.K., and Camilli, A. (2009). Experimental discovery of sRNAs in *Vibrio cholerae* by direct cloning, 5S/tRNA depletion and parallel sequencing. *Nucleic Acids Res.* **37**, e46.

Lo, I., Denef, V.J., VerBerkmoes, N.C., Shah, M.B., Goltsman, D., DiBartolo, G., Tyson, G.W., Allen, E.E., Ram, R.J., Detter, J.C., Richardson, P., Thelen, M.P., *et al.* (2007). Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. *Nature* **446**, 537–541.

Madsen, E.L. (2005). Identifying microorganisms responsible for ecologically significant biogeochemical processes. *Nat. Rev. Microbiol.* **3**, 439–446.

Makarieva, A.M., Gorshkov, V.G., Li, B.L., Chown, S.L., Reich, P.B., and Gavrilov, V.M. (2008). Mean mass-specific metabolic rates are strikingly similar across life's major domains: Evidence for life's metabolic optimum. *Proc. Natl. Acad. Sci. USA* **105**, 16994–16999.

Manefield, M., Whiteley, A.S., Griffiths, R.I., and Bailey, M.J. (2002). RNA stable isotope probing, a novel means of linking microbial community function to phylogeny. *Appl. Environ. Microbiol.* **68**, 5367–5373.

Marco, D. (2010). *Metagenomics: Theory, Methods and Applications*. Caister Academic Press, Norfolk, UK.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z.T., Dewell, S.B., Du, L., *et al.* (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380.

Martiny, A.C., Kathuria, S., and Berube, P.M. (2009). Widespread metabolic potential for nitrite and nitrate assimilation among *Prochlorococcus* ecotypes. *Proc. Natl. Acad. Sci. USA* **106**, 10787–10792.

McGeogh, M.A., and Gaston, K.J. (2002). Occupancy frequency distributions: Patterns, artefacts and mechanisms. *Biol. Rev.* **77**, 311–331.

McKernan, K.J., Peckham, H.E., Costa, G.L., McLaughlin, S.F., Fu, Y.T., Tsung, E.F., Clouser, C.R., Duncan, C., Ichikawa, J.K., Lee, C.C., Zhang, Z., Ranade, S.S., *et al.* (2009). Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res.* **19**, 1527–1541.

Medini, D., Donati, C., Tettelin, H., Masignani, V., and Rappuoli, R. (2005). The microbial pan-genome. *Curr. Opin. Genet. Dev.* **15**, 589–594.

Metzker, M.L. (2010). Sequencing technologies – The next generation. *Nat. Rev. Genet.* **11**, 31–46.

Meyer, M., Stenzel, U., Myles, S., Prufer, K., and Hofreiter, M. (2007). Targeted high-throughput sequencing of tagged nucleic acid samples. *Nucleic Acids Res.* **35**, e97.

Meyer, M., Stenzel, U., and Hofreiter, M. (2008). Parallel tagged sequencing on the 454 platform. *Nat. Protoc.* **3**, 267–278.

Miller, W., Drautz, D.I., Ratan, A., Pusey, B., Qi, J., Lesk, A.M., Tomsho, L.P., Packard, M.D., Zhao, F.Q., Sher, A., Tikhonov, A., Raney, B., *et al.* (2008). Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* **456**, 387–390.

Moran, N.A., McLaughlin, H.J., and Sorek, R. (2009). The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. *Science* **323**, 379–382.

Morin, R.D., Bainbridge, M., Fejes, A., Hirst, M., Krzywinski, M., Pugh, T.J., McDonald, H., Varhol, R., Jones, S.J.M., and Marra, M.A. (2008). Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *Biotechniques* **45**, 81–94.

Morris, R.M., Rappe, M.S., Connon, S.A., Vergin, K.L., Siebold, W.A., Carlson, C. A., and Giovannoni, S.J. (2002). SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420**, 806–810.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628.

Mou, X., Sun, S., Edwards, R.A., Hodson, R.E., and Moran, M.A. (2008). Bacterial carbon processing by generalist species in the coastal ocean. *Nature* **451**, 708–711.

Moya-Laraño, J. (2010). Can temperature and water availability contribute to the maintenance of latitudinal diversity by increasing the rate of biotic interactions? *Open Ecol. J.* **3**, 1–13.

Moya-Laraño, J. (in press) Genetic variation, predator-prey interactions and food web structure. *Philos. Trans. R. Soc. B-Biol. Sci.*

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344–1349.

Nelson, D.L., and Cox, M.M. (2009). *Lehninger's Principle of Biochemistry*. WH Freeman, New York.

Nolte, V., Pandey, R.V., Jost, S., Medinger, R., Ottenwälder, B., Boenigk, J., and Christian, S. (2010). Contrasting seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity. *Mol. Ecol.* **19**, 2908–2915.

Noonan, J.P., Hofreiter, M., Smith, D., Priest, J.R., Rohland, N., Rabeder, G., Krause, J., Detter, J.C., Paabo, S., and Rubin, E.M. (2005). Genomic sequencing of Pleistocene cave bears. *Science* **309**, 597–600.

Noonan, J.P., Coop, G., Kudaravalli, S., Smith, D., Krause, J., Alessi, J., Platt, D., Paabo, S., Pritchard, J.K., and Rubin, E.M. (2006). Sequencing and analysis of Neanderthal genomic DNA. *Science* **314**, 1113–1118.

Nusbaum, C., Ohsumi, T.K., Gomez, J., Aquadro, J., Victor, T.C., Warren, R.M., Hung, D.T., Birren, B.W., Lander, E.S., and Jaffe, D.B. (2009). Sensitive, specific polymorphism discovery in bacteria using massively parallel sequencing. *Nat. Methods* **6**, 67–69.

Oakley, B.B., Carbonero, F., van der Gast, C.J., Hawkins, R.J., and Purdy, K.J. (2010). Evolutionary divergence and biogeography of sympatric niche-differentiated bacterial populations. *ISME J.* **4**, 488–497.

Oren, A. (2004). Prokaryote diversity and taxonomy: Current status and future challenges. *Philos. Trans. R. Soc. Lond. B* **359**, 623–638.

Ossowski, S., Schneeberger, K., Clark, R.M., Lanz, C., Warthmann, N., and Weigel, D. (2008). Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Res.* **18**, 2024–2033.

Östman, O., Drakare, S., Kritzberg, E.S., Langenheder, S., Logue, J.B., and Lindstrom, E.S.Regional invariance among microbial communities. *Ecol. Lett.* **13**, 118–127.

Ouverney, C.C., and Fuhrman, J.A. (1999). Combined Microautoradiography-16S rRNA probe technique for determination of radioisotope uptake by specific microbial cell types in situ. *Appl. Environ. Microbiol.* **65**, 1746–1752.

Pace, N.R., Olsen, G.J., and Woese, C.R. (1986). Ribosomal RNA phylogeny and the primary lines of evolutionary descent. *Cell* **45**, 325–326.

Paine, R.T. (1966). Food web complexity and species diversity. *Am. Nat.* **100**, 65–75.
Passalacqua, K.D., Varadarajan, A., Ondov, B.D., Okou, D.T., Zwick, M.E., and Bergman, N.H. (2009). Structure and complexity of a bacterial transcriptome. *J. Bacteriol.* **191**, 3203–3211.
Pelletier, F., Garant, D., and Hendry, A.P. (2009). Eco-evolutionary dynamics. Introduction. *Philos. Trans. R. Soc. B-Biol. Sci.* **364**, 1483–1489.
Perkins, T.T., Kingsley, R.A., Fookes, M.C., Gardner, P.P., James, K.D., Yu, L., Assefa, S.A., He, M., Croucher, N.J., Pickard, D.J., Maskell, D.J., Parkhill, J., *et al*. (2009). A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella typhi*. *PLoS Genet.* **5**, e1000569.
Perkins, D.M., McKie, B.G., Malmqvist, B., Gilmour, S.G., Reiss, J., and Woodward, G. (2010). Environmental warming and biodiversity-ecosystem functioning in freshwater microcosms: Partitioning the effects of species identity, richness and metabolism. *Adv. Ecol. Res.* **43**, 177–209.
Pernthaler, A., Dekas, A.E., Brown, C.T., Goffredi, S.K., Embaye, T., and Orphan, V.J. (2008). Diverse syntrophic partnerships from-deep-sea methane vents revealed by direct cell capture and metagenomics. *Proc. Natl. Acad. Sci. USA* **105**, 7052–7057.
Petchey, O.L., McPhearson, P.T., Casey, T.M., and Morin, P.J. (1999). Environmental warming alters food-web structure and ecosystem function. *Nature* **402**, 69–72.
Petchey, O.L., Beckerman, A.P., Riede, J.O., and Warren, P.H. (2008). Size, foraging, and food web structure. *Proc. Natl. Acad. Sci. USA* **105**, 4191–4196.
Phillips, D.A., Ferris, H., Cook, D.R., and Strong, D.R. (2003). Molecular control points in rhizosphere food webs. *Ecology* **84**, 816–826.
Picardi, E., Horner, D.S., Chiara, M., Schiavon, R., Valle, G., and Pesole, G. (2010). Large-scale detection and analysis of RNA editing in grape mtDNA by RNA deep-sequencing. *Nucleic Acids Res.* **38**, 4755–4767.
Piganeau, G., Desdevises, Y., Derelle, E., and Moreau, H. (2008). Picoeukaryotic sequences in the Sargasso Sea metagenome. *Genome Biol.* **9**, r5.
Proulx, S.R., Promislow, D.E.L., and Phillips, P.C. (2005). Network thinking in ecology and evolution. *Trends Ecol. Evol.* **20**, 345–353.
Ptacnik, R., Moorthi, S.D., and Hillebrand, H. (2010). Hutchinson reversed, or why there need to be so many species. *Adv. Ecol. Res.* **43**, 1–43.
Purdy, K.J., Nedwell, D.B., Embley, T.M., and Takii, S. (1997). Use of 16S rRNA-targeted oligonucleotide probes to investigate the occurrence and selection of sulphate-reducing bacteria in response to nutrient addition to sediment slurry microcosms from a Japanese estuary. *FEMS Microbiol. Ecol.* **24**, 221–234.
Purdy, K.J., Embley, T.M., and Nedwell, D.B. (2002). The distribution and activity of sulphate reducing bacteria in estuarine and coastal marine sediments. *Antonie Van Leeuwenhoek* **81**, 181–187.
Purdy, K.J., Munson, M.A., Embley, T.M., and Nedwell, D.B. (2003). Use of 16S rRNA-targeted oligonucleotide probes to investigate function and phylogeny of sulphate-reducing bacteria and methanogenic archaea in a UK estuary. *FEMS Microbiol. Ecol.* **44**, 361–371.
Purdy, K.J., Nedwell, D.B., and Embley, T.M. (2003). Analysis of the sulphate-reducing bacterial and methanogenic archaeal populations in contrasting Antarctic sediments. *Appl. Environ. Microbiol.* **69**, 3181–3191.
Pushkarev, D., Neff, N.F., and Quake, S.R. (2009). Single-molecule sequencing of an individual human genome. *Nat. Biotechnol.* **27**, 847–850.
Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D.R., Li, J., *et al*. (2010). A human gut

microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65.

Quince, C., Lanzen, A., Curtis, T.P., Davenport, R.J., Hall, N., Head, I.M., Read, L.F., and Sloan, W.T. (2009). Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat. Methods* **6**, 639–641.

Radajewski, S., Ineson, P., Parekh, N.R., and Murrell, J.C. (2000). Stable isotope probing as a tool in microbial ecology. *Nature* **403**, 646–649.

Raffaelli, D. (2007). Food webs, body size and the curse of the latin binomial. In: *From Energetics to Ecosystems: The Dynamics and Structure of Ecological Systems* (Ed. by N. Rooney, K.S. McCann and D.L.G. Noakes), pp. 53–64. Springer, Netherlands.

Raffaelli, D., and Moller, H. (2000). Manipulative field experiments in animal ecology: Do they promise more than they can deliver? *Adv. Ecol. Res.* **30**, 299–338.

Ram, R.J., VerBerkmoes, N.C., Thelen, M.P., Tyson, G.W., Baker, B.J., Blake, R.C., Shah, M., Hettich, R.L., and Banfield, J.F. (2005). Community proteomics of a natural microbial biofilm. *Science* **308**, 1915–1920.

Raskin, L., Rittmann, B.E., and Stahl, D.A. (1996). Competition and coexistence of sulfate-reducing and methanogenic populations in anaerobic biofilms. *Appl. Environ. Microbiol.* **62**, 3847–3857.

Rasmussen, M., Li, Y.R., Lindgreen, S., Pedersen, J.S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R., Bertalan, M., Nielsen, K., *et al*. (2010). Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762.

Reiss, J., Bridle, J.R., Montoya, J.M., and Woodward, G. (2009). Emerging horizons in biodiversity and ecosystem functioning research. *Trends Ecol. Evol.* **24**, 505–514.

Reiss, J., Bailey, R.A., Cássio, F., Woodward, G., and Pascoal, C. (2010). Assessing the contribution of micro-organisms and macrofauna to biodiversity-ecosystem functioning relationships in freshwater microcosms. *Adv. Ecol. Res.* **43**, 151–176.

Reiss, J., Forster, J., Hirst, A.G., Cássio, F., Pascoal, C., Stewart, R., and Hirst, A.G. (2010). When microscopic organisms inform general ecological theory. *Adv. Ecol. Res.* **43**, 45–85.

Renaut, S., Nolte, A.W., and Bernatchez, L. (2010). Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. *Salmonidae*). *Mol. Ecol.* **19**, 115–131.

Rocap, G., Larimer, F.W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N.A., Arellano, A., Coleman, M., Hauser, L., Hess, W.R., Johnson, Z.I., Land, M., *et al*. (2003). Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**, 1042–1047.

Rodriguez-Brito, B., Li, L.L., Wegley, L., Furlan, M., Angly, F., Breitbart, M., Buchanan, J., Desnues, C., Dinsdale, E., Edwards, R., Felts, B., Haynes, M., *et al*. (2010). Viral and microbial community dynamics in four aquatic environments. *ISME J.* **4**, 739–751.

Rodriguez-Valera, F., Martin-Cuadrado, A.B., Rodriguez-Brito, B., Pasic, L., Thingstad, T.F., Rohwer, F., and Mira, A. (2009). Explaining microbial population genomics through phage predation. *Nat. Rev. Microbiol.* **7**, 828–836.

Ronaghi, M., Uhlen, M., and Nyren, P. (1998). A sequencing method based on real-time pyrophosphate. *Science* **281**, 363–365.

Rosset, R., Julien, J., and Monier, R. (1966). Ribonucleic acid composition of bacteria as a function of growth rate. *J. Mol. Biol.* **18**, 308–320.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., Wu, D.Y., Eisen, J.A., Hoffman, J.M., Remington, K., Beeson, K.,

Tran, B., *et al*. (2007). The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical pacific. *PLOS Biol.* **5**, 398–431.
Saiki, R.K., Bugawan, T.L., Horn, G.T., Mullis, K.B., and Erlich, H.A. (1986). Analysis of enzymatically amplified beta-globin and HLA-DQalpha DNA with allele-specific oligonucleotide probes. *Nature* **324**, 163–166.
Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H.A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487–491.
Sanger, F., and Coulson, A.R. (1975). Rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* **94**, 441–448.
Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchison, C.A., Slocombe, P.M., and Smith, M. (1977). Nucleotide sequence of bacteriophage $\psi\chi$174 DNA. *Nature* **265**, 687–695.
Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
Sarin, S., Prabhu, S., O'Meara, M.M., Pe'er, I., and Hobert, O. (2008). *Caenorhabditis elegans* mutant allele identification by whole-genome sequencing. *Nat. Methods* **5**, 865–867.
Schütte, U.M.E., Abdo, Z., Foster, J., Ravel, J., Bunge, J., Solheim, B., and Forney, L.J. (2010). Bacterial diversity in a glacier foreland of the high Arctic. *Mol. Ecol.* **19**, 54–66.
Sharma, V.K., Kumar, N., Prakash, T., and Taylor, T.D. (2010). MetaBioME: A database to explore commercially useful enzymes in metagenomic datasets. *Nucleic Acids Res.* **38**, D468–D472.
Shen, Y.F., Sarin, S., Liu, Y., Hobert, O., and Pe'er, I. (2008). Comparing platforms for *C. elegans* mutant identification using high-throughput whole-genome sequencing. *PLoS ONE* **3**, e4012.
Shi, Y.M., Tyson, G.W., and DeLong, E.F. (2009). Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* **459**, 266–269.
Shuster, S.M., Lonsdorf, E.V., Wimp, G.M., Bailey, J.K., and Whitham, T.G. (2006). Community heritability measures the evolutionary consequences of indirect genetic effects on community structure. *Evolution* **60**, 991–1003.
Sinigalliano, C.D., Kuhn, D.N., and Jones, R.D. (1995). Amplification of the amoA gene from diverse species of ammonium-oxidizing bacteria and from an indigenous bacterial population from seawater. *Appl. Environ. Microbiol.* **61**, 2702–2706.
Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B.H., and Hood, L.E. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature* **321**, 674–679.
Smith, C.J., Nedwell, D.B., Dong, L.F., and Osborn, A.M. (2007). Diversity and abundance of nitrate reductase genes (*narG* and *napA*), nitrite reductase genes (*nirS* and *nrfA*), and their transcripts in estuarine sediments. *Appl. Environ. Microbiol.* **73**, 3612–3622.
Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R., Arrieta, J.M., and Herndl, G.J. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl. Acad. Sci. USA* **103**, 12115–12120.
Srivatsan, A., Han, Y., Peng, J.L., Tehranchi, A.K., Gibbs, R., Wang, J.D., and Chen, R. (2008). High-precision, whole-genome sequencing of laboratory strains facilitates genetic studies. *PLoS Genet.* **4.**
Stanier, R.Y., and Smith, J.H.C. (1960). The chlorophylls of green bacteria. *Biochim. Biophys. Acta* **41**, 478–484.

Stanier, R.Y., and van Niel, C.B. (1962). The concept of a bacterium. *Arch. Microbiol.* **42**, 17–35.

Stevens, C.E., and Hume, I.D. (1998). Contributions of microbes in vertebrate gastrointestinal tract to production and conservation of nutrients. *Physiol. Rev.* **78**, 393–427.

Stewart, F.J., Ottesen, E.A., and DeLong, E.F. (2010). Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *ISME J.* **4**, 896–907.

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M.D.M., Breiner, H.W., and Richards, T.A. (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* **19**, 21–31.

Strous, M., Pelletier, E., Mangenot, S., Rattei, T., Lehner, A., Taylor, M.W., Horn, M., Daims, H., Bartol-Mavel, D., Wincker, P., Barbe, V., Fonknechten, N., *et al.* (2006). Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* **440**, 790–794.

Tettelin, H., Riley, D., Cattuto, C., and Medini, D. (2008). Comparative genomics: The bacterial pan-genome. *Curr. Opin. Microbiol.* **11**, 472–477.

Throbäck, I.N., Enwall, K., Jarvis, A., and Hallin, S. (2004). Reassessing PCR primers targeting *nirS*, *nirK* and *nosZ* genes for community surveys of denitrifying bacteria with DGGE. *FEMS Microbiol. Ecol.* **49**, 401–417.

Treusch, A.H., Leininger, S., Kletzin, A., Schuster, S.C., Klenk, H.P., and Schleper, C. (2005). Novel genes for nitrite reductase and *amo*-related proteins indicate a role of uncultivated mesophilic crenarchaeota in nitrogen cycling. *Environ. Microbiol.* **7**, 1985–1995.

Tringe, S.G., and Rubin, E.M. (2005). Metagenomics: DNA sequencing of environmental samples. *Nat. Rev. Genet.* **6**, 805–814.

Tringe, S.G., von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., Podar, M., Short, J.M., Mathur, E.J., Detter, J.C., Bork, P., Hugenholtz, P., *et al.* (2005). Comparative metagenomics of microbial communities. *Science* **308**, 554–557.

Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., and Gordon, J.I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031.

Tyler, H.L., Roesch, L.F.W., Gowda, S., Dawson, W.O., and Triplett, E.W. (2009). Confirmation of the sequence of '*Candidatus* Liberibacter asiaticus' and assessment of microbial diversity in huanglongbing-infected citrus phloem using a metagenomic approach. *Mol. Plant Microbe Interact.* **22**, 1624–1634.

Tyson, G.W., and Banfield, J.F. (2005). Cultivating the uncultivated: A community genomics perspective. *Trends Microbiol.* **13**, 411–415.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S., and Banfield, J. F. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37–43.

Tyson, G.W., Lo, I., Baker, B.J., Allen, E.E., Hugenholtz, P., and Banfield, J.F. (2005). Genome-directed isolation of the key nitrogen fixer *Leptospirillum ferrodiazotrophum* sp. nov. from an acidophilic microbial community. *Appl. Environ. Microbiol.* **71**, 6319–6324.

Urban, M.C., Leibold, M.A., Amarasekare, P., De Meester, L., Gomulkiewicz, R., Hochberg, M.E., Klausmeier, C.A., Loeuille, N., de Mazancourt, C., Norberg, J.,

Pantel, J.H., Strauss, S.Y., *et al.* (2008). The evolutionary ecology of metacommunities. *Trends Ecol. Evol.* **23**, 311–317.

Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., Zeng, K., Malek, J.A., Costa, G., McKernan, K., Sidow, A., Fire, A., *et al.* (2008). A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res.* **18**, 1051–1063.

Van Bers, N.E.M., Van Oers, K., Kerstens, H.H.D., Dibbits, B.W., Crooijmans, R., Visser, M.E., and Groenen, M.A.M. (2010). Genome-wide SNP detection in the great tit *Parus major* using high throughput sequencing. *Mol. Ecol.* **19**, 89–99.

Vandenkoornhuyse, P., Dufresne, A., Quaiser, A., Gouesbet, G., Binet, F., Francez, A.J., Mahe, S., Bormans, M., Lagadeuc, Y., and Couee, I. (2010). Integration of molecular functions at the ecosystemic level: Breakthroughs and future goals of environmental genomics and post-genomics. *Ecol. Lett.* **13**, 776–791.

van der Gast, C.J., Lilley, A.K., Ager, D., and Thompson, I.P. (2005). Island size and bacterial diversity in an archipelago of engineering machines. *Environ. Microbiol.* **7**, 1220–1226.

van Vliet, A.H.M. (2010). Next generation sequencing of microbial transcriptomes: Challenges and opportunities. *FEMS Microbiol. Lett.* **302**, 1–7.

VanGuilder, H.D., Vrana, K.E., and Freeman, W.M. (2008). Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques* **44**, 619–626.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., Fouts, D.E., Levy, S., *et al.* (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74.

Wagner, M., Roger, A.J., Flax, J.L., Brusseau, G.A., and Stahl, D.A. (1998). Phylogeny of dissimilatory sulfite reductases supports an early origin of sulfate respiration. *J. Bacteriol.* **180**, 2975–2982.

Wang, J., Wang, W., Li, R.Q., Li, Y.R., Tian, G., Goodman, L., Fan, W., Zhang, J.Q., Li, J., Zhang, J.B., Guo, Y.R., Feng, B.X., *et al.* (2008). The diploid genome sequence of an Asian individual. *Nature* **456**, 60–65.

Ward, B.B., and O'Mullan, G.D. (2005). Community level analysis: Genetic and biogeochemical approaches to investigate community composition and function in aerobic ammonia oxidation. *Environmental Microbiology*. Elsevier Academic Press, San Diegopp. 395–413.

Whitaker, R.J., Grogan, D.W., and Taylor, J.W. (2003). Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* **301**, 976–978.

Whittaker, R.H. (1969). New concepts of kingdoms of organisms. *Science* **163**, 150–159.

Wilhelm, B.T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C.J., Rogers, J., and Bahler, J. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* **453**, 1239–1243.

Woese, C.R. (1983). The primary lines of descent and the universal ancestor. In: *Evolution from Molecules to Men* (Ed. by D.S. Bendall), pp. 209–233. Cambridge University Press, Cambridge.

Woese, C.R., and Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci. USA* **74**, 5088–5090.

Wolf, J.B.W., Bayer, T., Haubold, B., Schilhabel, M., Rosenstiel, P., and Tautz, D. (2010). Nucleotide divergence vs. gene expression differentiation: Comparative transcriptome sequencing in natural isolates from the carrion crow and its hybrid zone with the hooded crow. *Mol. Ecol.* **19**, 162–175.

Woodward, G. (2009). Biodiversity, ecosystem functioning and food webs in fresh waters: Assembling the jigsaw puzzle. *Freshw. Biol.* **54**, 2171–2187.

Woodward, G., Benstead, J.P., Beveridge, O.S., Blanchard, J., Brey, T., Brown, L., Cross, W.F., Friberg, N., Ings, T.C., Jacob, U., Jennings, S., Ledger, M.E., *et al.* (2010). Ecological networks in a changing climate. *Adv. Ecol. Res.* **42**, 71–138.

Woodward, G., Perkins, D.M., and Brown, L.E. (2010). Climate change and freshwater ecosystems: Impacts across multiple levels of organization. *Philos. Trans. R. Soc. B-Biol. Sci.* **365**, 2093–2106.

Woodward, G., Blanchard, J., Lauridsen, R.B., Edwards, F.K., Jones, J.I., Figueroa, D., Warren, P.H., and Petchey, O.L. (2010). Individual-based food webs: Species identity, body size and sampling effects. *Adv. Ecol. Res.* **43**, 211–266.

Woyke, T., Teeling, H., Ivanova, N.N., Huntemann, M., Richter, M., Gloeckner, F.O., Boffelli, D., Anderson, I.J., Barry, K.W., Shapiro, H.J., Szeto, E., Kyrpides, N.C., *et al.* (2006). Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443**, 950–955.

Yoder-Himes, D.R., Chain, P.S.G., Zhu, Y., Wurtzel, O., Rubin, E.M., Tiedje, J.M., and Sorek, R. (2009). Mapping the *Burkholderia cenocepacia* niche response via high-throughput sequencing. *Proc. Natl. Acad. Sci. USA* **106**, 3976–3981.

Yoshida, T., Jones, L.E., Ellner, S.P., Fussmann, G.F., and Hairston, N.G. (2003). Rapid evolution drives ecological dynamics in a predator-prey system. *Nature* **424**, 303–306.

Yvon-Durocher, G., Montoya, J.M., Trimmer, M., and Woodward, G. (2010). Warming alters the size spectrum and shifts the distribution of biomass in freshwater ecosystems. *Glob. Chang. Biol.* (in press).

Yvon-Durocher, G., Reiss, J., Blanchard, J., Ebenman, B., Perkins, D.M., Reuman, D.C., Thierry, A., Woodward, G., and Petchey, O.L. (2010). Across ecosystem comparisons of size structure: Methods, approaches, and prospects. *Oikos* (in press).

Yvon-Durocher, G., Allen, A.P., Montoya, J.M., Trimmer, M., and Woodward, G. (2010). The temperature dependence of the carbon cycle in aquatic systems. *Adv. Ecol. Res.* **43**, 267–313.

Zak, D.R., Blackwood, C.B., and Waldrop, M.P. (2006). A molecular dawn for biogeochemistry. *Trends Ecol. Evol.* **21**, 288–295.

Zaura, E., Keijser, B.J.F., Huse, S.M., and Crielaard, W. (2009). Defining the healthy "core microbiome" of oral microbial communities. *BMC Microbiol.* **9**, e259.

Zuckerkandl, E., and Pauling, L. (1965). Molecules as documents of evolutionary history. *J. Theor. Biol.* **8**, 357–366.