We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

The Use of Bioinformatic Tools in Symbiosis and Co-Evolution Studies

Raúl Enrique Valle-Gough,

Blancka Yesenia Samaniego-Gámez,

Javier Eduardo Apodaca-Hernández
and Maria Leticia Arena-Ortiz

Abstract

Through millions of years, the multicellular organisms have coexisted and coevolved with the surrounding microorganisms, in an almost symbiotic relationship forming a complex entity known as holobiont. The composition and functions of these microbial communities were limited during many years to only a mere fraction, due to the use of culture-based techniques. The advent of molecular-based techniques allowed the identification of uncultured organisms in a culture-free manner. In recent years, the development of next generation sequencing techniques have allowed the high-throughput study of microbial communities allowing the identification and classification of otherwise uncultured microorganisms in a given environment, tissue or host through metagenomics. The next generation sequencing techniques have been used in the functional study of microbial assemblages and were able to identify the role of the microorganisms in biogeochemical cycles, pathogenic processes, metabolism and development, through metatranscriptomics. Taken together, the next generation sequencing based-studies have shown the existence of a complex metabolic network in different hosts and environments, with the microbial communities. This chapter will focus in different available bioinformatic tools that are suitable to study symbiosis and coevolution processes in a given sample.

Keywords: next generation sequencing, bioinformatics, metagenomics, metatranscriptomics, network analysis

1. Introduction

Higher multicellular organisms have coexisted and co-evolved with resident microorganisms in a relatively harmonious relationship over millions of years, forming a complex organism called holobiont. These processes of co-evolution have been documented by several studies carried out in the organization and composition of host microorganisms (microbiome) in different species [1–4]. The microbiome is currently considered a functional organ which is fundamental for the host organism, given that studies have shown that this organ is highly dynamic and adaptable, likewise plays an important role in physiological adaptation processes,

metabolism, and development [1, 2, 4–6]. The study of the role of the microbiome for years was limited to those organisms that could be susceptible to culture, the use of techniques based on molecular information (AFLP and RFLP) and denaturing gradient gel electrophoresis (DGGE) could reveal the presence of some species not cultured, but the gel resolution was being the main limitation since a single band could contain more than one sequence [7]. The development of genomic tools such as the new generation sequencing platforms (NGS) has allowed a better resolution of the diversity present in a given sample through what is known as metagenomics [2, 7–10]. These same sequencing platforms have allowed not only observing the diversity of a host or environment, but also the functional role of microorganisms as well as their possible interactions with physiological processes or biogeochemical cycles through metatranscriptomics [1, 2, 4, 9]. The understanding of the composition of microorganisms, functions, interactions, and other biological processes through NGS has been done through the development of different bioinformatics tools, which make use of large amounts of information and are able to compare them through different data bases [3, 9, 11]. The objective of this chapter is to provide information on the different existing bioinformatics tools that have been used in studies of co-evolution and symbiosis in different models.

2. Bioinformatic tools in metagenomics and metatranscriptomics in different samples

The first step of a NGS-based study involves the extraction of nucleic acids in sufficient quantity and quality to carry out the sequencing process in order to have an unbiased knowledge of the microbial diversity present in a sample [6, 12, 13]. The processing of DNA samples (environmental and host) can be performed by cell recovery by centrifugation gradients in differential media and the subsequent recovery of DNA by silica columns [6]. Another methodology used is the *in situ* lysis of the sample by the addition of enzymes (Proteinase K and lysozyme) with the subsequent separation of cell debris by centrifugation and recovery of DNA by solvent precipitation or by silica adsorption [14]. The main advantage of *in situ* lysis is that higher amounts of DNA are obtained when compared to cell recovery techniques; however, there is a risk of the presence of contaminants that may interfere with sequencing reactions [6].

In the case of RNA, the main methodologies perform *in situ* lysis of the sample under RNase-free conditions using different guanidine solvents and salts to avoid the presence of ribonucleases [15]. The samples should be placed at −80°C either in dry ice or liquid nitrogen to avoid their degradation.

Quality control of nucleic acids can be carried out by visualization on agarose gels, by spectrophotometric means (Nanodrop) and in microfluidic chambers (Bioanalyzers). This last system has been widely used since it allows the visualization and simultaneous quantification of nucleic acids [8, 16–18]. In the case of RNA, these systems have developed a scale known as RNA Integrity Number (RIN) which, based on the proportion between the major and minor subunits of the rRNA assigns a minimum value that must be greater than 8.0.

2.1 NGS shared tools for metagenomics and metatranscriptomics

Once a sample with sufficient quality and quantity was sent to sequencing, a series of files with the ".fastq" extension are obtained, which contains the information of the sequence and the quality for each base. This format is used by different programs (FASTQC and PRINSEQ) to perform the quality control of the

sequencing, showing basic statistics such as the total number of bases, read size, GC content, quality for each base in PHRED33 or PHRED64 scale, as well as the presence of overrepresented sequences [8, 19–23]. The files analyzed are introduced to different programs (Trimmomatic, TrimGalore, and CutAdapt) that trims the reads of the ".fastq" file, based on the quality for each nucleotide, eliminating sequences with a PHRED value below 20 and a minimum fragment size defined by the user [19, 20, 22, 24].

These programs are able to eliminate segments of initiators and sequencing adapters, which must be provided in a separate file. The output files of these programs are archives in ".fastq" format, where the sequences that are common for all samples are placed in one file, and the unique sequences for each individual sample are placed in several files [19, 20].

2.2 Metagenomic tools used in symbiosis and co-evolution studies

In recent years, the use of genomic approaches has revealed an unprecedented diversity and bacterial ubiquity in different types of samples (**Figure 1**), through the analysis of 16S ribosomal sequences [1, 2, 5, 6, 18, 19, 22, 25–27]. These techniques have allowed the molecular analysis of populations and how different biological processes have been established, controlled, and evolved [5, 28, 29].

The metagenomic composition analyzes have been carried out through the use of different programs (QIIME, QIIME2, and MOTHUR), that align the reads against a database of ribosomal genes (GreenGenes, SILVA, and RDP) and assign them operational taxonomic units (OTUs), using a distance of 3% and a confidence interval of 80% [29–33]. Once the OTUs have been assigned, the aforementioned programs allow the determination of diversity indices, richness, and main component analysis and perform the rarefaction of the samples [1, 2, 5, 19, 25, 29–32, 34, 35].

Other taxonomic classifiers are based on alignment of short sequences previously edited, by single or paired ends (Kraken, Kraken2, OneCodex) comparing them with the databases available in each program. In the case of Kraken, it makes
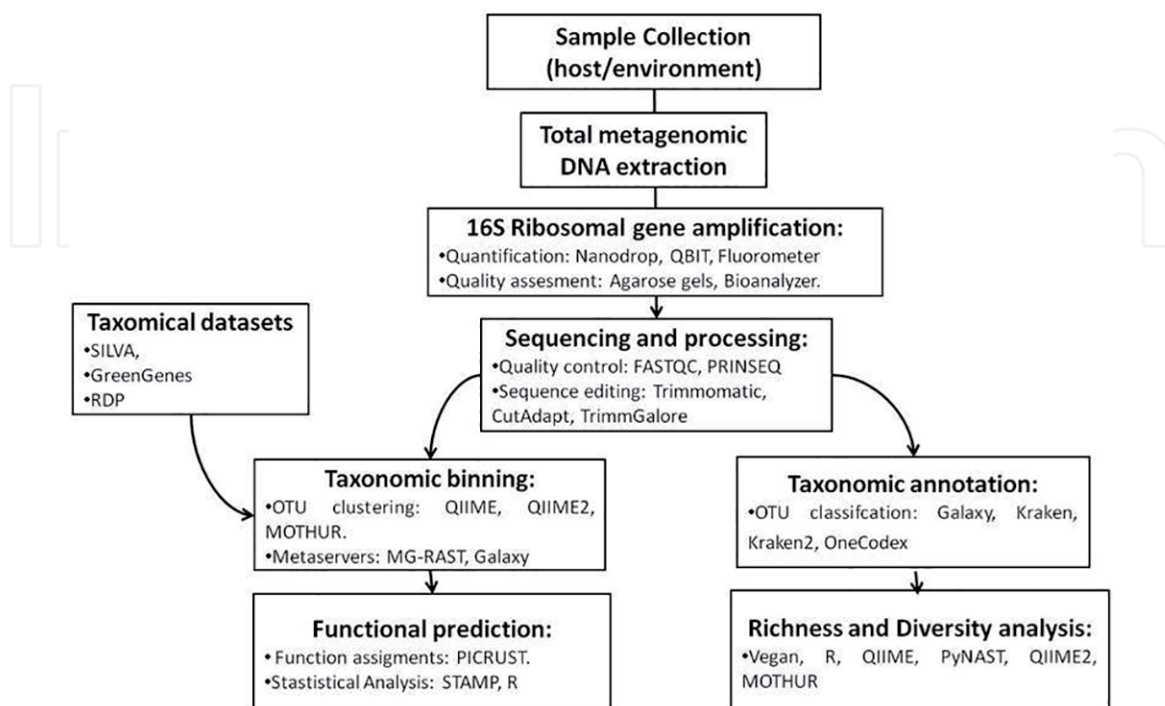

Figure 1.
Genomic and metagenomic techniques for the analysis in different samples.

use of the Ref-Seq database where the reads are divided into fragments known as k-mers and are compared with sequenced genomes [34, 36]. The resulting files of these programs are provided in tabular format (tsv), which facilitates their export and processing in other types of programs such as Vegan or R, where studies of richness, diversity, and rarefaction can be carried out [12, 34, 35, 37, 38].

The use of different taxonomic binning programs has been able to determine the presence of ubiquitous microbial phyla present in samples from arctic, temperate, and tropical environments such as: *Proteobacteria, Actinobacteria*, and *Cyanobacteria*, which are considered cosmopolite phyla. The main difference between each site is the proportion of each taxa, which reflects the conditions of each environment [6, 8–10, 39–41]. A similar behavior has been observed when studying the microbiome in different animal models where the phyla: *Proteobacteria, Acitnobacteria, Firmicutes*, and *Bacteroidetes* have been reported among those of greater relative abundance [4, 9, 39, 42–46]. This shows that microbial communities are highly dynamic where the physical-chemical factors of the site, health status, and nutrition shape the metagenome and can determine how reactive a microbial community is to environmental changes.

The use of genomic tools has made possible to identify the core microbiome of different organisms, given that, despite living in different habitats, they share similar bacterial communities, which implies the existence of biological filters that shape the bacterium-host interactions, resulting in a stable relationship with the holobiont [2, 28, 45–47]. In the case of *Apis mellifera*, a global core microbiome formed by *Proteobacteria, Firmicutes, Bacteroidetes,* and *Actinobacteria* has been identified, together with a high amount of lactic acid bacteria which have a beneficial activity in the health of the host organism due to their involvement in the immunomodulation of the intestinal microflora [25, 39, 48]. The presence of symbiont microorganisms within the intestinal tract in different animal species (*A. mellifera, Litopenaeus vannamei, Mus musculus, Homo sapiens*) have been reported as necessary for survival, since their cooperative behavior increases the vigor of a community [28, 39, 47, 49, 50]. Recent studies in fecal samples of farm animals have revealed the presence of intervening sequences (IVS), which are host-specific and provide a basis for the differentiation of the microorganisms derived from different hosts [51].

The role of microbial communities within a host is important. Given the existing delicate balance of these associations, any type of alterations in the microbiome composition could cause disease in the host organism [6, 12, 39, 45]. Previous studies have revealed that in diseased individuals of different species, the microbial diversity is significantly reduced. This could be due to the fact that alterations in the microbiome composition skew the association between the host and the microbiome producing dysbiosis and increasing the number of opportunistic pathogens [6, 12, 13, 16, 27, 39, 45, 52]. In marine environments, the continual presence of pathogens has been observed in environmental samples [16, 44] and in several marine organisms (*L. vannamei* and *M. nipponense*) [17, 18, 39, 45, 46]. The continual presence of pathogens in low proportion has been reported during the life cycle of these species, suggesting an active *in situ* infection in which the host has co-evolved with the parasitic organisms and developed mechanisms that cope with the pathogenic mechanisms of the parasites [13, 16, 17, 27, 45, 46, 52]. It has been observed that the developmental stage in *L. vannamei* influences the pathogenic response to *Vibrio*, where the proportion of protective commensal bacteria, *Bacteroides* and *Propionibacterium*, tend to decrease as the host aged in contrast the presence of *Vibrio* increases in diseased individuals [18]. Other mechanisms of coevolution have shown that processes of parasitism and predation can influence the global exchange of resources in an ecosystem. Studies conducted in *Escherichia coli* and the bacterial predator *Myxococcus xanthus* have shown that the genome evolution of the predator

and prey exhibited accelerated genome evolution when compared to controls, where the predator (*M. xanthus*) showed adaptations to cell mucoidy and the prey (*E. coli*) showed adaptations to outer membrane-proteases [7].

The functional analysis of the microbial communities has been carried out using the PICRUST program. This program estimates the families of genes present in a metagenome, by the phylogenetic comparison with sequences of gene families previously reported in databases. These predictions are pre-calculated for genes that code for proteins present in orthologous gene families (COG) or in the Kyoto Encyclopedia of Genes and Genomes (KEEG) [53]. The differential expression of these predicted functions could be assessed with the STAMP software which allows several statistical analysis, size effect, and sample corrections [54]. The use of the afore mentioned protocols have allowed the observation of various attributes in environmental samples related with carbon fixation, amino acid metabolism, and signal transduction in lakes, swamps, and other water bodies [9, 10, 16, 22, 44, 55]. These reports also showed the presence of several bacterial taxa (*Actinobacteria, Verrucomicrobia*, and *Proteobacteria*) who were able to synthesize several extracellular enzymes that digests the organic matter [9, 16, 24] or mineralize other nutrients [22, 44].

The influence of the microbiome on the host function have been proposed as a co-evolutionary process where the functionality and the composition of the microbiome can be influenced by the feeding habits of the host [4, 21], and the host can take advantage of the specialized microorganisms who are able to synthesize metabolites that are not present originally in the environment [6, 39]. The consumption of seaweeds by Japanese allows the introduction of algae associated bacteria, which transfer the genes involved in the degradation of the algal sulphated polysaccharides to competent gut resident bacteria with a process known as horizontal gene transfer [28]. Certain marine invertebrates (*Elysia chlorotica*) that feed on algae are able to maintain the algal plastids as photosynthetically symbionts which allow the use of photosynthates as food source [26]. These examples of coevolutionary processes show how the functionality of the microbiome could be influenced by the dietary habits of the host since; these metabolic add-ons allow the host to thrive in otherwise adverse environmental conditions (oligotrophic habitats).

2.3 Metatranscriptomic tools used in symbiosis and co-evolution studies

The metatranscriptomic allows the establishment of parallel relations between the host and the microbiome, but studies require a series of previous steps in order to obtain unbiased information such as the removal of rRNA and the microbial mRNA enrichment (**Figure 2**) [17, 19–21, 56].

The assembly of genomes and transcriptomes uses short sequences that are separated into fragments known as *k-mers*, which are aligned and compared graphically (De Brujin graphs) in order to perform *de novo* reconstruction of the genome or transcriptome. Several programs such as Velvet, SOAP, Trinity, and FLASH are capable of performing it by using a reference genome or transcriptome, if available [8, 57–63]. In the case of the Trinity platform, it is capable not only of assembling but also of mapping within the assembly (Bowtie1 and Bowtie2), basic statistical analysis of the assembly, quantifying transcripts (RSEM, Salmon, eXpress, and Kallisto), and performing differential expression of transcripts (edgeR, DESeq2, ROTS, and lima/voom) [8].

The metatranscriptomic studies have allowed to reveal the functions of the microorganisms within a host or in different environments and to identify, in both host and microbiome, transcripts related mainly to metabolic processes associated with the nutrient uptake. These observations suggest that the symbiotic chemoautotrophic bacteria provide organic compounds to the host organism that uses it for
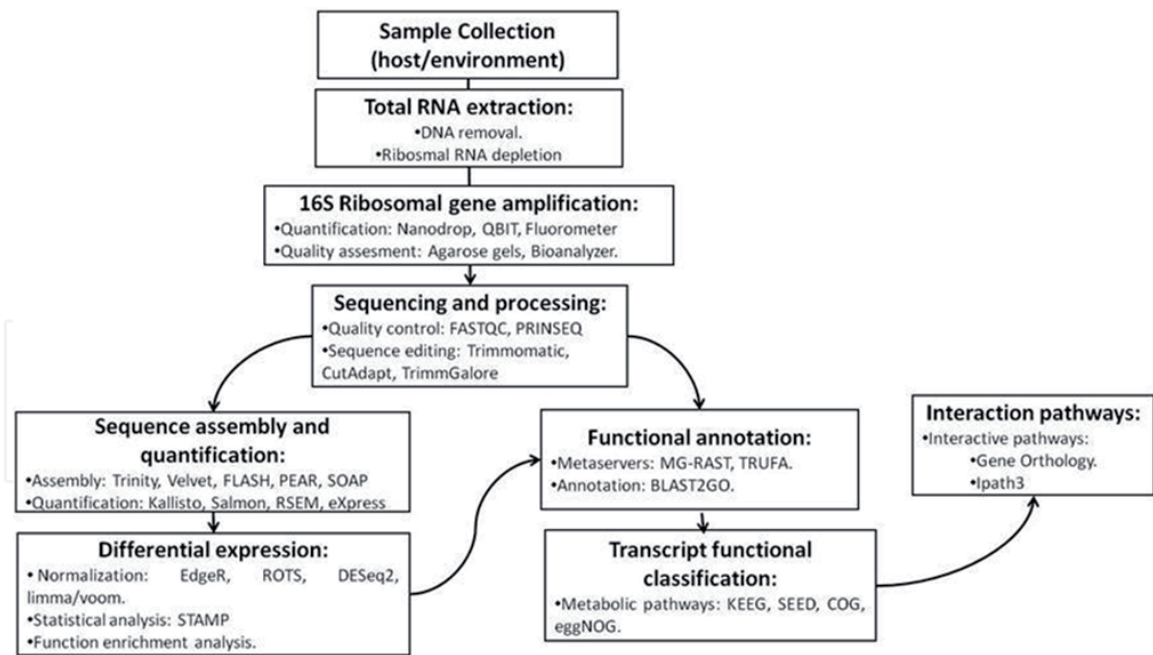
Figure 2.
Transcriptomic and metatranscriptomic techniques for the analysis in different samples.

its nutrition [11, 20, 45, 64]. In fact, recent studies have reported that more than a third of genes are shared among living organisms, especially to those related to the central metabolic pathways (Glucolysis, TCA, Oxidative phosphorylation, Purine and Pyrimidine metabolism) which could increase the efficiency for the digestion of several biomolecules [11, 17, 26, 45].

The use of bioinformatic tools in metatranscriptomics studies has allowed the visualization of the host-microbiome interactions, especially those related with the primary metabolism [13, 38, 52]. The visualization of the shared enzymatic modules is accomplished through the use of identifiers derived from KEGG orthology (KO) and Enzyme Codes (EC) on the iPath3 platform [65]. In this platform, it is possible to overlap metabolic functions (host-symbiont) using the EC and KO identifiers in different metabolic maps (general metabolic pathways, bacterial metabolism, and secondary metabolism), showing graphically the enzymatic modules of each individual and highlighting the enzymatic modules with a shared function.

Metatranscriptomic studies have been able to show that microorganisms are capable of generating complex trophic networks communicating with each other through chemical signals in a process known as quorum sensing [12, 26, 56]; however, this process is not restricted only to microorganisms; recent studies have suggested an interdominion quorum sensing [4, 17, 21].

2.4 Metaservers used in the metagenomic and metatranscriptomics studies

The bioinformatic tools mentioned in this chapter are open-source programs, requiring the user to have a UNIX or OSx operating system installed; a RAM memory greater than 16 GB, a hard disk greater than 500 GB of storage and knowledge about command lines in UNIX [16, 19, 20, 22, 24, 30–32, 35, 46, 48, 51]. These requirements can be complicated for those who want to initiate a bionformatic analysis; however, there are other options, such as the metaservers, which can allow the data processing in a graphical environment.

The metaservers are web service providers that assemble a series of programs and applications that otherwise are dispersed. Among the most used metaservers are Galaxy, TRUFA, and MG-RAST [22, 24, 34].

Galaxy: it is a collaborative initiative that provides a free set of tools and bioinformatics programs ranging from quality control of sequences (FASTQC), sequence editors, data grouping tools, tools for assembly (Trinity), sequence mapping (Bowtie), transcript quantification (Salmon and Kallisto), and metagenomic analysis programs (Mothur, Vegan, Kraken, and Krona) [34]. Being an open initiative, Galaxy presents a series of servers that offer different programs such as the functional prediction of a metagenome by PICRUST (Langille Lab and Huttentowe Lab) and servers dedicated to the functional annotation of transcriptomes (ANASTASIA).

TRUFA: Transcriptome User-Friendly Analysis [22], a program developed by the Institute of Physics of Cantabria, is a free server and contains several programs exclusively for transcriptomic (metatranscriptomic) analysis ranging from quality control (FASTQC and PRINSEQ), edited of sequences (CutAdapt), assembly of sequences (Trinity), quantification of transcripts (RSEM and eXpress) and functional annotation (BLAST2GO and HMMER). The files can be edited beforehand, and certain modules of the platform can be accessed, such as the functional annotation in case of already assembled sequences.

MG-RAST: Metagenomic Rapid Annotation-based on Subsystems Technology [24] is an open platform capable of analyzing sequences from different NGS platforms (Illumina, PacBio, and Nanopore). Unlike the aforementioned servers, MG-RAST has a pipeline that includes the quality control of the sequences, removal of adapters, detection of isoforms of transcripts, taxonomic comparison, and functional assignment. This server has several databases where the results can be analyzed regarding function (SEED, KEEG, COG, and NOG) and taxonomy (ITS, SILVA, RDP, and GreenGenes). It also has tools to export the data in tabular format, fasta, or in the form of BIOM type matrix.

BLAST2GO: it is a sequence annotator that is able to perform searches in the NCBI, which in its basic version has the BLAST algorithms to add taxonomic filters in order to accelerate the annotation. It also allows searches of Interprotein domains (InterProScan), allows the classification of proteins based on the Gene Orthology (GO) database, interaction maps between each GO term, function enrichment analysis (Fisher Exact Test) and the analysis of the metabolic modules present in the KEGG. The PRO version of this program allows making several annotations at the same time, using CLOUD-BLAST services and performing other types of analysis such as the differential expression of transcripts [37].

3. Conclusions

The study of how microbial communities contribute to environmental functions and the physiology of the host was limited to cultivable microorganisms. As, the free culture techniques based on molecular markers developed (AFLP, RFLP, and DGGE), this knowledge expanded. However, the knowledge obtained from these was quite limited. But now with the techniques of massive sequencing, it has been possible to obtain a better understanding of the role of microorganisms in different types of environments and hosts, both from the taxonomic and functional point of view.

The use of bioinformatic programs has allowed not only the reconstruction of the molecular phylogeny but also has allowed them to be studied from the functional point of view, showing the great potential for the future biotechnological exploitation of this microbial metabolic diversity such as enzymes with different catalytic activities from uncultivable organisms. Several combined methodologies that uses NGS techniques along with culture-based methods have been used to

obtain new bacterial strains using specific culture media or through the functional screening using specific primers in order to isolate the genes of interest.

Current research is now been carried out in order to obtain more precise metagenomic and metatranscriptomic assemblages, with new software specially designed (MetaVelvet, TriMetAss, and MetaAmos) to obtain complete genomes and transcriptomes of the bacterial communities. These protocols along with the integration of other "omic" techniques and systems biology could allow a better understanding of the complex metabolic and trophic networks that operates in an organism or environment.

Acknowledgements

Conflict of interest

The author declares that this chapter was conducted in the absence of any financial or commercial relationships that could be construed as a potential conflict of interest.

Author details

Raúl Enrique Valle-Gough[1], Blancka Yesenia Samaniego-Gámez[1], Javier Eduardo Apodaca-Hernández[2] and Maria Leticia Arena-Ortiz[3*]

1 Instituto Tecnológico Superior de Escarcega, Escarcega, Campeche, Mexico

2 Laboratory of Genomic Computing, Genomic Libraries (Bibliotecas Genómicas), Merida, Yucatan, Mexico

3 Universidad Nacional Autonoma de Mexico, Yucatan, Mexico

*Address all correspondence to: leticia.arena@ciencias.unam.mx

IntechOpen

References

[1] Aprill A. Marine animal microbiomes: Toward understanding host-microbiome interaction in a changing ocean. Frontiers in Marine Science. 2017;**4**:222. DOI: 10.3389/frmars.2017.00222

[2] Aires T, Yann M, Serrao E, Arnaud-Haond S. Hologenome theory supported by cooccurrence networks of species-specific bacterial communities in siphonous algae (*Caulerpa*). FEMS Microbiology Ecology. 2015;**91**(7):1-14

[3] Degli-Esposti M, Martinez-Romero E. The functional microbiome of arthropods. PLoS One. 2017;**12**(5):e0176573. DOI: 10.1371/journal.pone.0176573

[4] Killian M, Chapple ILC, Hanning M, Marsh PD, Meuric V, Pedersen AML, et al. The oral microbiome- an update for oral healthcare professionals. British Dental Journal. 2016;**221**(10):657-666

[5] Andersen R, Chapmann SJ, Artz RRE. Microbial communities in natural and disturbed peatlands: A review. Soil Biology and Biochemistry. 2015;(57):979-984

[6] Shi W, Syrenne R, Sun JZ, Yuab JS. Molecular approaches to study the insect gut symbiotic microbiota at the 'omics' age. Insect Sci. 2010;**17**:199-219

[7] Nair RR, Vasse M, Wiegloss S, Sun L, Nicco YT, Velicer GJ. Bacterial predator-prey coevolution selects on virulence-associated prey defenses. bioRxiv. 2018;**398495**:1-34. DOI: 10.1101/398495

[8] Grabher MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Trinity: Reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nature Biotechnology. 2011;**29**(7):644-652

[9] Jackson CR, Liew KC, Yule KM. Strucutral and functional changes in depht in microbial communities in a Tropical Malaysian Peat Swamp Forest. Microbial Ecology. 2009;**57**:402-412

[10] Kanokatrana P, Uengwetwanit T, Rattanachomsri U, Bunterngsook B, Nimchua T, Tangphatsornruang S, et al. Insights into the phylogeny and metabolic potential of a primary tropical peat swamp forest microbial community by metagenomic analysis. Microbial Ecology. 2011;**61**:518-528

[11] Jan C, Petersen JM, Werner J, Teeling H, Huan S, Glöckner FO, et al. The gill chamber epibiosis of deep-sea shrimp *Rimicaris exoculata*: An in-depth metagenomic investigation and discovery of *Zetaproteobacteria*. Environmental Microbiology. 2014;**16**(9):2723-2738

[12] Estrada-Peña A, Sprong H, Cabezas-Cruz A, De La Fuente J, Ramo A, Coipan EC. Nested coevolutionary networks shape the ecological relationships of ticks, hosts, and the Lyme disease bacteria of the Borrelia burgdorferi (s.l.) complex. Parasites & Vectors 2016;**9**(517): 1-15. DOI 10.1186/s13071-016-1803-z

[13] Kurtz J, Schulenburg H, Reusch TBH. Host–parasite coevolution—rapid reciprocal adaptation and its genetic basis. Zoology. 2016;**119**:241-243

[14] Rojas-Herrera R, Narváez-Zapata J, Zamudio-Maya M, Mena-Martínez ME. A Simple Silica-based method for metagenomic DNA extraction from soil and sediments. Molecular Biotechnology. 2008;**40**(1):13-17

[15] Chomzcsinsky P, Sacchi N. Single-step method for RNA isolation by acid guanidinium thyocyanate-phenol-chloroform extraction. Annals of Clinical Biochemistry. 1985;**162**(1): 156-159. DOI: 10.1006/abio.1987.9999

[16] Frias-López J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Crisholm SW, et al. Microbial community gene expression in ocean surface waters. Proceedings of the National Academy of Sciences of the United States of America. 2010;**105**(10):3805-3810

[17] Valle-Gough RE, Apodaca-Hernández JE, Rodriguez-Dorantes M, Arena-Ortiz ML. Metatrancriptomic analysis from the Hepatopancreas of adult white leg shrimp (Litopenaeus vannamei). Symbiosis. 2018;**76**(1):51-62. DOI: 10.1007/s13199-017-0534-z

[18] Xiong J, Dai W, Qiu Q, Zu J, Yang W, Li C. Response of host–bacterial colonization in shrimp to developmental stage, environment and disease. Molecular Ecology. 2017;**27**(18):3686-3699. DOI: 10.1111/mec.14822

[19] Bashiardes S, Zilberman-Schapira G, Elinav E. Use of metatranscriptomics in microbiome research. Bioinformatics and Biology Insights. 2016;**10**:19-25. DOI: 10.4137/BBI.S34610

[20] Feng F, Chen M, Xu C, Bai L, Yin X, Li X, et al. Transcriptomic analysis of Chinese bayberry (Myrica rubra) fruit development and ripening using RNA-Seq. BMC Genomics. 2012;**13**:19

[21] Frías-López J, Durán-Pinedo A. Effect of periodontal pathogens on the metatranscriptome of a healthy multispecies biofilm model. Journal of Bacteriology. 2012;**194**(8):2082-2095. DOI: 10.1128/JB.06328-11

[22] Komobis E, Cabellos L, Aguilar F, Frias-López C, Rpzas J, Marco ZR. TRUFA: A user-friendly web server for *de novo* RNA-seq analysis using cluster computing. Evolutionary Bioinformatics. 2015;**11**:97-104. DOI: 10.4137/EBO.S23873

[23] Schmieder R, Edwards E. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011;**27**(6):863-864

[24] Keegan KP, Glass EM, Meyer F. MG-RAST, a metagenomics service for analysis of microbial community structure and function. Methods in Molecular Biology. 2016;**1399**:207-233. DOI: 10.1007/978-1-4939-3369-3_13

[25] Babendreier D, Joller D, Romeis J, Bigler F, Widmer F. Bacterial community structures in honeybee intestines and their response to two insecticidal protein. FEMS Microbiology Ecology. 2007;**59**(3):600-610. DOI: 10.1111/j.1574-6941.2006.00249.x

[26] McFall-Ngai M, Hadfieldb MG, Thomas I, Boschc CG, Hannah VC, Domazet-Loso T, et al. Animals in a bacterial world, a new imperative for the life sciences. PNAS. 2013;**110**(9):3229-3236. DOI: 10.1073/pnas.1218525110

[27] Papkou A, Gokhale CS, Schulenburg H. Host–parasite coevolution: Why changing population size matters. Zoology. 2016;**119**:330-338

[28] Hamdi C, Balloi A, Essanaa CE, Gonella E, Raddadi N, Ricci I, et al. Gut microbiome dysbiosis and honey bee heatlth. Journal of Applied Entomology. 2011;**135**(7):524-533. DOI: 10.1111/j.1439-0418.2010.01609.x

[29] Schloss PD, Wescott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology. 2009;**75**(23):7537-7541. DOI: 10.1128/AEM.01541-09

[30] Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet C, Al-Ghalith GA, et al. QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. PeerJ.

2018;**6**:e27295v2. DOI: 10.7287/peerj.
preprints.27295v2

[31] Caporaso JG, Bittinger K, Bushman
FD, DeSantis TZ, Andersen GL, Knight
R. PyNAST: A flexible tool for aligning
sequences to a template alignment.
Bioinformatics. 2010;**26**(2):266-267

[32] Caporaso JG, Kuczynski J,
Stombaugh J, Bittinger K, Bushman
FD, Costello EK, et al. QIIME allows
analysis of high-throughput community
sequencing data. Nature Methods.
2010;**7**:335-336. DOI: 10.1038/
nmeth.f.303

[33] Edgar RC. Search and
clustering orders of magnitude
faster than BLAST. Bioinformatics.
2010;**26**(19):2460-2461. DOI: https://
doi.org/10.1093/bioinformatics/btq461

[34] Afgan E, Baker D, Batut B, van
den Beek M, Bouvier D, Čech M, et al.
The Galaxy platform for accessible,
reproducible and collaborative
biomedical analyses. Nucleic Acids
Research. 2018;**46**(1):537-W544. DOI:
10.1093/nar/gky379

[35] Bengtsson-Palme J, Boulund F, Fick
J, Kristiansson E, Larsson DGJ. Shotgun
metagenomics reveals a wide array
of antibiotic resistance genes and
mobile elements in a polluted lake
in India. Frontiers in Microbiology.
2014;**5**(648):1-14. DOI: 10.3389/
fmicb.2014.00648

[36] Wood DE, Salzberg SL. Kraken:
Ultrafast metagenomic sequence
classification using exact alignments.
Genome Biology. 2014;**15**:R46

[37] Conesa A, Götz S, García-
Gómez JM, Terol J, Talón M, Robles
M. Blast2GO: A universal tool for
annotation, visualization and analysis
in functional genomics research.
Bioinformatics. 2005;**21**(18):3674-3676.
DOI: 10.1093/bioinformatics/bti610

[38] Engel P, Moran N. The gut
microbiota of insects – diversity
in structure and function. FEMS
Microbiology Reviews. 2013;**37**(5):
699-735. DOI: 10.1111/1574-6976.12025

[39] Cornejo-Granados F, Gallardo-
Becerra L, Leonardo-Reza M,
Ochoa-Romo J, Ochoa-Leyva A. A meta-
analysis reveals the enviromental and
host factors shaping the structure and
function of shrimp microbiota. PeerJ.
2018;**6**:e5682

[40] Power P, Berlemont R.
Antarctic Soil Metagenome. In:
Highlander Sk, Rodriguez-Valera
F, White BA (eds) Encyclopedia of
Metagenomics. Boston, MA: Springer;
2015. pp. 1-7. DOI: https://doi.
org/10.1007/978-1-4899-7475-4

[41] Truu M, Juhansson J, Truu J.
Microbial biomass, actitvity and
community in constructed wetlands.
Science of the Total Environment.
2009;**407**:3958-3971

[42] Cui J, Xiao M, Liu M, Wang Z, Liu
F, Guo L, et al. Coupling metagenomics
with cultivation to select host-specific
probiotic micro-organisms for
subtropical aquaculture. Journal of
Applied Microbiology. 2017;**123**:
1274-1285. DOI: 10.1111/jam.13555

[43] Li E, Xu C, Wang X, Wang S, Zhao
Q, Zhang M, et al. Gut microbiota and
its modulation for healthy farming
of pacific white shrimp *Litopenaeus
vannamei*. Reviews in Fisheries Science
& Aquaculture. 2018;**26**(3):381-399.
DOI: 10.1080/23308249.2018.1440530

[44] Porchas-Cornejo MA, Martínez-
Porchas M, Vargas-Albores F, Gollas-
Galvan T, Martínez-Córdova LR,
Vazquez-Euan R, et al. High-resolution
detection of bacterial profile of
ocean water, before and after being
used by shrimp farms. Aquaculture
International. 2017;**25**(5):1833-1843

[45] Tzeng TD, Pao YY, Chen PC, Weng FCH, Jean WD, Wang D. Effects of host phylogeny and habitats on gut microbiomes of oriental river prawn (Macrobrachium nipponense). PLoS One. 2015;**10**(7):e0132860

[46] Vargas-Albores F, Porchas-Cornejo MA, Martínez-Porchas M, Villalpando-Canchola E, Gollas-Galván T, Martínez-Córdova LR. Bacterial biota of shrimp intestine is significantly modified by the use of a probiotic mixture: A high throughput sequencing approach. Helgoland Marine Research. 2017;**71**:5

[47] Crotti E, Baloi A, Hamdi C, Sansono L, Marzoarati M, Gonella E, et al. Microbial symbionts: A resource for the management of insect related problems. Microbial Biotechnology. 2012;**5**(3):307-317

[48] Mohr KI, Tebbe CC. Diversity and phylotype consistency of bacteria in the guts of three bee species (*Apoidea*) at an oilseed rape field. Environmental Microbiology. 2006;**8**(2):258-272. DOI: 10.1111/j.1462-2920.2005.00893.x

[49] Macfarlane S, Dillon JF. Microbial biofilms in the human gastrointestinal tract. Journal of Applied Microbiology. 2007;**102**(5):1187-1196. DOI: 10.1111/j.1365-2672.2007.03287.x

[50] Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, et al. A core gut microbiome in obese and lean twins. Nature. 2008;**457**(7228):480-484. DOI: 10.1038/nature07540

[51] Shen Z, Zhang N, Mustapha A, Xu D, Deng D, Reed M, et al. Identification of host-specific genetic markers within 16S rDNA intervening sequences of 743 genera of fecal bacteria. Journal of Data Mining in Genomics and Proteomics. 2016;**7**(1):1-9

[52] Purroy N, Wu CJ. Co-evolution of leukemia and host immune cells in chronic lymphocytic leukemia (CLL). Cold Spring Harbor Perspectives in Medicine. 2018;**7**(4):1-28. DOI: 10.1101/cshperspect.a026740

[53] Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes J, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nature Biotechnology. 2013;**1**(10):8

[54] Parks DH, Tyson GW, Hugenholtz P, Beiko RG. STAMP: Statistical analysis of taxonomic and functional profiles. Bioinformatics. 2014;**30**:3123-3124

[55] Koo H, Mojib N, Hakim JA, Hawes I, Tanabe Y, Andersen DT, et al. Microbial communities and their predicted metabolic functions in growth laminae of a unique large conical mat from Lake Untersee, East Antarctica. Frontiers in Microbiology. 2017;**8**:1347. DOI: 10.3389/fmicb.2017.01347

[56] Ezenwa VO, Gerardo NM, Inouye DW, Medina M, Xavier JB. Animal Behavior and Microbiome. Science. 2012;**338**(6104):198-199. DOI: 10.1126/science.1227412

[57] Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology. 2009;**10**(3):25.1-25.10

[58] Magoč T, Salzberg SL. FLASH: Fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 2011;**1**(21):2957-2963. DOI: 10.1093/bioinformatics/btr507

[59] Namiki T, Hachiya T, Tanaka H, Sakakibara Y. MetaVelvet: An extension of Velvet assembler to de novo metagenome assembly from short sequence reads. Nucleic Acids Research. 2012;**40**(20):e155

[60] Treangen TJ, Koren S, Sommer DD, Liu B, Astrovskaya I, Ondov B, et al. MetAMOS: A modular and open source metagenomic assembly and analysis pipeline. Genome Biology. 2013;**14**(1):R2. DOI: 10.1186/gb-2013-14-1-r2

[61] Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, et al. SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads. Bioinformatics. 2014;**30**(12):1660-1666

[62] Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Research. 2008;**18**(5):821-829. DOI: 10.1101/gr.074492.107

[63] Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics. 2014;**30**(5):614-620

[64] Thureborn P, Franzetti A, Lundin D, Sjölin S. Reconstructing ecosystem functions of the active microbial community of the Baltic Sea oxygen depleted sediments. PeerJ. 2016;**4**:e1593. DOI: 10.7717/peerj.1593

[65] Darzi Y, Letunic I, Bork P, Yamada T. iPath3.0: Interactive pathways explorer v3. Nucleic Acids Research. 2018;**46**(1):510-513