Chapter 10
Integrative Omics Approach for the Community Function Evaluation of Sponge and Coral Microbiomes

Fang Liu and Zhiyong Li

Contents

Abstract Sponges and corals are often found to be associated with symbionts that are phylogenetically diverse and ecologically important. Assessing the functions of symbionts in the context of host-microbe interaction and microbial community dynamics has emerged as a new frontier of sponge microbiology. Culture-independent molecular methods, such as 16S rRNA gene sequencing and metagenome sequencing, have been the main tools to tackle the diversity and function of sponge-/coral-associated microbiota. Nonetheless, using one or two methods might not be able to generate comprehensive insights into the cross talk between host and microbes, as well as the dynamics of symbiotic community. Here we present two hypothesized cases focusing on the integration of various high-throughput techniques to demonstrate the methodological framework and potential outcome of integrative omics approach. We chosed two existing researches as basis for design, raised new scientific questions, and developed integrative research plans using state-of-the-art techniques such as single-cell sequencing, metagenome binning, GeoChip, RNA-Seq, etc.

F. Liu · Z. Li (✉)
Marine Biotechnology Laboratory, State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, People's Republic of China
e-mail: zyli@sjtu.edu.cn

Z. Li (ed.), *Symbiotic Microbiomes of Coral Reefs Sponges and Corals*,
https://doi.org/10.1007/978-94-024-1612-1_10

Keywords Symbionts · High-throughput sequencing · Functional DNA array · Host-microbe interaction

10.1 Introduction

In the era of high-throughput omics study, not only the phylogenetic diversity but also the genomic features and ecological roles of sponge/coral microbiomes have been illustrated in an unprecedented way. In previous chapters, we have reviewed various aspects of the community function of sponge/coral microbiomes derived from omics studies, especially metagenomic studies, e.g., biogeochemical cycling, chemical defense, response to environmental stress, and host-symbiont recognition. Due to historical and technical reasons, most insights into sponge/coral microbiomes were derived from studies based on 16S rRNA gene amplicon sequencing. Nevertheless, the uncultured status of most microbial symbionts, the complexity of microbiomes, and the diverse evolutionary trajectories of microbes often resulted in gaps between the rRNA-based phylogenetic diversity and the functional annotations of metagenome/metatranscriptome assemblies [1–4]. Besides the challenges presented by natural microbiomes, scientists often have to face the constraints from funds, facilities, and personnel. As pointed out in a recent review, "Only by applying a diverse range of methodological approaches to a broad suite of model and non-model systems studied by a well-networked community of interdisciplinary researchers will we truly be able to reveal the extraordinary extent of symbiotic interactions that exist in nature" [5].

It is important to choose the omics tools wisely when they are available. In general, target gene sequencing (e.g., 16S rRNA gene, *amoA*, *nifH*) is most suitable for the in-depth exploration of microbial diversity or a specific function; shotgun metagenome sequencing serves the need for unveiling genetic diversity of unknown microbial communities and discovering novel genes; metatranscriptome sequencing and metaproteomics analysis are feasible for surveys of functional activities of unknown microbial communities and validation of metagenomic data; metabolomics can be used to describe the metabolic profiles of microbial communities and identify biomarkers; functional DNA array (e.g., GeoChip) can be applied to compare the functional diversity and activity of microbial communities across many samples [6, 7]. With the aim to enlighten readers on designing applicable and cost-effective integrative omics approaches, we herein present the possible strategies for integrating different approaches to explore the functional traits of sponge/coral microbiomes by constructing hypothesized cases based on previous studies. The omics approaches discussed below are not strange to most microbial ecologists, but we rarely see these approaches being integrated to study the sponge/coral microbiomes. The reason we choose certain studies [8, 9] as the basis to construct our hypothesized cases is because we believe these studies provide important materials and experiment foundation for further exploration of the sponge/coral microbiomes.

10.2 Case 1: Functional Analysis of Vertically Transmitted *Proteobacteria* throughout the Life Cycle of the Demosponge *Amphimedon queenslandica*

Amphimedon queenslandica is a low-microbial-abundance (LMA) sponge with a fully sequenced genome [10, 11]. As the first sponge species with its genome sequenced, *A. queenslandica* represents a good model for metazoan evolution study and innate immune research [12]. The LMA feature also makes *A. queenslandica* a good model for the study of host-microbe interaction. According to the research from Bayes [8], five putative vertically transmitted proteobacterial OTUs (hereafter *AqVTP*s) are present throughout the sponge life cycle. Notably, three putative vertically transmitted OTUs have low (< 90%) sequence similarities to their closest relatives and likely represent new orders of the *Proteobacteria*. Furthermore, a complete community shift from the *Proteobacteria*-dominated community present in sponges collected from their natural habitat to a community dominated by representatives of the *Chlamydiae* in the sponges maintained in aquaria was observed.

To further our understanding toward the function of vertically transmitted symbionts and the cross talk between sponges and symbionts, it is necessary to obtain the genomes of *AqVTP*s and analyze their activities along the sponge life cycle. We herein present a design of an integrative omics approach to study the subject topic.

In the beginning, we could obtain the genomes of those *AqVTP*s via either single-cell genome sequencing [13] or metagenome binning [14]. Then, the single-cell genomes or metagenome bins will serve as references for metatranscriptome analysis along the sponge life cycle. The genome-centered metatranscriptome analysis can reduce the requirement for computer resources and shorten the bioinformatics process. Since *A. queenslandica* genome and transcriptome are annotated extensively, it is possible to develop a functional DNA array to monitor the response of *A. queenslandica* during the establishment of symbiotic communities and the metabolism characteristics at different life stages. Using functional DNA array rather than transcriptome sequencing to monitor the sponge activities can reduce the research cost and accelerate the data analysis. In the past, functional DNA array data have already provided some insights into how corals and their dinoflagellate symbionts may communicate [15]. Thus, by incorporating single-cell genomics, metatranscriptomics, and microarray analysis, we can provide insights into (1) the genomic features and metabolism repertoire of *AqVTP*s; (2) the phylogenetic affiliation of *AqVTP*s based on multigene phylogeny; (3) the activity and roles of *AqVTP*s along the sponge life cycle; and (4) the mechanisms of *AqVTP*s – *A. queenslandica* metabolic interaction and immune recognition. In the next paragraphs, we will present a phase-wise description of this integrative omics approach. The key steps are illustrated in Fig. 10.1.

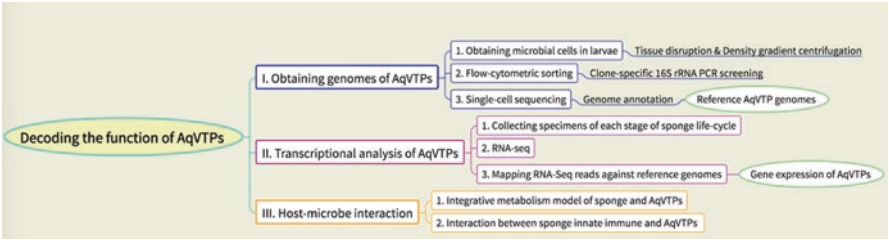

Fig. 10.1 The workflow of decoding the functions of vertically transmitted *Proteobacteria* (*AqVTP*s) throughout the life cycle of the demosponge *Amphimedon queenslandica*

Phase 1 Obtaining the *AqVTP* Genomes

A. queenslandica larvae are collected and disrupted in calcium-magnesium free artificial seawater followed by the density-gradient centrifugation to separate bacterial cells and sponge cells [16]. The fraction of bacterial cells is then submitted to flow cytometric sorting and whole-genome amplification [17]. According to the results from Bayes [8], specific primers, which will be used for screening *AqVTP* cells, can be designed based on the representative sequences of OTUs. After phylogenetic screening and verification, amplified genomes of *AqVTP*s are submitted to sequencing platform. Alternatively, if flow cytometric sorting is not available, we can sequence the metagenome of *A. queenslandica* and then extract the genomes of interest using metagenome binning approach [14, 18, 19]. Nonetheless, this method requires deep sequencing depth and considerable computer resources. There is a chance that not all the *AqVTP* genomes could be recovered from metagenomic datasets. Sometimes, the 16S rRNA operons might be missing in the binned genomes, which make it challenging to screen the *AqVTP*s.

Once the *AqVTP* genomes are available, we can perform the genome annotation and multigene phylogeny analysis. There are many automated platforms and software to make genome annotation and phylogenetic analysis easier for biologists, e.g., IMG 4 [20] and CheckM [21].

Phase 2 Transcriptional Activities of *AqVTPs*

A. queenslandica embryos, larvae, postlarvae, and adults are collected in biological replicates and preserved in RNA later. Unlike the transcriptome sequencing of sponges, sample preparation for prokaryotic symbiont mRNA remains to be complex and challenging [16, 22]. The yield of prokaryotic mRNA is generally lower than the optimal amount due to the loss during the removal of rRNA and eukaryotic mRNA. Given the LMA status of *A. queenslandica*, it is crucial to collect enough specimens for multiple rounds of RNA extraction. Once the RNA-Seq is finished, we can use the metatranscriptomic reads to estimate the gene expression of *AqVTP*s without de novo assembling of metatranscriptome. Then we can analyze the differential gene expression pattern of *AqVTP*s in the context of sponge life cycle.

Phase 3 Interaction Between *AqVTPs* and *A. queenslandica*
Based on the genomic and transcriptomic information of *AqVTP*s, it is possible to develop an integrative model for sponge-microbe symbiosis which also incorporates the expression of sponge genes (indicated by functional DNA array data). Furthermore, it is also possible to analyze the relation between sponge innate immune gene expression (NLRs, SRCRs, etc.) [12, 16] and the activities of *AqVTP*s along the sponge life cycle.

10.3 Case 2: Response of Coral Symbionts to the Cumulative Pressures of Climate Change

How coral-associated microbial communities respond to the global climate change has always been a focus of marine microbiology. Yet our understandings of how coral-associated microbes respond to elevated sea surface temperature (SST) and ocean acidification (OA) are limited, and we have barely any knowledge of how they respond to the interactive effects of these climate stressors. A recent study has conducted an 8-week experimental exposure to near-future climate change conditions and analyzed the bacterial community response of the corals *Acropora millepora* and *Seriatopora hystrix* using 16S rRNA gene amplicon pyrosequencing [9]. This study showed no visible signs of compromised host health, and coral-associated bacterial communities remained stable under the cumulative pressures. Nonetheless, it is still unknown how symbiotic communities respond to the cumulative pressures at functional gene level. Here we present an integrative omics approach that employs open (RNA-Seq) and closed format (GeoChip) high-throughput technologies [7] to extend the research and provide insights into the questions above.

The experimental design is based on the research from Webster et al. [9]. The workflow is showed in Fig. 10.2. In brief, *Acropora millepora* are deployed into triplicate aquaria for different pH/temperature treatments (8.1/28 °C, 8.1/31 °C, 7.9/28 °C, and 7.9/31 °C) and incubated for 8 weeks. We will focus on the gene expression of *Symbiodinium* symbionts (RNA-Seq) and the functional gene repertoire and expression of symbiotic bacteria, archaea, and fungi (GeoChip). The reason we use *Acropora millepora* as the material is because the genome of *Acropora digitifera* is available [23], and the *Acropora millepora genome* will be available soon in McDonnell Genome Institute. Therefore, we can extract the genetic information of dinoflagellate symbionts from the metatranscriptome datasets using *Acropora digitifera* genome and *Symbiodinium kawagutii* genome [24] as references. Considering the total RNA of *Acropora millepora* will probably be dominated by coral and dinoflagellate RNA, we will use GeoChip [25], which is insensitive to nontargeted nucleic acids, to detect the gene expression of other symbionts.

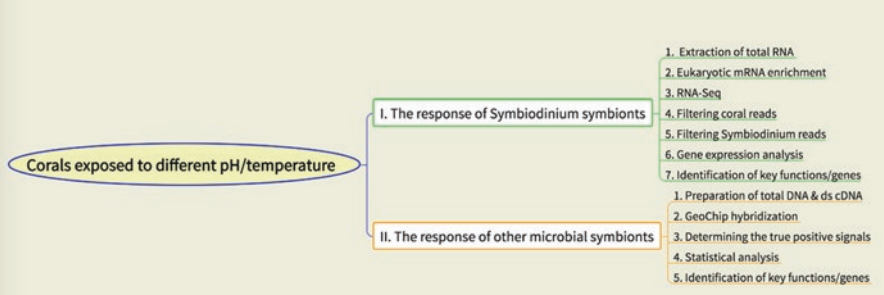

Fig. 10.2 The workflow of revealing the response of coral symbionts to the cumulative pressures of climate change

Part 1: The Response of *Symbiodinium* Symbionts

Although Webster et al. [9] did not observe any signs of compromised health in any coral hosts, investigating the molecular response of *Symbiodinium* is still beneficial for our understanding of the mechanisms related to the homeostasis of coral-associated dinoflagellate symbionts. Firstly, we will perform eukaryotic metatranscriptome sequencing. The oligo-d(T)-based eukaryotic mRNA enrichment is a mature approach available in many sequencing facilities. Thus, sequencing the eukaryotic metatranscriptome is less challenging than sequencing the prokaryotic metatranscriptome. Secondly, we map the reads against *Acropora digitifera* genome (or *Acropora millepora* genome when it is available) to filter out the host reads, then using *Symbiodinium kawagutii* genome as the reference to screen and quantify the *Symbiodinium* reads. The read count provides a measure of gene expression. Using DESeq [26], we can identify the differentially expressed genes (DEGs) across all the treatments. Thirdly, the enriched biological functions and pathways indicated by DEGs can be illustrated in several bioinformatics platforms, e.g., DAVID [27], PANTHER [28], REVIGO [29], etc.

Part 2: The Response of Other Microbial Symbionts

GeoChip, the most comprehensive functional gene array available, provides a rapid and cost-effective way to investigate the functional gene repertoire of microbiomes. Since its invention, GeoChip has been constantly updated and has been widely used in both terrestrial and ocean environments [7]. The latest version, GeoChip 5.0, contains 167,044 distinct probes, which are designed based on 395,894 coding sequences (CDS) from ~1500 functional gene families involved in microbial geochemical cycling, energy metabolism, metal homeostasis, organic remediation, secondary metabolism, stress responses, viruses, and virulence (http://glomics.com/gch-tech.html). Here we use GeoChip to assess the functional gene repertoire and transcriptional activities of *A. millepora* microbiomes. Firstly, the total DNA and RNA of *A. millepora* specimens are extracted simultaneously using Qiagen AllPrep DNA/RNA Kit or something similar. This operation can ensure the cohesiveness of the genetic information recovered between DNA and RNA samples. Then the total RNA are converted into ds cDNA for GeoChip hybridization. Secondly, the DNA

and ds cDNA samples are labeled and then loaded onto GeoChip for hybridization. This step is usually performed in Glomics Inc. along with data normalization as customer service. In the end, we will receive a data sheet recording the signal intensity of positive probes and the metadata of the positive probes (e.g., GenBank ID, gene name, functional category, phylogenetic affiliation). The data sheet can be directly imported into R environment (http://www.R-project.org/) or similar software for statistical analysis and visualization.

Determining positive signals is crucial for GeoChip analysis. In our case, a minimum of two positive spots out of three biological replicates is required for each probe/gene to be considered for data analysis. Only spots for which both DNA and RNA signals are detected are considered positive. Combining DNA and RNA data is a more conservative approach than analyzing DNA/RNA separately but will gain confidence that detected signals originate from sizable and active populations [30].

Analysis of GeoChip data usually can be divided into three levels: whole dataset (whole community structure), gene categories/functional genes, and probe/sequence levels. The statistical analysis should be carried out from high level (whole community) to low level (probe). Different statistical methods might be good for different levels. In our case, we can use clustering and ordination algorithms (e.g., UPGAMA, nMDS, PCA) to check if there is any grouping pattern in accordance with different treatments. Then using pairwise test (e.g., t-test, ANOVA), we can find the genes whose relative abundance or expression levels vary significantly between different treatments. Last but not least, we can analyze the genes of interest to investigate subcommunities that carry out special ecological functions (e.g., carbon degradation, stress response, secondary metabolism).

Acknowledgments We gratefully acknowledge the financial supports from the National Natural Science Foundation of China (NSFC) (31861143020, 41776138, 41742002, U1301131, 41176127, 41076077).

References

1. Burke C, Steinberg P, Rusch D, Kjelleberg S, Thomas T. Bacterial community assembly based on functional genes rather than species. Proc Natl Acad Sci U S A. 2011;108:14288–93.
2. Overballe-Petersen S, Willerslev E. Horizontal transfer of short and degraded DNA has evolutionary implications for microbes and eukaryotic sexual reproduction. BioEssays. 2014;36:1005–10.
3. Shapiro BJ, Polz MF. Microbial speciation. Cold Spring Harb Perspect Biol. 2015;7:a018143.
4. Bendall ML, Stevens SL, Chan LK, Malfatti S, Schwientek P, Tremblay J, et al. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. ISME J. 2016;10:1589–601.
5. Webster NS. Cooperation, communication, and co-evolution: grand challenges in microbial symbiosis research. Front Microbiol. 2014;5:16.
6. Ottman N, Smidt H, de Vos W, Belzer C. The function of our microbiota: who is out there and what do they do? Front Cell Infect Microbiol. 2012;2:104.

7. Zhou J, He Z, Yang Y, Deng Y, Tringe SG, Alvarez-Cohen L. High-throughput metagenomic technologies for complex microbial community analysis: open and closed formats. MBio. 2015;6:e02288–14.
8. Bayes JM. The characterisation of the microbiome of the demosponge Amphimedon queenslandica reveals host immune responses to changes in the symbiotic bacterial community. MPhil. The University of Queensland. 2013.
9. Webster NS, Negri AP, Botte ES, Laffy PW, Flores F, Noonan S, et al. Host-associated coral reef microbes respond to the cumulative pressures of ocean warming and ocean acidification. Sci Rep. 2016;6:19324.
10. Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier MEA, Mitros T, et al. The Amphimedon queenslandica genome and the evolution of animal complexity. Nature. 2010;466:720–6.
11. Fernandez-Valverde SL, Calcino AD, Degnan BM. Deep developmental transcriptome sequencing uncovers numerous new genes and enhances gene annotation in the sponge Amphimedon queenslandica. BMC Genomics. 2015;16:387.
12. Degnan SM. The surprisingly complex immune gene repertoire of a simple sponge, exemplified by the NLR genes: a capacity for specificity? Dev Comp Immunol. 2015;48:269–74.
13. Rinke C, Lee J, Nath N, Goudeau D, Thompson B, Poulton N, et al. Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. Nat Protoc. 2014;9:1038–48.
14. Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat Biotechnol. 2013;31:533–8.
15. Voolstra CR, Schwarz JA, Schnetzer J, Sunagawa S, Desalvo MK, Szmant AM, et al. The host transcriptome remains unaltered during the establishment of coral–algal symbioses. Mol Ecol. 2009;18:1823–33.
16. Ryu T, Seridi L, Moitinho-Silva L, Oates M, Liew YJ, Mavromatis C, et al. Hologenome analysis of two marine sponges with different microbiomes. BMC Genomics. 2016;17:1–11.
17. Siegl A, Kamke J, Hochmuth T, Piel J, Richter M, Liang C, et al. Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges. ISME J. 2011;5:61–70.
18. Burgsdorf I, Slaby BM, Handley KM, Haber M, Blom J, Marshall CW, et al. Lifestyle evolution in cyanobacterial symbionts of sponges. MBio. 2015;6:e00391–15.
19. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ. 2015;3:e1165.
20. Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Pillay M, et al. IMG 4 version of the integrated microbial genomes comparative analysis system. Nucleic Acids Res. 2014;42:D560–7.
21. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GWCM. Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043–55.
22. Moitinho-Silva L, Seridi L, Ryu T, Voolstra CR, Ravasi T, Hentschel U. Revealing microbial functional activities in the Red Sea sponge *Stylissa carteri* by metatranscriptomics. Environ Microbiol. 2014;16:3683–98.
23. Shinzato C, Shoguchi E, Kawashima T, Hamada M, Hisata K, Tanaka M, et al. Using the Acropora digitifera genome to understand coral responses to environmental change. Nature. 2011;476:320–32.
24. Lin S, Cheng S, Song B, Zhong X, Lin X, Li W, et al. The Symbiodinium kawagutii genome illuminates dinoflagellate gene expression and coral symbiosis. Science. 2015;350:691.
25. Gao Y, Wang S, Xu D, Yu H, Wu L, Lin Q, et al. GeoChip as a metagenomics tool to analyze the microbial gene diversity along an elevation gradient. Genomics Data. 2014;2:132–4.
26. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:1–21.

27. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2008;4:44–57.
28. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, et al. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. 2017;45:D183–9.
29. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of Gene Ontology terms. PLoS One. 2011;6:e21800.
30. Wawrik B, Boling WB, Van Nostrand JD, Xie J, Zhou J, Bronk DA. Assimilatory nitrate utilization by bacteria on the West Florida Shelf as determined by stable isotope probing and functional microarray analysis. FEMS Microbiol Ecol. 2012;79:400–11.