



# Metatranscriptomics: an approach for retrieving novel eukaryotic genes from polluted and related environments

Arkadeep Mukherjee<sup>1</sup> · M. Sudhakara Reddy<sup>1</sup>

Received: 29 March 2019 / Accepted: 6 January 2020 / Published online: 27 January 2020  
© King Abdulaziz City for Science and Technology 2020

## Abstract

Metatranscriptomics, a subset of metagenomics, provides valuable information about the whole gene expression profiling of complex microbial communities of an ecosystem. Metagenomic studies mainly focus on the genomic content and identification of microbes present within a community, while metatranscriptomics provides the diversity of the active genes within such community, their expression profile and how these levels change due to change in environmental conditions. Metatranscriptomics has been applied to different types of environments, from the study of human microbiomes, to those found in plants, animals, within soils and in aquatic systems. Metatranscriptomics, based on the utilization of mRNA isolated from environmental samples, is a suitable approach to mine the eukaryotic gene pool for genes of biotechnological relevance. Also, it is imperative to develop different bioinformatic pipelines to analyse the data obtained from metatranscriptomic analysis. In the present review, we summarise the metatranscriptomics applied to soil environments to study the functional diversity, and discuss approaches for isolating the genes involved in organic matter degradation and providing tolerance to toxic metals, role of metatranscriptomics in microbiome research, various bioinformatics pipelines used in data analysis and technical challenges for gaining biologically meaningful insight of this approach.

**Keywords** Metatranscriptomics · Eukaryotic genes · Toxic metal tolerance · Functional diversity · Microbiomes · Metagenomes

## Introduction

A complex ecosystem like soil bears a multitude of significant community-based operations and interactions occurring amongst microorganisms. There is thus, little or no probability of a microbe to live a lonely life in such an environment (Pace 1997). To define soil ecosystem, one should consider its biogeochemical makeup comprising physical (e.g., density, texture etc.), chemical (pH, available Phosphate, C:N ratio etc.), and biological (e.g., microbial biomass) properties (Dominati et al. 2010; Pointing and Belnap 2012). Maintenance and enhancement of soil properties are the primary roles of the inhabiting microbial population and involves interplay with other members of the same ecosystem (Doran and Zeiss 2000; Lavelle et al. 2006). Though

physical properties of soil have been extensively studied, the changing dynamic nature, biodiversity, and functionality by studied by holistic approaches and not by isolating the members, which accounts for the low success under laboratory condition (Kellner et al. 2011; Torsvik et al. 1996). Bacterial species of soil have been extensively studied, but the huge diversity of eukaryotes in soil and essential roles they play, is not well established. Various cultivation-based surveys have biased outcomes as soil eukaryotes are difficult to isolate because of the profuse growth rate of bacteria (Bending et al. 2007; Jumpponen and Johnson 2005; O'Brien et al. 2005).

To appreciate eukaryotic functional diversity in soil and the activities they express *in situ* in response to different environmental constraints, it is necessary to develop new experimental approaches adapted for studying these microorganisms (Nannipieri et al. 2003). One such approach developed in recent years is metagenomics. Metagenomic approach is based on direct isolation of nucleic acids from environmental samples and has proven to be a powerful tool for comparing and exploring soil ecology (Biddle et al. 2008). Though metagenomic studies provide valuable

✉ M. Sudhakara Reddy  
msreddy@thapar.edu

<sup>1</sup> Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala, Punjab 147004, India

information on the probable existence of microbial communities, it does not give a picture of the actual activities carried out by them at specific environment or changing environmental conditions. The metatranscriptomics approach connects community-based structures and function in single experiment, and reach beyond the community's genomic potential as assessed in DNA-based methods (Bailly et al. 2007). In this review, we summarise various methods used to study the microbial community structures through the metagenomic and metatranscriptomic approach, role of metatranscriptomics in functional diversity of eukaryotes and identification of novel genes involved in various biological processes from the environmental samples. An attempt has also been made to study the problems, challenges and future prospective of the metatranscriptomics.

## Microbial communities in soil

### PCR based methods

Various molecular methods have been developed which involves extraction of nucleic acid, directly or indirectly from soil and were able to detect species, genera, families or even higher taxonomic groups (Handelsman et al. 1998; Hirsch et al. 2010). Using techniques such as DNA re-association also provides information of diversity specially the variations in GC content for detecting changes in microbial community. This technique requires large amounts of DNA and resolution suffers as several taxonomic groups may have same G + C content. Hence it does not provide enough information about the composition and the role of microorganisms.

The format for achieving diversity information changed significantly with the advent of more complex molecular tools like the use of denaturing gradient gel electrophoresis (DGGE) (Deng et al. 2008) and eventually PCR-derived clone libraries of genes encoding the small subunit ribosomal RNA (rRNA) (Lara et al. 2007). Nucleic acids obtained from soil were subjected to different approaches of fingerprinting to assess the change in community structure with

respect to environment. Comparing profiles of low molecular weight RNAs (5S rRNA, tRNA) collected from environment to their pure cultures previously obtained, provided insights into the change in structure (Liu and Stahl 2007). Amplified Ribosomal DNA Restriction Analysis (ARDRA) DNA fingerprinting technique makes use of restriction enzyme digestion of amplified 16S rRNA genes followed by gel electrophoresis (Nüsslein and Tiedje 1998). Similar technique involving terminal restriction fragment length polymorphism (T-RFLP), measures only the terminal restriction fragment of every 16S rRNA gene thus reducing the complexity of ARDRA (Liu et al. 1997; Marsh 1999). PCR amplified intergenic spacer between 16S rRNA and 23S rRNA genes or ribosomal spacer analysis (RISA) was also used to assess the diversity of bacteria in certain soil type (Ranjard et al. 2000). Various PCR based fingerprinting techniques used to study the microbial community structures in different soil environments are depicted in Table 1.

Though PCR based studies are robust techniques for analyzing the metagenome they have many drawbacks mainly due to the biasness introduced by the primers and interference of stray environmental DNA. It will eventually include genetic material not only from the live (active/inactive) sources but also from dead organisms or plant (Venter et al. 2004; Gill et al. 2006). Another major disadvantage for such techniques is the presence of PCR-inhibiting substances like humic acids, bacterial exopolysaccharides and proteins usually associated with particular soil types (Gelsomino et al. 1999; Miller et al. 1999; Opel et al. 2010).

### Metagenomic approach

Metagenomic based methods rely on soil-extracted DNA to be directly inserted or cloned in vector systems like plasmids or bacterial artificial chromosomes (BACs) and then propagated in bacterial strains like *E. coli* thus creating a genomic library of all possible genetic content (Liles et al. 2003). The DNA library of the entire metagenome of soil has been estimated to be approximately  $10^6$  Bacterial Artificial Chromosome (BAC) clones with an insert size of 100 kb (Handelsman et al. 1998). The screening of metagenomic

**Table 1** Selected nucleic acid fingerprinting techniques for microbial community DNA analysis

Nucleic acid fingerprinting techniques	Contaminated soil	References
ARDRA	Copper	Smit et al. (1997), Weidner et al. (1996)
T-RFLP	Hydrocarbon	Clement et al. (1998)
RISA	Mercury, copper	Dell'Amico et al. (2008), Ranjard et al. (2000)
ERIC-PCR	Cadmium	Roane and Pepper (1999)
DGGE	Sewage, heavy metals	Kozdrój and van Elsas (2000), Muyzer and Smalla (1998), Sandaa et al. (1999)
TGGE	Zinc, phenol	Brim et al. (1999), Eichner et al. (1999)

libraries usually depend on two broad techniques (Schloss and Handelsman 2003). Firstly, they depend on the specific enzymatic activity of the recombinant cells. This is known as activity-based screening as identifying new bioactivity does not require information about previous known sequences. Secondly, sequence driven (Daniel 2002; Rondon et al. 2000) which makes use of conserved DNA sequences to design hybridization probes or PCR primers, which are derived from conserved regions of already known genes or protein families to screen metagenomic libraries for clones that contain sequences of interest. In this way, only novel variants of known functional classes of proteins can be identified (Knietzsch et al. 2003). A significant fraction of eukaryotes could not be encompassed by this method of creating gene library.

The metagenomic DNA library contains a very low proportion of clones of eukaryotic origin as prokaryotes outnumber eukaryotic organisms in soil (Lehembre et al. 2013). Also, owing to the huge size of the genome (e.g., 13.8 Mpb for the yeast *Schizosaccharomyces pombe* (Wood et al. 2002), eukaryotic genes could not be well represented within a functional genomic library, hence a significant number of their duties in soil still remain unconvincing (Yadav et al. 2014). Moreover, expression of eukaryotic protein-coding genes is prevented in most of the hosts (both prokaryotes and eukaryotes), because of the presence of introns and lack of conservation of transcription regulatory elements in promoter sequences (Bailly et al. 2007; Takasaki et al. 2013).

Function based study of environmental DNA holds a role of great importance as more and more data sets are generated by cloning methods creating knowledge of putative biocatalysts which were previously not achievable. Discovery of such novel biocatalysts require bioinformatic tools and available databases where sequence-based metagenomic study can be deciphered. A novel protein-coding gene can thus be categorized that often represents more than half of sequence similarity of the total identified ORF to a known functional protein. Warnecke et al. (2007) discovered more than 600 new putative glycosyl hydrolases from the termite hindgut. The major constraint for functional metagenomics is availability of suitable activity assays for high-throughput screens. Some enzymatic activities may require the collective action of several proteins that could happen to be dispersed elsewhere on the host's genome. Thus, the respective activities might not be detected when screening small insert metagenomic clone libraries. Moreover, the absence of chaperons in expression hosts might fail to fold the protein and hence cannot function in the host (Warnecke and Hess 2009). A major limitation of metagenomics is that it cannot reveal dynamic properties, such as impact of the environment on the activities of the community life. The only information obtained is the potential of the microbiome to exhibit functional properties associated with the presence

of certain genes but may overlook the expression pattern of the genes in a stressful environment. Hence the next subset of metagenomics which sheds light on the active functional profile of a microbial community is metatranscriptomics.

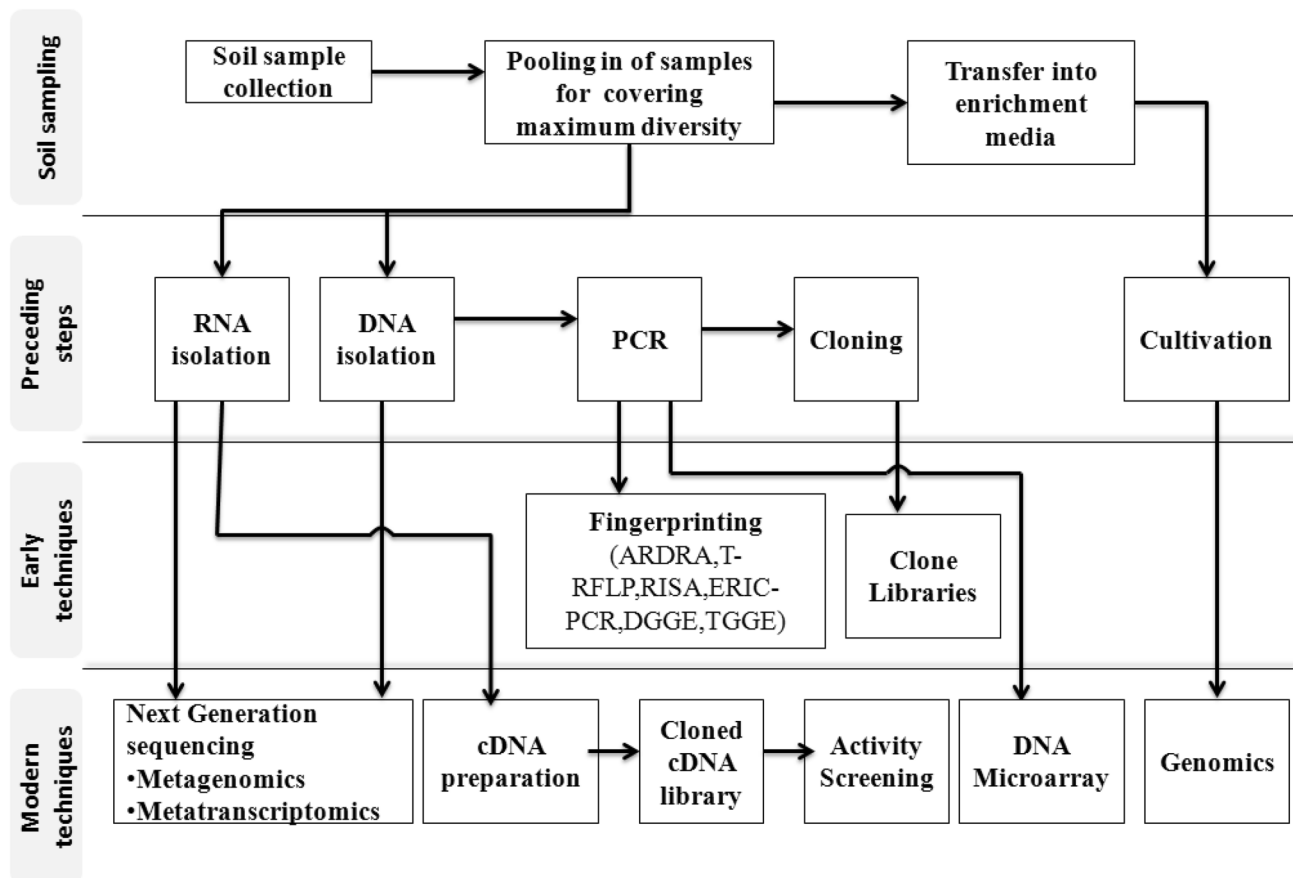
### Metatranscriptomics approach

Metatranscriptomics deal with those sets of genes which respond to a given environment at a given time to transcribe and exhibit activity (Chao-Rong and Zhang 2011; Moran 2009). Functional metatranscriptomics hence is a powerful tool which allows characterization of genes expressed by different eukaryotic microorganisms (e.g., fungi, protists, etc.) directly in the environment. This approach has a strong potential in biotechnology to discover novel genes of interest for the bioindustry, in bioremediation and as biomarkers. This approach allows characterization of genes implicated in adaptation to stressful conditions or involved in organic matter degradation. Functional metatranscriptomics applied to soil environment involves the extraction and analysis of mRNA which provides information on the regulation and expression profiles of complex communities. It is based upon extraction of environmental RNA instead of DNA and on the purification by affinity chromatography of the eukaryote-specific polyadenylated mRNA from the total environmental RNA mixture. These poly-A mRNAs can be converted into cDNA, which can be cloned in appropriate expression vectors such as plasmids which allow expression of the cloned genes in the eukaryotic hosts (Yadav et al. 2014). Different methodologies to study the soil diversity and function of microorganisms which involves both culture dependent and culture-independent techniques are depicted in Fig. 1.

### Metatranscriptomics of polluted soil

#### Functional diversity of eukaryotes

Maintaining community structure and diversity is a challenge faced by the residing microorganisms in polluted soil (Giller et al. 1998). Though diversity of community is negatively correlated to increase in stress levels, the effect of 'competitive exclusion' would enhance the removal of dominant organisms, and promote an increase in diversity of usually less abundant organisms. This is considered resilience as the affected soil microorganisms try to regain their size, composition and function after an initial disturbance (Allison and Martiny 2008). Studies pertaining to heavy metal cadmium and copper have indicated the presence of 'competitive exclusion' (Degens et al. 2001; Zhang et al. 2009). Eukaryotic life forms have been previously studied in extreme environments along with prokaryotes



**Fig. 1** Overview of methodologies for studying soil diversity and function of microorganisms which involves both culture dependent and culture-independent techniques

as they represent a significant fraction of the microbial biomass. They hold a repository of genes responsible for various activities in soil, hence, it is difficult to isolate and culture them under laboratory condition. Holistic approaches like metagenomics and metatranscriptomics have developed to identify genes and archive them as soil gene library. Information thus stored is a treasure house for novel biocatalysts and bioactive compounds which can be coded by these genes when suitably expressed (Daniel 2005; Tringe and Rubin 2005).

Soil fungi have been under investigation for centuries (Rossman 1998), but DNA-based analysis have paved their way to prove that phylogenetically, eukaryotes are as diverse as bacteria. A study by Fierer et al. (2007) dealt with grouping of sequences by Operational Taxonomic Units (OTU's) and revealed that archaeal or fungal OTUs appeared to be equal or exceed the number of unique bacterial OTUs in different soil types (Desert, Grassland, Rain forest) with minimum taxonomic overlap between the soil types. In depth study of spruce and beech forest soil showed dominance of fungi, metazoans and protists linked with enzyme activities for litter degradation for

maintenance of fertility of terrestrial ecosystems (Damon et al. 2012).

With the advent of high-throughput or next-generation sequencing (NGS), large volume of sequence data is generated, which enables efficient analysis of more complex expression profiles. The most common high-throughput sequencing platforms used in metatranscriptomic studies are the 454 Genome Sequencer FLX systems (Roche) and the HiSeq 2000 (Illumina, Inc.). Both the platforms facilitate large numbers of individual sequencing reactions to be performed in parallel. Direct sequencing of DNA or cDNA with/without any cloning step (Medini et al. 2008) increases the throughput in terms of numbers of base pairs sequenced per run and decreases cost per sequenced base, though the new technologies produced short reads (454 pyrosequencer produced ~ 400 bp). Developments in sequencing technologies have improved and read lengths are comparable to traditional dye-terminator sequencing technology.

Sequencing technologies played a major role in metagenomics and metatranscriptomics for studying the diversity as culture-based identification were slowly on the decline. This included species identification and functional annotation

of genes. A clear majority in culture-independent based sequence identification and submission with the development of NGS can be seen over the years. We have analyzed the gene sequences derived from soil eukaryotes (Fungi and Protists) deposited in GenBank with different search strategies employed in Entrez. The results indicated that by the year 2011, almost 99.9% of sequence data submitted to GenBank were uncultured soil eukaryotes and the submission of uncultured sequence data did not fall below 89% of the total sequences submitted to GenBank even in the year 2019.

The first report of using 454 pyrosequencing for studying the metatranscriptome of a soil microbial community was published by Leininger et al. (2006), where they reported an abundance of archaeal ammonia oxidizers than their bacterial counterparts. A study prior to this on activated sludge and samples collected from hot springs reported cDNA library preparation from extracted RNA exhibiting significant similarity to eukaryote mRNA-encoded protein sequences (Grant et al. 2006). Kit based RNA extraction from aquatic samples was carried out by Poretsky et al. (2005) and approximately 400 clones were analyzed after cDNA was amplified with random primers. They identified many novel genes associated to environmentally important processes such as sulfur oxidation from diverse taxonomic groups.

Pioneering work on functional diversity of communities of soil eukaryotic microorganisms, which included construction and screening of a cDNA library using polyadenylated mRNA extracted from a forest soil was carried out by researchers from University of Lyon 1, France (Bailly et al 2007). It was observed that fungi and especially unicellular eukaryotes (protists) dominated the microbial population with genes encoding for essential proteins involved in different biochemical processes. This was assessed by complementation of auxotrophic mutant yeast with soil cDNA. Studies were extended to investigate litter degradation in forest ecosystem thus making nutrients available in soil organic matter degradation (Damon et al. 2012). Major nitrogen assimilating and cell wall degrading enzymes (celluloses, pectin, lignin etc.) were identified (Damon et al. 2011). Metatranscriptomics received great impetus for studying different soil type especially nutrient-poor soil mainly because it avoids biasness when gauging community reaction to environmental changes. Along with high-throughput sequencing technology, structure and functioning of sandy soil microflora were studied and less frequent taxa Crenarchaeota were well represented using rRNA-tags with the discovery of specific enzymes for ammonia oxidation and CO<sub>2</sub> fixation (Urich et al. 2008).

Metatranscriptomic approach was therefore played an important role for studying polluted or stressful soil systems. First study on heavy metal contaminated soil using metatranscriptomics was reported by Lehembre et al. (2013), where

a huge metatranscriptomic library was screened with varied diversity of sequences which were screened for heavy metal tolerance mostly from unknown taxonomic groups. The isolated heavy metal tolerant genes were able to withstand from the toxicity of a variety of metals even at high concentrations (Thakur et al. 2018, 2019). This opened up a huge platform for investigating the proteins and linking them with different pathways for coping up with heavy metal tolerance.

## Discovery of novel eukaryotic genes

### Mining of industrially important enzymes

Enzymes from higher eukaryotic microorganisms especially fungi have been widely used as industrial important biocatalysts ranging from applications in food industries, pharmaceutical, textile etc. (Gudynaite-Savitch and White 2016; Guerriero et al. 2016). Although thorough discussion regarding fungal biocatalysts is beyond the scope of this review, certain important contributions of eukaryotic genes from environment are emphasized here. Role of eukaryotes in improving soil fertility by litter degradation involves various enzymatic processes to degrade lignocelluloses, hemicelluloses, polyphenols etc. (Štursová et al. 2012). Litter degradation rates depend on their physicochemical properties, which in turn vary with climatic factors, plant species which form the litter and the association of saprophytic fungal microhabitats (Goldmann et al. 2015). Besides production of useful metabolites, eukaryotes are also involved in biomass treatment in biorefineries and biofuel production. Essential enzymes acting on complex organic matter like lytic polysaccharide monoxygenases cleaves complex polysaccharides like cellulose and hemicelluloses resulting in oxidative cleavage. These eukaryotic microbes are extremely diverse and hence their function cannot be localized and neither can be said that they are ubiquitous in every ecosystem (Marmeisse et al. 2017). To trap the extensive enzyme resources from forest soil, the most convenient way is to create a metatranscriptomic library and perform activity screening which does not incorporate biasness towards specific species. Industrially important enzymes like acid phosphatase and laccase were identified by activity screening and semi-quantitative PCR, respectively, following RNA extraction and cDNA preparation (Kellner et al. 2011; Luis 2005). The only disadvantage of this process is the massive number of transformed clones which have to be screened for activity. As many as 30,000 recombinant yeast clones were screened for potential phosphate solubilizing enzyme, acid phosphatase and also an imidazoleglycerol-phosphate dehydratase by utilizing the concept of functional complementation and activity screening (Kellner et al. 2011). The metatranscriptomic library was ligated to a yeast secretion vector pTEF-MF-*Sfi*I A/B which was modified with



additional *SfiI* restriction sites to accommodate the library and was used to transform *Saccharomyces cerevisiae* mutant ML20C having *pho5*– (mutation in the repressible acid phosphatase gene) and a *his3*– mutation. Positive clones complemented by plasmid borne *his3*– gene and colony staining assays were further sequenced. Strong similarity with basidiomycete acid phosphatase from the sequence data confirmed the important role played by fungi in nutrient mobilization in the organic rich layer of forest soil. The availability of a large number of yeast strains which are mutated for a variety of metabolic pathways provide an opportunity to perform complementation with gene resources to screen and find out the gene of interest (Scherens and Goffeau 2004). Yeast can be genetically manipulated by plasmids having constitutive or regulated promoters for gene expression and rescuing the auxotrophs by complementation.

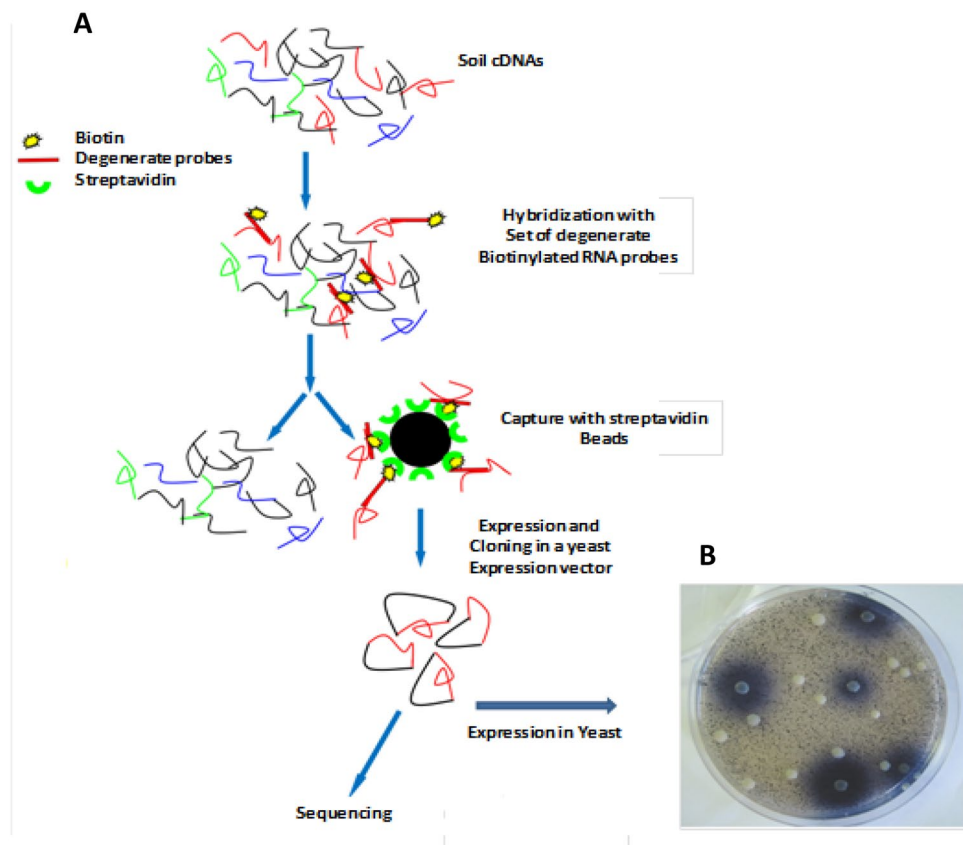
Along with mutant yeast strains, plasmids have also been modified to successfully clone environmental cDNA and effectively express the gene in the host. The transcription of foreign genes in mutant yeast will depend upon carefully placed compatible promoter in the plasmid upstream of the cloning site. Expression of environmental eukaryotic cDNA is efficiently cloned in pFL61 which has a Multiple Cloning Sites between Phosphoglycerate kinase (PGK) promoter and its terminator and also bearing a 2  $\mu$ m yeast origin of replication for yeast host and part of pUC19 *E. Coli* vector for replication in bacterial host (Minet et al. 1992).

An efficient technique to screen cDNA synthesized from soil-extracted polyadenylated mRNAs is using ‘Sequence capture by hybridization’ (SCH) (Bragalini et al. 2014). This method of pre-selection of target genes before activity screening increases the probability of exploring novel genes by reducing the number of random environmental cDNA. It was conducted using biotinylated degenerated RNA probes. The biotinylated probes hybridized to known and unknown sequences of members of a target gene family present in a metatranscriptomic library. They were further captured using streptavidin-coated paramagnetic beads. Tested on the glycoside hydrolase 11 gene family encoding endo-xylanases, the capture probes were able to relate > 90% of the cDNA to this gene family. Sequencing of cDNA revealed many phylogenetically diverse species, which were not yet included in public databases. The captured cDNA was cloned and expressed in Uracil-auxotrophic yeast strain DSY-5 also having non-functional sugar transporter protein. Yeast cells producing endo-xylanases developed a blue halo in the presence of xylan linked to a dye confirming hydrolysis of xylan. This method used the plasmid PDR 196-SfiI-Kan modified to contain two specific *SfiIA* and *SfiIB* sites for ligation of cDNA which also contain identical sites for directional cloning and placed under the control of a yeast *PMA1* promoter to facilitate constitutive expression of cDNA. Figure 2 depicts the overview of the sequence capture screening

activity by degenerate probes, transformation with cDNA clones on selective medium.

The enzymes retrieved from eukaryotic organisms have dominated the industrial processes especially where it has replaced traditional usage of chemical treatment of biomass, in textile as well as in food industries (Azizan et al. 2016). The recent advances to improve the commercial potential of industrially important enzymes have focused on the molecular techniques such as metagenomics and metatranscriptomics for accessing the environmental genes (Niehaus et al. 2011). The practice of collecting industrially important genes from the environmental metagenomic library has been discussed in detail by Ferrer et al. (2016). Sequence based as well as functional screening of environmental metagenome to achieve a stable, ‘industry-level’ enzyme is a time-consuming process. This requires efficient screening and expression systems. However, the approach of metatranscriptomics has been developed into a promising and practical technique for harvesting and screening protein-coding genes from an environmental library. Forest soil has been the primary source to harvest important and novel enzyme coding genes. Glycoside hydrolase that decompose cellulosic biomass was harvested from the fungal communities of forest soil (Takasaki et al. 2013). About 40% of genes involved in metabolism of carbohydrate and amino acid out of 9449 eukaryotic coding sequences were identified by metatranscriptomic approach. Functional metatranscriptomics was successfully applied to eukaryotes of forest soil where acid phosphatase and imidazoleglycerol-phosphate dehydratase of fungal origin was isolated and characterized using yeast mutants which could proficiently express the protein (Kellner et al. 2011). Many important eukaryotic oligopeptide transporters were discovered by functionally complementing yeast mutant defective in di/tripeptide uptake. These transporters were able to act upon a wide range of substrates and were further expressed in *Xenopus* oocytes (Damon et al. 2011). A significant number of non-fungal sources of industrially important enzymes in forest soils were explored by Damon et al. (2012), where metatranscriptomic library was probed for essential enzyme coding genes. Major plant cell wall and polymer (cellulose, hemicelluloses, pectin, lignin etc.) degrading enzymes were identified. Also, organic matter hydrolyzing enzymes like proteases, lipases, a phytase etc. were also identified. Forest soil is a host of carbohydrate active enzyme transcript sequences (CAZymes) and essentially about 74,000 active transcripts were identified in a maple forest where soil metatranscriptome was analyzed (Hesse et al. 2015). The use of metatranscriptomics in screening polyadenylated mRNAs from soil was implemented on endo-xylanases encoding gene glycoside hydrolase 11 gene family. An efficient functional screening mechanism of solution hybrid selection was used screening of enzyme encoding genes (Bragalini et al. 2014).

**Fig. 2 a** Overview of the Sequence capture screening activity by degenerate probes followed by **(b)** Transformation of yeast DSY-5 with cloned cDNA plated on a selective medium containing AZCL-xylan (Specific substrate to endoxylanase) and without Uracil. Positive clones degrade the substrate to release dark blue dye. Modified from Bragalini et al. (2014)



Metatranscriptomics has also been applied to polluted soil for discovering enzymes of industrial importance. Mukherjee et al. (2019a), aldehyde dehydrogenase gene belonging to the phylum *Ciliophora* was isolated from polluted soil, which when expressed in metal sensitive yeasts, conferred tolerance against various toxic metals. Aldehyde dehydrogenases are an important group of abiotic stress response enzymes which are expressed to eliminate toxic levels of aldehydes generated due to environmental stresses which includes metal stress. A novel serine protease inhibitor (serpin) belonging to the phylum *Tardigrada* was isolated, which showed increased expression levels in presence of metals and was able to confer metal tolerance. Similar to aldehyde dehydrogenase, serpins are also stress response proteins which act as defense system against biotic and abiotic stresses (Mukherjee et al. 2019b). Also, heterologous gene expression of characterized genes could be beneficial for improving stress tolerant capacity of crop plants as the genes are a part of important antioxidant systems for organisms.

The use of pesticides and chemical fertilizers in many agricultural lands has led to the accumulation of the persistent aromatic hydrocarbon pollutants in the soil. In a study by Sharma and Sharma (2018), pesticide and chemical fertilizer polluted soil was studied and the taxonomic and functional aspect of the microbial communities was investigated by metatranscriptomic approach. Transcripts

related to degradation of cypermethrin and related aromatic compounds were identified. Hence, using a single technique, both prokaryotic and eukaryotic enzymes were studied for pesticide degradation. The compound cypermethrin is degraded in a step wise process forming by products like carboxylate, benzoic acid, derivatives of aldehydes etc. It can thus be inferred that the process of removal of toxic compounds is an orchestration of a chain of functional proteins which may originate from any organisms in the soil ecosystem. Metatranscriptomic study of the wheat rhizosphere revealed various bacteria that are involved in the degradation of xenobiotics like aromatic compounds, carbazoles, naphthalene (Singh et al. 2018).

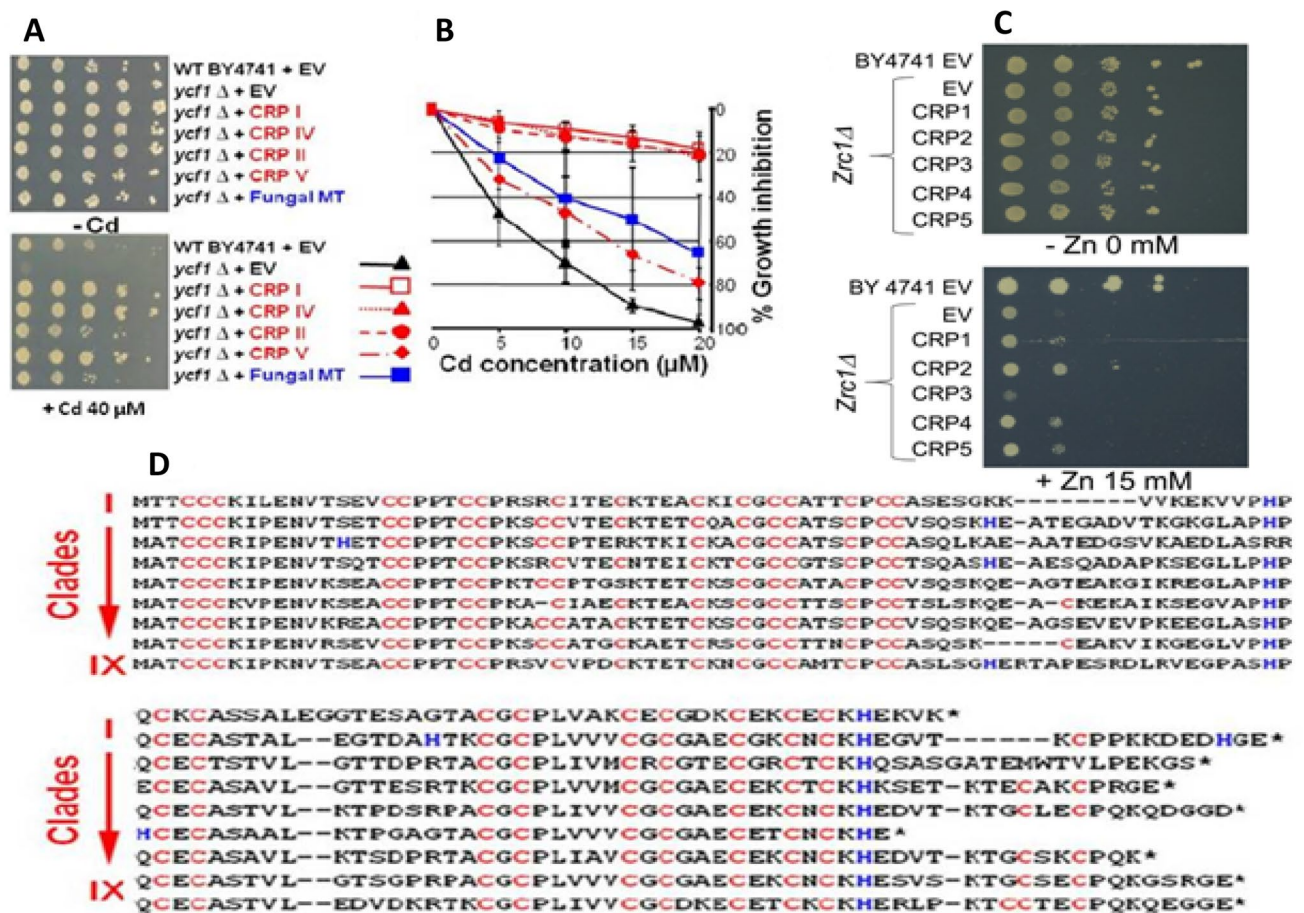
#### Mining of genes involved in metal tolerance

Though there are a limited number of published research papers, a significant amount of molecular characterization was performed with metal resistance genes obtained from the environment proving metatranscriptomics approach, a lucrative study method for bioremediation of heavy metals. Metallothioneins (MT), a family of ubiquitous cysteine-rich proteins have long been studied for their ability to coordinate the binding of metal ions by forming metal-thiolate complexes (Blindauer 2013; Mehra and Winge 1991). Though MTs have a defined classification system based on sequence

similarity analysis (Binz and Kägi 1999), they share less overall sequence conservation. This might be due to independent emergence of MT families throughout the course of evolution (Capdevila and Atrian 2011). Eukaryotic MTs, however, have not been covered fully as it has been studied only in specific species like vertebrates, plants, fungi etc. Thus, the search for new metal-binding proteins from environmental DNA or RNA will be a fruitful endeavor in terms of detection and bioremediation of heavy metals in soil. Soil-extracted mRNA converted to cDNA was screened on the basis of their functional role of resisting toxic amounts of heavy metals (Cervantes and Gutierrez-Corona 1994; Lehembre et al. 2013). This cDNA was cloned into suitable yeast expression vectors. Metal sensitive phenotypes of yeast were then transformed and the cDNAs were screened on their ability to rescue the yeast strain by functional complementation. Members of the cysteine-rich proteins (CRP) were identified and compared with the already existing

fungal metallothioneins (Fig. 3). The potential of these environmental genes to support the growth of sensitive yeast is much higher than existing fungal metallothioneins in media with cadmium metal stress (Lehembre et al. 2013). Further characterization of the CRPs was carried out by Ziller et al. (2017) (Fig. 3). Amino acid features were analyzed for these CRPs and their metal-binding abilities towards Cd, Zn and Cu metal ions. Previously isolated soil metatranscriptomes were sub-cloned and expressed as a fusion protein. These novel proteins or 'environmental metallothioneins' had the ability to chelate Cd(II), Zn(II), and Cu(I). Various yeast mutants (*ycf1 $\Delta$* , *zrc1 $\Delta$* ) derived from the wild type strain BY4741 (*MATa his3 $\Delta$ 1 leu2 $\Delta$ 0met15 $\Delta$ 0 ura3 $\Delta$ 0*) were used, thereby enabling efficient screening by functional complementation.

The potential of genes identified by functional metatranscriptomics can be extended to be used as biomarkers whose level of expression can be used to access the extent



**Fig. 3** **a** Functional complementation of the CRPs in Cadmium sensitive yeast phenotypes *ycf1 $\Delta$*  compared to the fungal metallothionein gene. **b** Inhibition of growth (%) in liquid broth spiked with cadmium. **c** CRP 1–5 expression in Zinc sensitive phenotypes *Zrc1 $\Delta$*  in presence and absence of Zinc metal exhibited multi-metal resistance potential.

**d** Nine cysteine-rich proteins (CRP2) were identified from soils of 12 different sites with their amino acids alignments showing conserved cysteine motifs. Adapted from Lehembre et al. (2013) and Ziller et al. (2017)



of toxicity of heavy metals in crop fields. A recent study on yeast-based biosensors for detection of environmental pollutants provides an explanation as to why eukaryotic models are now being utilized for this purpose. Yeast-based systems have an edge over bacteria based and enzyme-based biosensor systems as bacteria has low relevance for representing eukaryotic community and enzyme-based biosensors have lower relevance for the cell or the whole organisms (Jarque et al. 2016). Already existing yeast-based heavy metal detection are depicted in the Table 2. Toxic heavy metals, one of the most hazardous groups of contaminants, can be detected by both non-specific and specific biosensors. The sensing element in non-specific biosensors is based on the cAMP/PKA (cAMP-dependent protein kinase) signaling pathway which mediates gene expression in response to generic toxic substances in yeast. These biosensors have been shown to recognize  $As^{3+}$ ,  $Fe^{2+}$ ,  $Pb^{2+}$  and  $Cd^{2+}$ , but they seem to be nonselective to metals, responding to other generally toxic compounds (Radhika et al. 2005). With metatranscriptomics, a very specific metallothionein-based detection platform can be devised for biosensing heavy metal pollution in arable lands. A representation of a general biosensing detection system with yeasts is depicted in Fig. 4.

Critically designed experiments which capture microbial response reveal newer, underrepresented species, performing environmental activities like nutrient mobilization and breakdown of organic matter/hydrocarbons. Studies by Lehembre et al. (2013) and Ziller et al. (2017) have captured different roles of eukaryotes involved in metal chelation and homeostasis. This will greatly accelerate the tapping of natural resources for industrially important biocatalysts and also help to solve problems of soil fertility and pollution. Ubiquitin domain containing zinc finger protein was isolated and characterized by Thakur et al. (2018) which was screened in *S. cerevisiae* mutants and was able to impart tolerance towards different toxic metals. Similar study revealed a novel von Willebrand factor type D of vitellogenin protein from

polluted soil environment which depicted higher expression levels in presence of toxic metals and was able to confer tolerance towards toxic concentrations of metals (Thakur et al. 2019). Vitellogenin proteins are known for their antioxidant properties. From these studies it can be concluded that the technique to isolate enzyme coding genes from perturbed soil could serve as important biomarkers and genes suitable for bioremediation of environmental soil pollution. Table 3 summarizes all the published work on soil metatranscriptomics of various soil ecosystems.

## Metatranscriptomics in microbiome research

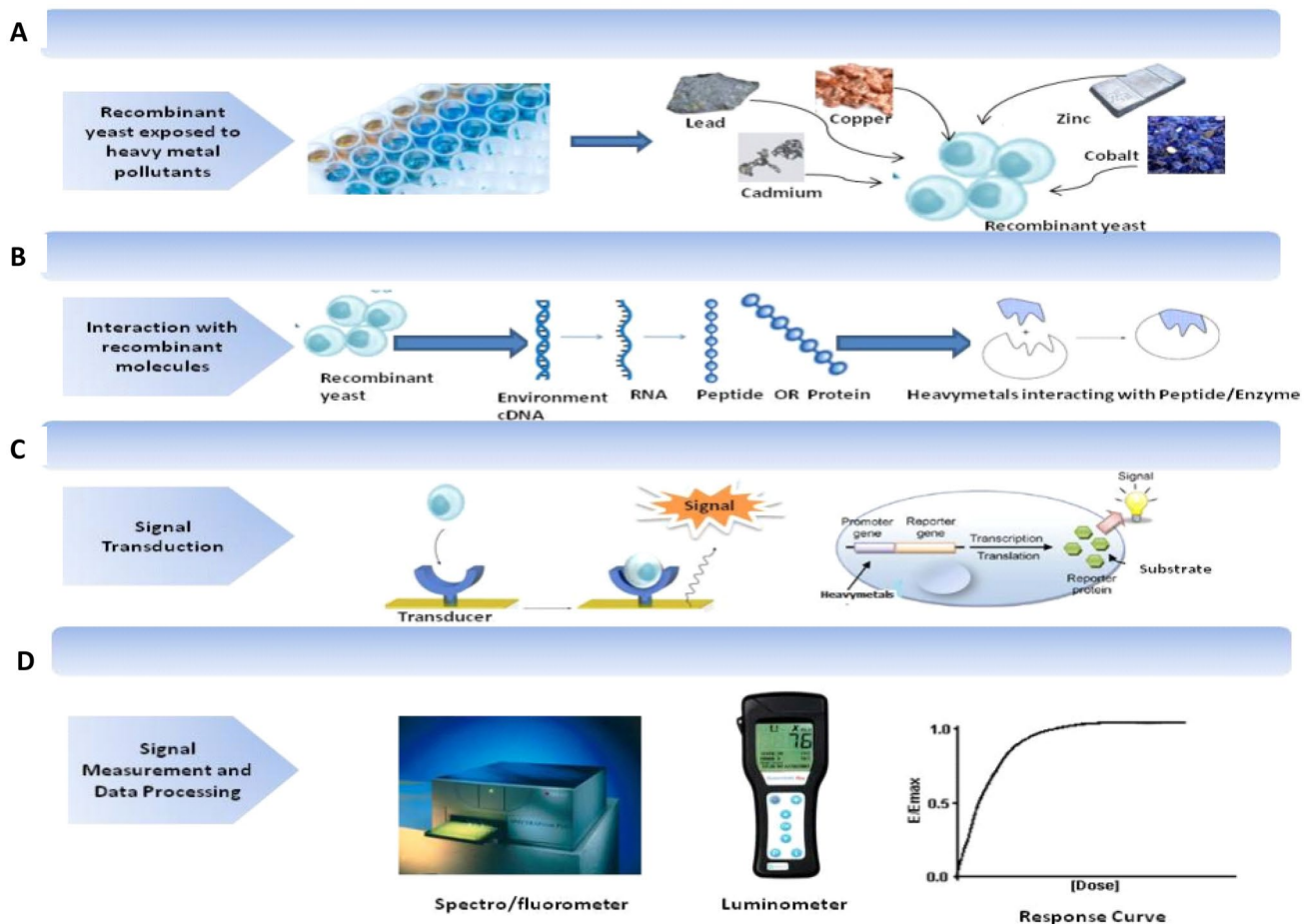
Metatranscriptomic approach has been employed in microbiome research majorly to characterize the human gut microflora and classifying the active population. It also helps to gain the insights into the host gene expression profile and also to study the gene expression of complex network of microbial communities at a given environment (gut, oral cavity, or respiratory tract), where culture-dependent techniques cannot assess their true functional roles. From such data, metabolic pathways in the microbial communities can thus be identified and be associated to particular environment. This can actively be associated with health, diseases and metabolism (Bashiardes et al. 2016).

Studies on the rumen microbiome of buffaloes, led to the understanding of the role of microbial carbohydrate assimilating enzymes present in the buffalo gut with the abundance of polysaccharide degrading bacteria (Patel et al. 2014). Moreover, the colonization of the bacteria in the rumen of cattle throughout its lifetime was studied using next-generation sequencing technology which has led to useful information of the digestive capabilities of the ruminant (Koringa et al. 2019). The rumen gut is particularly interesting from a study point as it holds a repertoire of carbohydrate assimilating active enzymes,

**Table 2** Yeast bio-detection systems of heavy metals in environment (Table modified from Jarque et al. (2016))

Detected metal(s)	Receptor/gene	Detection	Level of detection (range mM)	Ligand	References
Copper	CUP1	Amperometry	0.11–500	Copper	Lehmann et al. (2000)
		Colorimetry	1		Vopálenská et al. (2015)
		Fluorescence	0.5		Shetty et al. (2004)
		Luminescence	0.5		Roda et al. (2011)
Cadmium	AtPCS1	Fluorescence	0.2	Cadmium	Matsuura et al. (2013)
Heavy metals	cAMP-PKA	Fluorescence	12–45		Radhika et al. (2005)

*AtPCS1* phytochelatin synthase from *Arabidopsis thaliana*, *cAMP-PKA* cyclic adenosine monophosphate-protein kinase A



**Fig. 4** Yeast Biosensing Detection Principle. **a** Preparation of environmental samples and its exposure to yeast-based sensing system. **b** Recombinant product- environmental metal resistant gene reacts with

heavy metals inducing activity of **c** a reported gene or generation of electrical signal and **d** the signal associated with the response might be amplified and processed into valuable quantifiable data

which can degrade plant matter. It has been reported that breakdown of cellulose involves the action of three different types of cellulolytic enzymes endo- $\beta$ -1, 4-glucoanase, glycosyl hydrolase family, cellobiohydrolase and  $\beta$ -glucosidase (Terry et al. 2019). Several studies have employed the approach of metatranscriptomics in studying the eukaryotic population of rumen and their significant contribution towards plant fiber digestion. The presence of anaerobic fungi of phylum Neocallimastigomycota in the rumen has been associated with degradation of plant cell wall material where CAZyme families have been active in hemicelluloses digestion (Gruninger et al. 2018). High-throughput RNA sequencing technique for estimation of active population applied to rumen gut has found a total of 21.8% of the small subunit RNA belonging to the eukaryotes (Comtet-Marre et al. 2017). The protozoan population was also found to be an active member of the gut microflora and about 3–6% total non-rRNA reads were from protozoan species. However, *Piromyces* and *Neocallimastix*

accounted for the largest fungal cellulases identified from the rumen gut (Dai et al. 2015). Söllinger et al. (2018) correlated the activity of the transcripts along with the metabolites of rumen and cellulase transcripts originated from two eukaryotic groups, Neocallimastigaceae and Ciliophora.

Metatranscriptomic approach to study the rumen gut suggested the predominance of exoglucanases, which degrade crystalline cellulose. Many glycoside hydrolases, which were previously not identified by metagenomic studies were confirmed to be active in the metatranscriptomic study (Dai et al. 2015). Major cellulose binding modules were also found in the eukaryotic transcriptomes along with polysaccharide utilization loci which aid in the binding, transport and degradation of polymeric glycan structures (Seshadri et al. 2018). A detailed review by Terry et al. (2019) on the metagenomic and metatranscriptomic studies further describes the rumen gut microflora.

**Table 3** Published metatranscriptomics studies on different soil ecosystems

Soil Sample	RNA extraction method	mRNA enrichment	cDNA preparation	Sequencing/Screening	Size of dataset	Significant findings	Reference
Forest soil	Guanidine isothiocyanate, bead-beating	Poly(A) tail mRNA captured by affinity capture	Poly-dT primer (SMART cDNA kit,	Sanger sequencing	119 clone sequences	Eukaryotic diversity, novel enzymes	Bailly et al. (2007)
Sandy, Nutrient-poor soil	CTAB, P/C/I Lysing matrix E tube (Q-Biogene) PEG 8000	None	Random primer	Pyrosequencing	314,041	CO <sub>2</sub> fixation and ammonia oxidizers enzymes	Urich et al. (2008)
Sugar maple forest soil	Bead beating,, 8-hydroxyquinoline, RNA/ DNA Midi kit (Qiagen)	None	Selective amplification of eukaryotic mRNA by (SMART cDNA kit	Screening by auxotrophy-complementing genes in yeast	10 transformed yeast sequenced	Acid phosphatase, Imidazole glycerol-phosphate dehydratase	Kellner et al. (2011)
Spruce forest	Guanidine isothiocyanate, bead-beating	Affinity capture of polyadenylated mRNA on poly-dT coated paramagnetic beads	Poly-dT primer (SMART cDNA kit	Screening by auxotrophy-complementing genes in yeast	7.7 × 10 <sup>5</sup> screened transformants, 25 putative genes	Novel fungal oligopeptide transporter	Damon et al. (2011)
Beech and Spruce forest	Guanidine isothiocyanate, bead-beating	Poly-dT beads for separating eukaryotic polyadenylated RNA	SMART cDNA Library Construction Kit (Clontech)	Sanger sequencing	10,000 each	Eukaryotic Diversity, novel enzymes for organic matter degradation	Damon et al. (2012)
Soil near metal smelter	Guanidine isothiocyanate, bead-beating	Poly(A) tail mRNA captured by affinity capture	Poly-dT primer (SMART cDNA kit,	Sanger sequencing	30 k cDNA for each soil sample	Metal tolerant genes	Lehembre et al. (2013)
Agro forestry contaminated soil	RNA isolation Kit (Mo Bio labs)	None	Selective amplification of eukaryotic mRNA by (SMART cDNA kit	Screening by auxotrophy-complementing genes in yeast	0.5 kb- 1.0 kb size library	Ubiquitin fusion protein/Heavy metal tolerance gene	Thakur et al. (2018)
Agro forestry contaminated soil	RNA isolation Kit (Mo Bio labs)	None	Selective amplification of eukaryotic mRNA by (SMART cDNA kit	Screening by auxotrophy-complementing genes in yeast	1–4 kb size library	VWD like protein/ multi-metal tolerant	Thakur et al. (2019)
Metal contaminated forest soil	RNA isolation Kit (Mo Bio labs)	None	Selective amplification of eukaryotic mRNA by Mint2 cDNA kit	Screening by auxotrophy-complementing genes in yeast	1–4 kb size library	Aldehyde dehydrogenase conferring metal tolerance	Mukherjee et al. (2019a)
Metal contaminated forest soil	RNA isolation Kit (Mo Bio labs)	None	Selective amplification of eukaryotic mRNA by Mint2 cDNA kit	Screening by auxotrophy-complementing genes in yeast	1–4 kb size library	Serine protease inhibitor conferring metal tolerance	Mukherjee et al. (2019b)

## Bioinformatics used in metatranscriptomics

The gene expression profiling of the microbial population with respect to the environment is a major field of study. Metagenomic analysis along with metatranscriptomics allows a holistic approach to study a microbiome and its associated functions (Bashiardes et al. 2016). Several different metatranscriptomic pipelines have been developed for the efficient analysis of bioinformatics data concerning the functionality of the microbial community. MetaTrans is such a pipeline which is designed to handle RNA-Seq analyses concerning both taxonomic data and gene expression involving quality-control assessment, sorting of RNA into mRNA and non-mRNA, mapping the reads against functional databases and also associating differential gene expression database (Martinez et al. 2016). The Simple Annotation of Metatranscriptomes by Sequence Analysis (SAMSA) pipeline has been developed in conjunction with the Metagenomic Rapid Annotations using Subsystems Technology (MG-RAST) annotation server. The pipeline is validated with gut microbiome data and utilizes raw RNA-Seq data to form meaningful comparisons between the different species and their function. Following annotations and aggregation, the pipeline is designed to produce a graphical representation with custom R scripts to generate visual plots of data (Westreich et al. 2016). A suitable bioinformatic analysis platform applied to the human genome project is the HUMAnN, which involves high-throughput metagenomic functional reconstruction validated with 649 microbial communities from 7 body habitats. Several gene families and their respective pathways can be analyzed within a community as well as their relative abundances (Abubucker et al. 2012). Another bioinformatic pipeline concerning human gut microflora was developed by Leimena et al. (2013), which enabled study of symbiotic interactions within prokaryotic ecosystems. MetaQUBIC is designed as an integrated biclustering-based computational pipeline that can integrate metagenomic and metatranscriptomic data to study the multiple levels of human microbiome (Ma et al. 2019).

The use of robust bioinformatic analysis pipelines have been used to understand extreme environment microbiota as they are resistant or tolerant towards harsh environmental conditions. Integration of metagenomic and metatranscriptomic data was carried out successfully by Hua et al. (2015), where a ‘divide and conquer’ technique was employed to breakdown huge metagenomic data into subsets for the discovery of uncultured and/or rare taxa microbes and their associated functions. Another multi-omic pipeline IMP, a reproducible and modular analysis tool, which can integrate metagenomic and metatranscriptomic data was also developed (Narayanasamy et al. 2016).

Functional Mapping and Analysis Pipeline (FMAP) was developed that incorporates alignment of reads with reference databases, operon-level and pathways analysis and also the ability to process large data sets (Kim et al. 2016). SqueezeMeta developed by Tamames and Puente-Sanchez (2019) is an assembly-based bioinformatic pipeline dedicated to metagenomic/metatranscriptomic analysis and has a ‘fail-safe’ approach of assembly of metagenomes with internal checks and also has the feature for estimating the abundances of the genes in a metagenome. It also has the ability of retrieving the rRNA sequences and integrates a taxonomic assignment function. A similar work-flow pipeline which performs differential expression of genes and combines RNA-Seq data with Whole Genome sequencing for referencing, is the Meta-Omics Software for Community Analysis (MOSCA) (Sequeira et al. 2019). Both of the assembly-based workflows, filters out the rRNA data from the reads and then assemble the transcripts before assigning functions to them and have the provision of estimating relative abundance of genes (Sequeira et al. 2019). The development of a metatranscriptomic pipeline combining functional assignment and comprehensive statistics is the COMAN (Comprehensive Metatranscriptomics Analysis), which analyzes reads by first removing non-coding RNA. This uses a convenient web server, which has detailed instructions and is suitable for biologists with limited expertise in computational biology. For a more detailed analysis and comparison between different metatranscriptomic pipelines, the authors suggest referring to the article by Shakya et al. (2019).

## Problems and challenges in metatranscriptomics

Every method for studying natural habitats seems to have their own benefits and short comings. The first limitation of metatranscriptomic study, which has already been dealt upon, is the instability of RNA molecules. Unlike DNA, RNA is prone to degradation by various sources once they are isolated, average half-lives being seconds to minutes (Iost and Dreyfus 1995). There can be variations in RNA stability depending on the type of soil, microbial species and nutrients present (Redon et al. 2005). To save the RNA in the collected sample, liquid nitrogen must be used to freeze the samples post-collection or to shift them to RNA protective solutions (e.g., RNeasy Lysis Buffer, Qiagen, USA). Isolation of RNA from soil is incredibly challenging as it involves inefficient cell lysis due to the tough matrix soil provides. Downstream processing becomes difficult due to co-precipitation of complex molecules which inhibit PCR reactions. Nucleic acids are easily adsorbed on soil particles, which makes extraction very difficult, especially with low pH

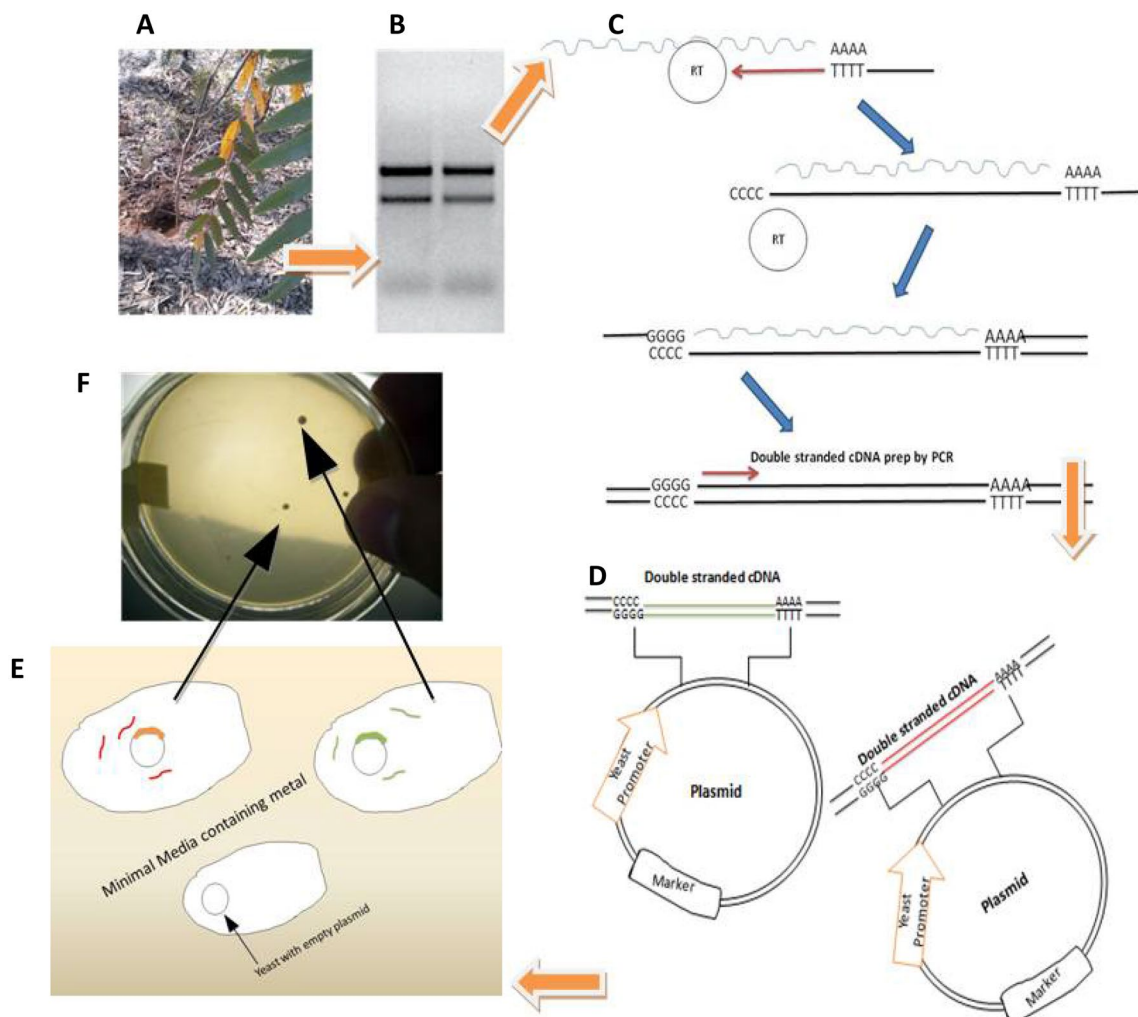


buffers, which specifically separate RNA from total nucleic acid. A detailed evaluation of challenges faced by scientists practicing metatranscriptomics as a tool to understand soil ecosystems has already been presented by Warnecke and Hess (2009) and Carvalhais et al. (2012).

A major step to bypass the mRNA enrichment stage and selectively amplify soil eukaryotic mRNA was first used by Bailly et al. (2007), where the technique of Switching Mechanism At the 5' end of RNA Template (SMART™) was used to synthesize cDNA. Traditional method of synthesizing cDNA has many shortcomings. They produce truncated cDNA due to premature termination of reverse transcription and as a result, most of the 5' terminal sequence information is lost resulting in underrepresentation of 5' end in cDNA library (Botero et al. 2005; Gubler and Hoffman

1983; D'Alessio and Gerard 1988). Moreover, use of ligation adapters for cloning cDNA generated undesirable byproducts (chimera).

SMART cDNA synthesis starts with as low as 2 nanogram of total RNA. A modified oligo(dT) primer containing specific restriction sites primes the first-strand synthesis reaction (Zhu et al. 2001). When the reverse transcriptase reaches the 5' end of the mRNA, the enzyme's terminal transferase activity adds a few additional nucleotides to the 3' end of the cDNA. The SMARTer Oligonucleotide base pairs with the non-template nucleotide stretch, create an extended template. The reverse transcriptase then switches templates and continues replicating to the end of the oligonucleotide. The overview of the process of metatranscriptomics from metal contaminated site is depicted in Fig. 5.



**Fig. 5** Overview of metatranscriptomics for mining heavy metal tolerant genes from Soil. **a, b** Soil collection and total RNA isolation **c** cDNA synthesis by SMART™, Reverse Transcriptase (RT) converts soil mRNA to cDNA. **d** Double stranded cDNA is cloned in plasmids having yeast promoter and suitable gene marker for selection in yeast

and *E. coli*. **e** Mutant/Auxotrophic yeast transformed with recombinant plasmids are able to survive metal stress due to the expression of plasmid-borne cDNA; control having empty plasmid cannot survive the stress. **f** Surviving yeast cells are selected on minimal media containing specific metal ions

The designed primer serves well as a priming site for the overall exponential amplification of the synthesized cDNA in contrast to incomplete cDNA or genomic DNA, which may otherwise be contaminating the cDNA library. However, the only downside of this method is the presence of truncated RNAs in poor quality RNA starting material which will eventually be amplified, and will contaminate the final cDNA library (Chenchik et al. 1998).

With the advancements in sequencing technologies, more and more data are being generated not only as metagenomes but also as gene expression data in the form of RNA-seq. However, the most important drawback of sequence analysis techniques is the deficiency of ample reference genomes, which results in inaccurate taxonomic as well as functional characterization. Also, the unavailability of experimental metadata to train the various pipelines is a major issue to understand the growing complexity of the deposited metatranscriptomic databases. The requirement of high-throughput in sequencing results in shorter reads, which are not deep enough to assign taxonomic and functional tags accurately. However, long-read sequencing (LRS) and de novo assembly of transcripts offer improvements in characterization of genes that are overlooked in existing NGS technologies (Clarke et al. 2009; Mantere et al. 2019).

The choice of bioinformatic pipelines is a major limitation as it is still unclear, as to which pipeline is the ideal for any kind of metatranscriptomic data. Most workflows merge various tools and techniques to achieve better analysis, but they lack a specific, benchmarked pipeline which can address all the problems together. Moreover, the bioinformatic pipelines are resource intensive and require a lot of computing as well as storage needs. Thus, development of workflows like the SqueezeMeta which is capable of running in small desktop/laptop computers and easy-to-use web-based pipelines like Galaxy, COMAN etc. are the need of the hour (Boekel et al. 2015; Ni et al. 2016).

## Concluding remarks

As discussed in this review, metatranscriptomics is a dependable and robust tool to discover novel functions and possible roles of microorganisms concerning soil ecosystem with an emphasis on eukaryotes. The survival strategies in heavy metal polluted soil have encouraged scientists to develop clear-cut holistic approaches to identify and discover new and underrepresented yet important microbes and functions. To understand the complex and dynamic nature of soil microbial community function and environmental impact due to metal contamination (Wasserman et al. 2008), it is essential to focus on the expressed subset of total genes and not work with the entire metagenome. Here metatranscriptomics helps us to understand the responses

of microbial population with environmental modification or perturbation, which will reveal the services performed pertaining to nutrient mobilization, breakdown of complex organic matter, imparting tolerance or resistance in response to heavy metals in environment etc. Sequence based discovery of participating organisms and their functional information from the RNA reveals a lot more than what isolation or cultivation-based techniques would have revealed. Heavy metal polluted soil can contain an enormous amount of information as stressed conditions trigger overexpression of a protein-coding genes (Mitchell et al. 2011) and increased activity of a community which would otherwise be dormant under ordinary conditions. This huge repository of information in the form of sequences is expanding due to high-throughput sequencing facilities that everybody has access to. The soil biota is well represented, more and more taxonomic annotations are assigned to previously unknown species especially eukaryotes. Protein-coding sequences of these environmental transcripts are validated through proper functional studies.

Mining and characterization of functional genes amongst millions of obtained sequences is a challenge faced by metatranscriptomics. Some of the problem is relieved by (i) Sizing the mRNA in suitable fractions (Yadav et al. 2014) (ii) Sequence capture by probe hybridization followed by activity screening for the desired function (Bragalini et al. 2014) and a more recent technique (iii) Ultra High-throughput screening by picodroplet microfluidics (Sjostrom et al. 2014; Marmeisse et al. 2017).

The eukaryotic genes recovered from polluted soil through metatranscriptomic approach thus can be used to study unknown taxonomic groups and their contribution towards metal tolerance. This will have great biotechnological importance such as biomarkers or organism based biosensors. It will also help us to manipulate the metal resistance level in plants which would be useful for remediation and revegetation of polluted soil (Khan 2005). Though these are lucrative options but they do not come hurdle-free. Finding suitable host for expression of environmental gene other than yeast can be a problem but there are ongoing studies carefully designed to express these genes in other living systems for example Damon et al. (2011) successfully introduced an environment oligopeptide transporter into *Xenopus* oocytes which could successfully transport dipeptides. Once they can be suitably expressed in different hosts, the novel genes and its products in the form of biocatalyst would thus open up numerous profitable paths in the area of biotechnology.

**Acknowledgements** Authors are thankful to Council of Scientific and Industrial Research, Govt. of India for sponsoring a research project on “Novel gene pool from copper polluted soil ecosystem using metatranscriptomic approach” under file No. 38(1425)/16/EMR-II.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

- Abubucker S, Segata N, Gol J, Schubert AM, Izard J, Cantarel BL, Rodriguez-Mueller B, Zucker J, Thiagarajan M, Henrissat B, White O (2012) Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput Biol* 8:e1002358. <https://doi.org/10.1371/journal.pcbi.1002358>
- Allison SD, Martiny JB (2008) Resistance, resilience, and redundancy in microbial communities. *Proc Natl Acad Sci USA* 105(Supplement 1):11512–11519. <https://doi.org/10.1073/pnas.0801925105>
- Azizan MS, Zamani AI, Stahmann KP, Ng CL (2016) Fungal metabolites and their industrial importance: a brief review. *Malays J Biochem Mol Biol* 19:15–23
- Bailly J, Fraissinet-Tachet L, Verner MC, Debaud JC, Lemaire M, Wésolowski-Louvel MR (2007) Soil eukaryotic functional diversity, a metatranscriptomic approach. *ISME J* 17:632–642. <https://doi.org/10.1038/ismej.2007.68>
- Bashiardes S, Zilberman-Schapira G, Elinav E (2016) Use of metatranscriptomics in microbiome research. *Bioinform Biol insights*. <https://doi.org/10.4137/BBI.S34610>
- Bending GD, Rodriguez-Cruz MS, Lincoln SD (2007) Fungicide impacts on microbial communities in soils with contrasting management histories. *Chemosphere* 69:82–88. <https://doi.org/10.1016/j.chemosphere.2007.04.042>
- Biddle JF, Fitz-Gibbon S, Schuster SC, Brenchley JE, House CH (2008) Metagenomic signatures of the Peru Margin subsurface biosphere show a genetically distinct environment. *Proc Natl Acad Sci USA* 105:10583–10588. <https://doi.org/10.1073/pnas.0709942105>
- Binz PA, Kägi JHR (1999) Metallothionein: molecular evolution and classification. In: Klaassen CD (ed) *Metallothionein IV. Advances in life sciences*. Birkhäuser, Basel, pp 7–13
- Blindauer CA (2013) Lessons on the critical interplay between zinc binding and protein structure and dynamics. *J Inorg Biochem* 121:145–155. <https://doi.org/10.1016/j.jinorgbio.2013.01.005>
- Boekel J, Chilton JM, Cooke IR, Horvatovich PL, Jagtap PD, Käll L, Lehtiö J, Lukasse P, Moerland PD, Griffin TJ (2015) Multi-omic data analysis using Galaxy. *Nat Biotechnol* 33:137–139. <https://doi.org/10.1038/nbt.3134>
- Botero LM, D'imperio S, Burr M, McDermott TR, Young M, Hassett DJ (2005) Poly (A) polymerase modification and reverse transcriptase PCR amplification of environmental RNA. *Appl Environ Microbiol* 71:1267–1275. <https://doi.org/10.1128/AEM.71.3.1267-1275.2005>
- Bragalini C, Ribiere C, Parisot N, Vallon L, Prudent E, Peyretailade E, Girlanda M, Peyret P, Marmeisse R, Luis P (2014) Solution hybrid selection capture for the recovery of functional full-length eukaryotic cDNAs from complex environmental samples. *DNA Res* 21:685–694. <https://doi.org/10.1093/dnares/dsu030>
- Brim H, Heyndrickx M, De Vos P, Wilmotte A, Springael D, Schlegel HG, Mergeay M (1999) Amplified rDNA restriction analysis and further genotypic characterisation of metal-resistant soil bacteria and related facultative hydrogenotrophs. *Syst Appl Microbiol* 22:258–268. [https://doi.org/10.1016/S0723-2020\(99\)80073-3](https://doi.org/10.1016/S0723-2020(99)80073-3)
- Capdevila M, Atrian S (2011) Metallothionein protein evolution: a mini-assay. *J Biol Inorg Chem* 16:977–989. <https://doi.org/10.1007/s00775-011-0798-3>
- Carvalho LC, Dennis PG, Tyson GW, Schenk PM (2012) Application of metatranscriptomics to soil environments. *J Microbiol Meth* 91:246–251. <https://doi.org/10.1016/j.mimet.2012.08.011>
- Cervantes C, Gutierrez-Corona F (1994) Copper resistance mechanisms in bacteria and fungi. *FEMS Microbiol Rev* 14:121–137. <https://doi.org/10.1111/j.1574-6976.1994.tb00083.x>
- Chao-Rong GE, Zhang QC (2011) Microbial community structure and enzyme activities in a sequence of copper-polluted soils. *Pedosphere* 21:164–169. [https://doi.org/10.1016/S1002-0160\(11\)60114-8](https://doi.org/10.1016/S1002-0160(11)60114-8)
- Chenckik A, Zhu YY, Diatchenko L, Li R, Hill J, Siebert PD (1998) Generation and use of high-quality cDNA from small amounts of total RNA by SMART PCR. In: Siebert PD, Larrick JW (eds) *RT-PCR Methods for gene cloning and analysis*. Eaton Publishing, Natick, MA, pp 305–319
- Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H (2009) Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol* 4:265–270. <https://doi.org/10.1038/nnano.2009.12>
- Clement BG, Kehl LE, DeBord KL, Kitts CL (1998) Terminal restriction fragment patterns (TRFPs), a rapid, PCR-based method for the comparison of complex bacterial communities. *J Microbiol Meth* 31:135–142. [https://doi.org/10.1016/S0167-7012\(97\)00105-X](https://doi.org/10.1016/S0167-7012(97)00105-X)
- Comtet-Marre S, Parisot N, Lepercq P, Chaucheyras-Durand F, Mosoni P, Peyretailade E, Bayat AR, Shingfield KJ, Peyret P, Forano E (2017) Metatranscriptomics reveals the active bacterial and eukaryotic fibrolytic communities in the rumen of dairy cow fed a mixed diet. *Front Microbiol* 8:67. <https://doi.org/10.3389/fmicb.2017.00067>
- D'Alessio JM, Gerard GF (1988) Second-strand cDNA synthesis with *E. coli* DNA polymerase I and RNase H: the fate of information at the mRNA 5' terminus and the effect of *E. coli* DNA ligase. *Nucleic Acids Res* 16:1999–2014. <https://doi.org/10.1093/nar/16.5.1999>
- Dai X, Tian Y, Li J, Su X, Wang X, Zhao S, Liu L, Luo Y, Liu D, Zheng H, Wang J, Dong Z, Hu S, Huang L (2015) Metatranscriptomic analyses of plant cell wall polysaccharide degradation by microorganisms in the cow rumen. *Appl Environ Microbiol* 81:1375–1386. <https://doi.org/10.1128/AEM.03682-14>
- Damon C, Vallon L, Zimmermann S, Haider MZ, Galeote V, Dequin S, Luis P, Fraissinet-Tachet L, Marmeisse R (2011) A novel fungal family of oligopeptide transporters identified by functional metatranscriptomics of soil eukaryotes. *ISME J* 5:1871–1880. <https://doi.org/10.1038/ismej.2011.67>
- Damon C, Lehenbre F, Oger-Desfeux C, Luis P, Ranger J, Fraissinet-Tachet L, Marmeisse R (2012) Metatranscriptomics reveals the diversity of genes expressed by eukaryotes in forest soils. *PLoS ONE* 7:e28967. <https://doi.org/10.1371/journal.pone.0028967>
- Daniel R (2002) Construction of environmental libraries for functional screening of enzyme activity. In: Brakmann S, Johnsson K (eds) *Directed molecular evolution of proteins: or how to improve enzymes for biocatalysis*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, pp 63–78. <https://doi.org/10.1002/3527600647.ch4>
- Daniel R (2005) The metagenomics of soil. *Nat Rev Microbiol* 3:470–478. <https://doi.org/10.1038/nrmicro1160>
- Degens BP, Schipper LA, Sparling GP, Duncan LC (2001) Is the microbial community in a soil with reduced catabolic diversity less resistant to stress or disturbance? *Soil Biol Biochem* 33:1143–1153. [https://doi.org/10.1016/S0038-0717\(01\)00018-9](https://doi.org/10.1016/S0038-0717(01)00018-9)



- Dell'Amico E, Mazzocchi M, Cavalca L, Allievi L, Andreoni V (2008) Assessment of bacterial community structure in a long-term copper-polluted ex-vineyard soil. *Microbiol Res* 163:671–683. <https://doi.org/10.1016/j.micres.2006.09.003>
- Deng W, Xi D, Mao H, Wanapat M (2008) The use of molecular techniques based on ribosomal RNA and DNA for rumen microbial ecosystem studies: a review. *Mol Biol Rep* 35:265–274. <https://doi.org/10.1007/s11033-007-9079-1>
- Dominati E, Patterson M, Mackay A (2010) A framework for classifying and quantifying the natural capital and ecosystem services of soils. *Ecol Eng* 69:1858–1868. <https://doi.org/10.1016/j.ecolecon.2010.05.002>
- Doran JW, Zeiss MR (2000) Soil health and sustainability: managing the biotic component of soil quality. *Appl Soil Ecol* 15:3–11. [https://doi.org/10.1016/S0929-1393\(00\)00067-6](https://doi.org/10.1016/S0929-1393(00)00067-6)
- Eichner CA, Erb RW, Timmis KN, Wagner-Döbler I (1999) Thermal gradient gel electrophoresis analysis of bioprotection from pollutant shocks in the activated sludge microbial community. *Appl Environ Microbiol* 65:102–109
- Ferrer M, Martínez-Martínez M, Bargiela R, Streit WR, Golyshina OV, Golyshin PN (2016) Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microb Biotechnol* 9:22–34. <https://doi.org/10.1111/1751-7915.12309>
- Fierer N, Breitbart M, Nulton J, Salamon P, Lozupone C, Jones R, Robeson M, Edwards RA, Felts B, Rayhawk S, Knight R (2007) Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Appl Environ Microbiol* 73:7059–7066. <https://doi.org/10.1128/AEM.00358-07>
- Gelsomino A, Keijzer-Wolters AC, Cacco G, van Elsas JD (1999) Assessment of bacterial community structure in soil by polymerase chain reaction and denaturing gradient gel electrophoresis. *J Microbiol Meth* 38:1–15. [https://doi.org/10.1016/S0167-7012\(99\)000548](https://doi.org/10.1016/S0167-7012(99)000548)
- Gill SR, Pop M, DeBoy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM, Nelson KE (2006) Metagenomic analysis of the human distal gut microbiome. *Science* 312:1355–1359. <https://doi.org/10.1126/science.1124234>
- Giller KE, Witter E, Mcgrath SP (1998) Toxicity of heavy metals to microorganisms and microbial processes in agricultural soils: a review. *Soil Biol Biochem* 30:1389–1414. [https://doi.org/10.1016/S0038-0717\(97\)00270-8](https://doi.org/10.1016/S0038-0717(97)00270-8)
- Goldmann K, Schöning I, Buscot F, Wubet T (2015) Forest management type influences diversity and community composition of soil fungi across temperate forest ecosystems. *Front Microbiol* 6:1300. <https://doi.org/10.3389/fmicb.2015.01300>
- Grant S, Grant WD, Cowan DA, Jones BE, Ma Y, Ventosa A, Heaphy S (2006) Identification of eukaryotic open reading frames in metagenomic cDNA libraries made from environmental samples. *Appl Environ Microbiol* 72:135–143. <https://doi.org/10.1128/AEM.72.1.135-143.2006>
- Gruninger RJ, Nguyen TTM, Reid ID, Yanke J, Wang P, Abbott DW, Tsang A, McAllister T (2018) Application of transcriptomics to compare the carbohydrate active enzymes that are expressed by diverse genera of anaerobic fungi to degrade plant cell wall carbohydrates. *Front Microbiol* 9:1581. <https://doi.org/10.3389/fmicb.2018.01581>
- Gubler U, Hoffman BJ (1983) A simple and very efficient method for generating cDNA libraries. *Gene* 25:263–269. [https://doi.org/10.1016/0378-1119\(83\)90230-5](https://doi.org/10.1016/0378-1119(83)90230-5)
- Gudynaite-Savitch L, White TC (2016) Fungal biotechnology for industrial enzyme production: focus on (Hemi) cellulase production strategies, advances and challenges. In: Schmoll M, Dattenböck C (eds) *Gene expression systems in fungi: advancements and applications*. Springer, Cham, pp 395–439. [https://doi.org/10.1007/978-3-319-27951-0\\_19](https://doi.org/10.1007/978-3-319-27951-0_19)
- Guerrero G, Hausman JF, Strauss J, Ertan H, Siddiqui KS (2016) Lignocellulosic biomass: biosynthesis, degradation, and industrial utilization. *Eng Life Sci* 16:1–16
- 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. [https://doi.org/10.1016/S1074-5521\(98\)90108-9](https://doi.org/10.1016/S1074-5521(98)90108-9)
- Hesse CN, Mueller RC, Vuyisich M, Gallegos-Graves LV, Gleasner CD, Zak DR, Kuske CR (2015) Forest floor community metatranscriptomes identify fungal and bacterial responses to N deposition in two maple forests. *Front Microbiol* 6:337. <https://doi.org/10.3389/fmicb.2015.00337>
- Hirsch PR, Mauchline TH, Clark IM (2010) Culture-independent molecular techniques for soil microbial ecology. *Soil Biol Biochem* 42:878–887. <https://doi.org/10.1016/j.soilbio.2010.02.019>
- Hua ZS, Han YJ, Chen LX, Liu J, Hu M, Li SJ, Kuang JL, Chain PS, Huang LN, Shu WS (2015) Ecological roles of dominant and rare prokaryotes in acid mine drainage revealed by metagenomics and metatranscriptomics. *ISME J* 9:1280. <https://doi.org/10.1038/ismej.2014.212>
- Iost I, Dreyfus M (1995) The stability of *Escherichia coli* lacZ mRNA depends upon the simultaneity of its synthesis and translation. *EMBO J* 14:3252
- Jarque S, Bittner M, Blaha L, Hilscherova K (2016) Yeast biosensors for detection of environmental pollutants: current state and limitations. *Trends Biotechnol* 34:408–419. <https://doi.org/10.1016/j.tibtech.2016.01.007>
- Jumpponen A, Johnson LC (2005) Can rDNA analyses of diverse fungal communities in soil and roots detect effects of environmental manipulations—a case study from tallgrass prairie. *Mycologia* 97:1177–1194. <https://doi.org/10.1080/15572536.2006.11832728>
- Kellner H, Luis P, Portetelle D, Vandenberg M (2011) Screening of a soil metatranscriptomic library by functional complementation of *Saccharomyces cerevisiae* mutants. *Microbiol Res* 166:360–368. <https://doi.org/10.1016/j.micres.2010.07.006>
- Khan AG (2005) Role of soil microbes in the rhizospheres of plants growing on trace metal contaminated soils in phytoremediation. *J Trace Elem Med Bio* 18:355–364. <https://doi.org/10.1016/j.jtemb.2005.02.006>
- Kim J, Kim MS, Koh AY, Xie Y, Zhan X (2016) FMAP: functional mapping and analysis pipeline for metagenomics and metatranscriptomics studies. *BMC Bioinform* 17:420. <https://doi.org/10.1186/s12859-016-1278-0>
- Knietsch A, Bowien S, Whited G, Gottschalk G, Daniel R (2003) Identification and characterization of coenzyme B12-dependent glycerol dehydratase and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. *Appl Environ Microbiol* 69:3048–3060. <https://doi.org/10.1128/AEM.69.6.3048-3060.2003>
- Koringa PG, Thakkar JR, Pandit RJ, Hinsu AT, Parekh MJ, Shah RK, Jakhesara SJ, Joshi CG (2019) Metagenomic characterisation of ruminal bacterial diversity in buffaloes from birth to adulthood using 16S rRNA gene amplicon sequencing. *Funct Integr Genomics* 19:237–247. <https://doi.org/10.1007/s10142-018-0640-x>
- Kozdroj J, van Elsas JD (2000) Response of the bacterial community to root exudates in soil polluted with heavy metals assessed by molecular and cultural approaches. *Soil Biol Biochem* 32:1405–1417. [https://doi.org/10.1016/S0038-0717\(00\)00058-4](https://doi.org/10.1016/S0038-0717(00)00058-4)
- Lara E, Berney C, Harms H, Chatzinotas A (2007) Cultivation-independent analysis reveals a shift in ciliate 18S rRNA gene diversity in a polycyclic aromatic hydrocarbon-polluted soil.



- FEMS Microbiol Ecol 62:365–373. <https://doi.org/10.1111/1j.1574-6941.2007.00387.x>
- Lavelle P, Decaens T, Aubert M, Barot S, Blouin M, Bureau F, Margerie P, Mora P, Rossi JP (2006) Soil invertebrates and ecosystem services. *Eur J Soil Biol* 42:S3–S15. <https://doi.org/10.1016/j.ejsobi.2006.10.002>
- Lehembre F, Doillon D, David E, Perrotto S, Baude J, Foulon J, Harfouche L, Vallon L, Poulain J, Da Silva C, Wincker P (2013) Soil metatranscriptomics for mining eukaryotic heavy metal resistance genes. *Environ Microbiol* 15:2829–2840. <https://doi.org/10.1111/1462-2920.12143>
- Lehmann M, Riedel K, Adler K, Kunze G (2000) Amperometric measurement of copper ions with a deputy substrate using a novel *Saccharomyces cerevisiae* sensor. *Biosens Bioelectron* 15:211–219. [https://doi.org/10.1016/S0956-5663\(00\)00060-9](https://doi.org/10.1016/S0956-5663(00)00060-9)
- Leimena MM, Ramiro-Garcia J, Davids M, van den Bogert B, Smidt H, Smid EJ, Boekhorst J, Zoetendal EG, Schaap PJ, Kleerebezem M (2013) A comprehensive metatranscriptome analysis pipeline and its validation using human small intestine microbiota datasets. *BMC Genom* 14:530. <https://doi.org/10.1186/1471-2164-14-530>
- Leininger S, Urlich T, Schlöter M, Schwark L, Qi J, Nicol GW, Prosser JI, Schuster SC, Schleper C (2006) Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* 442:806–809. <https://doi.org/10.1038/nature04983>
- Liles MR, Manske BF, Bintrim SB, Handelsman J, Goodman RM (2003) A census of rRNA genes and linked genomic sequences within a soil metagenomic library. *Appl Environ Microbiol* 69:2684–2691. <https://doi.org/10.1128/AEM.69.5.2684-2691.2003>
- Liu W, Stahl D (2007) Molecular approaches for the measurement of density, diversity, and phylogeny. In: Hurst C, Crawford R, Garland J, Lipson D, Mills A, Stetzenbach L (eds) *Manual of environmental microbiology*, 3rd edn. ASM Press, Washington DC, pp 139–156. <https://dx.doi.org/10.1128/9781555815882.ch12>
- Liu WT, Marsh TL, Cheng H, Forney LJ (1997) Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl Environ Microbiol* 63:4516–4522
- Luis P, Kellner H, Martin F, Buscot F (2005) A molecular method to evaluate basidiomycete laccase gene expression in forest soils. *Geoderma* 128:18–27. <https://doi.org/10.1016/j.geoderma.2004.12.023>
- Ma A, Sun M, McDermid A, Liu B, Ma Q (2019) MetaQUBIC: a computational pipeline for gene-level functional profiling of metagenome and metatranscriptome. *Bioinformatics* 35:4474–4477. <https://doi.org/10.1093/bioinformatics/btz414>
- Mantere T, Kersten S, Hoischen A (2019) Long-read sequencing emerging in medical genetics. *Front Genet* 10:426. <https://doi.org/10.3389/fgene.2019.00426>
- Marmeisse R, Kellner H, Fraissinet-Tachet L, Luis P (2017) Discovering protein-coding genes from the environment: time for the eukaryotes? *Trends Biotechnol* 35:824–835. <https://doi.org/10.1016/j.tibtech.2017.02.003>
- Marsh TL (1999) Terminal restriction fragment length polymorphism (T-RFLP): an emerging method for characterizing diversity among homologous populations of amplification products. *Curr Opin Microbiol* 2:323–327. [https://doi.org/10.1016/S1369-5274\(99\)80056-3](https://doi.org/10.1016/S1369-5274(99)80056-3)
- Martinez X, Pozuelo M, Pascal V, Campos D, Gut I, Gut M, Azpiroz F, Guarner F, Manichanh C (2016) MetaTrans: an open-source pipeline for metatranscriptomics. *Sci Rep* 6:26447. <https://doi.org/10.1038/srep26447>
- Matsuura H, Yamamoto Y, Muraoka M, Akaishi K, Hori Y, Uemura K, Tsuji N, Harada K, Hirata K, Bamba T, Miyasaka H (2013) Development of surface-engineered yeast cells displaying phytochelatin synthase and their application to cadmium biosensors by the combined use of pyrene-excimer fluorescence. *Biotechnol Prog* 29:1197–1202. <https://doi.org/10.1002/btpr.1789>
- Medini D, Serruto D, Parkhill J, Relman DA, Donati C, Moxon R, Falkow S, Rappuoli R (2008) Microbiology in the post-genomic era. *Nat Rev Microbiol* 6:419–430. <https://doi.org/10.1038/nrmicro1901>
- Mehra RK, Winge DR (1991) Metal ion resistance in fungi: molecular mechanisms and their regulated expression. *J Cell Biochem* 45:30–40. <https://doi.org/10.1002/jcb.240450109>
- Miller DN, Bryant JE, Madsen EL, Ghiorse WC (1999) Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Appl Environ Microbiol* 65:4715–4724
- Minet M, Dufour ME, Lacroute F (1992) Complementation of *Saccharomyces cerevisiae* auxotrophic mutants by *Arabidopsis thaliana* cDNAs. *Plant J* 2:417–422. <https://doi.org/10.1046/j.1365-313X.1992.t01-38-00999.x>
- Mitchell E, Frisbie S, Sarkar B (2011) Exposure to multiple metals from groundwater—a global crisis: geology, climate change, health effects, testing, and mitigation. *Metallomics* 3:874–908. <https://doi.org/10.1039/C1MT00052G>
- Moran MA (2009) Metatranscriptomics: eavesdropping on complex microbial communities—large-scale sequencing of mRNAs retrieved from natural communities provides insights into microbial activities and how they are regulated. *Microbe* 4:329. <https://doi.org/10.1128/microbe.4.329.1>
- Mukherjee A, Yadav R, Marmeisse R, Fraissinet-Tachet L, Reddy MS (2019a) Heavy metal hypertolerant eukaryotic aldehyde dehydrogenase isolated from metal contaminated soil by metatranscriptomics approach. *Biochimie* 160:183–192. <https://doi.org/10.1016/j.biochi.2019.03.010>
- Mukherjee A, Yadav R, Marmeisse R, Fraissinet-Tachet L, Reddy MS (2019b) Detoxification of toxic heavy metals by serine protease inhibitor isolated from polluted soil. *Int Biodeterior Biodegrad* 143:104718. <https://doi.org/10.1016/j.ibiod.2019.104718>
- Muyzer G, Smalla K (1998) Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie Van Leeuwenhoek* 73:127–141. <https://doi.org/10.1023/A:1000669317571>
- Nannipieri P, Ascher J, Ceccherini M, Landi L, Pietramellara G, Renella G (2003) Microbial diversity and soil functions. *Eur J Soil Sci* 54:655–670. <https://doi.org/10.1046/j.1351-0754.2003.0556.x>
- Narayanasamy S, Jarosz Y, Muller EE, Heintz-Buschart A, Herold M, Kaysen A, Laczny CC, Pinel N, May P, Wilmes P (2016) IMP: a pipeline for reproducible reference-independent integrated metagenomic and metatranscriptomic analyses. *Genome Biol* 17:260. <https://doi.org/10.1186/s13059-016-1116-8>
- Ni Y, Li J, Panagiotou G (2016) COMAN: a web server for comprehensive metatranscriptomics analysis. *BMC Genom* 17:622. <https://doi.org/10.1186/s12864-016-2964-z>
- Niehaus F, Gabor E, Wieland S, Siegert P, Maurer KH, Eck J (2011) Enzymes for the laundry industries: tapping the vast metagenomic pool of alkaline proteases. *Microb Biotechnol* 4:767–776. <https://doi.org/10.1111/j.1751-7915.2011.00279.x>
- Nüsslein K, Tiedje JM (1998) Characterization of the dominant and rare members of a young Hawaiian soil bacterial community with small-subunit ribosomal DNA amplified from DNA fractionated on the basis of its guanine and cytosine composition. *Appl Environ Microbiol* 64:1283–1289
- O'Brien HE, Parent JL, Jackson JA, Moncalvo JM, Vilgalys R (2005) Fungal community analysis by large-scale sequencing of environmental samples. *Appl Environ Microbiol* 71:5544–5550. <https://doi.org/10.1128/AEM.71.9.5544-5550.2005>
- Opel KL, Chung D, McCord BR (2010) A study of PCR inhibition mechanisms using real time PCR. *J Forensic Sci* 55:25–33. <https://doi.org/10.1111/j.1556-4029.2009.01245.x>

- Pace NR (1997) A molecular view of microbial diversity and the biosphere. *Science* 276:734–740. <https://doi.org/10.1126/science.276.5313.734>
- Patel DD, Patel AK, Parmar NR, Shah TM, Patel JB, Pandya PR, Joshi CG (2014) Microbial and carbohydrate active enzyme profile of buffalo rumen metagenome and their alteration in response to variation in the diet. *Gene* 545:88–94. <https://doi.org/10.1016/j.gene.2014.05.003>
- Pointing SB, Belnap J (2012) Microbial colonization and controls in dryland systems. *Nat Rev Microbiol* 10:551–562. <https://doi.org/10.1038/nrmicro2831>
- Poretzky RS, Bano N, Buchan A, LeClerc G, Kleikemper J, Pickering M, Pate WM, Moran MA, Hollibaugh JT (2005) Analysis of microbial gene transcripts in environmental samples. *Appl Environ Microbiol* 71:4121–4126. <https://doi.org/10.1128/AEM.71.7.4121-4126.2005>
- Radhika V, Milkevitch M, Audigé V, Proikas-Cezanne T, Dhanasekaran N (2005) Engineered *Saccharomyces cerevisiae* strain BioS-1, for the detection of water-borne toxic metal contaminants. *Biotechnol Bioeng* 90:29–35. <https://doi.org/10.1002/bit.20344>
- Ranjard L, Nazaret S, Gourbière F, Thioulouse J, Linet P, Richaume A (2000) A soil microscale study to reveal the heterogeneity of Hg (II) impact on indigenous bacteria by quantification of adapted phenotypes and analysis of community DNA fingerprints. *FEMS Microbiol Ecol* 31:107–115. <https://doi.org/10.1111/j.1574-6941.2000.tb00676.x>
- Redon E, Loubière P, Coccain-Bousquet M (2005) Role of mRNA stability during genome-wide adaptation of *Lactococcus lactis* to carbon starvation. *J Biol Chem* 280:36380–36385. <https://doi.org/10.1074/jbc.M506006200>
- Roane TM, Pepper IL (1999) Microbial responses to environmentally toxic cadmium. *Microb Ecol* 38:358–364. <https://doi.org/10.1007/s002489901001>
- Roda A, Roda B, Cevenini L, Michelini E, Mezzanotte L, Reschiglian P, Hakkila K, Virta M (2011) Analytical strategies for improving the robustness and reproducibility of bioluminescent microbial bioreporters. *Anal Bioanal Chem* 401:201–211. <https://doi.org/10.1007/s00216-011-5091-3>
- Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C, Tiong CL (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* 66:2541–2547. <https://doi.org/10.1128/AEM.66.6.2541-2547.2000>
- Rossmann AY (1998) Protocols for an all taxa biodiversity inventory of fungi in a Costa Rican conservation area. Parkway Publishers, Inc., USA. <https://doi.org/10.2307/3761287>
- Sandaa RA, Torsvik V, Enger Ø, Daee FL, Castberg T, Hahn D (1999) Analysis of bacterial communities in heavy metal-contaminated soils at different levels of resolution. *FEMS Microbiol Ecol* 30:237–251. <https://doi.org/10.1111/j.1574-6941.1999.tb00652.x>
- Scherens B, Goffeau A (2004) The uses of genome-wide yeast mutant collections. *Genome Biol* 5:229. <https://doi.org/10.1186/gb-2004-5-7-229>
- Schloss PD, Handelsman J (2003) Biotechnological prospects from metagenomics. *Curr Opin Biotechnol* 14:303–310. [https://doi.org/10.1016/S0958-1669\(03\)00067-3](https://doi.org/10.1016/S0958-1669(03)00067-3)
- Sequeira JC, Rocha M, Alves MM, Salvador AF (2019) MOSCA: an automated pipeline for integrated metagenomics and metatranscriptomics data analysis. *Practical Applications of Computational Biology and Bioinformatics, 12th International Conference (Springer International Publishing)*. [https://doi.org/10.1007/978-3-319-98702-6\\_2](https://doi.org/10.1007/978-3-319-98702-6_2)
- Seshadri R, Leahy SC, Attwood GT, Teh KH, Lambie SC, Cookson AL, Eloie-Fadros EA, Pavlopoulos GA, Hadjithomas M, Varghese NJ, Paez-Espino D (2018) Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nat Biotechnol* 36:359. <https://doi.org/10.1038/nbt.4110>
- Shakya M, Lo CC, Chain PS (2019) Advances and challenges in metatranscriptomic analysis. *Front Genet* 10:904. <https://doi.org/10.3389/fgene.2019.00904>
- Sharma R, Sharma PK (2018) Metatranscriptome sequencing and analysis of agriculture soil provided significant insights about the microbial community structure and function. *Ecol Genet Genomics* 6:9–15. <https://doi.org/10.1016/j.egg.2017.10.001>
- Shetty RS, Deo SK, Liu Y, Daunert S (2004) Fluorescence-based sensing system for copper using genetically engineered living yeast cells. *Biotechnol Bioeng* 88:664–670. <https://doi.org/10.1002/bit.20331>
- Singh DP, Prabha R, Gupta VK, Verma MK (2018) Metatranscriptome analysis deciphers multifunctional genes and enzymes linked with the degradation of aromatic compounds and pesticides in the wheat rhizosphere. *Front Microbiol* 9:1331. <https://doi.org/10.3389/fmicb.2018.01331>
- Sjostrom SL, Bai Y, Huang M, Liu Z, Nielsen J, Joensson HN, Svahn HA (2014) High-throughput screening for industrial enzyme production hosts by droplet microfluidics. *Lab Chip* 14:806–813. <https://doi.org/10.1039/c3lc51202a>
- Smit E, Leeflang P, Wernars K (1997) Detection of shifts in microbial community structure and diversity in soil caused by copper contamination using amplified ribosomal DNA restriction analysis. *FEMS Microbiol Ecol* 23:249–261. <https://doi.org/10.1111/j.1574-6941.1997.tb00407.x>
- Söllinger A, Tveit AT, Poulsen M, Noel SJ, Bengtsson M, Bernhardt J, Hellwing ALF, Lund P, Riedel K, Schleper C, Højberg O (2018) Holistic assessment of rumen microbiome dynamics through quantitative metatranscriptomics reveals multifunctional redundancy during key steps of anaerobic feed degradation. *MSystems* 3:e00038–e118. <https://doi.org/10.1128/mSystems.00038-18>
- Štursová M, Žifčáková L, Leigh MB, Burgess R, Baldrian P (2012) Cellulose utilization in forest litter and soil: identification of bacterial and fungal decomposers. *FEMS Microbiol Ecol* 80:735–746. <https://doi.org/10.1111/j.1574-6941.2012.01343.x>
- Takasaki K, Miura T, Kanno M, Tamaki H, Hanada S, Kamagata Y, Kimura N (2013) Discovery of glycoside hydrolase enzymes in an avicel-adapted forest soil fungal community by a metatranscriptomic approach. *PLoS ONE* 8:e55485. <https://doi.org/10.1371/journal.pone.0055485>
- Tamames J, Puente-Sanchez F (2019) SqueezeM, a highly portable, fully automatic metagenomic analysis pipeline. *Front Microbiol* 9:3349. <https://doi.org/10.3389/fmicb.2018.03349>
- Terry SA, Badhan A, Wang Y, Chaves AV, McAllister TA (2019) Fibre digestion by rumen microbiota—a review of recent metagenomic and metatranscriptomic studies. *Can J Anim Sci* 99:678–692. <https://doi.org/10.1139/cjas-2019-0024>
- Thakur B, Yadav R, Fraissinet-Tachet L, Marmesse R, Reddy MS (2018) Isolation of multi-metal tolerant ubiquitin fusion protein from metal polluted soil by metatranscriptomic approach. *J Microbiol Methods* 152:119–125. <https://doi.org/10.1016/j.mimet.2018.08.001>
- Thakur B, Yadav R, Vallon L, Marmesse R, Fraissinet-Tachet L, Reddy MS (2019) Multi-metal tolerance of von Willebrand factor type D domain isolated from metal contaminated site by metatranscriptomics approach. *Sci Total Environ* 661:432–440. <https://doi.org/10.1016/j.scitotenv.2019.01.201>
- Torsvik V, Sørheim R, Goksøyr J (1996) Total bacterial diversity in soil and sediment communities—a review. *J Ind Microbiol Biotechnol* 17:170–178. <https://doi.org/10.1007/BF01574690>
- Tringe SG, Rubin EM (2005) Metagenomics: DNA sequencing of environmental samples. *Nat Rev Genet* 6:805–814. <https://doi.org/10.1038/nrg1709>

- Urich T, Lanzén A, Qi J, Huson DH, Schleper C, Schuster SC (2008) Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS ONE* 3:e2527. <https://doi.org/10.1371/journal.pone.0002527>
- Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304:66–74. <https://doi.org/10.1126/science.1093857>
- Vopálenská I, Váchová L, Palková Z (2015) New biosensor for detection of copper ions in water based on immobilized genetically modified yeast cells. *Biosens Bioelectron* 72:160–167. <https://doi.org/10.1016/j.bios.2015.05.006>
- Warnecke F, Hess M (2009) A perspective: metatranscriptomics as a tool for the discovery of novel biocatalysts. *J Biotech* 142:91–95. <https://doi.org/10.1016/j.jbiotec.2009.03.022>
- Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, McHardy AC, Djordjevic G, Aboushadi N, Sorek R (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* 450:560–565. <https://doi.org/10.1038/nature06269>
- Wasserman GA, Liu X, Factor-Litvak P, Gardner JM, Graziano JH (2008) Developmental impacts of heavy metals and undernutrition. *Basic Clin Pharmacol Toxicol* 102:212–217. <https://doi.org/10.1111/j.1742-7843.2007.00187.x>
- Weidner S, Arnold W, Puhler A (1996) Diversity of uncultured microorganisms associated with the seagrass *Halophila stipulacea* estimated by restriction fragment length polymorphism analysis of PCR-amplified 16S rRNA genes. *Appl Environ Microbiol* 62:766–771
- Westreich ST, Kor I, Mills DA, Lemay DG (2016) SAMSA: a comprehensive metatranscriptome analysis pipeline. *BMC Bioinform* 17:399. <https://doi.org/10.1186/s12859-016-1270-8>
- Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, Stewart A, Sgouros J, Peat N, Hayles J, Baker S, Basham D (2002) The genome sequence of *Schizosaccharomyces pombe*. *Nature* 415:871–880. <https://doi.org/10.1038/nature724>
- Yadav RK, Barbi F, Ziller A, Luis P, Marmeisse R, Reddy MS, Fraissinet-Tachet L (2014) Construction of sized eukaryotic cDNA libraries using low input of total environmental metatranscriptomic RNA. *BMC Biotechnol* 14:80. <https://doi.org/10.1186/1472-6750-14-80>
- Zhang Y, Zhang X, Zhang H, He Q, Zhou Q, Su Z, Zhang C (2009) Responses of soil bacteria to long-term and short-term cadmium stress as revealed by microbial community analysis. *Bull Environ Contam Toxicol* 82:367–372. <https://doi.org/10.1007/s00128-008-9613-4>
- Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD (2001) Reverse transcriptase template switching: A SMART™ approach for full-length cDNA library construction. *Biotechniques* 30:892–897. <https://doi.org/10.2144/01304pf02>
- Ziller A, Yadav RK, Capdevila M, Reddy MS, Vallon L, Marmeisse R, Atrian S, Palacios Ó, Fraissinet-Tachet L (2017) Metagenomics analysis reveals a new metallothionein family: sequence and metal-binding features of new environmental cysteine-rich proteins. *J Inorg Biochem* 167:1–11. <https://doi.org/10.1016/j.jinorgbio.2016.11.017>