

## **Presentation abstracts**

Jim Prosser, The University of Aberdeen

### **Which scientific questions can be addressed using soil metagenomics?**

'Traditional' molecular ecology techniques have now been used for 20 years to characterise natural microbial communities. They have demonstrated, and continue to demonstrate high diversity in these communities and the existence of previously unsuspected, high-level, taxonomic groups.

Early descriptive studies were soon complemented by those aimed at identification of links between community composition, environmental characteristics and environmental factors and attempts to determine mechanisms driving microbial community composition. Attempts were also made to link community composition to ecosystem function, and these increased with the development of qPCR to estimate abundance of both 16S rRNA and functional genes and techniques for estimation of activity of specific groups, e.g. stable isotope probing and analysis of transcriptional activity. Again, these new techniques provided useful information, but studies often followed techniques rather than determining critically which techniques were required for specific ecological questions. More positively, they activated interest in broader ecological questions and their relevance to microbial diversity, activity and ecosystem function, and awakened long-standing debates, such as those surrounding bacterial and archaeal species concepts.

High-throughput sequencing techniques are now advancing and, increasingly, replacing first-generation molecular techniques and should significantly improve our ability to address these questions. Metagenomics, metatranscriptomics and, inevitably and inexorably, metaproteomics and metabolomics, introduce something different and something new. They provide an holistic view of, e.g. gene or transcript content, independent of environmental structure or even cellular structure.

While current emphasis is placed on technical and bioinformatics challenges, the greatest challenges are arguably conceptual, in determining which, if any, fundamental ecological questions can be addressed using metagenomics and other omics approaches; whether new concepts are required to direct meaningful experimental studies; and the extent to which knowledge of microbial ecology gained during the past 130 years can inform such conceptual challenges.

Thomas Freitag, James Hutton Institute, Aberdeen

### **Soils perspective**

Microbes are the main agents of biogeochemical cycles in soil and affect their development and fertility. Nucleic acid second generation sequencing technologies of the collective microbial genomes (metagenomes) now allow analysing the diversity and function of these communities, independent of cultivation. Directed towards specific taxonomic marker genes (e.g. small

subunit ribosomal RNA; rRNA), metagenomic sequencing is frequently used to demonstrate and compare microbial community composition in different soils or to investigate effects of experimental treatment or environmental change. Ultimately, metagenomic shotgun sequencing approaches aim at analysing the total genomic content and at predictions of physiological functions. The advent of second generation sequencing technologies in the last few years has revealed a staggering microbial diversity and functional complexity. Advantages and limitations resulting from these technologies will be discussed.

Andy Taylor, James Hutton Institute, Aberdeen

### **Fungal perspective**

The cryptic nature of fungi growing within opaque substrates and the general uniformity of their vegetative forms has greatly hindered our ability to detect, characterise and monitor fungal communities. With few exceptions, all of our current knowledge of species distributions, and consequently any conservation strategy, is based upon observations made on macroscopic fruiting structures produced by fungi. This situation is rapidly changing with the application of metagenomic approaches to determining the phylogenetic and functional diversity of fungi in a wide range of ecosystems. In this talk I will highlight two examples where whole community approaches have been applied to the evaluation of fungal diversity. The first study focussed on the tooth fungi, which are covered by a group Biodiversity Action Plan. Many of these fungi are considered threatened or near threatened and often regarded as indicative of old growth Scots pine forests, to which many are thought to be restricted. Our recent work has demonstrated that they may in fact much more frequent than previously thought and their conservation status is likely to require a re-evaluation. The other study examined the mutualistic fungi associated with four declining montane plant hosts in the Scottish arctic-alpine zone. The fungi form obligate mycorrhizal associations with the roots of the host plants and are essential for the growth and survival of the plants. High species richness was encountered in these extreme environments with 257 species detected, of which 34 were new to the UK. Remarkably, an additional 23 species are considered to be new to science. This figure is even more surprising since it refers only to species found as fruiting structures - including the below ground sequence data increases this number to ca. 101. In addition, the data are derived from only 23 arctic-alpine sites - i.e. a small fraction of potential sites. The continuing decline of the host plants in montane habitats means that these fungi are in danger of being lost even before they are known. The application of community level approaches to assessing fungal communities is still very much work in progress but there is clearly huge potential. The main current constraints are with handling the extreme small-scale spatial heterogeneity in taxon distribution and in the severe lack of reference sequence data in most fungal groups.

Harry J Flint, Rowett Institute of Nutrition and Health, University of Aberdeen

### **Applications of metagenomics to gut microbial communities**

Complex anaerobic microbial communities are responsible for the breakdown of otherwise non-digestible plant material in the gut of farm animals and man. In herbivores, especially ruminants, these communities provide most of the energy obtained from the diet, while in humans the large intestinal community plays an important role in mediating the effects of diet upon health. The extreme diversity and oxygen sensitivity of micro-organisms within these communities make them obvious targets for metagenomic approaches. Major sequence-based projects on human gut metagenomes, funded by the EU and in the USA, have produced new information on the relationship between microbiota diversity, metabolic health and gut health. Meanwhile functional screening approaches, although more demanding of time, have the potential to reveal new genes and activities, while metagenome data can also be linked with targeted analyses and analyses of gene products (eg. metaproteomics). The decreasing costs and increasing speed of nucleic acid sequencing offer great opportunities for the analysis of gut samples, but it should be noted that interpretation and annotation of sequence data continues to rely critically on knowledge obtained from cultural microbiology and biochemistry.

Ian Toth, James Hutton Institute, Dundee

### **Environmental and Agricultural Metagenomics: Crops Perspective**

Around 40% of crops worldwide are lost to pests and diseases and, even using modern farming techniques that include the use of pesticides and herbicides, this figure can reach 15-20% in Europe. However, over the next 5 years the use of pesticides and herbicides is set to reduce under new legislation. New and improved methods of crop protection are thus needed to maintain current food supplies and increase them into the future. Microbials, the use of microorganisms in crop production, is one area that is being actively pursued as a replacement for chemicals (including fertilizers to assist plant growth). We know that some microbes act together in what are called suppressive soils to reduce disease incidence of the growing crop. Amplicon sequencing of specific sections of nucleic acid within a microbe's genome (e.g. ribosomal RNA) can provide information about the different microbes present in a soil; groups that can then, in some cases, be isolated and tested for their suppressive abilities. This method can also be used to identify different groups of microbes on a range of plant parts, allowing us to investigate a variety of factors, including pathogen spread. While amplicon sequencing has allowed us to understand the makeup and distribution of microbial communities in the plant environment, there is the potential for much more information to become available through the use of metagenomics. Metagenomics examines nucleic acids from whole genomes of microbes within a community, and allows us to study changes in gene function in addition to microbial community composition. As well as crop protection and plant growth enhancement, such an approach has many other potential applications on plants. For example, it is being used to investigate changes in microbial communities, together with gene function related to metabolic processes, involved in the fermentation of rice and grapes. Investigating microbial populations and their cognate gene functions provides new

possibilities for studying microbial communities on plants, and is set to allow rapid progress towards improved, chemical free, crop production.

Mick Watson (1) and John Wallace (2) 1. ARK-Genomics, The Roslin Institute, University of Edinburgh. 2. Rowett Institute of Nutrition and Health, University of Aberdeen.

### **Deep metagenomic sequencing of multiple ruminant guts reveals species-specific microbiomes**

Ruminant species such as cattle, sheep, deer, goats and reindeer represent a fascinating opportunity to study gut function and metagenomics. With a solely vegetarian diet, the gut microbiome of ruminants is perfectly adapted to digest and gain energy from plant material. However, different animals have different diets (e.g. grass vs lichen) which require different microbes to digest. We have deep-sequenced the rumen gut microbiome of several ruminant species using the Illumina HiSeq 2000 system, generation several hundred gigabases of sequence data in the process. Here we describe bioinformatic investigations into these data, including assembly into large contigs, prediction and annotation of genes and taxon assignment. We show that the vast majority of sequence data comes from as-yet uncharacterized bacterial species, representing huge potential for discovery of novel enzymes. Finally, we show that ruminants can be clustered together according to their gut microbiome structure and species abundance. Bio would be: Mick Watson is an established bioinformatician, with 15 years' experience in industry and academia. He was involved in the implementation and management of pipelines for functional genomics at GlaxoWellcome, SNP discovery at Incyte Genomics and Target Discovery at Paradigm Therapeutics, before joining the Institute for Animal Health as Head of Bioinformatics in 2002. He has been Director of ARK-Genomics since July 2010, and has won several grants to create a bioinformatics research group allied to the ARK-Genomics facility. His group's research focuses on the use of computational and mathematical techniques to understand genome function with an emphasis on systems of relevance to animal health and food security. Publications include both primary bioinformatics research papers and collaborative research in a variety of technical and scientific journals. The outputs from his research have included novel algorithm development, as well as the application of bioinformatics techniques to microbial and meta-genomics. Amongst other projects, he is Roslin PI on a Technology Strategy Board project to use NGS to discover biomass-degrading genes from rumen gut microbiota. To date, the project has generated 300Gb of sequence data, the assembly of which has produced over 3.6 million novel proteins.

### **Six Questions to be addressed during the Table Discussion period**

1. What practical applications will emerge in the short term to the benefit of Scottish agriculture?
2. How can we move towards harmonisation across biological kingdoms?
3. How can we harness the lead Scotland has in high performance computing?

4. How can metagenomics help in understanding climate change and its mitigation?
5. Can we quantify unknown biodiversity using metagenomics?
6. Have we done enough of the old stuff (bacteriology/virology culture)?