

REVIEW

**OPEN ACCESS**

Full open access to this and thousands of other papers at <http://www.la-press.com>.

## Sequencing and Computational Approaches to Identification and Characterization of Microbial Organisms

Brijesh Singh Yadav<sup>1,3</sup>, Venkateswarlu Ronda<sup>2,3</sup>, Dinesh P. Vashista<sup>1</sup> and Bhaskar Sharma<sup>3</sup>

<sup>1</sup>Department of Botany and Microbiology, H.N.B. Garhwal University, Srinagar, India. <sup>2</sup>Central Institute of Fisheries Technology, Cochin, India. <sup>3</sup>Division of Biochemistry, Indian Veterinary Research Institute, Izatnagar, India. Corresponding author email: [venkatvet@gmail.com](mailto:venkatvet@gmail.com)

---

**Abstract:** The recent advances in sequencing technologies and computational approaches are propelling scientists ever closer towards complete understanding of human-microbial interactions. The powerful sequencing platforms are rapidly producing huge amounts of nucleotide sequence data which are compiled into huge databases. This sequence data can be retrieved, assembled, and analyzed for identification of microbial pathogens and diagnosis of diseases. In this article, we present a commentary on how the metagenomics incorporated with microarray and new sequencing techniques are helping microbial detection and characterization.

**Keywords:** computational method, sequencing, microbial pathogen, diagnosis

---

*Biomedical Engineering and Computational Biology* 2013:5 43–49

doi: [10.4137/BECB.S10886](https://doi.org/10.4137/BECB.S10886)

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article published under the Creative Commons CC-BY-NC 3.0 license.



## Introduction

Microbes have evolved to survive in every type of condition on our planet, including human and animal bodies. Although many are not harmful, a few cause life threatening diseases. Traditionally these are identified by culturing in appropriate media and biochemical or serological testing. However, a large number of microbes have yet to be characterized than are known. They play vital roles in all of their ecosystems including animal bodies.<sup>1</sup> In spite of being extremely small, the sheer numbers of microbes living on the planet have large and critical effects on the cycling of nutrients and compounds essential for the survival of all organisms.<sup>2</sup> To survive in so many types of habitats, microbes have evolved a great number of mechanisms to find energy, digest food, and reproduce. These mechanisms are being applied in a number of ways in agriculture, energy production, medicine, and warfare. All this is possible by thousands of protein molecules and the nucleic acid sequences that encode them.<sup>3</sup>

Microbes are encountered in all walks of human life. The vast majority of the bacteria in the body are rendered harmless by the protective effects of the immune system and a few are indeed beneficial. The relationship between microbes and humans is delicate and complex. Ten times as many microbes live on or inside our body as we have cells. The microbes living in our digestive system break down food and produce useful vitamins. The millions of microbes that coat our skin and intestinal lumen form a protective barrier against more dangerous microbes.<sup>4</sup> In spite of the benefits, a relatively small number of microbes are harmful to humans. Many diseases and epidemics are caused by microbes including the plague during the middle Ages, smallpox, AIDS, influenza, food poisoning, and anthrax. These diseases result in severe illness or even death in humans. As scientists learn more about bacteria, fungi and viruses, they are better able to treat and prevent these diseases. Common treatments include antimicrobials which kill bacteria and fungi as well as vaccines that help the body fight off viruses.

## Traditional methods for detecting and identifying pathogens

The traditional methods for detecting and identifying pathogens require culturing of bacteria or

viruses and detecting them using phenotypic, biochemical, or serological tests. These methods have proven very successful for many types of microorganisms and are standardized for most of the known human and domesticated animal pathogens. However, one important limitation is that only a small fraction of the estimated number of microbial species have been described, and so are applicable to relatively a few species. Many microorganisms are very difficult to culture or may not grow at all and therefore they cannot be identified using traditional techniques. As many as 99% of all microbial species are estimated to fall into this category. This is particularly problematic when unknown diseases wreak havoc in human and animal populations.<sup>5</sup>

## Recent advances in identification and characterization of microbes

Over the last two decades, sequence analysis of conserved genes has become a reliable, accurate, inexpensive, and scalable method of microbial identification in health and environmental sciences. These advantages have resulted in routine use of sequencing methods to complement and sometimes replace traditional phenotypic methods of identification. Various molecular identification techniques have emerged offering speed combined with specific and sensitive detection. They are simple, rapid, reliable, and dependent on the presence of nucleic acids, both DNA and RNA, which code for proteins. These methods include polymerase chain reaction (PCR), DNA microarrays, metagenomic analysis, and next-generation sequencing, among others. Detection of DNA is now possible on a single molecule, and high-throughput analysis allows thousands of detection reactions to be performed at once, thus allowing a range of characteristics to be rapidly and simultaneously determined. Some of the recent molecular detection methods can be performed in the laboratory or clinical settings as well as at the farm site. Although some of these techniques provide immediate results, many require extensive computational approaches for analysis and interpretation of the data. The last two decades witnessed explosion in the genome sequencing of microbes and other life forms alike which lead to misleading results and inconclusive



interpretations. Therefore, the reintroduction of biologically inspired computational methods was needed to enhance the understanding of biological systems as information processing systems. Recent technologies like DNA microarray, metagenomics, and next generation sequencing are dependent on computational methods and algorithms for handling the sequence data.

## DNA Microarray Chip: Broad Range Microbial Detection

The DNA microarray, also called DNA chip or microarray chip, is one of the emerging research techniques used in the field of clinical microbiology for identification of human, animal, plant, and insect pathogens which are difficult to identify through the afore mentioned techniques. It is a sequence-based hybridization dependent pathogen identification method with massive multiplexing ability. With the availability of genome or partial genome sequences of almost all pathogens, genome based diagnostic methods such as PCR and real time PCR are finding increasing application in diagnosis. Both the Food and Agricultural Organization of the United Nations (FAO) and the United States Food and Drug Administration (USFDA) have approved many diagnostic assays based on PCR. One of the restrictions of PCR based methods is limited multiplexing capability though Mass Tag PCR has to some extent can overcome this liability. DNA microarrays offer more flexibility than PCR or other gene based methods as it has the capability to screen all the pathogens simultaneously.<sup>6</sup>

Many online oligonucleotide design programs are available for designing oligonucleotide probes. Most of them are available as freeware. eArray (<http://www.genomics.agilent.com>) is a web-based application for microarray probe designing. Roche NimbleGen (<http://www.nimblegen.com>) provides long-oligo probes (70 mers) and like eArray, has flexible design capability for advanced gene expression analysis. Primer3 (<http://www.broad.mit.edu/genomesoftware/other/primer3.html>) is commonly used software for designing primers and probes in the development of microarrays. In this program, probe selection is based on three criteria: oligonucleotide melting point, specificity to a single target or at least to the shortest list of possible targets, and

the inability to fold into a stable secondary structure at the hybridization temperature. Array Designer (<http://www.premierbiosoft.com/dnamicroarray/dnamicroarray.html>) and Visual OMP (<http://www.dnasoftware.com/vo-microcase.html>) are also softwares for probe designing which are optimized to maximize the specificity of the probes.

The microarray chip used for diagnostic purposes contains oligonucleotides specific to target pathogens. When attached to a glass substrate, these oligonucleotides are called features or probes and contain picomoles of DNA sequences. The probes designed for diagnostic assays should be unique to a specific pathogen. Moreover, they should not bind to all the other pathogen genomes, host genome, and other non-specific genome sequences present in clinical samples. Achieving this specificity requires computationally extensive comparison of target genomes with all known non-target sequences.<sup>7</sup> The number of DNA spots can be many thousands or even hundreds of thousands. These probes are used to interrogate target (labeled cDNA or cRNA prepared from clinical samples) sequences under high-stringency hybridization conditions. Probe-target hybridization is detected and quantified by fluorescence-based detection of fluorophore-labeled targets to determine relative abundance of nucleic acid sequences in the target. Specific hybridization makes detection of disease causing pathogens possible with high speed, sensitivity, and specificity. DNA microarrays have the capability to screen all the known pathogens and also yet to be identified pathogens simultaneously for control and prevention of microbial diseases.<sup>8</sup> For example, the ViroChip was the first microarray chip designed for broad range of viral pathogens. This chip identified the then unknown severe acute respiratory syndrome (SARS) virus as a coronavirus.<sup>9</sup> GreeneChipPm that contained approximately 30,000 probes targeting vertebrate viruses and rRNA sequences of fungi, bacteria, and protozoa successfully identified viruses at the species level and was used to implicate *Plasmodium falciparum* for an unexplained death.<sup>10</sup> The most comprehensive microarray chip containing almost all the published sequences of viruses was recently developed at LLNL (Lawrence Livermore National Laboratory) and is called Pan-Microbial Detection Array.<sup>11</sup> PathoChip™ is a microarray chip for the detection of 44 highly prevalent and fastidious



pathogenic bacteria. It was used for screening a variety of clinical isolates collected from blood, sputum, stool, cerebral spinal fluid, pus, and urine to evaluate the technique.<sup>12</sup> Leblanc and co-researchers<sup>13</sup> reported a novel method of magnetic bead microarray for the rapid detection and identification of the four recognized species in the pestivirus genus of the Flaviviridae family (ie, classical swine fever virus, border disease virus, BVDV1 and 2) which allowed specific and sensitive virus detection. They concluded that based on the simplicity of the assay, the protocols for hybridization and magnetic bead detection offer an emerging application for molecular diagnoses in virology that is amenable for use in modestly equipped laboratories.

### **Metagenomics—Panmicrobial Analysis**

Metagenomics has been recently introduced to study the genomic content of an environmental sample of microbes. It is a derivation of conventional microbial genomics, with the key difference being that it bypasses the requirement for obtaining pure cultures for sequencing. Since the samples are obtained from communities rather than isolated populations, metagenomics may serve to establish hypotheses concerning interactions between microbial community members. This process begins with sample and metadata collection and proceeds to DNA extraction, library construction, sequencing, read preprocessing, and assembly. Community composition analysis is employed at several stages of this workflow, and databases and computational tools are used to facilitate the analysis. Advances in throughput and cost-efficiency of sequencing technology are fueling a rapid increase in the number and size of metagenomic datasets being generated. However, bioinformaticists are faced with the problem of how to handle and analyze these datasets in an efficient and useful way.<sup>14</sup>

Information from metagenomic studies will be fully exploited only if appropriate data-management and data-analysis methods are in place. One requirement is that the data should be immediately accessible in a form suitable for computer analysis; another is that it be freely available without impediment to any researchers, be they in academia or industry. The three nucleic acid sequence archives GenBank, EMBL-Bank, and DDBJ have spearheaded the cause of free

availability of sequence information. In the process, sequences of a large numbers of fragments have been registered in the international DNA databanks. However, the details of function of the sequence are not available and are of limited use. Analysis and comparison of complex metagenomic data is driving the development of a new class of bioinformatics and visualization software. The field is moving forward rapidly, driven by enormous improvements in sequencing technology and the availability of many complementary technologies. Analysis and clustering of metagenomic sequences with the help of bioinformatics tools according to phenotypes and genomes might in future help in environmental preservation.<sup>15</sup> Once DNA sequence data are generated, sequences must be analyzed with special considerations in mind to facilitate accurate bacterial identification. First, different taxonomic classifications can be used for identification, and different species identifications may be generated depending on the taxonomic scheme.

Bioinformatics resources, such as MEGAN (<http://ab.inf.uni-tuebingen.de/software/megan/>), allow analysis of large metagenomic data sets using laptop computers. In a preprocessing step, the set of DNA sequences is compared against databases of known sequences using BLAST or another comparison tool. MEGAN is then used to compute and explore the taxonomical content of the data set, employing the NCBI taxonomy to summarize and order the results. The metagenomics RAST server (<http://metagenomics.nmpdr.org>) is an open-source metagenomics service providing a new paradigm for the annotation and analysis of metagenomes. With built-in support for multiple data sources and a back end that houses abstract data types, the metagenomics RAST is stable, extensible, and freely available to all researchers. MetaABC (<http://bits2.iis.sinica.edu.tw/MetaABC/>) is a metagenomic platform that integrates several binning tools coupled with methods for removing artifacts, analyzing unassigned reads, and controlling sampling biases. It allows users to arrive at a better interpretation via series of distinct combinations of analysis tools. After execution, MetaABC provides outputs in various visual formats such as tables, pie and bar charts as well as clustering result diagrams. ERGO (<http://ergo.integratedgenomics.com/ERGO/>) version of the ERGOTM database contains 618 complete or



nearly complete genomes, of which 319 are bacteria, 116 are eukarya, 34 are archaea, and 149 are viruses. In total, these genomes contain over 1,300,000 Open Reading Frames (ORFs), more than 60% of which have a functional annotation. This percentage of annotated genes is actually much higher for the bacterial genomes, reaching an average of 70%. Every genome that goes into the ERGO system is annotated from scratch, whether it has been sequenced at Integrated Genomics or at another sequencing center. More than 450 of the genomes are available for subscription or as part of a stand-alone ERGO server package from Integrated Genomics. IMG/M (<http://img.jgi.doe.gov/m>) is an integrated microbial genomes and metagenomes (IMG/M) system providing support for comparative analysis of microbial community aggregate genomes (metagenomes) in a comprehensive integrated context.<sup>16</sup> CAMERA (<http://camera.calit2.net>) database includes environmental metagenomic and genomic sequence data, associated environmental parameters (metadata), precomputed search results, and software tools to support powerful cross-analysis of environmental samples.

Recently a number of important issues have emerged with respect to metagenomics analysis of microbes. As an example, the colonic microbiota is a vast ecosystem with approximately 800–1000 species per individual, but these estimates are rapidly undergoing revision because the science of metagenomics and microbial pan-arrays is so new. Approximately 62% of the bacteria identified from the human intestine were previously not known and 80% of the bacteria identified by metagenomic sequencing were considered not cultivable using existing techniques. To date, fewer than 20 *Lactobacillus* species have been found consistently in the mammalian gastrointestinal tract. These findings indicate that membership in indigenous communities is restricted to a limited subset of all bacteria and that bacterial populations are not randomly distributed in and on the human body. Preliminary studies suggest that the predominant species in the genitourinary tract and on skin sites are fundamentally different from the populations predominant in the gastrointestinal tract.<sup>17</sup> Extensive studies, with respect to the regional microbiomes of the healthy and diseased individuals using sequencing and computational approaches, might provide some clues regarding the role of microbial communities in health and disease.

## Next-Generation Sequencing: Gateway to the Future

Sequencing is one technique that transformed biology from qualitative to a quantitative science and led to the emergence of bioinformatics as an important discipline. Initially, sequencing started with radio-isotope labeled sequencing products analyzed on slab gels. This slow process was overtaken by fluorescent labeling and capillary electrophoresis that improved speed and data quality of sequencing. Recently the next generation sequencing platforms have made possible massively parallel sequencing without the need for lengthy electrophoresis. There are two different approaches for next-generation sequencing. One utilizes sequencing by hybridization while the other uses sequencing by synthesis. Both of these approaches are superior in their ability to perform the task on massively parallel arrays that is made possible by high performance lasers and high resolution detectors. For example, the recent tSMS (true Single Molecule Sequencing-Helicos Biosciences) utilizes bright fluorescent labels and high resolution cameras that can detect fluorescence from single nucleotide incorporation. Similarly, SMRT (Single Molecule Real Time) sequencing technology developed by Pacific Biosciences utilizes an innovative zero-mode waveguide (ZMW) arrays to obtain very high signal to noise ratio for capturing fluorescence from single nucleotide incorporation. All the above mentioned sequencing technologies utilize complementary base pairing and enzymes to read the sequence. Recently, new technologies based on physical sequencing, like nanopore sequencing (Oxford Nanopore Technologies) and a single molecule DNA image based sequencing that is claimed to generate 20 kb plus reads (ZS Genetics) are under development. These diverse approaches and sophistication of next-generation sequencing has brought great challenges for bioinformaticists to tackle alignment, sequence scoring, data assembly, storage, and release of huge amounts of data.<sup>18</sup>

The ability to simultaneously acquire huge amounts of sequence data when applied to clinical and environmental samples helps in identification of pathogenic microbes. Moreover, genome variability and evolution within the host can be tracked over short periods of time. These approaches were already being used in diagnostic virology for detection of novel pathogenic viruses and for mapping of resistance to antiviral drugs.<sup>19</sup>



In recent years, several pathogens of veterinary importance have been sequenced world over. Whole-genome sequencing of microbes has revolutionized the methods by which these organisms are studied and has heightened expectations regarding the ability to predict potential targets for antimicrobial agents and vaccines. It is now possible to sequence entire bacterial and viral genomes or sample entire transcriptomes more efficiently and in greater depth than ever before. It is less expensive, quicker, and more efficient to access gene sequences by whole genome sequencing than traditional gene-by-gene approaches. It is desirable to sequence hundreds or even thousands of related genomes to sample genetic diversity within and between bacterial and viral populations. Molecular epidemiology using whole genome sequences of pathogens will reveal more precise phylogenetic relationships as compared to gene or partial sequences, thus giving an exact picture of geographical and evolutionary origin of the bacterial or viral isolates. The number of complete genomes of viral/bacterial pathogens has increased dramatically in recent years with submission of enormous sequence data in sequence repositories such as GenBank. Currently, there are thousands of bacterial and viral whole genome sequences available in the public domain at NCBI.

Deep Sequencing is a method of NGS technology based on highly parallel DNA sequence analysis, yielding thousands to millions of sequence reads per run. The instruments, now typically used for identifying and analyzing pathogens, include the Roche 454 pyrosequencing system and the Illumina Genetic Analyzer. This method has been used to identify several novel viruses, for example hemorrhagic fever Lujovirus from South Africa, the Dandenong virus—an arenavirus associated with fatal disease in transplant recipients, the Merkel cell polyomavirus associated with a rare skin cancer, and several viruses associated with gastroenteritis such as cosavirus and klassevirus/salivirus. Because of the relatively short read lengths, DNA-pyrosequencing for microbial identification has focused on hypervariable regions within small ribosomal-subunit RNA genes, especially 16S rRNA genes. Specific hypervariable regions have preferentially been used to identify different classes of bacteria via pyrosequencing.<sup>20</sup>

## Database and Computational Approach Evolution

Since the first assembly of protein sequence database in 1960s, numerous databases for nucleic acids and proteins have been created. This was largely driven by ever evolving sequencing technologies and sophisticated computational programs for assembling, annotating, and retrieving. As the number of base pairs that can be sequenced in one single reaction reached billions of bases, the development of software programs for assembling and comparing were needed. The development of powerful alignment and annotation programs lead to identification of inaccurate sequence assemblies and helped to refine the existing databases. For example, retro-analysis of 202 published bacterial and viral metagenomes using a recently developed computational program DeconSeq, revealed the presence of human DNA contaminations in 64% of the metagenomes.<sup>21</sup> This will definitely help in improving the accuracy of draft genome sequences. With improvements in draft genome sequences, better computational approaches will be developed for accurate identification of pathogens. For example, the recent development of PathSeq, a comprehensive computational tool for the identification or discovery of microorganisms by deep sequencing of human tissue could accurately detect the positional integration of human papillomavirus (HPV) type 18 in HeLa cell lines.<sup>22</sup> This reciprocal improvement in databases and computational programs is going to bring us very close to accurate identification of microbes in any type of sample in the least amount of time ever possible.

## Conclusion

Microbes have evolved to survive in every type of condition on our planet including human and animal bodies. Although many are not harmful, a few cause life threatening diseases. Traditionally these are identified by culturing in appropriate media and biochemical or serological testing. However, more numbers of microbes have yet to be characterized than are known. The recent advances in sequencing technologies and computational approaches are raising the possibility of complete understanding of microbial and environmental interactions. The massive multiplexing ability of microarrays is already being used for diagnosis of viral disease in humans and animals. Metagenomic analysis is also flooding data



banks with uncharacterized sequence information. Besides these, the powerful next generation sequencing platform is rapidly transforming the landscape of microbial identification and characterization. New computational approaches are being developed for analysis of these huge databases for accurate identification of pathogens using sequence information. The reciprocal improvements in quality of databases and computational approaches are going to deliver tools for accurate identification and characterization of microbes in future.

### Author Contributions

Wrote the first draft of the manuscript: BSY, VR. Contributed to the writing of the manuscript: BSY, VR, DPV, BS. Jointly developed the structure and arguments for the paper: BSY, VR, DPV, BS. Made critical revisions and approved final version: BS. All authors reviewed and approved the final manuscript.

### Funding

Author(s) disclose no funding sources.

### Competing Interests

Author(s) disclose no potential conflicts of interest.

### Disclosures and Ethics

As a requirement of publication the authors have provided signed confirmation of their compliance with ethical and legal obligations including but not limited to compliance with ICMJE authorship and competing interests guidelines, that the article is neither under consideration for publication nor published elsewhere, of their compliance with legal and ethical guidelines concerning human and animal research participants (if applicable), and that permission has been obtained for reproduction of any copyrighted material. This article was subject to blind, independent, expert peer review. The reviewers reported no competing interests. Provenance: the authors were invited to submit this paper.

### References

1. Hentges DJ. The anaerobic microflora of the human body. *Clin Infect Dis.* 1993;16(4):S175.
2. Jetten MS. The microbial nitrogen cycle. *Environ Microbiol.* 2008;10(11):2903–9.
3. Noel J. All About Microbes and Microbiology. microbes.org.
4. Harmon K. Bugs inside: what happens when the microbes that keep us healthy disappear? *Sci Amer.* 2009.
5. Lazcka O, Del Campo FJ, Munoz FX. Pathogen detection: a perspective of traditional methods and biosensors. *Biosens Bioelectron.* 2007;22(7):1205–17.
6. Kaderali L, Schliep A. Selecting signature oligonucleotides to identify organisms using DNA arrays. *Bioinformatics.* 2002;18(10):1340–9.
7. He Z, Wu L, L Xi, Fields MW, Zhou J. Empirical establishment of oligonucleotide probe design criteria. *Appl Environ Microbiol.* 2005;71:3753–60.
8. Wilson WJ, Strout CL, De Santis TZ, Stilwell JL, Carrano AV, Andersen GL. Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. *Mol Cell Probes.* 2002;16(2):119–27.
9. Wang D, Urisman A, Liu YT, et al. Viral discovery and sequence recovery using DNA microarrays. *PLoS Biol.* 2003;1(2):E2.
10. Quan PL, Palacios G, Jabado OJ, et al. Detection of respiratory viruses and subtype identification of influenza A viruses by GreeneChipResp oligonucleotide microarray. *J Clin Microbiol.* 2007;45(8):2359–64.
11. Gardner SN, Jaing CJ, McLoughlin KS, Slezak TR. A microbial detection array (MDA) for viral and bacterial detection. *BMC Genomics.* 2010;11:668.
12. Lee Y, Lee EK, Cho YW, et al. ProteoChip: a highly sensitive protein microarray prepared by a novel method of protein immobilization for application of protein-protein interaction studies. *Proteomics.* 2003;3(12):2289–304.
13. Leblanc N, Gantelius J, Schwenk JM, et al. Development of a magnetic bead microarray for simultaneous and simple detection of four pestiviruses. *J Virol Methods.* 2009;155(1):1–9.
14. Tringe SG, Rubin EM. Metagenomics: DNA sequencing of environmental samples. *Nature Rev Gen.* 2005;6(11):805–14.
15. Kunin V, Copeland A, Lapidus A, Mavromatis K, Hugenholtz P. A bioinformatician's guide to metagenomics. *Microbiol Mol Biol Rev.* 2008;72(4):557–78.
16. Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC. Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* 2011;21(9):1552–60.
17. Thies FL, Konig W, Konig B. Rapid characterization of the normal and disturbed vaginal microbiota by application of 16S rRNA gene terminal RFLP fingerprinting. *J Med Microbiol.* 2007;56(Pt 6):755–61.
18. Mardis ER. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet.* 2008;9:87–402.
19. Barzon L, Lavezzo E, Militello V, Toppo S, Palù G. Applications of next-generation sequencing technologies to diagnostic virology. *Int J Mol Sci.* 2011;12(11):7861–84.
20. Dolled-Filhart MP, Lee M Jr, Ou-Yang CW, Haraksingh RR, Lin JC. Computational and bioinformatics frameworks for next-generation whole exome and genome sequencing. *Scientific World Journal.* 2013;2013:730210.
21. Schmieder R, Edwards R. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS ONE.* 2011;6(3):e17288.
22. Kostic AD, Ojesina AI, Pedamallu CS, Jung J, Getz G, Meyerson M. PathSeq: A comprehensive computational tool for the identification or discovery of microorganisms by deep sequencing of human tissue. *Nat Biotechnol.* 2011;29(5):393–6.