

Research Highlight

Genomics and Metagenomics in Microbial Ecology: Recent Advances and Challenges

TAKASHI NARIHIRO^{1*}, and YOICHI KAMAGATA¹

¹*Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Central 6, Tsukuba, Ibaraki 305-8566, Japan*

In the book “*What is Life?*,” physicist Erwin Schrödinger (24) expected that:

“*To give the statement life and colour, let me anticipate what will be explained in much more detail later, namely, that the most essential part of a living cell—the chromosome fibre may suitably be called an aperiodic crystal.*”

Ten years later, Watson and Crick discovered that the “aperiodic crystal,” named deoxyribonucleic acid (DNA), had a double helix structure consisting of nucleotides (adenine, thymine, cytosine, and guanine) and a sugar-phosphate backbone (34). A gene is a sequence of DNA that encodes functional molecules (e.g., RNAs and proteins). A genome is the entire set of genes of an organism and a blueprint of physiological and biochemical functions that are expressed by an organism. To date, the genes and genomes of diverse living organisms have been sequenced by a large number of research teams. As of February 2017, more than 100,000 genomic sequences including those of 89,843 prokaryotes, 3,976 eukaryotes, and 7,088 viruses, have been deposited in the NCBI GenBank database (Fig. 1) (1). The widespread use of high-throughput sequencing technologies has increased the number of genomes of prokaryotes that have been sequenced in the last two years (>25,000 genomes per year). In addition, more than 1,000 metagenomic studies are available through the JGI Integrated Microbial Genomes & Microbiomes (IMG/M) web server (Fig. 2) (9). These metagenomic sequences are prompting researchers to attain a clearer understanding of the ecophysiology of yet-to-be cultivated organisms in environments (e.g., marine, freshwater, and soil), engineered systems (e.g., waste/wastewater treatment process), and host-associated microbiomes (e.g., humans, mammals, and Arthropoda) (4, 12).

Some of the most recent outcomes are (meta)genome-based insights into syntrophy in anaerobic methanogenic ecosystems. Syntrophic substrate metabolizers (syntrophs) play an important role in the oxidation of by-products (e.g., volatile fatty acids, primary alcohols, and amino acids) from the fermentative degradation of high-molecular-weight organic compounds in methanogenic ecosystems (6, 13, 14, 23). Extensive efforts have been made to elucidate the mechanisms underlying syntrophic degradation pathways and energy conservation systems based on the genomes of the cultured representatives of syntrophs (26). Nobu *et al.* (17, 19) constructed the draft genome of *Syntrophorhabdus aromaticivorans* strain UI,

isolated from a methanogenic isophthalate-degrading enrichment culture (22), and identified syntrophic degradation pathways for benzoate, phenol, and terephthalate. This achievement was the first explicit genome-based analysis of anaerobic phenol and terephthalate metabolism, which had previously remained unclear. The strain UI genome encodes flavin oxidoreductase-heterodisulfide reductase (Flox-Hdr), electron transfer flavoproteins (ETFs), hydrogenases, and formate dehydrogenases, which appear to be closely associated with the generation of H₂ and formate, important intermediates for the interspecies electron transfer underpinning anaerobic syntrophy. More recently, comparative genomics of three syntrophic bacteria: “*Syntrophomonas wolfei* subsp. *methylbutyratica*” strain 4J5 (15), *Syntrophothermus lipocalidus* strain TGB-C1 (2), and *Syntrophomonas wolfei* subsp. *wolfei* strain Göttingen (25) elucidated the metabolic pathways of branched-chain fatty acids (BCFAs) such as 2-methylbutyrate and isobutyrate (16). Anaerobic BCFA metabolism remained unclear even though BCFAs are important products from amino acids. These findings indicate that the genomes of BCFA-degrading strains 4J5 and TGB-C1 encode unique β -oxidation-related enzymes for syntrophic BCFA oxidation as well as various energy conservation systems including ETF-linked acyl-CoA dehydrogenase, ETF-linked iron-sulfur (Fe-S) binding reductase, ETF dehydrogenase (FixABCX), and Flox-Hdr. Manzoor *et al.* (7) constructed the draft genome of the acetate-oxidizing syntroph *Syntrophaceticus schinkii* strain Sp3. The whole transcriptome profiling of co-cultures of strain Sp3 and *Methanoculleus bourgensis* fed with acetate revealed that genes involved in the Wood-Ljungdahl pathway, hydrogenases, ATP synthase, formate dehydrogenases, and acetate transporters, were clearly expressed, supporting its unique physiology for syntrophic acetate oxidation (8). A metagenomic approach may lead to the identification of methanogenic microbiota that contain, to a large extent, fastidious syntrophs that have never been cultivated. Wawrik *et al.* (35) performed metagenomic sequencing of an uncultured *Smithella*-predominated methanogenic enrichment culture capable of mineralizing long-chain *n*-paraffins (C28–C50). They demonstrated for the first time that the uncultured *Smithella* sp. SDB metagenome encoded the alkylsuccinate synthase gene (*assA*) for the degradation of *n*-paraffins, while ETF-linked Fe-S oxidoreductase, periplasmic formate dehydrogenase, and hydrogenase were used for energy conservation. By using metagenomics together with metatranscriptomics and single-cell genomics, Nobu *et al.* (18) demonstrated that phylogenetically diverse microbes cooperate in the synergistic degradation of terephthalate, catabolic by-products (acetate, butyrate, and H₂), and detritus compounds (e.g., amino acids and glycerol) in a methano-

* Production Editor, Microbes and Environments:
E-mail: t.narihiro@aist.go.jp;
Tel: +81-29-861-6591; Fax: +81-29-861-6587.

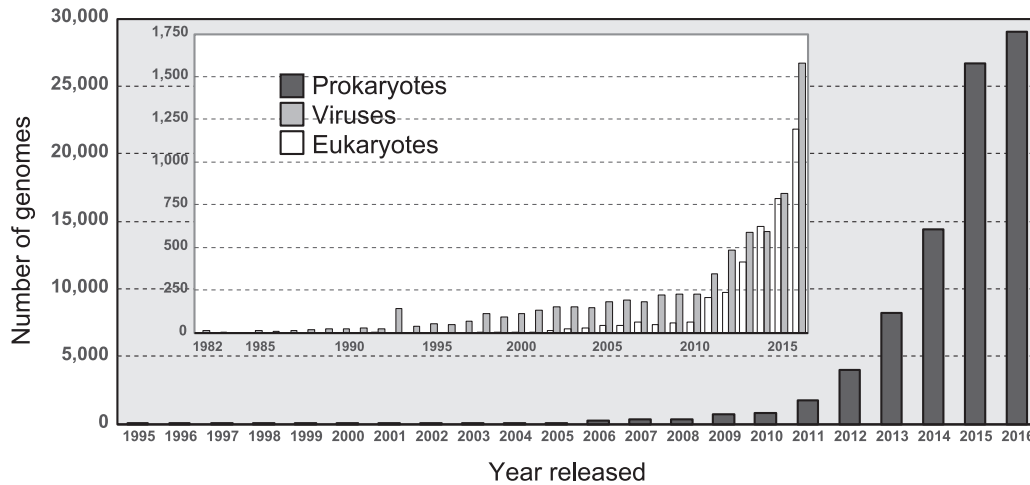


Fig. 1. Number of genomes of prokaryotes, eukaryotes, and viruses sequenced each year based on the NCBI GenBank database (1).

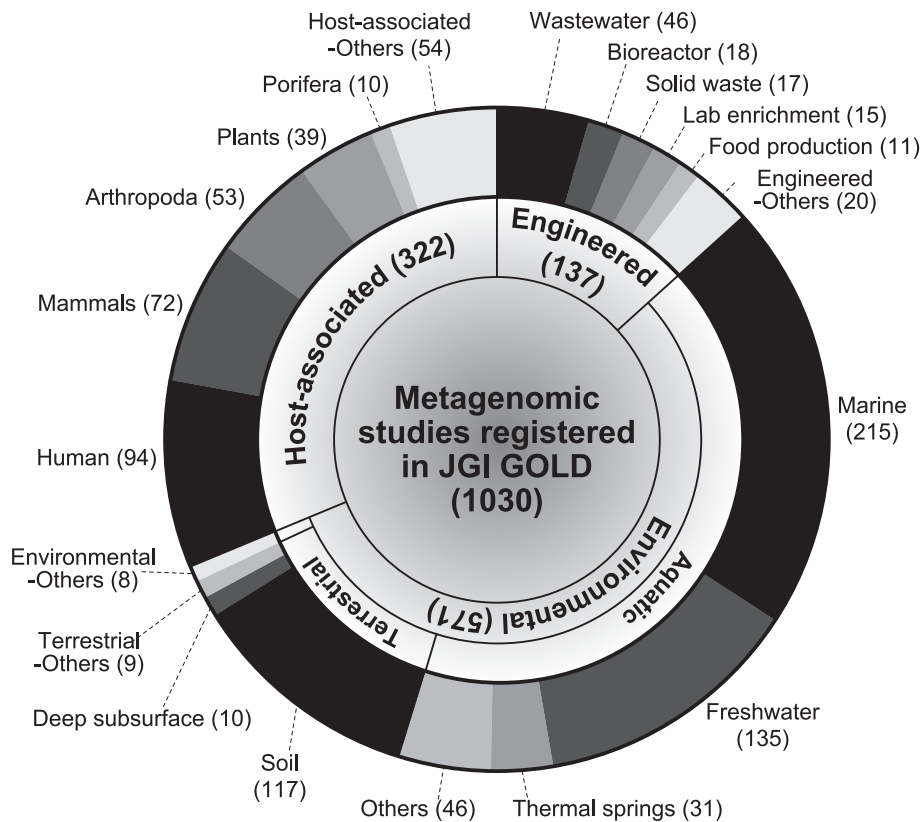


Fig. 2. Number of metagenomic studies in the JGI Integrated Microbial Genomes & Microbiomes (IMG/M) web server (9) as of February 2017. Category and subcategory names are according to the IMG/M database.

genic bioreactor; *i.e.*, *Pelotomaculum* and *Syntrophorhabdus* degrade terephthalate to acetate; *Thermotogae* members syntrophically oxidize acetate; *Syntrophaceae*, “*Ca. Marinimicrobia*” (SAR406), and “*Ca. Hydrogenedentes*” (NKB19) may play roles in the degradation of butyrate/BCFAs, amino acids, and glycerol, respectively; “*Ca. Atribacteria*” (OP9) and “*Ca. Cloacimonetes*” (WWE1) are identified as syntrophic propionate oxidizers. This study successfully provided a comprehensive description of a complex syntrophy-associated microbiota in order to more clearly understand the metabolic capacities of as-yet uncultivated organisms, named “microbial dark matter.”

A metagenomic approach is frequently used to analyze host-microbe interactions in mammals, plants, and other living organisms (Fig. 2) (3). Minami *et al.* (10) conducted a comparative analysis of 29 genomes of *Methylobacterium* spp. belonging to three 16S rRNA gene-based phylogeny groups (Groups I, II, and III) derived from soybean stem samples, and found that Group I *Methylobacterium* members were dominant in soybean-associated *Methylobacterium* and possessed a complete set of genes for the degradation of ureide compounds (allantoin and allantoinate), implying competition between *Methylobacterium* spp. and soybean plants in the

utilization of ureide compounds as the nitrogen source. In contrast, the genomes of two strains associated with Group II *Methylobacterium* lacked the genes of ureidoglycorate urealase, the last step of ureide degradation. This genome-based evidence suggests that the survival strategy of *Methylobacterium* spp. in the rhizosphere and phyllosphere is metabolically diverse. Metagenomic analyses of the bacterial community of the taproot of sugar beet revealed that the genes involved in plant growth-promoting traits (PGPTs) (*i.e.*, plant disease suppression, phosphate solubilization, methanol utilization, and 1-aminocyclopropane-1-carboxylic acid [ACC] deaminase) were frequently assigned to the genera *Mesorhizobium*, *Bradyrhizobium*, and *Streptomyces*, suggesting their important roles in the taproot of sugar beet (31).

Viral genome data has been accumulated due to the importance in clinical and biological aspects (Fig. 1) (5, 21, 27). Takemura (28, 29) discovered a novel DNA virus, named *Tokyovirus*, of the family *Marseilleviridae* from water/soil samples collected from Arakawa River in Tokyo. A total of 487 coding sequences (CDSs) were identified from the *Tokyovirus* genome, and most CDSs exhibited high homology with other *Marseilleviridae* family members, while several putative CDSs were found to be functionally unknown *Tokyovirus*-specific CDSs, implying that *Tokyovirus* evolved separately from the previously identified members of the *Marseilleviridae* family. Oh *et al.* (20) reconstructed the draft genomes of virophages of inland lakes from metagenomic sequences and found a novel virophage population from a planktonic microbial community of the Tibetan mountain lake, Lake Qinghai. The Qinghai Lake virophage (QLV) genome harbors 25 open reading frames including ATPase, cysteine protease, major and minor capsid proteins, and DNA helicase/primase/polymerase, which are commonly conserved in other virophage genomes. They successfully developed a strategy to detect virophage populations using genes encoding major capsid proteins as a molecular marker. In order to facilitate studies on viral (meta) genomics, a recovery technique of viral genomes using tandem column chromatography (32) and a sequencing method for intracellular RNA viruses, so-called fragmented and loop primer ligated dsRNA sequencing (FLDS) (33), have been developed.

Morohoshi *et al.* (11) recently reported the complete genome sequence of *Pseudomonas chlororaphis* subsp. *aurantiaca* strain StFRB508, a bacterium that regulates the production of the antifungal agent phenazine through *N*-acyl-L-homoserine lactone (AHL)-mediated quorum sensing. They revealed that the StFRB508 genome encodes three quorum-sensing systems comprising a set of AHL-synthase and AHL-receptor genes (*i.e.*, *phzI/phzR*, *aurI/aurR*, and *csaI/csaR*), which may play an important role in the biosynthesis of phenazine. Tashiro *et al.* (30) performed genomic analyses on small colony variants (SCVs) of *Escherichia coli* with phenotypic differences such as slower growth, less biofilm formation, and greater resistance to aminoglycoside antibiotics than the parent strain. The findings obtained indicated that the transposon *Tn1000*, which was originally located in the conjugative F plasmid, was inserted into the *hemB* gene locus in the chromosome, suggesting that SCVs are caused by inactivation of the heme biosynthetic pathway. These findings strongly suggest that the F plasmid has a role

in the appearance of genetic variants as a result of DNA recombination within a single cell.

In “*What is Life?*,” Schrödinger discussed the order of the mechanisms of life as follows (24):

“*The orderliness encountered in the unfolding of life springs from a different source. It appears that there are two different ‘mechanisms’ by which orderly events can be produced: the ‘statistical mechanism’ which produces order from disorder and the new one, producing order from order.*”

In recent years, vast amounts of genomic and metagenomic sequencing data have been generated and applied to elucidate the novel ecophysiological functions of living organisms; however, researchers are inclined to ignore sequences that have neither conserved regions, motifs, nor known arrangements, and consider them to be disordered or disorganized. Since a large amount of omics-derived data is accumulating, finding order from disorder will represent the next challenge for researchers.

References

1. Coordinators, N.R. 2016. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 44:D7–19.
2. Djao, O.D.N., X.J. Zhang, S. Lucas, *et al.* 2010. Complete genome sequence of *Syntrophothermus lipocalidus* type strain (TGB-C1(T)). *Stand. Genomic Sci.* 3:267–275.
3. Hacquard, S., R. Garrido-Oter, A. Gonzalez, *et al.* 2015. Microbiota and host nutrition across plant and animal kingdoms. *Cell Host Microbe.* 17:603–616.
4. Hiraoka, S., C.C. Yang, and W. Iwasaki. 2016. Metagenomics and bioinformatics in microbial ecology: Current status and beyond. *Microbes Environ.* 31:204–212.
5. Houldcroft, C.J., M.A. Beale, and J. Breuer. 2017. Clinical and biological insights from viral genome sequencing. *Nat. Rev. Microbiol.* 15:183–192.
6. Kato, S. 2015. Biotechnological aspects of microbial extracellular electron transfer. *Microbes Environ.* 30:133–139.
7. Manzoor, S., B. Muller, A. Niazi, A. Schnurer, and E. Bongcam-Rudloff. 2015. Working draft genome sequence of the mesophilic acetate oxidizing bacterium *Syntrophaceticus schinkii* strain Sp3. *Stand. Genomic Sci.* 10:99.
8. Manzoor, S., E. Bongcam-Rudloff, A. Schnurer, and B. Muller. 2016. Genome-guided analysis and whole transcriptome profiling of the mesophilic syntrophic acetate oxidizing bacterium *Syntrophaceticus schinkii*. *PLoS One.* 11:e0166520.
9. Markowitz, V.M., I.M.A. Chen, K. Chu, *et al.* 2014. IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res.* 42:D568–D573.
10. Minami, T., M. Anda, H. Mitsui, *et al.* 2016. Metagenomic analysis revealed methylamine and ureide utilization of soybean-associated *Methylobacterium*. *Microbes Environ.* 31:268–278.
11. Morohoshi, T., T. Yamaguchi, X. Xie, W.Z. Wang, K. Takeuchi, and N. Someya. 2017. Complete genome sequence of *Pseudomonas chlororaphis* subsp. *aurantiaca* reveals a triplicate quorum-sensing mechanism for regulation of phenazine production. *Microbes Environ.* 32:47–53.
12. Narihiro, T., and Y. Kamagata. 2013. Cultivating yet-to-be cultivated microbes: The challenge continues. *Microbes Environ.* 28:163–165.
13. Narihiro, T., M.K. Nobu, N.K. Kim, Y. Kamagata, and W.T. Liu. 2015. The nexus of syntrophy-associated microbiota in anaerobic digestion revealed by long-term enrichment and community survey. *Environ. Microbiol.* 17:1707–1720.
14. Narihiro, T., M.K. Nobu, R. Mei, and W.T. Liu. 2015. Microbial community involved in anaerobic purified terephthalic acid treatment process, p. 31–48. *In* H.H.P. Fang, and T. Zhang (ed.), *Anaerobic Biotechnology: Environmental Protection and Resource Recovery*. Imperial College Press, London.

15. Narihiro, T., M.K. Nobu, H. Tamaki, Y. Kamagata, and W.T. Liu. 2016. Draft genome sequence of *Syntrophomonas wolfei* subsp. *methylbutyrate* strain 4J5 (JCM 14075), a mesophilic butyrate and 2-methylbutyrate degrading syntroph. *Genome Announc.* 4:e00047-16.
16. Narihiro, T., M.K. Nobu, H. Tamaki, Y. Kamagata, Y. Sekiguchi, and W.T. Liu. 2016. Comparative genomics of syntrophic branched-chain fatty acid degrading bacteria. *Microbes Environ.* 31:288–292.
17. Nobu, M.K., T. Narihiro, H. Tamaki, *et al.* 2014. Draft genome sequence of *Syntrophorhabdus aromaticivorans* strain UI, a mesophilic aromatic compound-degrading syntroph. *Genome Announc.* 2:e01064-13.
18. Nobu, M.K., T. Narihiro, C. Rinke, Y. Kamagata, S.G. Tringe, T. Woyke, and W.T. Liu. 2015. Microbial dark matter ecogenomics reveals complex synergistic networks in a methanogenic bioreactor. *ISME J.* 9:1710–1722.
19. Nobu, M.K., T. Narihiro, H. Tamaki, *et al.* 2015. The genome of *Syntrophorhabdus aromaticivorans* strain UI provides new insights for syntrophic aromatic compound metabolism and electron flow. *Environ. Microbiol.* 17:4861–4872.
20. Oh, S., D. Yoo, and W.T. Liu. 2016. Metagenomics reveals a novel virophage population in a Tibetan Mountain Lake. *Microbes Environ.* 31:173–177.
21. Paez-Espino, D., E.A. Elie-Fadrosh, G.A. Pavlopoulos, A.D. Thomas, M. Huntemann, N. Mikhailova, E. Rubin, N.N. Ivanova, and N.C. Kyrpides. 2016. Uncovering Earth's virome. *Nature.* 536:425–430.
22. Qiu, Y.L., S. Hanada, A. Ohashi, H. Harada, Y. Kamagata, and Y. Sekiguchi. 2008. *Syntrophorhabdus aromaticivorans* gen. nov., sp. nov., the first cultured anaerobe capable of degrading phenol to acetate in obligate syntrophic associations with a hydrogenotrophic methanogen. *Appl. Environ. Microbiol.* 74:2051–2058.
23. Schink, B., and A.J.M. Stams. 2013. Syntrophism among prokaryotes, p. 471–493. *In* E. Rosenberg, E. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (ed.), *The Prokaryotes*. 4th ed. Springer-Verlag Berlin Heidelberg.
24. Schrödinger, E. 1944. *What is Life?*, Cambridge University Press, Cambridge.
25. Sieber, J.R., D.R. Sims, C. Han, *et al.* 2010. The genome of *Syntrophomonas wolfei*: new insights into syntrophic metabolism and biohydrogen production. *Environ. Microbiol.* 12:2289–2301.
26. Sieber, J.R., M.J. McInerney, and R.P. Gunsalus. 2012. Genomic insights into syntrophy: the paradigm for anaerobic metabolic cooperation. *Annu. Rev. Microbiol.* 66:429–452.
27. Takai, K. 2016. Virologists are “symbionts” in microbial ecology. *Microbes Environ.* 31:367–368.
28. Takemura, M. 2016. Draft genome sequence of *Tokyovirus*, a member of the family *Marseilleviridae* isolated from the Arakawa River of Tokyo, Japan. *Genome Announc.* 4:e00429-16.
29. Takemura, M. 2016. Morphological and taxonomic properties of *Tokyovirus*, the first *Marseilleviridae* member isolated from Japan. *Microbes Environ.* 31:442–448.
30. Tashiro, Y., H. Eida, S. Ishii, H. Futamata, and S. Okabe. 2017. Generation of small colony variants in biofilms by *Escherichia coli* harboring a conjugative F plasmid. *Microbes Environ.* 32:40–46.
31. Tsurumaru, H., T. Okubo, K. Okazaki, *et al.* 2015. Metagenomic analysis of the bacterial community associated with the taproot of sugar beet. *Microbes Environ.* 30:63–69.
32. Urayama, S., Y. Yoshida-Takashima, M. Yoshida, Y. Tomaru, H. Moriyama, K. Takai, and T. Nunoura. 2015. A new fractionation and recovery method of viral genomes based on nucleic acid composition and structure using tandem column chromatography. *Microbes Environ.* 30:199–203.
33. Urayama, S., Y. Takaki, and T. Nunoura. 2016. FLDS: a comprehensive dsRNA sequencing method for intracellular RNA virus surveillance. *Microbes Environ.* 31:33–40.
34. Watson, J.D., and F.H. Crick. 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature.* 171:737–738.
35. Wawrik, B., C.R. Marks, I.A. Davidova, M.J. McInerney, S. Pruitt, K.E. Duncan, J.M. Suflita, and A.V. Callaghan. 2016. Methanogenic paraffin degradation proceeds via alkane addition to fumarate by ‘*Smithella*’ spp. mediated by a syntrophic coupling with hydrogenotrophic methanogens. *Environ. Microbiol.* 18:2604–2619.