JOHNS HOPKINS
UNIVERSITY

Zanvyl Krieger School of Arts and Sciences
Advanced Academic Programs

410.734.81 and 410.734.82
Practical Introduction to Metagenomics

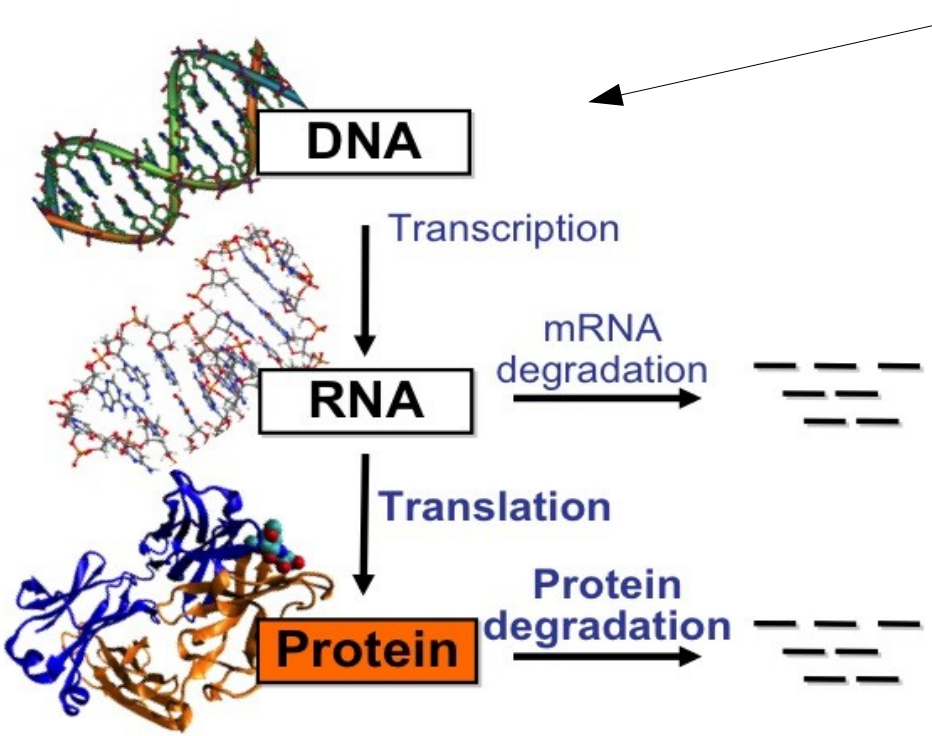
Topic: Metatranscriptomics and metaproteomics

Instructor: Joshua Orvis

Introduction

Our material to this point has been about metagenomics as the focus of diversity and functional studies in different habitats. This is the most global view available, but it doesn't (and can't) answer many questions about the complete activity of an environment. Why?

Central dogma of molecular biology



Of these three molecular tiers, we've really only covered the direct study of genomic sequence. Why does this distinction matter?

Studies of the genomic sequence are great for taxonomic studies, but can really only provide answers about the functional *potential* of a community. This is critical, of course, but it doesn't address what functions that community is actually expressing at any given time.

For that, we can study the expressed RNAs (**metatranscriptomics**) and/or their translated products (**metaproteomics**)

Intro to metatranscriptomics

From your molecular biology courses you know that cells employ innumerous methods for regulation of gene expression.

From histone modifications to transcription factor binding, there is a host of things which distinguish the genes which are present in the genome from those which are being expressed at any given time. By extracting and sequencing the mRNAs we get a more direct view of the active genes in a sample. This type of data also has the potential to reveal much more than gene to transcript relationships.

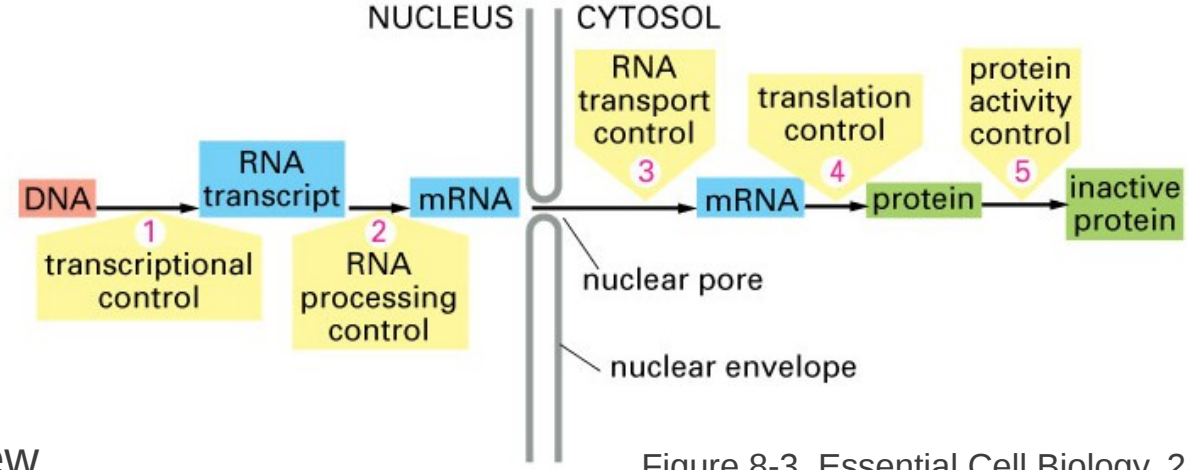


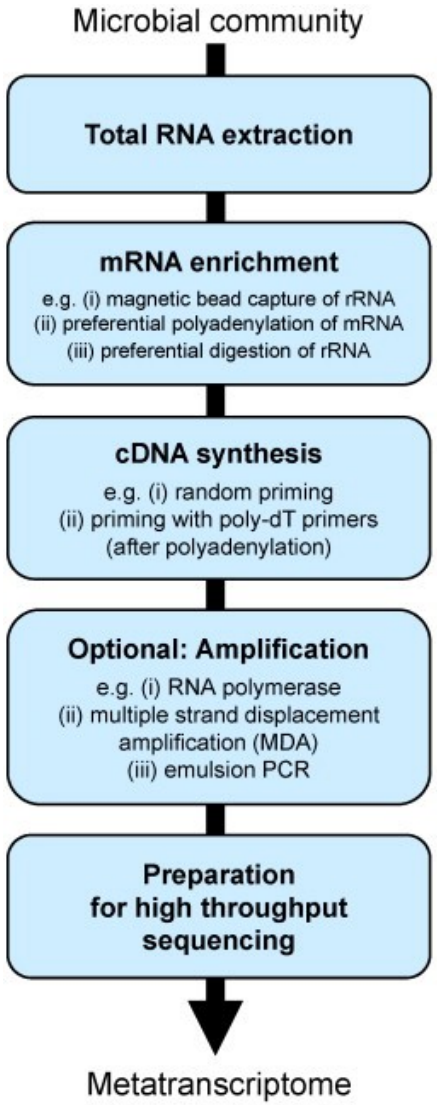
Figure 8-3 Essential Cell Biology, 2/e (2004, Garland Science)

Normalization of detected transcripts to their corresponding gene abundance suggests that numerically less abundant microorganisms may nevertheless contribute actively to ecologically relevant processes. [1]

Along the same lines, it is a recurrent observation that many transcripts are of unknown function or phylogenetic origin, and have not been detected in genomic/metagenomic data sets. These novel sequences may be derived from less abundant species or variable genomic regions that are not represented in sequenced genomes. [1]

[1] http://hdl.handle.net/1721.1/64570

Intro to metatranscriptomics



Initially limited to microarray experiments, metatranscriptomics have shifted to the use of shotgun sequencing, but there are still issues with sequencing and analysis of both mRNAs and rRNAs from transcript sampling. From your reading this week:

The expression of prokaryotic genes remains difficult to study mainly due to problems related to the isolation of mRNA [40,41,42,43,44]. The half-life of mRNA is short and it is usually a small fraction of the total RNA. In addition, mRNA enrichment is challenging in prokaryotes, as prokaryotic mRNA lacks the 39-end poly (A) tail that marks mature mRNA in eukaryotes.

Furthermore, it is important to take into account that metabolically active bacteria contain more ribosomal RNA than latent or starved cells [45].

Because of this fact analyzing the ribosomal RNA transcripts of an ecosystem identifies the active members of the microbiota and provides a general picture of their differential activity levels.

[see the paper on the next slide for the reference identifiers]

http://dx.doi.org/10.1016/j.jbiotec.2009.03.022

Metatranscriptomics of the GI tract

Metatranscriptomic Approach to Analyze the Functional Human Gut Microbiota

Gosalbes MJ, Moya A, et al.
PLoS One: 2011 Mar 8;6(3):e17447.

Published rather recently but before the availability of the Human Microbiome Data, this study of 10 healthy subjects was the largest of the GI tract to date. From the paper:

"Here, we report the metatranscriptomic study of the human GI tract microbiota in ten healthy individuals to elucidate a functional profile. We applied large scale pyrosequencing of the RNA community and used 16S rRNA transcripts as a marker of the phylogenetic structure of the active bacterial community. We also analyzed the protein-coding fraction to characterize the functions present in this habitat and the microorganisms involved in them. Additionally, this RNA-based approach allowed us to find, for the first time, untranslated regulatory elements in the gut microbial community."

Metatranscriptomic Approach to Analyze the Functional Human Gut Microbiota

María José Gosalbes[1,2], Ana Durbán[1,2], Miguel Pignatelli[1,2¤], Juan José Abellan[1,2], Nuria Jiménez-Hernández[1,2], Ana Elena Pérez-Cobas[1], Amparo Latorre[1,2], Andrés Moya[1,2*]

1 Unidad Mixta de Investigación en Genómica y Salud-Centro Superior Investigación en Salud Pública (Generalitat Valenciana)/Instituto Cavanilles de Biodiversidad y Biología Evolutiva (Universitat de València), València, Spain, **2** CIBER en Epidemiología y Salud Pública, València, Spain

Abstract

The human gut is the natural habitat for a large and dynamic bacterial community that has a great relevance for health. Metagenomics is increasing our knowledge of gene content as well as of functional and genetic variability in this microbiome. However, little is known about the active bacteria and their function(s) in the gastrointestinal tract. We performed a metatranscriptomic study on ten healthy volunteers to elucidate the active members of the gut microbiome and their functionality under conditions of health. First, the microbial cDNAs obtained from each sample were sequenced using 454 technology. The analysis of 16S transcripts showed the phylogenetic structure of the active microbial community. *Lachnospiraceae, Ruminococcaceae, Bacteroidaceae, Prevotellaceae,* and *Rickenellaceae* were the predominant families detected in the active microbiota. The characterization of mRNAs revealed a uniform functional pattern in healthy individuals. The main functional roles of the gut microbiota were carbohydrate metabolism, energy production and synthesis of cellular components. In contrast, housekeeping activities such as amino acid and lipid metabolism were underrepresented in the metatranscriptome. Our results provide new insights into the functionality of the complex gut microbiota in healthy individuals. In this RNA-based survey, we also detected small RNAs, which are important regulatory elements in prokaryotic physiology and pathogenicity.

Citation: Gosalbes MJ, Durbán A, Pignatelli M, Abellan JJ, Jiménez-Hernández N, et al. (2011) Metatranscriptomic Approach to Analyze the Functional Human Gut Microbiota. PLoS ONE 6(3): e17447. doi:10.1371/journal.pone.0017447

Editor: Lluis Quintana-Murci, Institut Pasteur, France

Received September 21, 2010; **Accepted** February 3, 2011; **Published** March 8, 2011

Funding: This work has been funded by grant BFU2008-04501-E/BMC and SAF2009-13032-C02-01 from Ministerio de Ciencia e Innovación, Spain. AD and AEPC are recipients of a fellowship from the Instituto de Salud Carlos III, Spain. MJG is recipient of a post-doctoral contract from 'CIBER en Epidemiología y Salud Pública' (CIBEResp), Spain. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: andres.moya@uv.es

¤ Current address: EMBL-EBI, Cambridge, United Kingdom

Introduction

The gastro-intestinal (GI) tract is an essential metabolic organ that is populated with a huge number of microbes. The intestinal microbiota is important for human health because of nutrient processing, development of the immune system, colonization resistance and stimulation of a variety of other host activities [1,2,3,4].

Our knowledge about bacterial diversity in the human GI tract has increased concomitantly with the development of different molecular approaches such as fingerprinting techniques of 16S rDNA amplicons, sequencing of 16S rDNA clones, fluorescent in situ hybridization, DNA microarrays or, more recently, high-throughput sequencing [5,6,7,8,9,10,11,12,13,14,15,16,17]. All these studies have shown that the composition of the intestinal microbiota varies between individuals due to host genotype, age, health status and diet, though the predominant population is fairly stable under normal conditions. We also know that the predominant bacterial groups in the human GI tract are *Bacteroidetes, Firmicutes* and *Actinobacteria*, and that substantial variability exists in the particular bacterial lineages carried by an individual [9,10,12,14,16,17,18]. Since the GI microbiota is highly diverse and variable across individuals, it is difficult to establish the relationship between particular microorganisms and health status. The stability of the GI microbiome is a function, not only of its composition, but also of the gene expression of its members. It is therefore essential to explore the gene expression of the microorganisms in the GI tract.

Recently, metagenomics applied in a variety of microbial habitats, including the GI tract, have led to the discovery and characterization of new genes from uncultivated microorganisms, assembly of whole genomes from community DNA sequence data and comparison of microbial community composition between different environments [9,14,17,19,20,21,22,23,24,25,26,27,28]. Although metagenomic data provide extensive information about microbiota diversity, gene content and their potential functions, there is no indication on whether DNA comes from viable cells or whether the predicted genes are expressed at all and, if so, under what conditions and to what extent.

Environmental metatranscriptomics retrieves and sequences environmental mRNAs from a microbial ecosystem to assess what genes may be expressed in that community. To date, metatranscriptomic studies have been applied mainly to samples from water and soil environments [29,30,31,32,33,34,35]. In the GI ecosystem, the diversity of gut microbiota has been the subject of many metagenomics studies but only a few have focused on the active

Intro to metaproteomics

As shown on the previous slides there are still regulatory steps that prevent the transcripts from being translated.

The full set of expressed proteins in a habitat is known as the metaproteome, a term first coined by Rodriguez-Valera in 2004.

The figure on the right is from the 2nd of your assigned readings for this week, and describes a general process flow for metaproteomic research. The key portion is extraction of protein from the sample followed by chromatography techniques for separation and fractionation.

This is followed by mass spec analysis for peptide identification.

Now let's consider the reading, which is a more thorough review.

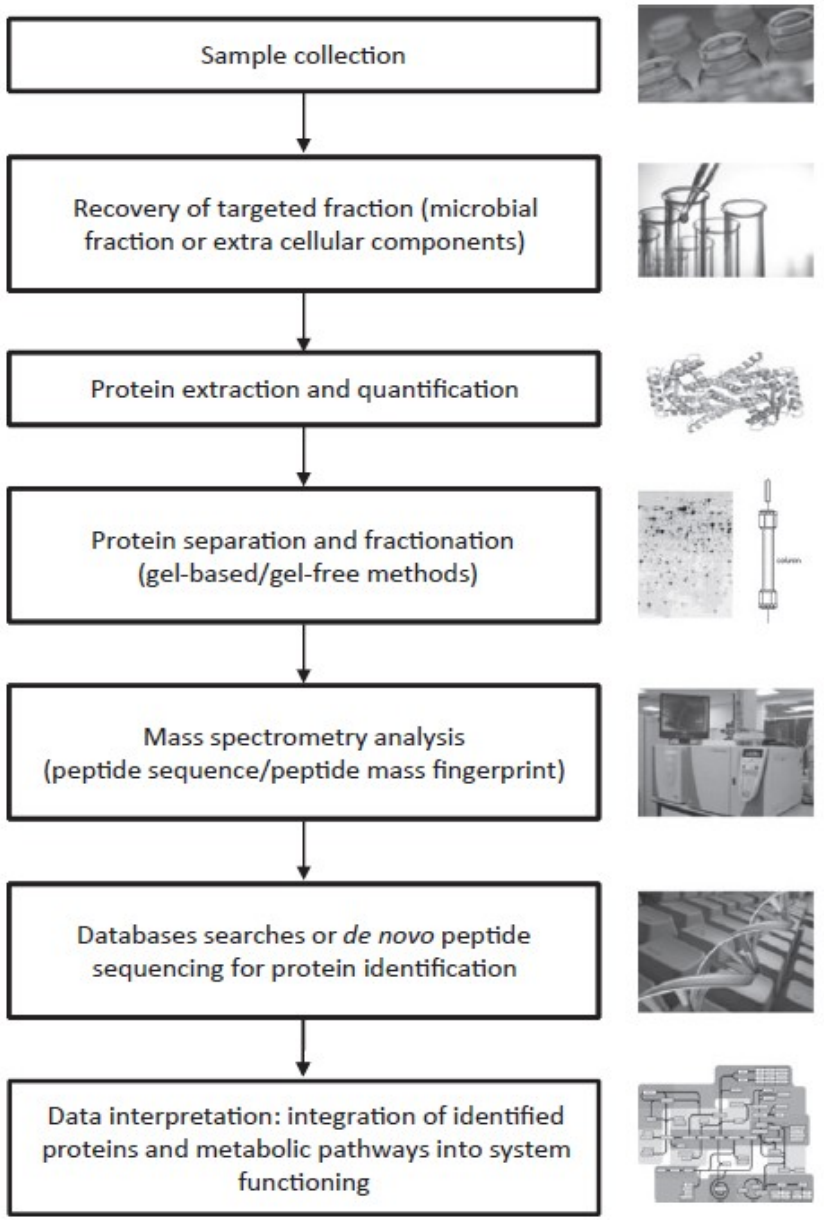


Fig. 2. Typical workflow for metaproteomics analysis.

Metaproteomics review

Exploring mixed microbial community functioning: recent advances in metaproteomics

Siggins, Gunningle, & Abram
FEMS Microbiol Ecol. 2012 May;80(2):265-80

This paper discusses the relationship, strengths and weaknesses of the different '-omics' approaches in general.

It then goes on to cover metaproteomics approaches specifically, and associated issues in a range of habitat types including humans, soil, marine and freshwaters, and even bioengineered systems.

“The feasibility of metaproteomic studies has been successfully demonstrated in very diverse natural and engineered environments.

Future metaproteomic studies should aim to progress from proof of concept approaches to experimental designs leading to practical applications.”

FEMS MICROBIOLOGY ECOLOGY

MINIREVIEW

Exploring mixed microbial community functioning: recent advances in metaproteomics

Alma Siggins[1], Eoin Gunnigle[1] & Florence Abram[2]

[1]Microbial Ecology Laboratory, Department of Microbiology and Ryan Institute, National University of Ireland, Galway (NUI, Galway), Galway, Ireland; and [2]Functional Environmental Microbiology, Department of Microbiology, National University of Ireland, Galway (NUI, Galway), Galway, Ireland

Correspondence: Florence Abram, Functional Environmental Microbiology, Department of Microbiology, National University of Ireland, Galway (NUI, Galway), University Road, Galway, Ireland. Tel.: +353 0 91 492390; fax: +353 0 91 494598; e-mail: florence.abram@nuigalway.ie

Received 16 May 2011; revised 7 October 2011; accepted 13 December 2011.
Final version published online 16 January 2012.

DOI: 10.1111/j.1574-6941.2011.01284.x

Editor: Ian Head

Keywords
environmental proteomics; human gut microbiota; marine and freshwater environment; soil; bioengineered systems.

Abstract

System approaches to elucidate ecosystem functioning constitute an emerging area of research within microbial ecology. Such approaches aim at investigating all levels of biological information (DNA, RNA, proteins and metabolites) to capture the functional interactions occurring in a given ecosystem and track down characteristics that could not be accessed by the study of isolated components. In this context, the study of the proteins collectively expressed by all the microorganisms present within an ecosystem (metaproteomics) is not only crucial but can also provide insights into microbial functionality. Overall, the success of metaproteomics is closely linked to metagenomics, and with the exponential increase in the availability of metagenome sequences, this field of research is starting to experience generation of an overwhelming amount of data, which requires systematic analysis. Metaproteomics has been employed in very diverse environments, and this review discusses the recent advances achieved in the context of human biology, soil, marine and freshwater environments as well as natural and bioengineered systems.

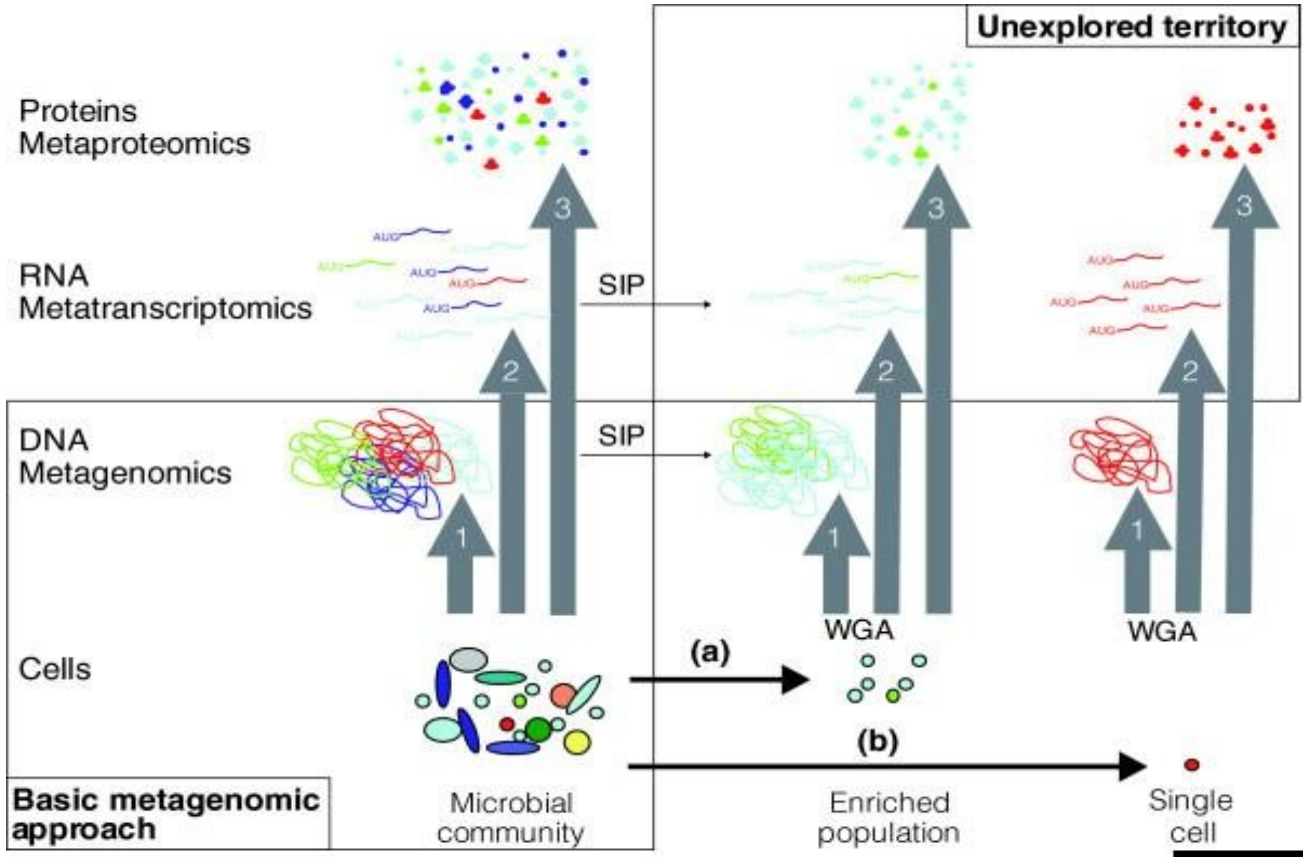
Introduction

Microorganisms occupy virtually every habitat on our planet, and their activities largely determine the environmental conditions of today's world. Indeed, microorganisms are heavily involved in biogeochemistry, ensuring the recycling of elements such as carbon and nitrogen (Madsen, 2011). In addition, microorganisms are extensively used to degrade anthropogenic waste prior to release into the environment (Hussain *et al.*, 2010; Park *et al.*, 2011). In their natural habitat, microorganisms coexist in mixed communities, the complexity of which is specific to each environment, for example from six estimated individual taxa for an acid mine drainage biofilm (Ram *et al.*, 2005), up to 10^6 estimated taxa per gram of soil (Wilmes & Bond, 2006). As most of the microorganisms present in the environment have not been cultured, their investigation requires the use of molecular techniques that bypass the traditional isolation and cultivation of individual species (Amann *et al.*, 1995). Moreover, even when isolation is possible, a single species removed from its natural environment might not necessarily display the same characteristics under laboratory conditions as it does within its ecological niche. Therefore, the study of mixed microbial communities within their natural environment is key to the investigation of the diverse roles played by microorganisms, and to the identification of the microbial potential for biotechnological application, including but not limited to: pharmaceutical, diagnostics, waste treatment, bioremediation and renewable energy generation. An emerging field of research in microbial ecology encompasses system approaches (Fig. 1), whereby all levels of biological information are investigated (DNA, RNA, proteins and metabolites) to capture the functional interactions occurring in a given ecosystem and identify characteristics that could not be accessed by the study of isolated components (Raes & Bork, 2008; Röling *et al.*, 2010). Recent technological advances, including the development of high-throughput 'omics' methods, make such system approaches possi-

Synthesis required

I really like this figure from Phil Hugenholtz illustrating the basic metagenomic approach through complementary technologies. Though created in 2007, it summarizes the different approaches here well.

The take-home idea is that no singular approach will be the definitive one. It is really in the correlation of the data from all these approaches and with the environmental metadata associated with them that the complex identity and functional activity of microbial systems will be better elucidated.



Warnecke and Hugenholtz Genome Biology 2007 8:231

Thank you

You have reached the end of the material for this course.

I want to take a minute to thank all you brave volunteers for the first offering of this course. It was a rewarding experience for me, and I can see from the caliber of many of your submissions (and private communications you've sent) that you took a lot from it too.

I've enjoyed the enthusiasm in which many of you tackled your projects and drove discussions in the forums. I've been quite impressed. Thank you for putting up with silly pictures within lectures, joining late-notice office hours, Q/A sessions with speakers who only agreed to join the day before, and my backlog of grading!

Please feel free to e-mail me any questions you have even after the course is over. I've kept in communication with a large number of students over the years, even hiring a few.

I asked you to begin the semester by creating a quote file on the server, and I'll end with a favorite of my own:

> "I don't know half of you half as well as I should like and I like less than half of you, half as well as you deserve." – Bilbo Baggins

Good luck in your careers and future studies, and may they always drive each other forward.

Further reading

Metagenomic analysis: Past and future trends
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3067235/

Building on basic metagenomics with complementary technologies
http://genomebiology.com/2007/8/12/231