Exploration of Microbial Cells: The Storehouse of Bio-wealth Through Metagenomics and Metatranscriptomics

2

Ravi Ranjan, Asha Rani, and Rajesh Kumar

Abstract

Microbes are omnipresent, most abundant, versatile, have been studied exhaustively, and used historically for the human welfare. This enormous diversity of microbes serves as tiny cellular factories and has been used to generate "bio" energy, gas, fuel, and polymers and in waste management. However, to date majority of these microbes remain unexplored and thus remain unexploited for bioprospecting. Recent advancements in molecular biology techniques, next-generation high-throughput sequencing, and bioinformatics have aided to circumvent this caveat by providing insights in the genomes and biological process of these microbes. These recent developments in scientific research have advantages but also present the researchers with economic and computational challenges, but none the less it has led to tremendous discoveries compared to decades ago. In this review, we focus on the technological advancements and the recent studies using metagenomics and metatranscriptomics for exploration of "microbial cell factories" – the storehouse of biological wealth.

2.1 Introduction

Microbes are ubiquitous and found in diverse habitats and in a colossal diversity. Soil, sediments, and aquatic (marine) habitats are probably the most complex natural niches with respect to the microbial community size and species diversity. The number of prokaryotes is estimated to be over $4–6 \times 10^{30}$, exceeding, by various orders of magnitude, all plant and animal diversity (Woese 1987; Amann et al. 1995; Whitman et al. 1998; Curtis et al. 2002; Ward 2002; Schloss and Handelsman 2004). These tiny microorganisms represent the richest repertoire of molecular and

RR and AR have contributed equally, and considered as co-first and co-contributing authors

R. Ranjan, Ph.D. (✉) • A. Rani, Ph.D. (✉)
Department of Medicine, University of Illinois, Chicago, IL 60612, USA
e-mail: ranjan.du@gmail.com; aasha511@gmail.com

R. Kumar
Department of Botany, Hindu College, University of Delhi, Delhi 110007, India
e-mail: rajesh_bot@yahoo.com

V.C. Kalia (ed.), *Microbial Factories*, DOI 10.1007/978-81-322-2598-0_2

chemical diversity and possess huge potential for bioprospecting (Madigan et al. 2010). These microbes have been used to generate bioenergy, biogas, biofuel (ethanol and diesel), biopolymers, biocatalysts, and antibiotics and in waste management (Kalia et al. 2003a, b; Ni et al. 2007; Kalia and Purohit 2008; Gulder and Moore 2009; Kumar et al. 2009, 2013, 2014a, b; Singh et al. 2009; Williams 2009; Weiland 2010; Kuhad et al. 2011; Patel et al. 2011, 2015; Soares et al. 2012; Vermelho et al. 2012; Patel and Kalia 2013; Singh et al. 2013; Kalia 2014; Koch et al. 2014; Latif et al. 2014; Ling et al. 2015). However, only a miniscule fraction (~1–5 %) can be readily cultured using the routine laboratory culture techniques – which is appropriately termed as the "great plate count anomaly" (Staley and Konopka 1985; Kellenberger 2001; Tanaka et al. 2014). This indicates that the majority (~95–99 %) of the microbes remain "unculturable" and thus unexploited for commercial applications and bioprospecting. With the advancement in culture-independent techniques (metagenomics and metatranscriptomics) and advent of high-throughput next-generation DNA sequencing, the limitation of culturing the microbes has been circumvented, and peeking in their genome is feasible than before (Handelsman 2004; Streit and Schmitz 2004; Sharma et al. 2005; Warnecke and Hess 2009; Metzker 2010).

Over the past decade, these advances have yielded a massive amount of genomic, transcriptomic, metagenomic, and metatranscriptomic data of microbes and microbial communities. Analyzing the functional potential has become a common but a challenging task, and this has led to the generation of bioinformatics and computational, statistical, and analytical software packages to make sense of the huge genomic data (Meyer et al. 2008; Gilbert and Hughes 2011; Mitra et al. 2011; Kuczynski et al. 2012a; Luo et al. 2013; Gifford et al. 2014; Ladoukakis et al. 2014; Nilakanta et al. 2014). These developments have widely transformed our capacity to investigate and exploit the "microbial cell factories" for biotechnological applications.

2.2 Sequencing the New Generation Way

Since the inception of chain termination-based DNA sequencing (Sanger et al. 1977), which was the major source of automated sequencing offered by Applied Biosystems (www.appliedbiosystems.com/), there has been tremendous advancements in the sequencing techniques – "next generation" and development of "third generation" by various companies using different technologies (Schadt et al. 2010). Improvements in next-generation sequencing (NGS) technologies in the last decade have significantly doubled the data output, reduced the cost per genome, and surpassed the analogous improvement in computer technologies as predicted by Moore's law (www.genome.gov/sequencingcosts/). In recent years, the Illumina Inc. (http://www.illumina.com/index.html) has emerged as one of the leading manufactures of the NGS instruments. It acquired the Solexa's Genome AnalyzerIIx, and its technology is based on "sequencing by synthesis" (SBS), which uses fluorescently labeled reversible-terminator dyes. It offers a wide variety of sequencers, ranging from its benchtop versions [MiSeq and NextSeq 500] for small research lab to commercial versions [HiSeq 2500, 3000, and 5000, xFive, and xTen]. The company launched a multimillion dollar product, HiSeqX Ten, which would provide large-scale whole-genome sequencing for $1,000 per genome. Roche Diagnostics Corporation (www.roche.com/index.htm) is also one of the vendors for NGS instruments. It acquired the 454 Life Sciences, which uses the pyrosequencing technology, in which pyrophosphates are generated, while the DNA polymerase adds nucleotides to the template DNA (King and Scott-Horton 2008). Roche offers the GS Junior Plus (benchtop version) and GS-FLX+ System. The Junior boasts up to 700-base read length, and the FLX+ boasts 1,000-base read length. Life Technologies Inc. (www.lifetechnologies.com/), which was acquired by Thermo Fisher Scientific, has two NGS platforms, SOLiD and Ion Torrent. The SOLiD was earlier owned by Applied Biosystems

and is based on the Sequencing by Oligonucleotide Ligation and Detection. Currently, the 5500xl SOLiD and 5500 SOLiD (benchtop version) are available. This advanced sequencers have enabled researchers to sequence more samples (genomic, metagenomic, and metatranscriptomic) in less time with high output. Many bacterial genomes and metagenomes have been sequenced using the NGS technologies, which are tabulated in Table 2.1. One has to take into consideration factors like amount of data output, read length and number of reads, number of samples multiplexed in one sequencing run, sequencing time, and cost. These factors are important as it determines the sequencing depth and coverage for the samples. The more number of samples multiplexed in one sequencing run, the more savings on cost per sample, but it results in low number of reads or data per sample (Sims et al. 2014).

Currently, there are few emerging companies with novel sequencing technologies offering the Third-Generation Sequencers (TGS). Oxford Nanopore Technologies Limited (https://nanoporetech.com/) is developing "strand sequencing" technology that passes intact DNA polymers through a protein nanopore, sequencing in real time as the DNA translocates through the pore (Branton et al. 2008; Metzker 2010; Guy et al. 2012; Benowitz 2014; Laszlo et al. 2014). It is expected to achieve long reads, low cost, and high speed with minimal sample preparation and instrumentation. One of the advantages of the nanopore technology is that it is designed to perform the direct analysis of RNA strands and direct, electronic analysis of proteins. Pacific Biosciences of California, Inc. (PacBio) (http://www.pacificbiosciences.com) is an emerging TGS venture, which projects a new method to sequence DNA cheaply and rapidly by watching an array of single DNA molecules being replicated in real time, called as "SMRT" – for Single-Molecule sequencing in Real Time (Eid et al. 2009). Currently it offers the PacBio RS II system, which is capable of extraordinarily long reads (depending upon starting library, half of the data are in reads more than 14,000 base pairs long with the longest reads over 40,000 base pairs), extremely high accuracy, and exquisite sensitivity. Helicos BioSciences Corporation platform was the first DNA-sequencing instrument to operate by imaging individual DNA molecules. Helicos technology is based on a "Virtual Terminator" technology. The reversible terminators are tethered inhibitors and are efficiently incorporated with high fidelity while preventing incorporation of additional nucleotides (Bowers et al. 2009). Complete Genomics (http://www.completegenomics.com) is offering a proprietary novel technology termed as "DNA nanoball or DNB" (Drmanac et al. 2010; Carnevali et al. 2011). This technology causes each long single molecule to consolidate, or ball up, into a small DNB particle. These DNBs are approximately 200 nm in diameter, and a library contains millions of DNBs that together represent the complete genome. Further, the Combinatorial Probe-Anchor Ligation (cPAL) chemistry attaches one of four possible fluorescent-labeled probes to the DNB anchor, depending on the sequence. This allows the fluorescence signals to be detected and read very efficiently by the sequencer. While none of the emerging TGS technologies have been thoroughly applied and tested with metagenomics and metatranscriptomics, nonetheless it offers promising alternatives, even further cost reduction, higher data output, and reduced time for sequencing.

2.3 Culture Dependent *vs* Culture Independent (Metagenomics)

Pure culture of microbes and the single-cell genomics are powerful tools for exploring for microbes for their biotechnological potential. In culture-dependent methods, bacteria are isolated from environmental samples by growing them in growth medium and under optimal physiological conditions. While many efforts and advances have been made in microbiological culture techniques, it is still challenging to culture a majority of bacterial species using the available laboratory culturing techniques (Staley and Konopka 1985; Stewart 2012; Tanaka et al. 2014). This limitation has severely impacted the limits for commercial

Table 2.1 List of bacterial genomes sequenced by NGS

Organism	Isolation source	Bioprospecting				NGS platform	[a]Genome size (Mbp)	References
		PHA	Hydrogen	Bioethanol	Other products			
Azotobacter vinelandii strain CA6	nr	nr	Yes	nr	Yes	454 GS-FLX Titanium Ion Torrent PGM	nr	Noar and Bruno-Barcena (2013)
Bacillus cereus strain A1	Activated sludge	nr	Yes	Yes	Yes	Illumina HiSeq 2000	5.6	Zhang et al. (2014)
Bacillus cereus strain tsu1	Agar-cellulose plate	Yes	nr	nr	Yes	Illumina HiSeq 2000	5.8	Li et al. (2014)
Brevundimonas naejangsanensis strain B1	Activated sludge	nr	Yes	nr	Yes	Illumina HiSeq 2000	2.9	Su et al. (2014)
Caloramator celer strain JW/YL-NZ35	Hot spring sediments	nr	Yes	Yes	Yes	Illumina HiSeq 2000	2.6	Ciranna et al. (2013)
Clostridium bifermentans strain WYM	Landfill leachate sludge	nr	Yes	Yes	Yes	Illumina MiSeq	3.4	Wong et al. (2014a)
Clostridium intestinale strain URNW	nr	nr	Yes	Yes	Yes	Roche/454 GS-FLX Illumina HiSeq 2000	4.6	Lal et al. (2013b)
Clostridium pasteurianum NRRL B-598	nr	nr	Yes	Yes	Yes	Roche GS Junior	6	Kolek et al. (2014)
Clostridium pasteurianum strain ATCC 6013	nr	nr	Yes	Yes	Yes	454 GS-FLX Illumina MiSeq	4.4	Pyne et al. (2014)
Clostridium perfringens strain JJC	Landfill leachate sludge	nr	Yes	nr	Yes	Illumina MiSeq	3.2	Wong et al. (2014b)
Clostridium sp. strain Ade.TY	Landfill leachate sludge	nr	Yes	nr	Yes	Illumina MiSeq	3.1	Wong et al. (2014c)
Clostridium termitidis strain CT1112	Termite gut (*Nasutitermes lujae*)	nr	Yes	Yes	Yes	Roche/454 GS-FLX	6.4	Lal et al. (2013a)
Erythrobacter litoralis strain DSM 8509	Cyanobacterial mat	Yes	nr	nr	Yes	Illumina mate pair	3.1	Wang et al. (2014)

Erythrobacter longus strain DSM 6997	Surface of green seaweed	Yes	nr	nr	Yes	Illumina mate pair	3.5	Wang et al. (2014)
Halanaerobium saccharolyticum strain DSM 6643T	nr	nr	Yes	Yes	Yes	Illumina Roche 454	2.8	Kivisto et al. (2013)
Haloarcula hispanica strain N601	Solar saltern	Yes	nr	nr	Yes	Roche GS-FLX Illumina GAIIx	3.9	Ding et al. (2014)
Halomonas hydrothermalis MTCC 5445	Seacoast	Yes	nr	nr	Yes	Roche 454 GS-FLX	nr	Bharadwaj Sv et al. (2015)
Pseudomonas mediterranea strain CFBP 5447^T	Tomato plants	Yes	nr	nr	Yes	Illumina GAIIx	6.3	Licciardello et al. (2014)
Rhizobium lupini HPC(L)	Saline desert soil	Yes	nr	nr	Yes	Ion torrent	5.2	Agarwal and Purohit (2013)
Streptomyces exfoliatus DSMZ 41693	Shotgun library	Yes	nr	nr	Yes	454 GS-FLX	8.8	Martinez et al. (2014)
Sulfurospirillum sp. strain MES	Microbial electrosynthesis system metagenome	nr	Yes	nr	Yes	Illumina MiSeq Pacific Biosciences	2.6	Ross et al. (2015)
Thermoanaerobacterium aotearoense SCUT27	Hot spring	nr	Yes	Yes	Yes	Illumina HiSeq 2000	2.8	Ai et al. (2014)

nr not reported, *PHA* polyhydroxyalkanoate

[a]Genome size Mbp: estimated genome size (megabase pairs)

utilization of these microbes. The 16S ribosomal RNA (rRNA) gene is routinely used to explore the microbial diversity and determine evolutionary and phylogenetic relationships between microbes (Woese and Fox 1977; Pace et al. 1986; Woese 1987). The present census of the microbial diversity, based on 16S rRNA genes of pure cultures and "unculturable" populations in environmental samples, includes at least 52 phylum-level bacterial and 20 phylum-level archaeal phylogenetic lineages; however, majority of the environmental microbes are not available as pure culture (Rappe and Giovannoni 2003; Schloss and Handelsman 2004). Interestingly, the human gastrointestinal tract and other body parts also harbor various microbes (Koren et al. 2011; Human Microbiome Project 2012; Schommer and Gallo 2013; Kliman 2014; Rajilic-Stojanovic and de Vos 2014). This recent surge of research in molecular microbial ecology based on 16S rRNA provides a convincing evidence for the existence of many diverse lineages of bacterial phyla encompassing novel unculturable bacteria (Rani 2008). However, the 16S rRNA gene information helps to access the microbial diversity, and it is limited to offer any genomic insights and functional potential of the microbe.

Metagenomics (also referred to as environmental and community genomics) is the culture-independent genomic analysis of assemblage of microorganisms of any environmental sample (Handelsman et al. 1998; Handelsman 2004). Metagenome analyses are initiated by the isolation of pure quality, high molecular weight environmental DNA, or metagenomic DNA from the environmental sample without culturing the microbes. The DNA isolation protocol should be such that maximum cells are lysed to maximize the yield with minimal shearing of DNA. Precaution should be taken to avoid the degradation of released nucleic acids because of the ubiquitous presence of environmental nucleases and also concomitant release of cellular nucleases during cell lysis. Simultaneously high molecular weight contaminants (humic acid and fulvic acid) also coprecipitate along with DNA, which hinder the downstream applications (Ranjan 2008). Many commercial DNA isolation kits are available from various manufacturers and have been compared to isolate DNA from different environmental samples (Dineen et al. 2010; Mahmoudi et al. 2011; Claassen et al. 2013; Vishnivetskaya et al. 2014). To date there is no accepted universal method for DNA isolation of different environmental samples; however, the major emphasis of the DNA isolation should be an unbiased genomic representation of all the microbial species. The metagenomic DNA can be used to amplify the 16S rRNA and sequenced to access the microbial diversity – called as amplicon sequencing metagenomics. The metagenomic DNA can be cloned in various cloning vectors [such as plasmids, cosmids, fosmids, bacterial artificial chromosome (BAC) vectors] and screened for functional genes, called as functional metagenomics, or can be sequenced using the next-generation sequencing, called as shotgun metagenomics (Fig. 2.1) (Ranjan et al. 2005; Ranjan 2008; Vester et al. 2015). Since the availability of the first two human pathogenic bacterial genome sequences (Fleischmann et al. 1995; Fraser et al. 1995) and the first two monumental shotgun metagenome sequencing of the Sargasso Sea and acid mine drainage, the paradigm in microbial genomics has been revolutionized (Tyson et al. 2004; Venter et al. 2004). In an effort to characterize the Earth microbial communities, The Earth Microbiome Project was formed. It includes collaborations among scientists worldwide and proposes to analyze 200,000 samples from various environmental communities using metagenomics, metatranscriptomics, and amplicon sequencing. This will help to generate a global Gene Atlas, environmental metabolic models for each biome, and approximately 500,000 reconstructed microbial genomes (www.earthmicrobiome.org/). To date there are more than 5000 bacterial genome sequences available and metagenomes sequenced from various environmental samples (www.ncbi.nlm.nih.gov/genome/browse/, http://metagenomics.anl.gov/).

The next-generation sequencing (NGS) technology has enabled us to perform massive parallel ultra-deep sequencing and transformed the landscape of microbial genomics through their ability to generate hundreds of megabases in a

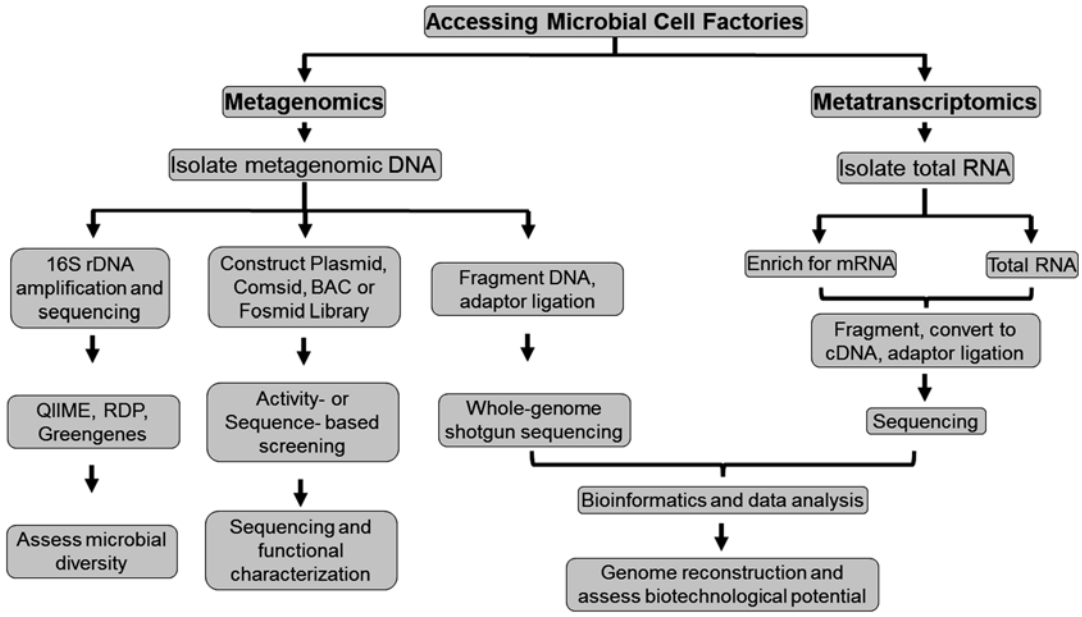

Fig. 2.1 Schematic overview of the metagenomic and metatranscriptomic approach to explore the "microbial cell factories"

single run. However, the use of these instruments requires the metagenomic DNA to be processed (called as library preparation) uniquely for different sequencers, as these are based on different sequencing technologies. The library preparation involves random fragmentation of the metagenomic DNA, using Covaris instrument [which is based on Adaptive Focused Acoustics (AFA) technology] or the transposome-mediated fragmentation technique or nebulized or enzymatically fragmented. The fragments are modified with the ligation of an adapter or barcode (sequence of nucleotides) and amplified. These double-stranded DNA are then denatured, and the nucleotide base composition on the single-stranded DNA molecules is detected by the sequencers. These procedures are dependent on the sequencing platform/instrument (Head et al. 2014). Additionally, the commercial DNA library prep kits (Illumina, Nextera XT DNA Sample Preparation Kit; Bioo Scientific Corporation, NEXTflex Rapid DNA-Seq Kit; and other similar kits) have eliminated the requirement of high amount of input DNA required for library preparation. Recently, Parkinson et al. described the process of preparation of DNA libraries from 20 pg of DNA from *Escherichia coli* (*E. coli*) using a modified transposome-mediated fragmentation technique (Parkinson et al. 2012).

2.4 Metatranscriptomics: Getting to the Functionality

Next-generation high-throughput sequencing technologies have enabled researchers to directly explore the microbial processes at transcriptional or the messenger RNA (mRNA) level, termed as "metatranscriptomics." The process basically involves three steps: (1) isolating the high-quality total RNA directly from the environmental sample and enriching for mRNA, (2) converting to cDNA, and (3) sequencing using NGS (Fig. 2.1). However, there are few options that a researcher can tailor based on their desired applications. The total RNA in a cell contains majority (~95–98 %) of ribosomal RNA (rRNA) and transfer RNA (tRNA), and the remaining (1–4 %) is mRNA. Since the mRNA has very short half-life and is prone to degradation, precaution should be taken at the time of collection and storage, as this can alter the microbial tran-

scriptional profile (Deutscher 2006). Commercial RNA preservation and storage reagents [RNAlater (Life Technologies, Inc.), LifeGuard Soil Preservation Solution (MO Bio Laboratories Inc), etc.] should be added to alleviate the degradation. Most of the RNA isolation from different environmental samples involves the bead beating or homogenization using glass or zirconium beads for efficient lysis of microbial cells. Other methods involving microwave-based rupture, liquid nitrogen grinding, and enzymatic lysis have been shown to be less efficient than those involving bead beating. However, till now, there has been no universal method for RNA extraction for different environmental samples (Wang et al. 2012; Carvalhais and Schenk 2013). Currently, there are few commercial RNA extraction kits like PowerSoil Total RNA Isolation Kit (MO Bio Laboratories, Inc.), FastRNA Pro Soil-Direct kit (MP Biomedicals), E.Z.N.A. Soil RNA kit (Omega Bio-Tek), etc. These kits boast of the high purity and yield. As there is only a tiny fraction of mRNA in the total RNA pool, it may be enriched by rRNA subtractive hybridization. The contaminating rRNA can be subtracted by using the MICROBExpress Bacterial mRNA Enrichment kit (Life Technologies Inc) and Ribo-Zero rRNA Removal Kit – Bacteria (Illumina, Inc.). Another alternative is an Exonuclease (5′-phosphate-dependent exonuclease) treatment of total RNA, which is based on the principle to degrade the bacterial rRNA as it possesses a 5′-monophosphate. In few instances, the total RNA can be subjected to subtractive hybridization followed by exonuclease treatment, which results in highly enriched mRNA and maximal removal of the rRNA (Zoetendal et al. 2006; Mettel et al. 2010; Carvalhais and Schenk 2013). Once the enriched mRNA is obtained or the total RNA can be converted to cDNA, it must be noted that in case if the mRNA or the RNA is low in concentration, it can be converted to cDNA and amplified to increase the yield using multiple displacement amplification [MDA] or multiple annealing and looping-based amplification cycles [MALBAC] (Gilbert et al. 2008; Zong et al. 2012; de Bourcy et al. 2014; Motley et al. 2014). Recently, researchers have reported the method to prepare whole-transcriptome cDNA libraries from a minute (500 pg) amount of total RNA (Tariq et al. 2011). Similar steps as DNA libraries can then be followed to prepare cDNA libraries based on the NGS platform.

2.5 Challenge: Generating to Analysis

The NGS data contains random microbial community DNA (metagenomics) and mRNA/RNA sequences (metatranscriptomics), and the question is to know which microbes are present, understand what they are doing, and how can we use them for biotechnological applications. Regardless of the NGS platform used for sequencing, there are three basic steps involved for analysis: (1) trimming of poor quality of sequences based on Phred scores or the ASCII characters, (2) assembly of reads and binning, and (3) gene prediction, annotation, and data interpretation. Several labs have established bioinformatic resources and computational and programming (Linux based, Perl, R) capabilities can configure their own modules and tools for data analysis, canonically termed as "pipeline." Many of the research labs have limited resources and are unable to develop their own data analysis tools. There are few online high-throughput pipelines (MG-RAST server, EBI Metagenomics server) that a researcher can use to upload raw sequence data and obtain the phylogenetic, taxonomic, and functional information (Meyer et al. 2008; Hunter et al. 2014). Commercial software such as CLC Genomics Workbench (Qiagen – CLC Genomics) is available as a complete package and has all the major software modules for performing the data analysis. The comparison of different pipelines and tools for metagenomic data analysis is reviewed by Wooley and Ye (2009), Carvalhais et al. (2012), Thomas et al. (2012), Luo et al. (2013), Ladoukakis et al. (2014), and Nilakanta et al. (2014) and tabulated in Table 2.2.

Table 2.2 List of software, online pipeline, bioinformatics, and statistical tools for NGS data analysis

Software/pipeline/online tool	Application	References/web-page link
AliView (ALignment Viewer)	Fast alignment viewer and editor for large datasets	Larsson (2014)
		www.github.com/AliView
BaCoCa (BAse COmposition Calculator)	Software tool to identify sequence biases in large datasets of gene and taxon	Kuck and Struck (2014)
		http://software.zfmk.de
BRENDA (BRaunschweig ENzyme Database)	Enzymes information for a metagenome data with the use of functional biochemical and molecular data	Schomburg et al. (2013)
		http://www.brenda-enzymes.org
CAMERA (The Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis)	Online computing service and high-performance computing infrastructure for metagenomic data analysis	Seshadri et al. (2007)
		http://camera.calit2.net
CARMA/WebCARMA	Pipeline for short, unassembled reads for species composition and genetic potential of microbial datasets	Krause et al. (2008) and Gerlach et al. (2009)
		http://webcarma.cebitec.uni-bielefeld.de
CLC Genomics workbench	Commercial pipeline, for almost all the NGS platforms (SOLiD, Ion Torrent, Illumina, and Roche 454). NGS data analysis for genomics, transcriptomics, and epigenomics	http://www.clcbio.com/products/clc-genomics-workbench/
CoMET (Comparative METagenomics)	Platform for comparative metagenomics of big datasets for metagenomic short reads	Lingner et al. (2011)
		http://comet.gobics.de
DAMBE5 (Data Analysis for Molecular Biology and Evolution)	Software package for descriptive and comparative sequence manipulation and analysis	Xia (2013)
		http://dambe.bio.uottawa.ca
DOTUR (Distance-Based OTU and Richness)	Operational taxonomic units and species richness estimation	Schloss and Handelsman (2005)
		www.plantpath.wisc.edu/fac/joh/dotur.html
EnGenIUS (Environmental Genome Informational Utility System)	Metagenome research toolset for large (>250,000) sequence reads	Kaplarevic et al. (2008)
EBI metagenomics	Raw sequence data upload for taxonomic and functional analysis	Hunter et al. (2014)
		http://www.ebi.ac.uk/metagenomics/

(continued)

Table 2.2 (continued)

Software/pipeline/online tool	Application	References/web-page link
EnvDB	Database for most complete and updated catalog of environmental distribution of prokaryotes based on 16S rDNA sequences currently stored in GenBank	Pignatelli et al. (2009) http://metagenomics.uv.es/envDB
GAAS (Genome relative Abundance and Average Size)	Software package for metagenomic sequences to estimate relative genome abundance and average genome length	Angly et al. (2009) http://sourceforge.net/projects/gaas/
Greengenes	Chimera detection, sequence alignment, and taxonomy identification using multiple published taxonomies	McDonald et al. (2012) http://greengenes.lbl.gov
IMG/M (the Integrated Microbial Genomes)	Comparative analysis of publicly available genomes, a community resource	Markowitz et al. (2012) http://img.jgi.doe.gov
KEGG (Kyoto Encyclopedia of Genes and Genomes)	Functional database for genomic, chemical, and systemic functional analysis	Kanehisa et al. (2014) http://www.genome.jp/kegg/kegg1.html
MEGAN (MEtaGenome Analyzer)	Comparative metagenomic and metatranscriptomic data analysis using NCBI taxonomy, and SEED, COG, and KEGG for functional classifications	Huson and Weber (2013) http://www-ab.informatik.uni-tuebingen.de
MetaBioME	Identification of novel homologues for known enzymes (of high commercial use) in completed bacterial genomes and metagenomic datasets	Sharma et al. (2010) http://metasystems.riken.jp/metabiome/
METAGENassist	Comparative metagenomic web server for amplified 16S rRNA and shotgun metagenomic data	Arndt et al. (2012) http://www.metagenassist.ca
MG-RAST (Metagenomics Rapid Annotations using Subsystems Technology)	Freely available and commonly used, automated pipeline for phylogenetic and functional analysis of metagenomes	Meyer et al. (2008) http://metagenomics.anl.gov/
MoCAT	Metagenomic sequence assembly and gene prediction pipeline for small and large metagenomic data produced by the Illumina sequencing	Kultima et al. (2012) http://www.bork.embl.de/mocat/
mothur	Comprehensive software to analyze microbial community sequence data	Schloss et al. (2009) http://www.mothur.org/

NCBI (the National Center for Biotechnology Information)	Community server for gene and function annotation of the metagenome and metatranscriptome data	http://www.ncbi.nlm.nih.gov/
PAST (PAleontological Statistics)	Statistical software for univariate and multivariate statistics	Hammer et al. (2001) http://folk.uio.no/ohammer/past/
PICRUSt (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States using 16S)	Functional composition prediction of metagenome data using gene data and reference genome database	Langille et al. (2013) http://picrust.github.com
QIIME (Quantitative Insights Into Microbial Ecology)	Comparative data analysis of microbial communities using 16S and metagenomic data	Kuczynski et al. (2012b) http://qiime.org
RDP (the Ribosomal Database Project)	Website for rRNA gene sequence data analysis	Cole et al. (2009) http://rdp.cme.msu.edu/
RITA (Rapid Identification of High-Confidence Taxonomic Assignments)	Accurate classification of very short sequences using different classification groups and varying degrees of confidence	MacDonald et al. (2012) http://kiwi.cs.dal.ca/Software/RITA
SEED	Genome annotations for multiple genomes for developing de novo annotations	Overbeek et al. (2005) http://pubseed.theseed.org/
STAMP (STatistical Analysis of Metagenomic Profiles)	Statistical software for statistical tests and analyzing taxonomic and functional profiles	Parks et al. (2014) http://kiwi.cs.dal.ca/Software/STAMP
WebMGA	Fast metagenomic data analysis web server for taxonomic and functional data analysis	Wu et al. (2011) http://weizhongli-lab.org/metagenomic-analysis/

2.6 Microbial Genomes, Metagenomics, and Metatranscriptomics

Many draft genomes, metagenomes, and metatranscriptome are now available because of these advancements and have enabled researchers to explore microbes at both the genomic and functional levels. *Pseudomonas mediterranea* strain CFBP 5447T produces cyclic lipopeptides and medium-chain-length polyhydroxyalkanoates (PHA) from various carbon sources and is known to convert biodiesel-derived glycerol to PHA. Cyclic lipopeptides such as cormycin A and corpeptins are produced by this strain, with antimicrobial and biosurfactant activities. The genome sequence of this species will enable applications at industrial level with cost-effective strategies for bioprospecting (Licciardello et al. 2014; Solaiman et al. 2005). *Halomonas hydrothermalis* MTCC 5445, a halophilic, Gammaproteobacteria, can grow in high salt (5 % NaCl) concentrations and produces polyhydroxybutyrate (PHB). *H. hydrothermalis* accumulates PHB intracellularly (75 % of dry weight) and can utilize waste glycerol from biodiesel to produce PHB. It can ferment glucose, sucrose, maltose, fructose, and ribose sugars. The whole-genome sequence of this species will aid in the efforts of high industrial production of PHAs (Bharadwaj Sv et al. 2015). Due to high commercial value of *Streptomyces*, multiple genomes of this group have been completely sequenced, and numerous are ongoing. Streptomycetes produce antibiotics and bioactive and biological compounds for industrial and biotechnological applications. *Streptomyces exfoliatus*, DSMZ 41693, contains genes for poly-3-hydroxyoctanoate depolymerase, used for the synthesis of (*R*)-3-hydroxyalkanoic acids. The genome sequence of this strain has revealed many genes for biosynthesis of different metabolites including polyketides and terpenes. *S. exfoliatus* can degrade poly-3-hydroxyalkanoates, poly-3-hydroxybutyrate, and poly-3-hydroxyoctanoate. This strain is a potential novel producer of (*R*)-3-hydroxyalkanoic acids and aids in degradation of bioplastics. The genes encoding for cellulases, amylases, xylanases, chitinases, proteases, lipases, and esterases have been detected in this strain. This strain can serve as a suitable candidate for bioremediation processes due to the presence of heavy-metal-resistance genes (Martinez et al. 2014). *Haloarcula hispanica* is a halophilic archaean, isolated from a solar saltern, and has been used for isolation of novel haloviruses. The genome sequence of strain N601 can improve our understanding of the physiology, genome organization, and virus-host interactions within different species of the group. Many haloarchaeal species (*Haloferax*, *Haloarcula*, *Natrialba*, and *Haloquadratum*) can synthesize short-chain-length PHAs (SCL-PHAs), a large family of biopolymers with desirable biodegradability, biocompatibility, and thermoplastic features (Ding et al. 2014). Two PHA-producing strains of genus *Erythrobacter* are sequenced recently, *E. longus* strain DSM 6997 and *E. litoralis* strain DSM 8509. They represent aerobic anoxygenic phototrophic bacteria (AAPB), and the draft genomes are available as type strain of *Erythrobacter* genus. These strains are characterized as Gram-negative rods and have the ability to store PHA as a carbon source. Both the strains are slightly halophilic and can play a significant role as potential industrial PHA producers (Wang et al. 2014). Genome sequence of *Bacillus cereus* strain tsu1, isolated on an agar-cellulose plate, contains genes for cellulose degradation and biosynthesis pathways of PHB. *B. cereus* tsu1 can degrade cellulose and can produce valuable biopolymers. Genome sequence of this strain will provide significant strategies for sustainable bioenergy developments and reduced reliance on petroleum-based plastics (Li et al. 2014). *B. cereus* strain A1 is capable of hydrogen production and utilizing starch and starch wastewater. Strain A1 is a facultative anaerobe isolated from anaerobic digestion reactor. This strain can hydrolyze starch and can ferment glucose into hydrogen, and end products are acetate and ethanol (Zhang et al. 2014). *Rhizobium lupini* strain HPC(L) was isolated from saline desert soil. This train HPC(L) belongs to Alphaproteobacteria class, Gram-negative soil-inhabiting organism, and can grow in minimal media supplemented

with $CaCO_3$ carbon source. Identification of the PHB synthesis gene cluster supports the carbon-limiting stress under desert conditions (Agarwal and Purohit 2013). *Brevundimonas naejangsanensis* strain B1 is a facultative anaerobic bacterium, can ferment sugars, and is capable of high-efficiency hydrogen production. The genome of this strain can provide insights into mechanisms of high-yield hydrogen production in this strain (Su et al. 2014). *Halanaerobium saccharolyticum* subsp. *saccharolyticum* strain DSM 6643T is a halophilic anaerobic fermentative bacterium capable of efficient hydrogen production (Kivisto et al. 2013). *Clostridium intestinale* strain URNW is a Gram-positive, mesophilic, anaerobic, spore-forming bacterium closely related to the butyrate-producing hydrogen producers, such as *C. intestinale*, *C. acetobutylicum*, *C. perfringens*, *C. butyricum*, and *C. beijerinckii*. This strain is a potential candidate for the production of hydrogen or ethanol (Lal et al. 2013b). *Caloramator celer* strain JW/YL-NZ35 is a strictly anaerobic bacterium capable of producing hydrogen and ethanol and converts sugars to H_2, CO_2, acetate, ethanol, and formate. This strain is able to produce hydrogen at high yields in a natural microbial community and in pure culture (Ciranna et al. 2013). *Clostridium pasteurianum* NRRL B-598 is a heterofermentative, rod-shaped bacterium with versatile sugar-fermenting and proteolytic abilities with hydrogen and ethanol production (Kolek et al. 2014). *Clostridium perfringens* is a Gram-positive strict anaerobe that ferments a vast range of carbohydrates and produces acetate, butyrate, lactate, ethanol, hydrogen, and carbon dioxide of vast industrial applications. The genome sequence will help to identify genes that inhibit and promote hydrogen production in this species (Wong et al. 2014b). *Sulfurospirillum* sp. strain MES was isolated from a metagenome of microbial electrosynthesis system (MES) producing acetate and hydrogen. The reported genome predicts the potential of denitrification in this species. Phylogenetically *Sulfurospirillum* sp. strain MES is closely related to the cultured *S. cavolei* strain Phe91 and uncultured/enrichment culture clones from wastewater-activated sludge and petroleum reservoirs (Ross et al. 2015). *Azotobacter vinelandii* mutant strain CA6 displays different characteristics compared to its originating wild-type parent strain, CA (such as altered molybdate uptake, slow growing, tungstate tolerance, and production of hydrogen gas). The complete genome sequences of these strains may provide a genetic basis for these distinct mutant phenotypes (Noar and Bruno-Barcena 2013). *Clostridium* sp. strain Ade.TY is a new biohydrogen-producing species, accompanied by the production of acetate, butyrate, lactate, formate, ethanol, and butanol of industrial applications. The genome sequence of this strain may provide insights for efficient biohydrogen production and gene interactions, if any, involved in the process (Wong et al. 2014c). *Clostridium bifermentans* strain WYM is an effective biohydrogen producer, and its genome annotation may provide insights into the metabolic pathways involved in efficient biohydrogen production. This strain ferments a wide range of carbohydrates, glucose, fructose, maltose, glycerol, and sorbitol, and produces acetate, lactate, ethanol, hydrogen, and carbon dioxide (Wong et al. 2014a). *Clostridium pasteurianum* strain ATCC 6013 is an anaerobic Gram-positive model organism for the study of nitrogen fixation and clostridial ferredoxins. This strain can ferment waste glycerol and produce biodiesel, bioethanol, and hydrogen gas. The genome sequence of this strain was generated using a modified (hybrid) next-generation sequencing method (Pyne et al. 2014). *Clostridium termitidis* strain CT1112 is a cellulolytic bacterium that can utilize sugars and cellulose and can produce hydrogen, carbon dioxide, acetate, formate, lactate, and ethanol (Lal et al. 2013a). *Thermoanaerobacterium aotearoense* SCUT27 is a thermophilic, strict anaerobe and can utilize xylan, dextran, glucose, cellobiose, xylose, mannose, galactose, and arabinose. This strain has been metabolically engineered as a biocatalyst for the ethanol, hydrogen, and L-lactic acid production (Ai et al. 2014). Recently, the NGS technology using SOLiD system was applied to characterize the biogas-producing microbial community and explore the functional and taxonomical complexity from a

composite microbial consortium developing in a biogas fermenter. They identified that both the microbiological diversity and the regulatory role of the hydrogen metabolism appear to be the driving forces optimizing biogas-producing microbial communities. They further suggested a biogas-producing consortium can be determined through the use of metagenomic approach, which can contribute to significant progress in the efficacy and economic improvement of biogas generation (Wirth et al. 2012). Similar studies were carried out to study methane-producing microbial community in solid-state biogas reactor using the Roche/454 GS-FLX Titanium pyrosequencing platform. They identified several novel microbes with varied functional capabilities, which affects the biogas reactor performance (Li et al. 2013). In another study, metagenomes of four parallel biogas reactors digesting fish waste and cow manure was studied, comparing the initial inoculum at day 0 with day 59. During the start phase it operated stably, and important Archaeal and Bacterial species degrading the protein-rich substrate were identified, and in particular microbes involved syntrophic in methane production seemed to be important for the operation of the biogas plant (Solli et al. 2014). Recently, 12 uncultured bacterial near complete genomes with relative abundance as low as 0.06 % were reconstructed from activated sludge metagenomic datasets, thus highlighting the power of metagenomics to discover novel microorganisms (Albertsen et al. 2013).

Metatranscriptomics has offered a new insight into how the microbial community (metagenome) responds to changes in environmental conditions at a functional level. One of the first studies using a metatranscriptome approach of a biogas-producing microbial community from a production-scale biogas plant identified Euryarchaeota and Firmicutes as dominant phyla (Zakrzewski et al. 2012). A metatranscriptomic study has identified that *Micromonospora* species dominates the expression of lignocellulolytic enzymes in the thermophilic community, and this genus is a promising source of lignocellulolytic enzymes for industrial-scale production (Simmons et al. 2014). Metatranscriptomic analysis of *Alviniconcha* (genus of deepwater sea snails) symbionts revealed key differences among symbiont types and in the expression of genes relating to energy metabolism, hydrogen, and sulfur oxidation (Sanders et al. 2013). Metabolic pathways and cellulose-degrading enzymes, endo-β-1,4-glucosidase and β-1,4-glucosidase, were identified by transcriptome analysis (Leonardo et al. 2011). Some studies have coupled metagenomics and metatranscriptomics to complement "microbial cell factories'" abundance to functions. Analyzing metatranscriptomes of microbial cell factories for biofuel production is a powerful technique for discovering potentially robust lignocellulolytic enzymes. Comparative metatranscriptomics have been previously performed on lignocellulose degradation with a focus on termite gut, soil microbiota, and lignocellulolytic enzymes (He et al. 2013). Using these approaches with metagenomic data, novel genes and genomes can be discovered for exploration of "microbial cell factories."

2.7 Conclusion

The culture-independent techniques (metagenomics and metatranscriptomics) coupled with the advanced high-throughput sequencing technologies and data analysis have revolutionized the field of microbiology. It has enabled researchers to explore the microbial genomes and biological process and bioprospecting the "microbial cell factories" for biotechnological applications and human welfare.

Acknowledgments The authors would like to sincerely thank Dr. V. C. Kalia, Chief Scientist, Institute of Genomics and Integrative Biology (CSIR-IGIB), Delhi, India, for helpful and constructive comments that greatly contributed to improving the final version of the chapter. The motivation behind this review stems from the Ph.D., research undertaken at IGIB and Delhi University by RR, AR, and RK.

References

Agarwal L, Purohit HJ (2013) Genome sequence of *Rhizobium lupini* HPC(L) isolated from saline desert soil, Kutch (Gujarat). Genome Announc 1:00071-12. doi:10.1128/genomeA.00071-12

Ai H, Zhang J, Yang M, Yu P, Li S, Zhu M, Dong H, Wang S, Wang J (2014) Draft genome sequence of an anaerobic, thermophilic bacterium, *Thermoanaerobacterium aotearoense* SCUT27, isolated from a hot spring in China. Genome Announc 2:00041-14. doi:10.1128/genomeA.00041-14

Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH (2013) Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat Biotechnol 31:533–538. doi:10.1038/nbt.2579

Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev 59:143–169

Angly FE, Willner D, Prieto-Davo A, Edwards RA, Schmieder R, Vega-Thurber R, Antonopoulos DA, Barott K, Cottrell MT, Desnues C et al (2009) The GAAS metagenomic tool and its estimations of viral and microbial average genome size in four major biomes. PLoS Comput Biol 5, e1000593. doi:10.1371/journal.pcbi.1000593

Arndt D, Xia J, Liu Y, Zhou Y, Guo AC, Cruz JA, Sinelnikov I, Budwill K, Nesbo CL, Wishart DS (2012) METAGENassist: a comprehensive web server for comparative metagenomics. Nucleic Acids Res 40:W88–W95. doi:10.1093/nar/gks497

Benowitz S (2014) Nanopore DNA sequencing: new approaches to an old challenge. National Human Genome Research Institute. Report number: 27555651. www.genome.gov/27555651

Bowers J, Mitchell J, Beer E, Buzby PR, Causey M, Efcavitch JW, Jarosz M, Krzymanska-Olejnik E, Kung L, Lipson D et al (2009) Virtual terminator nucleotides for next-generation DNA sequencing. Nat Methods 6:593–595. doi:10.1038/nmeth.1354

Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, Di Ventra M, Garaj S, Hibbs A, Huang X et al (2008) The potential and challenges of nanopore sequencing. Nat Biotechnol 26:1146–1153. doi:10.1038/nbt.1495

Bharadwaj Sv V, Shrivastav A, Dubey S, Ghosh T, Paliwal C, Maurya R, Mishra S (2015) Draft genome sequence of *Halomonas hydrothermalis* MTCC 5445, isolated from the west coast of India. Genome Announc 3:01419-14. doi:10.1128/genomeA.01419-14

Carnevali P, Baccash J, Halpern AL, Nazarenko I, Nilsen GB, Pant KP, Ebert JC, Brownley A, Morenzoni M, Karpinchyk V et al (2011) Computational techniques for human genome resequencing using mated gapped reads. J Comput Biol 19:279–292. doi:10.1089/cmb.2011.0201

Carvalhais LC, Schenk PM (2013) Sample processing and cDNA preparation for microbial metatranscriptomics in complex soil communities. Methods Enzymol 531:251–267. doi:10.1016/B978-0-12-407863-5.00013-7

Carvalhais LC, Dennis PG, Tyson GW, Schenk PM (2012) Application of metatranscriptomics to soil environments. J Microbiol Methods 91:246–251. doi:10.1016/j.mimet.2012.08.011

Ciranna A, Larjo A, Kivisto A, Santala V, Roos C, Karp M (2013) Draft genome sequence of the hydrogen- and ethanol-producing anaerobic alkalithermophilic bacterium *Caloramator celer*. Genome Announc 1:00471-13. doi:10.1128/genomeA.00471-13

Claassen S, du Toit E, Kaba M, Moodley C, Zar HJ, Nicol MP (2013) A comparison of the efficiency of five different commercial DNA extraction kits for extraction of DNA from faecal samples. J Microbiol Methods 94:103–110. doi:10.1016/j.mimet.2013.05.008

Cole J, Wang Q, Cardenas E, Fish J, Chai B, Farris R, Kulam-Syed-Mohideen A, McGarrell D, Marsh T, Garrity G et al (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res 37:D141–D145. doi:10.1093/nar/gkn879

Curtis TP, Sloan WT, Scannell JW (2002) Estimating prokaryotic diversity and its limits. Proc Natl Acad Sci U S A 99:10494–10499. doi:10.1073/pnas.142680199

de Bourcy CF, De Vlaminck I, Kanbar JN, Wang J, Gawad C, Quake SR (2014) A quantitative comparison of single-cell whole genome amplification methods. PLoS ONE 9, e105585. doi:10.1371/journal.pone.0105585

Deutscher MP (2006) Degradation of RNA in bacteria: comparison of mRNA and stable RNA. Nucleic Acids Res 34:659–666. doi:10.1093/nar/gkj472

Dineen SM, Aranda R, Anders DL, Robertson JM (2010) An evaluation of commercial DNA extraction kits for the isolation of bacterial spore DNA from soil. J Appl Microbiol 109:1886–1896. doi:10.1111/j.1365-2672.2010.04816.x

Ding JY, Chiang PW, Hong MJ, Dyall-Smith M, Tang SL (2014) Complete genome sequence of the extremely halophilic archaeon *Haloarcula hispanica* strain N601. Genome Announc 2:00178-14. doi:10.1128/genomeA.00178-14

Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, Carnevali P, Nazarenko I, Nilsen GB, Yeung G et al (2010) Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science 327:78–81. doi:10.1126/science.1181498

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B et al (2009) Real-time DNA sequencing from single polymerase molecules. Science 323:133–138. doi:10.1126/science.1162986

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF,

Dougherty BA, Merrick JM et al (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269:496–512. doi:10.1126/science.7542800

Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM et al (1995) The minimal gene complement of *Mycoplasma genitalium*. Science 270:397–403. doi:10.1126/science.270.5235.397

Gerlach W, Junemann S, Tille F, Goesmann A, Stoye J (2009) WebCARMA: a web application for the functional and taxonomic classification of unassembled metagenomic reads. BMC Bioinformatics 10:430. doi:10.1186/1471-2105-10-430

Gifford S, Satinsky B, Moran M (2014) Quantitative microbial metatranscriptomics. In: Paulsen IT, Holmes AJ (eds) Environmental microbiology. Humana Press, Totowa, pp 213–229

Gilbert JA, Hughes M (2011) Gene expression profiling: metatranscriptomics. Methods Mol Biol 733:195–205. doi:10.1007/978-1-61779-089-8_14

Gilbert J, Field D, Huang Y, Edwards R, Li W, Gilna P, Joint I (2008) Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. PLoS ONE 3, e3042. doi:10.1371/journal.pone.0003042

Gulder TA, Moore BS (2009) Chasing the treasures of the sea – bacterial marine natural products. Curr Opin Microbiol 12:252–260. doi:10.1016/j.mib.2009.05.002

Guy AT, Piggot TJ, Khalid S (2012) Single-stranded DNA within nanopores: conformational dynamics and implications for sequencing; a molecular dynamics simulation study. Biophys J 103:1028–1036. doi:10.1016/j.bpj.2012.08.012

Hammer O, Haper D, Ryan P (2001) PAST: paleontological statistics software package for education and data analysis. Palaeontol Electron 4:9

Handelsman J (2004) Metagenomics: application of genomics to uncultured microorganisms. Microbiol Mol Biol Rev 68:669–685. doi:10.1128/MMBR.68.4.669-685.2004

Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol 5:R245–R249. doi:10.1016/S1074-5521(98)90108-9

He S, Ivanova N, Kirton E, Allgaier M, Bergin C, Scheffrahn R, Kyrpides N, Warnecke F, Tringe S, Hugenholtz P (2013) Comparative metagenomic and metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites. PLoS ONE 8, e61126. doi:10.1371/journal.pone.0061126

Head SR, Komori HK, LaMere SA, Whisenant T, Van Nieuwerburgh F, Salomon DR, Ordoukhanian P (2014) Library construction for next-generation sequencing: overviews and challenges. Biotechniques 56:61–77. doi:10.2144/000114133

Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. Nature 486:207–214. doi:10.1038/nature11234

Hunter S, Corbett M, Denise H, Fraser M, Gonzalez-Beltran A, Hunter C, Jones P, Leinonen R, McAnulla C, Maguire E et al (2014) EBI metagenomics-a new resource for the analysis and archiving of metagenomic data. Nucleic Acids Res 42:D600–D606. doi:10.1093/nar/gkt961

Huson DH, Weber N (2013) Microbial community analysis using MEGAN. Methods Enzymol 531:465–485. doi:10.1016/B978-0-12-407863-5.00021-6

Kalia VC (2014) Microbes, antimicrobials and resistance: the battle goes on. Indian J Microbiol 54:1–2. doi:10.1007/s12088-013-0443-7

Kalia VC, Purohit HJ (2008) Microbial diversity and genomics in aid of bioenergy. J Ind Microbiol Biotechnol 35:403–419. doi:10.1007/s10295-007-0300-y

Kalia VC, Chauhan A, Bhattacharyya G, Rashmi (2003a) Genomic databases yield novel bioplastic producers. Nat Biotechnol 21:845–846. doi:10.1038/nbt0803-845

Kalia VC, Lal S, Ghai R, Mandal M, Chauhan A (2003b) Mining genomic databases to identify novel hydrogen producers. Trends Biotechnol 21:152–156. doi:10.1016/s0167-7799(03)00028-3

Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M (2014) Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res 42:D199–D205. doi:10.1093/nar/gkt1076

Kaplarevic M, Murray AE, Cary SC, Gao GR (2008) EnGenIUS-environmental genome informational utility system. J Bioinform Comput Biol 6:1193–1211. doi:10.1142/S0219720008003850

Kellenberger E (2001) Exploring the unknown. The silent revolution of microbiology. EMBO Rep 2:5–7. doi:10.1093/embo-reports/kve014

King C, Scott-Horton T (2008) Pyrosequencing: a simple method for accurate genotyping. J Vis Exp 11:630. doi:10.3791/630

Kivisto A, Larjo A, Ciranna A, Santala V, Roos C, Karp M (2013) Genome sequence of *Halanaerobium saccharolyticum* subsp. *saccharolyticum* strain DSM 6643T, a halophilic hydrogen-producing bacterium. Genome Announc 1:e00187-13. doi:10.1128/genomeA.00187-13

Kliman HJ (2014) Comment on 'the placenta harbors a unique microbiome'. Sci Transl Med 6:254le4. doi:10.1126/scitranslmed.3009864

Koch C, Muller S, Harms H, Harnisch F (2014) Microbiomes in bioenergy production: from analysis to management. Curr Opin Biotechnol 27:65–72. doi:10.1016/j.copbio.2013.11.006

Kolek J, Sedlar K, Provaznik I, Patakova P (2014) Draft genome sequence of *Clostridium pasteurianum* NRRL B-598, a potential butanol or hydrogen producer. Genome Announc 2:e00192-14. doi:10.1128/genomeA.00192-14

Koren O, Spor A, Felin J, Fak F, Stombaugh J, Tremaroli V, Behre CJ, Knight R, Fagerberg B, Ley RE et al (2011) Human oral, gut, and plaque microbiota in patients with atherosclerosis. Proc Natl Acad Sci U S A 108:4592–4598. doi:10.1073/pnas.1011383107

Krause L, Diaz NN, Goesmann A, Kelley S, Nattkemper TW, Rohwer F, Edwards RA, Stoye J (2008) Phylogenetic classification of short environmental DNA fragments. Nucleic Acids Res 36:2230–2239. doi:10.1093/nar/gkn038

Kuck P, Struck TH (2014) BaCoCa-a heuristic software tool for the parallel assessment of sequence biases in hundreds of gene and taxon partitions. Mol Phylogenet Evol 70:94–98. doi:10.1016/j.ympev.2013.09.011

Kuczynski J, Lauber CL, Walters WA, Parfrey LW, Clemente JC, Gevers D, Knight R (2012a) Experimental and analytical tools for studying the human microbiome. Nat Rev Genet 13:47–58. doi:10.1038/nrg3129

Kuczynski J, Stombaugh J, Walters WA, Gonzalez A, Caporaso JG, Knight R (2012b) Using QIIME to analyze 16S rRNA gene sequences from microbial communities. Curr Protoc Bioinformatics Chapter 1:Unit 1E.5. doi:10.1002/9780471729259.mc01e05s27

Kuhad RC, Gupta R, Singh A (2011) Microbial cellulases and their industrial applications. Enzyme Res 2011:280696. doi:10.4061/2011/280696

Kultima JR, Sunagawa S, Li J, Chen W, Chen H, Mende DR, Arumugam M, Pan Q, Liu B, Qin J et al (2012) MOCAT: a metagenomics assembly and gene prediction toolkit. PLoS ONE 7, e47656. doi:10.1371/journal.pone.0047656

Kumar T, Singh M, Purohit HJ, Kalia VC (2009) Potential of *Bacillus* sp. to produce polyhydroxybutyrate from biowaste. J Appl Microbiol 106:2017–2023. doi:10.1111/j.1365-2672.2009.04160.x

Kumar P, Patel SK, Lee JK, Kalia VC (2013) Extending the limits of *Bacillus* for novel biotechnological applications. Biotechnol Adv 31:1543–1561. doi:10.1016/j.biotechadv.2013.08.007

Kumar P, Pant DC, Mehariya S, Sharma R, Kansal A, Kalia VC (2014a) Ecobiotechnological strategy to enhance efficiency of bioconversion of wastes into hydrogen and methane. Indian J Microbiol 54:262–267. doi:10.1007/s12088-014-0467-7

Kumar P, Singh M, Mehariya S, Patel SK, Lee JK, Kalia VC (2014b) Ecobiotechnological approach for exploiting the abilities of *Bacillus* to produce co-polymer of polyhydroxyalkanoate. Indian J Microbiol 54:151–157. doi:10.1007/s12088-014-0457-9

Ladoukakis E, Kolisis FN, Chatziioannou AA (2014) Integrative workflows for metagenomic analysis. Front Cell Dev Biol 2:70. doi:10.3389/fcell.2014.00070

Lal S, Ramachandran U, Zhang X, Munir R, Sparling R, Levin DB (2013a) Draft genome sequence of the cellulolytic, mesophilic, anaerobic bacterium *Clostridium termitidis* strain CT1112 (DSM 5398). Genome Announc 1:e00281-13. doi:10.1128/genomeA.00281-13

Lal S, Ramachandran U, Zhang X, Sparling R, Levin DB (2013b) Draft genome sequence of the hydrogen- and ethanol-producing bacterium *Clostridium intestinale* strain URNW. Genome Announc 1:e00871-13. doi:10.1128/genomeA.00871-13

Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega Thurber RL, Knight R et al (2013) Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol 31:814–821. doi:10.1038/nbt.2676

Larsson A (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30:3276–3278. doi:10.1093/bioinformatics/btu531

Laszlo AH, Derrington IM, Ross BC, Brinkerhoff H, Adey A, Nova IC, Craig JM, Langford KW, Samson JM, Daza R et al (2014) Decoding long nanopore sequencing reads of natural DNA. Nat Biotechnol 32:829–833. doi:10.1038/nbt.2950

Latif H, Zeidan AA, Nielsen AT, Zengler K (2014) Trash to treasure: production of biofuels and commodity chemicals via syngas fermenting microorganisms. Curr Opin Biotechnol 27:79–87. doi:10.1016/j.copbio.2013.12.001

Leonardo FC, da Cunha AF, da Silva MJ, Carazzolle MF, Costa-Leonardo AM, Costa FF, Pereira GA (2011) Analysis of the workers head transcriptome of the asian subterranean termite, *Coptotermes gestroi*. Bull Entomol Res 101:383–391. doi:10.1017/s0007485310000556

Li A, Chu Y, Wang X, Ren L, Yu J, Liu X, Yan J, Zhang L, Wu S, Li S (2013) A pyrosequencing-based metagenomic study of methane-producing microbial community in solid-state biogas reactor. Biotechnol Biofuels 6:3. doi:10.1186/1754-6834-6-3

Li H, Zhou S, Johnson T, Vercruysse K, Ropelewski AJ, Thannhauser TW (2014) Draft genome sequence of new *Bacillus cereus* strain tsu1. Genome Announc 2:e01294-14. doi:10.1128/genomeA.01294-14

Licciardello G, Bella P, Devescovi G, Strano CP, Sarris PF, Catara AF, Venturi V, Catara V (2014) Draft genome sequence of *Pseudomonas mediterranea* strain CFBP 5447T, a producer of filmable medium-chain-length polyhydroxyalkanoates. Genome Announc 2:e01260-14. doi:10.1128/genomeA.01260-14

Ling LL, Schneider T, Peoples AJ, Spoering AL, Engels I, Conlon BP, Mueller A, Schaberle TF, Hughes DE, Epstein S et al (2015) A new antibiotic kills pathogens without detectable resistance. Nature 517:455–459. doi:10.1038/nature14098

Lingner T, Asshauer KP, Schreiber F, Meinicke P (2011) CoMet-a web server for comparative functional profiling of metagenomes. Nucleic Acids Res 39:W518–W523. doi:10.1093/nar/gkr388

Luo C, Rodriguez RL, Konstantinidis KT (2013) A user's guide to quantitative and comparative analysis of metagenomic datasets. Methods Enzymol 531:525–547. doi:10.1016/B978-0-12-407863-5.00023-X

MacDonald NJ, Parks DH, Beiko RG (2012) Rapid identification of high-confidence taxonomic assignments for metagenomic data. Nucleic Acids Res 40, e111. doi:10.1093/nar/gks335

Madigan MT, Clark DP, Stahl D, Martinko JM (2010) Brock biology of microorganisms, 13th edn. Benjamin Cummings, San Francisco

Mahmoudi N, Slater GF, Fulthorpe RR (2011) Comparison of commercial DNA extraction kits for isolation and purification of bacterial and eukaryotic DNA from PAH-contaminated soils. Can J Microbiol 57:623–628. doi:10.1139/w11-049

Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P et al (2012) IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Res 40:D115–D122. doi:10.1093/nar/gkr1044

Martinez V, Hormigo D, Del Cerro C, Gomez de Santos P, Garcia-Hidalgo J, Arroyo M, Prieto A, Garcia JL, de la Mata I (2014) Genome sequence of *Streptomyces exfoliatus* DSMZ 41693, a source of poly(3-hydroxyalkanoate)-degrading enzymes. Genome Announc 2:e01272-13. doi:10.1128/genomeA.01272-13

McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J 6:610–618. doi:10.1038/ismej.2011.139

Mettel C, Kim Y, Shrestha PM, Liesack W (2010) Extraction of mRNA from soil. Appl Environ Microbiol 76:5995–6000. doi:10.1128/AEM.03047-09

Metzker ML (2010) Sequencing technologies – the next generation. Nat Rev Genet 11:31–46. doi:10.1038/nrg2626

Meyer F, Paarmann D, D'Souza M, Olson R, Glass E, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A et al (2008) The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9:386. doi:10.1186/1471-2105-9-386

Mitra S, Rupek P, Richter DC, Urich T, Gilbert JA, Meyer F, Wilke A, Huson DH (2011) Functional analysis of metagenomes and metatranscriptomes using SEED and KEGG. BMC Bioinformatics 12(Suppl 1):S21. doi:10.1186/1471-2105-12-s1-s21

Motley ST, Picuri JM, Crowder CD, Minich JJ, Hofstadler SA, Eshoo MW (2014) Improved multiple displacement amplification (iMDA) and ultraclean reagents. BMC Genomics 15:443. doi:10.1186/1471-2164-15-443

Ni M, Leung DYC, Leung MKH (2007) A review on reforming bio-ethanol for hydrogen production. Int J Hydrogen Energy 32:3238–3247. doi:10.1016/j.ijhydene.2007.04.038

Nilakanta H, Drews KL, Firrell S, Foulkes MA, Jablonski KA (2014) A review of software for analyzing molecular sequences. BMC Res Notes 7:830. doi:10.1186/1756-0500-7-830

Noar JD, Bruno-Barcena JM (2013) Complete genome sequences of *Azotobacter vinelandii* wild-type strain CA and tungsten-tolerant mutant strain CA6. Genome Announc 1:e00313-13. doi:10.1128/genomeA.00313-13

Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crecy-Lagard V, Diaz N, Disz T, Edwards R et al (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res 33:5691–5702. doi:10.1093/nar/gki866

Parkinson NJ, Maslau S, Ferneyhough B, Zhang G, Gregory L, Buck D, Ragoussis J, Ponting CP, Fischer MD (2012) Preparation of high-quality next-generation sequencing libraries from picogram quantities of target DNA. Genome Res 22:125–133. doi:10.1101/gr.124016.111

Pace N, Stahl D, Lane D, Olsen G (1986) The analysis of natural microbial populations by ribosomal RNA sequences. In: Marshall KC (ed) Advances in microbial ecology. Springer US, New York, pp 1–55

Parks DH, Tyson GW, Hugenholtz P, Beiko RG (2014) STAMP: statistical analysis of taxonomic and functional profiles. Bioinformatics 30:3123–3124. doi:10.1093/bioinformatics/btu494

Patel SK, Kalia VC (2013) Integrative biological hydrogen production: an overview. Indian J Microbiol 53:3–10. doi:10.1007/s12088-012-0287-6

Patel SK, Singh M, Kalia VC (2011) Hydrogen and polyhydroxybutyrate producing abilities of *Bacillus* spp. from glucose in two stage system. Indian J Microbiol 51:418–423. doi:10.1007/s12088-011-0236-9

Patel SK, Kumar P, Singh M, Lee JK, Kalia VC (2015) Integrative approach to produce hydrogen and polyhydroxybutyrate from biowaste using defined bacterial cultures. Bioresour Technol 176:136–141. doi:10.1016/j.biortech.2014.11.029

Pignatelli M, Moya A, Tamames J (2009) EnvDB, a database for describing the environmental distribution of prokaryotic taxa. Environ Microbiol Rep 1:191–197. doi:10.1111/j.1758-2229.2009.00030.x

Pyne ME, Utturkar S, Brown SD, Moo-Young M, Chung DA, Chou CP (2014) Improved draft genome sequence of *Clostridium pasteurianum* strain ATCC 6013 (DSM 525) using a hybrid next-generation sequencing approach. Genome Announc 2:e00790-14. doi:10.1128/genomeA.00790-14

Rajilic-Stojanovic M, de Vos WM (2014) The first 1000 cultured species of the human gastrointestinal microbiota. FEMS Microbiol Rev 38:996–1047. doi:10.1111/1574-6976.12075

Rani A (2008) Studies for identification, distribution and diversity of microbes through 16S rDNA sequences. Ph.D. thesis, University of Delhi, Delhi, India

Ranjan R (2008) Functional metagenomics to identify novel genes for biocatalysts. Ph.D. thesis, University of Delhi, Delhi, India

Ranjan R, Grover A, Kapardar RK, Sharma R (2005) Isolation of novel lipolytic genes from uncultured bacteria of pond water. Biochem Biophys Res Commun 335:57–65. doi:10.1016/j.bbrc.2005.07.046

Rappe MS, Giovannoni SJ (2003) The uncultured microbial majority. Annu Rev Microbiol 57:369–394. doi:10.1146/annurev.micro.57.030502.090759

Ross DE, Marshall CW, May HD, Norman RS (2015) Draft genome sequence of *Sulfurospirillum* sp. strain MES, reconstructed from the metagenome of a micro-

bial electrosynthesis system. Genome Announc 3:e01336-14. doi:10.1128/genomeA.01336-14
Sanders JG, Beinart RA, Stewart FJ, Delong EF, Girguis PR (2013) Metatranscriptomics reveal differences in *in situ* energy and nitrogen metabolism among hydrothermal vent snail symbionts. ISME J 7:1556–1567. doi:10.1038/ismej.2013.45
Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74:5463–5467
Schadt EE, Turner S, Kasarskis A (2010) A window into third-generation sequencers. Hum Mol Genet 19:R227–R240. doi:10.1093/hmg/ddq416
Schloss PD, Handelsman J (2004) Status of the microbial census. Microbiol Mol Biol Rev 68:686–691. doi:10.1128/MMBR.68.4.686-691.2004
Schloss PD, Handelsman J (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl Environ Microbiol 71:1501–1506. doi:10.1128/AEM.71.3.1501-1506.2005
Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ et al (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541. doi:10.1128/aem.01541-09
Schomburg I, Chang A, Placzek S, Sohngen C, Rother M, Lang M, Munaretto C, Ulas S, Stelzer M, Grote A et al (2013) BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. Nucleic Acids Res 41:D764–D772. doi:10.1093/nar/gks1049
Schommer NN, Gallo RL (2013) Structure and function of the human skin microbiome. Trends Microbiol 21:660–668. doi:10.1016/j.tim.2013.10.001
Seshadri R, Kravitz SA, Smarr L, Gilna P, Frazier M (2007) CAMERA: a community resource for metagenomics. PLoS Biol 5, e75. doi:10.1371/journal.pbio.0050075
Sharma R, Ranjan R, Kapardar RK, Grover A (2005) 'Unculturable' bacterial diversity: an untapped resource. Curr Sci 89:72–77
Sharma VK, Kumar N, Prakash T, Taylor TD (2010) MetaBioME: a database to explore commercially useful enzymes in metagenomic datasets. Nucleic Acids Res 38:D468–D472. doi:10.1093/nar/gkp1001
Simmons C, Reddy A, D'haeseleer P, Khudyakov J, Billis K, Pati A, Simmons B, Singer S, Thelen M, VanderGheynst J (2014) Metatranscriptomic analysis of lignocellulolytic microbial communities involved in high-solids decomposition of rice straw. Biotechnol Biofuels 7:495. doi:10.1186/s13068-014-0180-0
Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP (2014) Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet 15:121–132. doi:10.1038/nrg3642
Singh M, Patel SK, Kalia VC (2009) *Bacillus subtilis* as potential producer for polyhydroxyalkanoates. Microb Cell Fact 8:38. doi:10.1186/1475-2859-8-38
Singh M, Kumar P, Patel SK, Kalia VC (2013) Production of polyhydroxyalkanoate co-polymer by *Bacillus thuringiensis*. Indian J Microbiol 53:77–83. doi:10.1007/s12088-012-0294-7
Soares I, Tavora Z, Patera R, Baroni S (2012) Microorganism-produced enzymes in the food industry. In: Valdez DB (ed) Food industry, scientific, health and social aspects of the food industry. InTech, Rijeka
Solaiman DY, Catara V, Greco S (2005) Poly(hydroxyalkanoate) synthase genotype and PHA production of *Pseudomonas corrugata* and *P. mediterranea*. J Ind Microbiol Biotechnol 32:75–82. doi:10.1007/s10295-005-0205-6
Solli L, Havelsrud O, Horn S, Rike A (2014) A metagenomic study of the microbial communities in four parallel biogas reactors. Biotechnol Biofuels 7:146. doi:10.1186/s13068-014-0146-2
Staley JT, Konopka A (1985) Measurement of *in situ* activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annu Rev Microbiol 39:321–346. doi:10.1146/annurev.mi.39.100185.001541
Stewart EJ (2012) Growing unculturable bacteria. J Bacteriol 194:4151–4160. doi:10.1128/jb.00345-12
Streit WR, Schmitz RA (2004) Metagenomics-the key to the uncultured microbes. Curr Opin Microbiol 7:492–498. doi:10.1016/j.mib.2004.08.002
Su H, Zhang T, Bao M, Jiang Y, Wang Y, Tan T (2014) Genome sequence of a promising hydrogen-producing facultative anaerobic bacterium, *Brevundimonas naejangsanensis* strain B1. Genome Announc 2:e00542-14. doi:10.1128/genomeA.00542-14
Tanaka T, Kawasaki K, Daimon S, Kitagawa W, Yamamoto K, Tamaki H, Tanaka M, Nakatsu CH, Kamagata Y (2014) A hidden pitfall in the preparation of agar media undermines microorganism cultivability. Appl Environ Microbiol 80:7659–7666. doi:10.1128/aem.02741-14
Tariq MA, Kim HJ, Jejelowo O, Pourmand N (2011) Whole-transcriptome RNAseq analysis from minute amount of total RNA. Nucleic Acids Res 39, e120. doi:10.1093/nar/gkr547
Thomas T, Gilbert J, Meyer F (2012) Metagenomics – a guide from sampling to data analysis. Microb Inform Exp 2:3. doi:10.1186/2042-5783-2-3
Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428:37–43. doi:10.1038/nature02340
Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W et al (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science 304:66–74. doi:10.1126/science.1093857
Vermelho AB, Supuran CT, Guisan JM (2012) Microbial enzyme: applications in industry and in bioremedia-

tion. Enzyme Res 2012:980681. doi:10.1155/2012/980681
Vester J, Glaring M, Stougaard P (2015) Improved cultivation and metagenomics as new tools for bioprospecting in cold environments. Extremophiles 19:17–29. doi:10.1007/s00792-014-0704-3
Vishnivetskaya TA, Layton AC, Lau MCY, Chauhan A, Cheng KR, Meyers AJ, Murphy JR, Rogers AW, Saarunya GS, Williams DE et al (2014) Commercial DNA extraction kits impact observed microbial community composition in permafrost samples. FEMS Microbiol Ecol 87:217–230. doi:10.1111/1574-6941.12219
Wang Y, Hayatsu M, Fujii T (2012) Extraction of bacterial RNA from soil: challenges and solutions. Microbes Environ 27:111–121. doi:10.1264/jsme2.ME11304
Wang Y, Zhang R, Zheng Q, Jiao N (2014) Draft genome sequences of two marine phototrophic bacteria, *Erythrobacter longus* strain DSM 6997 and *Erythrobacter litoralis* strain DSM 8509. Genome Announc 2:e00677-14. doi:10.1128/genomeA.00677-14
Ward BB (2002) How many species of prokaryotes are there? Proc Natl Acad Sci U S A 99:10234–10236. doi:10.1073/pnas.162359199
Warnecke F, Hess M (2009) A perspective: metatranscriptomics as a tool for the discovery of novel biocatalysts. J Biotechnol 142:91–95. doi:10.1016/j.jbiotec.2009.03.022
Weiland P (2010) Biogas production: current state and perspectives. Appl Microbiol Biotechnol 85:849–860. doi:10.1007/s00253-009-2246-7
Whitman WB, Coleman DC, Wiebe WJ (1998) Prokaryotes: the unseen majority. Proc Natl Acad Sci U S A 95:6578–6583
Williams PG (2009) Panning for chemical gold: marine bacteria as a source of new therapeutics. Trends Biotechnol 27:45–52. doi:10.1016/j.tibtech.2008.10.005
Wirth R, Kovacs E, Maroti G, Bagi Z, Rakhely G, Kovacs KL (2012) Characterization of a biogas-producing microbial community by short-read next generation DNA sequencing. Biotechnol Biofuels 5:41. doi:10.1186/1754-6834-5-41
Woese CR (1987) Bacterial evolution. Microbiol Rev 51:221–271. doi:0146-0749/87/020221-51$02.00/0
Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A 74:5088–5090. doi:10.1073/pnas.74.11.5088
Wong YM, Juan JC, Gan HM, Austin CM (2014a) Draft genome sequence of *Clostridium bifermentans* strain WYM, a promising biohydrogen producer isolated from landfill leachate sludge. Genome Announc 2:e00077-14. doi:10.1128/genomeA.00077-14
Wong YM, Juan JC, Gan HM, Austin CM (2014b) Draft genome sequence of *Clostridium perfringens* strain JJC, a highly efficient hydrogen producer isolated from landfill leachate sludge. Genome Announc 2:e00064-14. doi:10.1128/genomeA.00064-14
Wong YM, Juan JC, Ting A, Wu TY, Gan HM, Austin CM (2014c) Draft genome sequence of *Clostridium* sp. strain Ade.TY, a new biohydrogen- and biochemical-producing bacterium isolated from landfill leachate sludge. Genome Announc 2:e00078-14. doi:10.1128/genomeA.00078-14
Wooley JC, Ye Y (2009) Metagenomics: facts and artifacts, and computational challenges. J Comput Sci Technol 25:71–81. doi:10.1007/s11390-010-9306-4
Wu S, Zhu Z, Fu L, Niu B, Li W (2011) WebMGA: a customizable web server for fast metagenomic sequence analysis. BMC Genomics 12:444. doi:10.1186/1471-2164-12-444
Xia X (2013) DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. Mol Biol Evol 30:1720–1728. doi:10.1093/molbev/mst064
Zakrzewski M, Goesmann A, Jaenicke S, Junemann S, Eikmeyer F, Szczepanowski R, Al-Soud W, Sorensen S, Puhler A, Schluter A (2012) Profiling the metabolically active community from a production scale biogas plant by means of high throughput metatranscriptome sequencing. J Biotechnol 158:248–258. doi:10.1016/j.jbiotec.2012.01.020
Zhang T, Bao M, Wang Y, Su H, Tan T (2014) Genome sequence of *Bacillus cereus* strain A1, an efficient starch-utilizing producer of hydrogen. Genome Announc 2:e00494-14. doi:10.1128/genomeA.00494-14
Zoetendal EG, Booijink CC, Klaassens ES, Heilig HG, Kleerebezem M, Smidt H, de Vos WM (2006) Isolation of RNA from bacterial samples of the human gastrointestinal tract. Nat Protoc 1:954–959. doi:10.1038/nprot.2006.143
Zong C, Lu S, Chapman AR, Xie XS (2012) Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science 338:1622–1626. doi:10.1126/science.1229164

Ravi Ranjan Ph.D. completed his PhD in Molecular Microbiology from University of Delhi and CSIR-Institute of Genomics and Integrative Biology (IGIB), Delhi. He is a life member of the Association of Microbiologists of India (AMI). Presently he is a Research Assistant Professor at the University of Illinois at Chicago. His research interests are in the field of environmental microbiology and genomics, microbiome, and host-pathogen interactions.

Asha Rani Ph.D. received MSc in Biomedical Sciences and PhD in Microbial Biotechnology and Genomics from ACBR University of Delhi and CSIR-IGIB, Delhi. She is presently a Research Assistant Professor at the University of Illinois at Chicago. Her research interests are in the field of environmental and microbial genomics and microbiome in health and disease.

Rajesh Kumar received his MSc and PhD from Department of Botany, University of Delhi. He is presently the Head of the Department of Botany at Hindu College, University of Delhi. His current interests are biological control of *Aspergillus flavus* and identification and characterization of antagonist microbes using 16S rRNA gene sequence analysis.

http://www.springer.com/978-81-322-2597-3

Microbial Factories
Biofuels, Waste treatment: Volume 1
Kalia, V.C. (Ed.)
2015, XI, 353 p. 50 illus., 13 illus. in color., Hardcover
ISBN: 978-81-322-2597-3