Metabarcoding and Metagenomics in Discovery of Strains of Interest

Subjects: Genetics
Created by: Souvik Kusari , Reuben Maghembe

Contemporary research on microbial diversity has endured enormous advances due to improved approaches with state-of-the-art technologies and methods incorporating the application of metabarcoding and metagenomics through next-generation sequencing. Metabarcoding and metagenomics entail comprehensive methods for the characterization of microbial communities in terms of diversity and taxonomic affiliations. While metabarcoding utilizes molecular markers conserved and shared across various taxonomic groups, metagenomics involves whole community genome sequencing, which allows the identification of individual species at the strain level. Metabarcoding markers such as 16S, 18S, or ITS, coupled to bioinformatics analyses of molecular operational taxonomic units (OTU) and amplicon sequence variants (ASV), give rise to identifiable groups of closely related members of the microbiome. This can aid in pinpointing groups of interest in the community for further analysis. Integrating metabarcoding with metagenomics offers an understanding of the community's ecological diversity at the molecular, biochemical, and system levels. Here we provide a short snapshot of the use and significance of metabarcoding and metagenomics in the discovery of strains of interest from nature's virtually inexhaustible resource of microorganisms.

This article is related to our exhaustive review entitled "*Omics for Bioprospecting and Drug Discovery from Bacteria and Microalgae*" published recently in the journal *Antibiotics*[1]. Focusing typically on bacteria and microalgae, in the review, we provide examples and summarize recent success-stories on the efficacy of metabarcoding and metagenomics in microbiome research geared towards the identification of strains of interest. We further explain and highlight the significance of other correlated "omics" approaches as well as bioinformatics tools required for valid strain identification, detailed characterization, and modulation.

Metabarcoding

Metabarcoding has emerged as one of the quickest and robust methods for microbiome analysis in free communities as well as in host organisms. In a target community, metabarcoding targets a conserved specific molecular marker for diversity as well as phylogenetic mapping of the microbes constituting the respective community. The small subunit (SSU) ribosomal RNA gene 16S for prokaryotic DNA, and 18S for eukaryotic DNA, are the frequently employed marker regions. Numerous studies ranging from the field to the clinical laboratory setting have ratified 16S as the broadly used molecular marker in studying microbiome diversities. The gene contains variable regions ranging from V1 to V9, which are targeted for various interests. Amplification of sub-regions such as V1-V3, V3-V4, V3-V5, or V4-V5, among others, have attained considerable application in the identification of operational taxonomic units (OTUs) and amplicon sequence variants (ASV) in microbiomes employing 16S rRNA sequencing. However, the substantially precise and extensively used sub-regions are V4 and V6. These have not only been used to identify numerous axenic species isolated from nature, but also in the characterization of the microbiome harboring a plethora of ecological niches. Sequencing technology and downstream analysis workflows requiring sequence reads, quality control, assembly, and annotation account for the prevalent challenges in next-generation sequencing (NGS). Therefore, it is crucially valuable to diversify markers, regions within a given marker sequence, as well as the choice of sequencing platforms. Besides, suitable analytical software and bioinformatics tools play a significant role in NGS. For partial 16S rRNA gene sequencing from Illumina sequencing technology, Mothur (https://nephele.niaid.nih.gov/details_mothur/) has been one of the commonly used pipelines for microbiome analysis, integrating several variable options for OUT picking. Other pipelines at the OTU-level include QIIME-uclust and USEARCH-UPARSE. For ASV, the currently used packages typically include DADA2, Qiime2-Deblur, and USEARCH-UNOISE3, displaying variable sensitivities and specificities. However, despite the rapid development of software packages with user-friendly

interfaces and accuracy, metabarcoding by partial sequencing cannot always lead to adequate community characterization and occasionally taxonomic assignment of some closely related taxa. Dramatic advents in molecular taxonomy with high throughput sequencing have furthered metabarcoding to effective analysis with high-resolution through sequencing of the entire 16S rRNA gene. This has yielded robust taxonomic placement and characterization of diverse microbiomes. Pyrosequencing of the full-length gene (ranging from 1500 – 1700 bp) coupled to extensive downstream bioinformatics has ratified OTUs closer to the species down to strain level compared to the usual partial sequencing of sub-regions.

The internal transcribed spacer (ITS) is another marker for metabarcoding, which is located between the 16S and 23S RNA genes in prokaryotes and is used as a target marker for intragenomic variations. ITS1 and ITS2 are the two types of ITS markers in eukaryotes. While ITS1 is principally situated between 18S and 5.8S-rRNA, ITS2 resides between the 5.8S-rRNA and 26S regions, or between 18S and 28S for opisthokonts. Recently, up to 32 microalgal strains from culture to different taxa and concomitant screening for their ice nucleation active (INA) compounds was achieved by using these markers. ITS has been the most potent marker for the characterization of fungi, including soil and endophytic genera.

ITS can be pooled with additional markers for attaining a better resolution. In a recent barcoding study of freshwater green microalgae, ITS1 and ITS2 of the nuclear rRNA gene (nuITS1 and nuITS2) were combined with the ribulose bisphosphate carboxylase large (*rbcl*) subunit gene, which established dominant resolution in the screening of the microalgae. The identification of the cyanobacterium *Arthrospira* and green microalga *Dunaliella* has also been guided by *Rbcl*. In another study, 36 strains of green microalgae were identified by 18S rRNA sequencing and were clustered into their respective genera, which guided further analysis of relevant protein and lipid profiles. Similar to 16S, full-length 18S sequence retrieval has been shown to offer a comprehensive characterization of community members.

Taken together, an overview of community structure, composition, diversity, and taxonomic positions of different groups within and between diverse ecosystems are unraveled by metabarcoding. The identification and characterization of bacterial and microalgal communities can be streamlined by metabarcoding so that promising “bioproducts” can be targeted in a preliminary snapshot.

Even though metabarcoding is a swift and emergently inexpensive method, it has a considerably limited resolution and can barely distinguish closely-related species or strains. The polymerase chain reaction (PCR) short length sequencing associated with guanine-cytosine (GC) content bias sequencing errors, and the assignment of OTUs, poses significant challenges when employing metabarcoding. Metabarcoding is relatively restricted to at most the genus level, though it provides improved OTU picking. Thus far, metabarcoding cannot establish the molecular function of each microbe in an ecosystem. The structures and functions of several genes in a community remain untapped since metabarcoding targets only one portion of the metagenome that leaves some genes untapped. Bacterial and microalgal communities are typically complex; therefore, it is necessary to employ methods having extensive coverage in order to yield a holistic account of the communities towards bioprospecting and drug discovery.

Metagenomics

Metagenomics encompasses decoding information interlocked into the DNA of the entire microbial community in a target. Whole metagenome sequencing provides a more thorough discernment into community diversity and function than does metabarcoding. Notable progress in microbial community characterization is the generation of genome and protein databases. High-throughput sequencing of metagenomes is acquiescent to downstream analysis, the entire community structure analysis, comparative differences among ecosystems, precise descriptions of strains of interest as well as novel genes. Moreover, interactions among microbes in their natural environment or a laboratory setting, as well as the analysis of genes involved in several biochemical pathways, can be accomplished by this approach. Numerous studies have unraveled the molecular adaptation of the microorganisms to their environment (natural or artificial), using metagenomic screening, through cluster analysis and dereplicating their

metabolic links.

Modern NGS studies have recognized an expansive scope of bacterial genomes and their associated genes. For example, the full genome sequences of 74 strains belonging to seven orders of the phylum Cyanobacteria is available in the Cyanobacterial Knowledge Base (CKB). Together with other databases such as genome databases in the National Center for Biotechnology Information (NCBI), this database is central to structural and functional annotation of the bacteria of interest. The assortment of species or strains apposite for the goal of a given biotechnological study can be guided by genome annotation, based on molecular blueprints unraveled from the database before proceeding with microbial selection for cultivation and advanced screening.

Further reading

Ana V. Lasa; Antonio J. Fernández-González; Pablo J. Villadas; Nicolás Toro; Manuel Fernández-López; Metabarcoding reveals that rhizospheric microbiota of Quercus pyrenaica is composed by a relatively small number of bacterial taxa highly abundant. *Scientific Reports* **2019**, *9*, 1695, 10.1038/s41598-018-38123-z.

Andrei Prodan; Valentina Tremaroli; Harald Brolin; Aeilko H. Zwinderman; Max Nieuwdorp; Evgeni Levin; Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. *PLOS ONE* **2020**, *15*, e0227434, 10.1371/journal.pone.0227434.

Arul Prakasam Peter; Karthick Lakshmanan; Shylajanaciyar Mohandass; Sangeetha Varadharaj; Sivasudha Thilagar; Kaleel Ahamed Abdul Kareem; Prabaharan Dharmar; Subramanian Gopalakrishnan; Uma Lakshmanan; Cyanobacterial KnowledgeBase (CKB), a Compendium of Cyanobacterial Genomes and Proteomes. *PLOS ONE* **2015**, *10*, e0136262, 10.1371/journal.pone.0136262.

Cecile Monard; Stephan Gantner; Jan Stenlid; Utilizing ITS1 and ITS2 to study environmental fungal diversity using pyrosequencing. *FEMS Microbiology Ecology***2012**, *84*, 165-175, 10.1111/1574-6941.12046.

Chee Kuan Kwei; David Lewis; Keith King; William Donohue; Brett Neilan; Molecular classification of commercial Spirulina strains and identification of their sulfolipid biosynthesis genes. *Journal of Microbiology and Biotechnology* **2011**, *21*, 359–365, .

Denis L.J. Lafontaine; David Tollervey; The function and synthesis of ribosomes. *Nature Reviews Molecular Cell Biology* **2001**, *2*, 514-520, 10.1038/35080045.

Dominique Chã¨neby; Laurent Philippot; Alain Hartmann; Catherine Hã©Nault; Jean-Claude Germon; Dominique Chèneby; Catherine Hénault; 16S rDNA analysis for characterization of denitrifying bacteria isolated from three agricultural soils. *FEMS Microbiology Ecology* **2000**, *34*, 121-128, 10.1111/j.1574-6941.2000.tb00761.x.

Filipa Godoy-Vitorino; Josefina Romaguera; Chunyu Zhao; Daniela Vargas-Robles; Gilmary Ortiz Morales; Frances Vázquez-Sánchez; Maria Sanchez-Vázquez; Manuel De La Garza-Casillas; Magaly Martinez-Ferrer; James Robert White; et al. Cervicovaginal Fungi and Bacteria Associated With Cervical Intraepithelial Neoplasia and High-Risk Human Papillomavirus Infections in a Hispanic Population. *Frontiers in Microbiology* **2018**, *9*, 2533, 10.3389/fmicb.2018.02533.

Ines Krohn-Molt; Bernd Wemheuer; Malik Alawi; Anja Poehlein; Simon Güllert; Christel Schmeisser; Andreas Pommerening-Röser; Adam Grundhoff; Rolf Daniel; Dieter Hanelt; et al. Metagenome Survey of a Multispecies and Alga-Associated Biofilm Revealed Key Elements of Bacterial-Algal Interactions in Photobioreactors. *Applied and Environmental Microbiology* **2013**, *79*, 6196-6206, 10.1128/AEM.01641-13.

Jethro S. Johnson; Daniel J Spakowicz; Bo-Young Hong; Lauren M. Petersen; Patrick Demkowicz; Lei Chen; Shana R. Leopold; Blake M. Hanson; Hanako O. Agresta; Mark Gerstein; et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nature Communications* **2019**, *10*, 5029-

11, 10.1038/s41467-019-13036-1.

Juan Jovel; Jordan Patterson; Weiwei Wang; Naomi Hotte; Sandra O'keefe; Troy Mitchel; Troy Perry; Dina Kao; Andrew Mason; Karen L. Madsen; et al. Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics. *Frontiers in Microbiology* **2016**, *7*, 253, 10.3389/fmicb.2016.00459.

Krishna Preetha; Lijo John; Cherampillil Sukumaran Subin; Koyadan Kizhakkedath Vijayan; Phenotypic and genetic characterization of Dunaliella (Chlorophyta) from Indian salinas and their diversity. *Aquatic Biosystems* **2012**, *8*, 27-27, 10.1186/2046-9063-8-27.

Lo'Ai Alanagreh; Caitlin Pegg; Amritha Harikumar; Mark Buchheim; Assessing intragenomic variation of the internal transcribed spacer two: Adapting the Illumina metagenomics protocol. *PLOS ONE* **2017**, *12*, e0181491, 10.1371/journal.pone.0181491.

Matthew D. Di Guglielmo; Karl R. Franke; Courtney Cox; Erin L. Crowgey; Whole genome metagenomic analysis of the gut microbiome of differently fed infants identifies differences in microbial composition and functional genes, including an absent CRISPR/Cas9 gene in the formula-fed cohort. *Human Microbiome Journal* **2019**, *12*, 100057, 10.1016/j.humic.2019.100057.

Oscar Franzén; Jianzhong Hu; Xiuliang Bao; Steven H. Itzkowitz; Inga Peter; Ali Bashir; Improved OTU-picking using long-read 16S rRNA gene amplicon sequencing and generic hierarchical clustering. *Microbiome* **2015**, *3*, 43, 10.1186/s40168-015-0105-6.

Patel, A.; Chaudhary, S.; Syed, B.A.; Gami, B.; Patel, P.; Rbcl marker based approach for molecular identification of Arthrospira and Dunaliella isolates from non-axenic cultures. *J. Genet. Genet. Eng.* **2018**, *2*, 24-34.

Pedro González-Torres; Leszek P Pryszcz; Fernando Santos; Manuel Martinez-Garcia; Toni Gabaldón; Josefa Anton; Interactions between Closely Related Bacterial Strains Are Revealed by Deep Transcriptome Sequencing. *Applied and Environmental Microbiology* **2015**, *81*, 8445-8456, 10.1128/AEM.02690-15.

Sámed I. I. A. Hadi; Hugo Santana; Patrícia P. M. Brunale; Taísa G. Gomes; Márcia D. Oliveira; Alexandre Matthiensen; Marcos E. C. Oliveira; Flávia C. P. Silva; Bruno Brasil; DNA Barcoding Green Microalgae Isolated from Neotropical Inland Waters. *PLOS ONE* **2016**, *11*, e0149284, 10.1371/journal.pone.0149284.

Sophie Groendahl; Maria Kahlert; Patrick Fink; The best of both worlds: A combined approach for analyzing microalgal diversity via metabarcoding and morphology-based methods. *PLOS ONE* **2017**, *12*, e0172808, 10.1371/journal.pone.0172808.

Søren Michael Karst; Morten Simonsen Dueholm; Simon J. McIlroy; Rasmus Hansen Kirkegaard; Per Halkjær Nielsen; Mads Albertsen; Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene sequences without primer bias. *Nature Biotechnology* **2018**, *36*, 190-195, 10.1038/nbt.4045.

Sylvie Tesson; Tina Šantl-Temkiv; Ice Nucleation Activity and Aeolian Dispersal Success in Airborne and Aquatic Microalgae. *Frontiers in Microbiology* **2018**, *9*, e2681, 10.3389/fmicb.2018.02681.

Tuan Noraida Tuan Hamzah; Shiou Yih Lee; Asep Hidayat; Razak Terhem; Ibrahim Faridah-Hanum; Rozi Mohamed; Diversity and Characterization of Endophytic Fungi Isolated From the Tropical Mangrove Species, Rhizophora mucronata, and Identification of Potential Antagonists Against the Soil-Borne Fungus, Fusarium solani. *Frontiers in Microbiology* **2018**, *9*, 1707, 10.3389/fmicb.2018.01707.

Van Thang Duong; Faruq Ahmed; Skye R. Thomas-Hall; Simon Quigley; Ekaterina Nowak; Peer M. Schenk; High Protein- and High Lipid-Producing Microalgae from Northern Australia as Potential Feedstock for Animal Feed and Biodiesel. *Frontiers in Bioengineering and Biotechnology* **2015**, *3*, 53, 10.3389/fbioe.2015.00053.

Vímac Nolla-Ardèvol; Miriam Peces; Marc Strous; Halina E. Tegetmeyer; Metagenome from a Spirulina digesting biogas reactor: analysis via binning of contigs and classification of short reads. *BMC Microbiology* **2015**, *15*, 277, 10.1186/s12866-015-0615-1.

Yu. S. Bukin; Yu. P. Galachyants; I. V. Morozov; S. V. Bukin; A. S. Zakharenko; T. I. Zemskaya; The effect of 16S rRNA region choice on bacterial community metabarcoding results. *Scientific Data* **2019**, *6*, 190007, 10.1038/sdata.2019.7.

References

1. Reuben Maghembe; Donath Damian; Abdalah Makaranga; Stephen Nyandoro; Sylvester Lyantagaye; Souvik Kusari; Rajni Hatti-Kaul; Omics for Bioprospecting and Drug Discovery from Bacteria and Microalgae.. *Antibiotics* **2020**, *9*, 229, 10.3390/antibiotics9050229.

Keywords

Metabarcoding;metagenomics;microbiome;diversity;strains;bioinformatics