Indian J. Microbiol. (June 2008) 48:163–172

REVIEW

A biological treasure metagenome: pave a way for big science

So-Youn Park · Geun-Joong Kim

Received: 20 February 2008 / Accepted: 12 June 2008

Abstract The trend of recent researches, in which synthetic biology and white technology through system approaches based on "Omics technology" are recognized as the ground of biotechnology, indicates the coming of the 'metagenome era' that accesses the genomes of all microbes aiming at the understanding and industrial application of the whole microbial resources. The remarkable advance of technologies for digging out and analyzing metagenome is enabling not only practical applications of metagenome but also system approaches on a mixed-genome level based on accumulated information. In this situation, the present review is purposed to introduce the trends and methods of research on metagenome and to examine big science led by related resources in the future.

Keywords Metagenome · Gene mining · Novel metabolites · Systems approach · Biological treasure

S-Y Park[1] · G-J Kim[1] (✉)
[1]Department of Biological Sciences,
College of Natural Sciences,
Chonnam National University,
Yongbong-Dong, Buk-Gu, Gwangju 500-757,
Korea

e-mail: gjkim@chonnam.ac.kr

Introduction

Microbes have existed on earth for over 3.8 billion years, and have the most extensive bio-diversity, occupying over 60% of the total biomass. They have proliferated in the ecosystem on earth for a long age and evolved fittingly to various habitats including extreme environments. Accordingly, their taxonomical species and metabolic functions are also very diverse, and therefore they are a treasure of resources with infinite value in researches and applications [1]. Since the appearance of mankind on earth, microbes and mankind have maintained a close relationship through thousands of years. The existence of microbes was first found by Anton van Leeuwenhoek in 1673, and a large numbers of prototypes have been presented to provide bio-resources and clues for the applications of microbes in human health, agriculture and industry as well as for scientific researches on ecosystem function, global biogeochemical cycle, the origin and evolution of life [2, 3]. However, microbial species known to have led the bio-industry and the functions of ecosystem are only culturable ones estimated to be less than 1% of microbes existing in the natural niches, and most of microbes are recognized as VBNC (viable but non-culturable) species. Accordingly, new attempts to overcome the limitation of pure culture and to explore the total microbial resource have formed a research area, and the potential of related researches have induced a new paradigm shift [4, 5].

Metagenomics is a research area that studies metagenome, the mixed genomes of all microorganisms existing in a certain environment. It extracts DNA from various microbial communities and analyzes the structure and function of genomes using molecular techniques based on cloning. With the coming of the 'Omics' age in the 21 [st] century

as mentioned above, the science of analyzing and utilizing natural ecological communities is called metagenomics, ecological genomics or environmental genomics [6]. That is, metagenomics is a microbial community analysis method to approach microbial genomes, which goes beyond the limitation of pure culture technologies and thus does not rely on pure culture. Through research in this area, we can measure microbial diversity, dig out unidentified functional genes and understand bio-ecosystem through a window to microbial genome contents in a specific environment, and ultimately can explore useful genetic resources and apply them industrially and medically.

Since Torvisk [7] proposed environmental genome extraction without sorting out cells in 1980, cloning using the genomes of uncultured organisms was attempted in 1991 and related technologies have made remarkable progress. Thus, research on metagenome has been settled as a part of biological studies [8]. Metagenomic research has been activated by the development of genome manipulating technologies, and despite its short history, new functional genes, proteins and bio-active substances have been mined and comprehensive understanding has been attained on the ecology and physiology of microbes. Accordingly, it has become inevitable to readjust the scope of application in the entire range of biotechnology based on potential values of metagenome [9]. During its early period, metagenomics was mainly focused on phylogenetic analysis for measuring biodiversity, but with the development of screening and cloning techniques for obtaining genes from uncultivable strains, researchers began to produce results for various purposes. Based on such results, researches for getting more useful genes (genomes) have become common and related information has been increasing explosively [6]. In addition, huge amount of metagenome sequence information from the researches is integrated using bioinformatic research tools [10, 11].

The present study is purposed to review progress in research methods related to metagenome and the trends of research for the application of metagenome, and to discuss the movements of big science formed through approaches to the hidden microbial world.

I. Current status of metagenome researches

The diversity of biological species in metagenome is measured usually through sequence-driven analysis, and research on microbial ecological functions and screening of novel biocatalysts are performed through sequence homology based- and activity-based approaches (Fig. 1). Research on metagenome begins with the extraction of DNA from environmental samples and goes through the analysis of libraries prepared by cloning in appropriate vectors. Library preparation is the most basic means of metagenome research but, at the same time, this stage provides many hurdles and restrictions on the discovery of useful genes or metabolites. Existing strategies for library construction use a limited range of host cells based on available information, so they cannot suitably access metagenome of extensive diversity [9]. In particular, problems in expression and regulation resulting from the structural differences between known ORF and unassigned URF result very unfavorable condition in finding the activity of new resources [12]. Nevertheless, metagenome is the most promising candidate for exploring new biological resources and therefore there will be continuous studies for refining strategies and developing new methods and technologies. Given that the possibility and frequency of finding novel genes, bio-catalysts and metabolites through traditional pure culture technology are decreasing gradually, exploration of resources hidden in metagenome as a treasure of new resources is expected to provide a new breakthrough.

a. Sequence-driven analysis

Homology-based screening, which is the basis for useful sequence-based gene exploration, is a process of screening target genes based on conserved residues, motifs and domains using PCR or hybridization [13, 14]. It has found many genes that encode individual enzymes. As it relies on conserved sequence patterns, this method cannot screen non-homologous enzymes and enzymes with specific functions as a result of convergent evolution. However, when compared to activity-based screening, it is advantageous in that it can screen regardless of the gene expression in the host, namely, regardless of the genetic regulation and folding landscape of the host. Recently, beyond the probing of individual genes, attention is paid to the whole sequencing of a large clone with high probability of containing useful ORF, which also provided for the simultaneous finding of operon or gene cluster [15].

The results of phylogenetic analysis, gene prediction and annotation on the sequence of genomes extracted from environmental samples and, furthermore, the results of metabolic assembly through genome reconstruction give not only the understanding of microbial ecology and physiology but also the expectation that we can explore useful genetic resources and the whole pathway of specific compounds *in-vivo* [16]. With the development of the cloning of large genome fragments (>3-500 kb) and technology related to automated sequencing including PCR, it is now

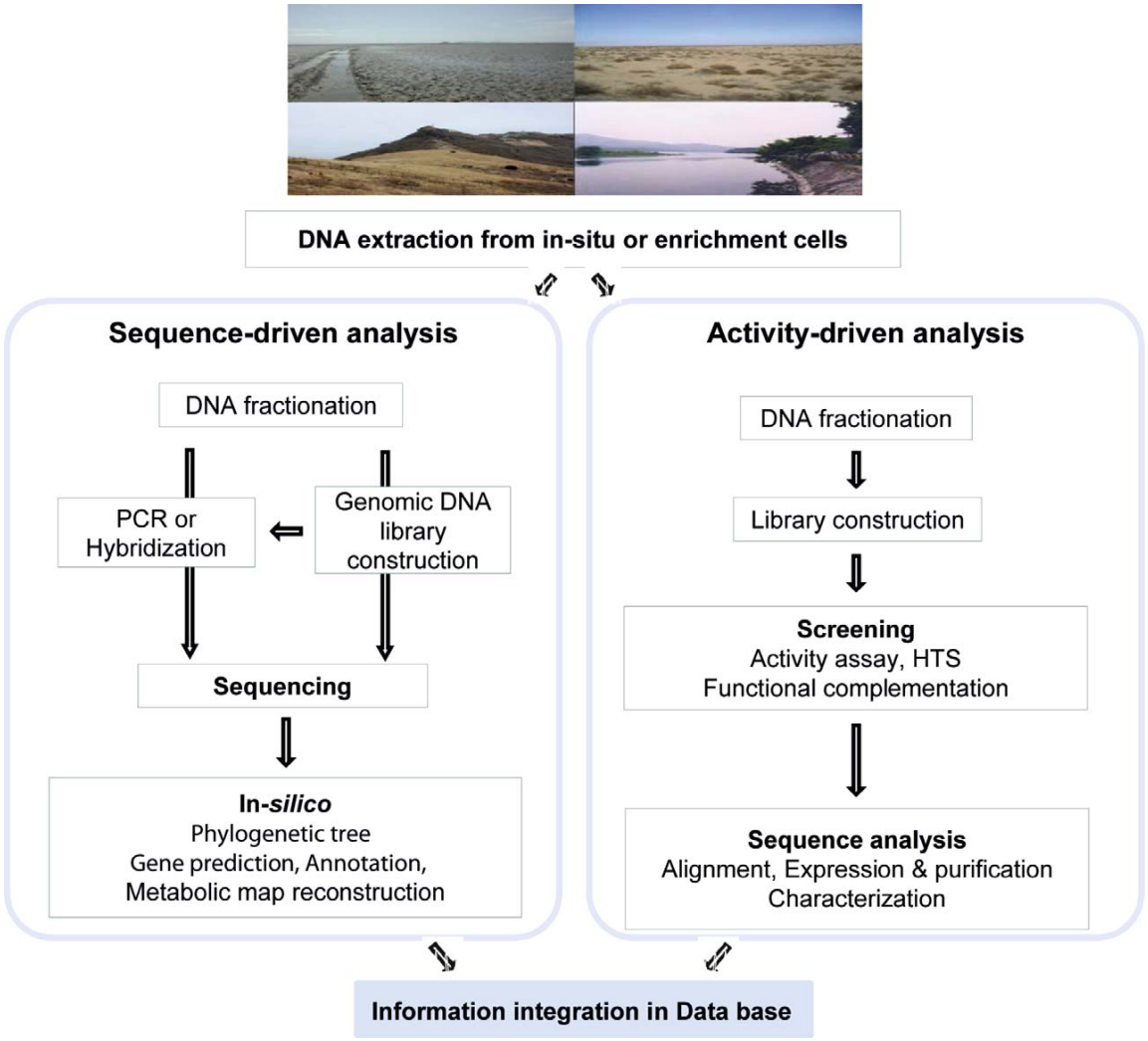

Fig. 1 Schematic representation of the steps involved in sequence- and activity-driven analyses of metagenome libraries with target sequence or activity from environmental niches. In the sequence-driven method, the probe sequence is DNA from known or putative sequence from sequence data bases such as GenBank

possible to analyze individual species in the whole community of uncultivable strains and to understand the functions of microbial communities in habitats [2, 16]. In particular, the finding of proteorhodopsin by Beja et al. clarified a major provider of energy flow in ecosystem [17], and consequently, a fundamental revision of the geochemical cycle is demanded. What is more, beginning with Tyson [18] and Venter [19], environmental shotgun sequencing such as human gut [20] and waste water sludge [21] stimulated interest in the diversity of microbial genomes and indwelling metabolic gene families, thus enabled the fundamental understanding of population diversity and community functions in specific environmental niches [22]. However, genome reconstruction in the state of mixed genomes through sequence-driven analysis has still many difficulties due to technological restrictions on the assembly system for thousands of contigs [10]. Accordingly, we need to develop new methods for gene assigning and phylogenetic classification such as Kohonen's self-organizing map (SOM) [23] and PhyloPythia [24], and to build efficient on-line public databases for storing, comparing and analyzing extensive information.

b. Activity-driven analysis

In the activity-driven method, expression is a prerequisite to screening from the shotgun or cosmid clone libraries. This method has the shortcoming that desired genetic resources have to be explored through heterologous expression, for example, the genetic trait and organization of inserts should be compatible with the host system. The codon usage and/or folding machinery should be also similar. Still, however,

it is widely attempted because it solves some of problems in the sequence-driven approach relying on the known sequence information [9]. Considering functional expression, a precondition for the industrial application of genes discovered, it is obviously an advantage that genes expressed in specific cells can be screened. However, it is also partly inevitable to screen some genes from uncultivable cells not much different from screened ones form cultivable cells in terms of novelty, sequence and functional space. With the introduction of high throughput screening (HTS) technique that can detect even extremely weak activity at the maximum efficiency, new methods are developed for fast detection of target libraries with a small amount of sample and time [25, 26], but it is worthy of consideration to use, as an alternative, defective strains from which activity itself or similar activity to be screened has been removed [27]. In this context, we may consider the use of resources from several institutions producing and supplying knock-out mutants for ORF of *E. coli* (http://cgsc.biology.yale.edu) or yeast (http://med.stanford.edu/sgtc) as hosts1. The advantage of functional complementation related to the viability of hosts is that it is possible to induce growth only with very low activity. Besides, it can be a meaningful alternative to dig out genetic resources in the concept of library (host) to library (metagenomic DNA) by cloning DNA fragments in a broad spectrum range of plasmids and using various hosts belonging to the spectrum. Another useful method is the exploration of libraries by using *in-vitro* transcription and translation kits originating from various prokaryotic/eukaryotic cells. In this case, it is possible to use codons obtained from the partial sequencing of metagenomic resources to be explored and to adjust tRNA, amino acid, ribosome and chaperone (chemical or biological) artificially [12]. Depending on cases, libraries can be constructed using the synthesis of prokaryotic cDNA that excludes the co-cloning of promoter/operator, which is a major restriction factor of protein expression in a host.

Despite the potentially infinite value of metagenome, results are achieved limitedly, for example, in enzymes and antibiotics that can be screened easily through phenotype. This suggests that there are many constraints in the screening process. Existing typical screening processes or exploration processes relying on hosts lower the expectancy of novelty in finding new genes. The discovery of new biological resources will be achievable not with existing traditional processes but with new screening processes that have novel spaces for sequence and function. However, even if the novelty or potential of unit processes developed is high, we need to use both of the two types of screening processes, namely, sequence-driven and function-driven approach.

c. Analyses and interpretation of metagenome data

Analysis of information on the genetic function and metabolic capacity of metagenome sequence demands interpretation and comprehensive data management systems (Fig. 2). Metagenome data are assembled and annotated using integrating management tools within the sequencing center. In addition, from integrated comparative analysis, we can infer the characteristics of dominant organisms in microbiomes. The specific functions and interactions between microbiomes are also understood. Data gathered in each assembler or fragmentary metagenome sequence data go through processes such as sequence annotation using a gene finder like Glimmer or Fgeresb and functional annotation like COG, Pfam, InterPro or KEGG [28]. Genes captured through these processes are compared with other known proteins in conserved motifs, domains and fold, which deriving the functional roles of the gene *in-silico*. Again with the predicted functional roles, a metabolic pathway is constructed and the metabolic capacity of the whole microbiome is estimated [29, 30]. In addition, we can analyze various aspects of a specific ecosystem such as species population and gene family through building the genome data of dominant organisms representing the microbiome in specific environment [31, 32]. After all, metagenome data will give the understanding of complex biological systems through on-line public databases and data integration using bioinformatics tools, which in turn will lead to a leap into systems biology. Although it is in its early stage, assembled genome analyzed in integrated environment is providing a window for forming more complete genomes, and is expected to reduce time and cost in finding and annotating new resources considerably [32]. In other words, complete analysis of metagenome data is partly impossible up to date, partly due to lack of interpretation systems but it is highly probable in the near future.

d. Limitation and overcome

The metagenomic DNA extraction process, which was known to be a hurdle for making metagenome libraries in the early stage, is being optimized into individual or combined processes [33] such as enzymatic lysis, bead-beating and chemo-physical method. Different from aquatic environment, activated sludge and enrichment community where it is relatively easy to isolate pure genomes, there is no generalized protocol for the extraction of genomes from soils containing complex physico-chemical elements. The reason is due to that the kind and composition of soil particles are varied according to area or ecosystem and that the amount of clay and organic humus detrimental to

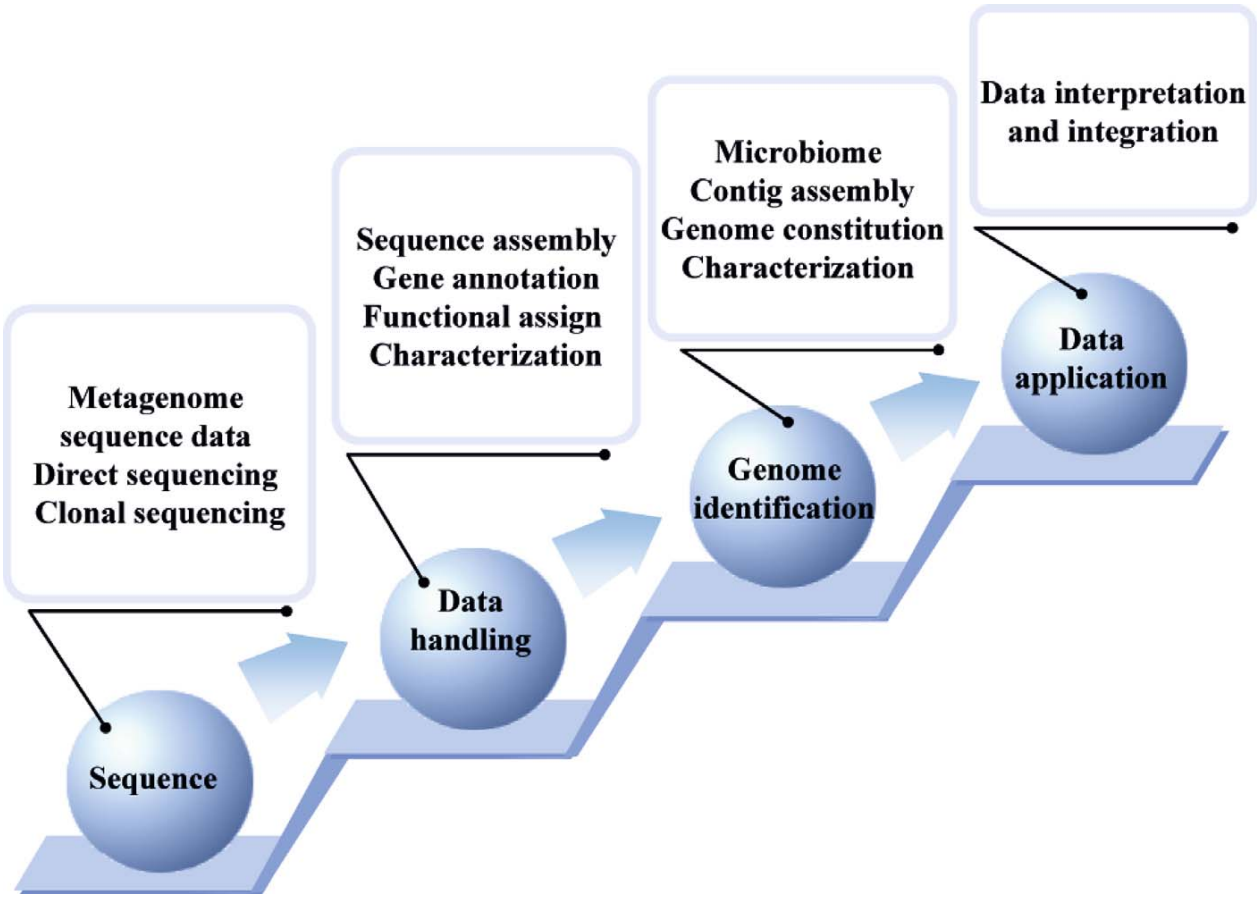

Fig. 2 The hierarchy of metagenome data processing. Information concerning the present and possible future states of the metagenomes from various sources is gathered and processed by systematically integrative systems. This information results in various fields of applications that include, but are not limited to, environmental, ecological and industrial needs, some of which also provide a clue for the origin and minimal genome of living organisms

genome extraction is relatively high in samples with desirable microbial diversity. Furthermore, in case of gram positive or minority strains, it is known hard to extract genomes through a general process of cell lysis [34, 35]. As a solution, stable isotope labeling [36, 37] or single molecule PCR [38] has been introduced. Gel filtration, agarose plugs and/or electro-elution are also being incorporated into the extraction process of metagenome but there are still many difficult problems to be solved [39]. One of them is related to the library construction process using low-purity genomes and this problem must be solved in order to increase positive hits. The existence of organic humus and phenolic compounds has an inhibition effect on PCR or the cleavage efficiency of restriction enzymes. If the sample is diluted to solve this problem, the probability of finding DNA fragments from minor strains will drop further. Thus, researchers try to obtain pure and large genomes first and then do cloning them using fosmid [40] or BAC vector [41], and subsequently do exploiting for valuable genes linked with phylogenetic or metabolic markers under condition with a reduced number of clones, but such a method should still raise the percentage of clones with insert. As a directed indicator of research on soil metagenome, the outline of the super-contig of individual genomes has begun to appear through shotgun sequencing and/or paired-end sequencing for some aquatic environments and enrichment strain pools, and the whole genome sequence of individuals in metagenome has begun to be completed [42]. Despite these technological advances, however, efforts are still continued to isolate pure genomes and make the libraries using them, to overcome the low positive hits of useful genetic resources.

As enrichment of mixed cell from ecological niches, which sacrifices diversity to some degree, is attempted as a method for enhancing the frequency of positive hits from metagenome library, we can access to active clones or target genes easily depending on cases [37,43]. Occasionally, in order to improve homologous gene probing and prediction methods relying on available sequence information,

researchers have tried RT-PCR, gene cassette PCR and microarrays for chasing a limited region of whole gene without restricting the sequence space of the other region [44,45]. In addition, the improvement of the vector system and the extent of the host spectrum range are being attempted for the efficient construction of libraries [46], which are very important in increasing the frequency of finding useful genetic resources. These efforts will increase the use of the approach based on gene expression prerequisite to screening. Recently, a HTS method has been reported, which use a reporter protein and FACS through the introduction of SIGEX (substrate-induced gene-expression screening) linked with the finding of functional promoters in specific genes or operons [47], but some problems are recognized [48].

II. The value of metagenome resources

Besides the physiological and ecological values including population diversity, dynamics and their functional roles in ecosystem that have been explained with existing methods, there is an obvious reason for obtaining useful genes or physiologically active substances from metagenome. It is because existing methods for library screening relying on pure-cultured cells has been exhausted and it is very difficult to sustain the novelty of resources originating from culturable strains. According to what is known, biological degradation and synthesis is possible for almost every organic compound that can be synthesized chemically. However, regardless of the existence of related enzymes that have been identified, the functional- and sequence spaces of resources obtainable from screening the whole of living organisms in ecosystem are still left mostly unexplored [49]. What is more, it has been proved that not only the functional protein has been resulted by divergent evolution but various active beings also resulted from convergent evolution [28]. Therefore, if hurdles in the screening process can be overcome, it will be possible to find resources in new area. Of course, it is generally known that the approaches of screening from the natural niches are compete with protein engineering technologies that mutated or fine-tuned existing genetic resources *in-vitro* or induce forced evolution [50]. However, the limitation of engineering processes in exploration of sequence space and the problem of the stepwise screening process that cannot gather effectively the concerted effect of beneficial mutation in alternative landscape may explain reasonably the strength of the exploration of metagenome originating from living organisms that already adapted biologically functional space (optimized in various landscapes) by evolutionary experience. That is, metagenome can play a significant role as a method to access new genetic resources and get desired products from the highly precise, specific and selective enzyme reaction of thousands of substrates used in industry [51].

One of major trends of research on biologically-mediated process is white biotechnology, which is to find alternatives to petrochemical compounds using renewable resources [32]. Therefore, attention is paid to the production of fossil fuels by bio-conversion or fermentation using biomass. Thus, the acquisition of regulatory genes, metabolic enzymes and gene clusters related to the production of organic acids, alcohols and solvents are also obtainable from metagenome [52, 53]. What is more, organic compounds, which have been out of people's attention for economical reasons, are again spotlighted along with their application to improved price competitiveness, low risk of environmental pollution and innovative tools of genetic engineering. We also expect a large room for the role of metagenome in increasing agricultural productivity and the utilization and reformulation of biomass [54, 55]. Besides, research on human metagenome can derive the causes of diseases and new treatment methods through the analyses of qualitative and quantitative dynamics of microbial communities [56, 57]. Also, in response to the serious side effects of drugs, increasing drug-resistant pathogens, and inefficient prevention and treatment of diseases by existing drugs, we need to find new natural inhibitors or suppressors in metagenome as a kind of antibiotics [58]. In this respect, there are many attempts to approach the new potential of metagenome resources through analyzing resistome formed naturally by biological species existing around the ecological producers of these substances (Fig. 3). It is generally believed that such an expectation can be realized by research on metagenome remained in ecosystem through countless mutations, mutual antagonism, suppression and extinction in tens of millions of microbial species for billions of years [59-61].

III. Big science to be led by metagenomics

The ecological and economical potentiality discussed above contains facts sufficient to give a role to metagenomics as a part of big science in the future. The understanding of microbial diversity gives an insight into physiological and ecological characteristics, and research on genes and their functions using complete genomes in microbial communities can give the understanding of the functions of microbial communities in ecosystem and differences among microbial communities according to habitat [3, 62]. These efforts will establish system biology as a complete set of genes over individual gene targets and allow us to go beyond

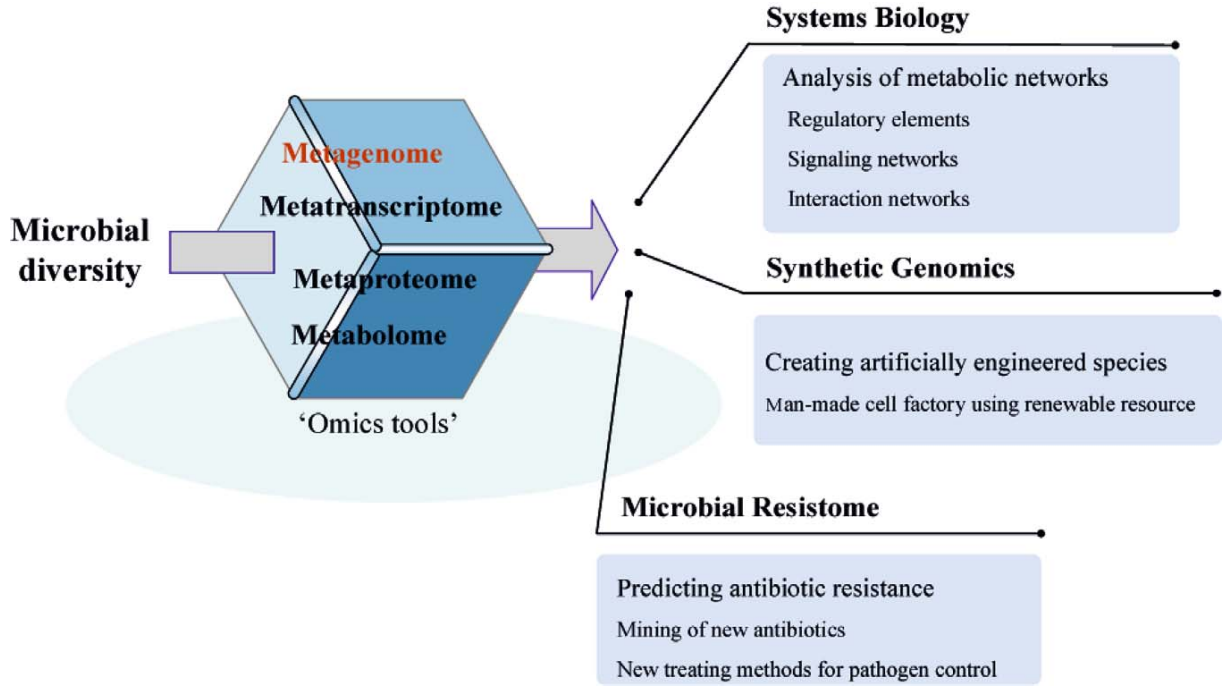

Fig. 3 Schematic representation showing a branched way for big-science of metagenome. Integrating subdivision of research areas of metagenome provides the potential of interest related with systems biology, synthetic genomics and microbial resistome

research in individual cells and to compare and analyze genome functions among individuals. The expansion from metagenome to metatranscriptome [63], metaproteome [64] and metabolom [65] will make it possible to characterize environmental qualities and find new functional genes or metabolic pathways, and give an insight into the understanding of microbial resistance, resilience and functional redundancy (Fig. 3).

Genomic data collected through metagenome will be used ultimately in creating synthetically engineered species and solving global problems such as medical services and energies. The goal of synthetic genomics is to produce cell-level bio-factories, aiming at the biological production of clean energies such as ethanol and hydrogen as future energy resources [66]. As one of such efforts, Dr. Venter created bacteriophage artificially using chemo-synthesized DNA in 2003, and 'minimal genome' is currently under research [67]. Recently it has been reported that genome transplantation between mycoplasma species was made successfully, and this shows we are in the early stage of artificial cell creation [68]. It is expected that useful substances can be synthesized efficiently by the method of making microbes as cell factories equipped with minimal but plentiful or suitable genome and then adding one or more genes for specific purpose. The assignment of speciality may partly be attainable through genes and metabolic pathways to be explored through metagenome.

Through metagemomics, scientists have obtained a new view to the microbial world different from traditional concepts and are working to overcome difficulties in future society [69]. The exhaustion of natural resources such as fossil fuels will increase people's interest in biological resources using renewable resources [70], and even just with this, metagenome is highly worthy of being studied. Microbial diversity is so extensive that it is impossible to estimate their history in the ecosystem of the planet, and even now at all of ecosystem they may continue to mutate in order to resist or adapt themselves to unceasing changes. In the aspect of the understanding of microbes, we may not be able to access the entire microbe resource even if a century passes and mankind advances deep into the universe through remarkable development of technologies. That is, even after the lapse of the 21 [st] century, metgenomics will still be a current science rather than a past science. With this fact, we may be destined to pay respect to organic matters, especially to living things.

Acknowledgement This study was financially supported by Chonnam National University.

References

1. Hunter-Cevera JC (1998) The value of microbial diversity. Curr Opin Microbiol 1:278–285

2. Eisen JA (2007) Environmental shotgun sequencing: its potential and challenges for studying the hidden world of microbes. PLoS Biol 5:e82
3. Zwolinski MD (2007) DNA sequencing: Strategies for soil microbiology. Soil Sci Am J 71:592–600
4. Handelsman J (2004) Metagenomics: application of genomics to uncultured microorganisms. Microbiol Mol Biol Rev 68:669–685
5. Bull AT, Ward AC and Goodfellow M (2000) Search and discovery strategies for biotechnology: The paradigm shift. Microbiol Mol Biol Rev 64:573–606
6. Xu JP (2006) Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. Mol Ecol 15:1713–1731
7. Torsvik VL (1980) Isolation of bacterial DNA from soil. *Soil Biol Biochem* 12:15–21
8. Ferrer M, Beloqui A and Golyshin PN (2007) Microbial metagenomes: moving forward industrial biotechnology. J Chem Technol Biotechnol 82: 421–423
9. Ferrer M, Martinez-Abarca F and Golyshin PN (2005) Mining genomes and 'metagenomes' for novel catalysts. Curr Opin Biotechnol 16:588–593
10. Kowalchuk GA, Speksnijder AG, Zhang K, Goodman RM and van Veen JA (2007) Finding the needles in the metagenome haystack. Microb Ecol 53:475–485
11. Field D and Kyrpides N (2007) The positive role of the ecological community in the genomic revolution. Microb Ecol 53:507–511
12. Park SH, Cheong DE, Lee JY, Han SS, Lee JH and Kim GJ (2007) Analyses of the structural organization of unidentified open reading frames from metagenome. Biochem Biophys Res Commun 356:961–967
13. Kim BS, Kim SY, Park J, Park W, Hwang KY, Yoon YJ, Oh WK, Kim BY and Ahn JS (2007) Sequence-based screening for self-sufficient P450 monooxygenase from a metagenome library. J Appl Microbiol 102:1392–400
14. Jiang C and Wu B (2007) Molecular cloning and functional characterization of a novel decarboxylase from uncultured microorganisms. Biochem Biophys Res Commun 357: 421–426
15. Brady SF, Bauer JD, Clarke-Pearson MF and Daniels R (2007) Natural products from isnA-containing biosynthetic gene clusters recovered from the genomes of cultured and uncultured bacteria. J Am Chem Soc 129:12102–12103
16. Price ND, Papin JA, Schilling CH and Palsson BO (2003) Genome-scale microbial *in silico* models: the constraints-based approach. Trends Biotechnol 21:162–169
17. Beja O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP, Jovanovich SB, Gates CM, Feldman RA, Spudich JL, Spudich EN and DeLong EF (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. Science 289:1902–1906
18. Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS and Banfield JF (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428:37–43
19. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers YH and Smith HO (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science 304:66–74
20. Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM and Nelson KE (2006) Metagenomic analysis of the human distal gut microbiome. Science 312:1355–1359
21. Garcia Martin H, Ivanova N, Kunin V, Warnecke F, Barry KW, McHardy AC, Yeates C, He S, Salamov AA, Szeto E, Dalin E, Putnam NH, Shapiro HJ, Pangilinan JL, Rigoutsos I, Kyrpides NC, Blackall LL, McMahon KD and Hugenholtz P (2006) Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. Nat Biotechnol 24:1263–1269
22. Koonin EV (2007) Metagenomic sorcery and the expanding protein universe. Nat Biotechnol 25:540–542
23. Abe T, Sugawara H, Kinouchi M, Kanaya S and Ikemura T (2005) Novel phylogenetic studies of genomic sequence fragments derived from uncultured microbe mixtures in environmental and clinical samples. DNA Res 12:281–290
24. McHardy AC, Martin HG, Tsirigos A, Hugenholtz P and Rigoutsos I (2007) Accurate phylogenetic classification of variable-length DNA fragments. Nat Methods 4:63–72
25. Jean-Philippe G and Jean-Louis R (2004) Enzyme assays for high-throughput screening. Current Opinion in Biotechnology 15:314–322
26. Hu Y, Chen GYJ and Yao SQ (2005) Activity-based high-throughput screening of enzymes by using a DNA microarray. Angew Chem Int Ed 44:1048–1053
27. Li YG, Wexler M, Richardson DJ, Bond PL and Johnston AW (2005) Screening a wide host-range, waste-water metagenomic library in tryptophan auxotrophs of *Rhizobium leguminosarum* and of *Escherichia coli* reveals different classes of cloned trp genes. Environ Microbiol 7: 1927–1936
28. Lee D, Redfern O and Orengo C (2007) Predicting protein function from sequence and structure. Nat Rev Mol Cell Biol 8:995–1005
29. Markowitz VM, Ivanova N, Palaniappan K, Szeto E, Korzeniewski F, Lykidis A, Anderson I, Mavromatis K, Kunin V, Garcia Martin H, Dubchak I, Hugenholtz P and Kyrpides NC (2006) An experimental metagenome data management and analysis system. Bioinformatics 22:e359–367
30. Mavromatis K, Ivanova N, Barry K, Shapiro H, Goltsman E, McHardy AC, Rigoutsos I, Salamov A, Korzeniewski F, Land M, Lapidus A, Grigoriev I, Richardson P, Hugenholtz P and Kyrpides NC (2007) Use of simulated data sets to evaluate the fidelity of metagenomic processing methods. Nat Methods 4:495–500
31. Havre SL, Webb-Robertson BJ, Shah A, Posse C, Gopalan B and Brockman FJ (2005) Bioinformatic insights from metagenomics through visualization. Proc IEEE Comput Syst Bioinform Conf, pp 341–350
32. Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, McHardy AC, Djordjevic G, Aboushadi N, Sorek R, Tringe SG, Podar M, Martin HG, Kunin V, Dalevi D, Madejska J, Kirton E, Platt D, Szeto E, Salamov A, Barry K, Mikhailova N, Kyrpides NC, Matson EG, Ottesen EA, Zhang X, Hernández M,

Murillo C, Acosta LG, Rigoutsos I, Tamayo G, Green BD, Chang C, Rubin EM, Mathur EJ, Robertson DE, Hugenholtz P and Leadbetter JR (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature 450:560–565

33. Krsek M and Wellington EM (1999) Comparison of different methods for the isolation and purification of total community DNA from soil. J Microbiol Methods 39:1–16
34. Frostegard A, Courtois S, Ramisse V, Clerc S, Bernillon D, Le Gall F, Jeannin P, Nesme X and Simonet P (1999) Quantification of bias related to the extraction of DNA directly from soils. Appl Environ Microbiol 65:5409–5420
35. Miller DN, Bryant JE, Madsen EL and Ghiorse WC (1999) Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. Appl Environ Microbiol 65:4715–4724
36. Neufeld JD, Dumont MG, Vohra J and Murrell JC (2006) Methodological considerations for the use of stable isotope probing in microbial ecology. Microb Ecol 53: 435–442
37. Schloss PD and Handelsman J (2003) Biotechnological prospects from metagenomics. Curr Opin Biotechnol 14: 303–310
38. Ohuchi S, Nakano H and Yamane T (1998) *In vitro* method for the generation of protein libraries using PCR amplification of a single DNA molecule and coupled transcription/translation. Nucleic Acids Res 26:4339–4346
39. Yu WH, Su SC and Lee CY (2008) A novel retrieval system for nearly complete microbial genomic fragments from soil samples. J Microbiol Methods. 72:197–205
40. Nesbø CL, Boucher Y, Dlutek M and Doolittle WF (2005) Lateral gene transfer and phylogenetic assignment of environmental fosmid clones. Environ Microbiol 7:2011–2026
41. Berry AE, Chiocchini C, Selby T, Sosio M and Wellington EM (2003) Isolation of high molecular weight DNA from soil for cloning into BAC vectors. FEMS Microbiol Lett 223:15–20
42. Huson DH, Auch AF, Qi J and Schuster SC (2007) MEGAN analysis of metagenomic data. Genome Res 17:377–386
43. Gabor EM, de Vries EJ and Janssen DB (2004) Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. Environ Microbiol 6:948–958
44. Cowan D, Meyer Q, Stafford W, Muyanga S, Cameron R and Wittwer P (2005) Metagenomic gene discovery: past, present and future. Trends Biotechnol 23:321–329
45. Galvão TC, Mohn WW and de Lorenzo V (2005) Exploring the microbial biodegradation and biotransformation gene pool. Trends Biotechnol 23:497–506
46. Terpe K (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol 72:211–222
47. Uchiyama T, Abe T, Ikemura T and Watanabe K (2005) Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. Nat Biotechnol 23:88–93
48. Yun J and Ryu S (2005) Screening for novel enzymes from metagenome and SIGEX, as a way to improve it. Microb Cell Fact 4:8
49. Lorenz P, Liebeton K, Niehaus F and Eck J (2002) Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. Curr Opin Biotechnol 13:572–577
50. Petri R and Schmidt-Dannert C (2004) Dealing with complexity: evolutionary engineering and genome shuffling. Curr Opin Biotechnol 15:298–304
51. Oh HJ, Cho KW, Jung IS, Kim WH, Hur BK and Kim GJ (2003) Expanding functional spaces of enzymes by utilizing whole genome treasure for library construction. J Mol Catal B-enzymatic 26: 241–250
52. Lorenz P and Eck J (2005) Metagenomics and industrial applications. Nat Rev Microbiol 3:510–516
53. Ferrer M, Beloqui A and Golyshin PN (2007) Microbial metagenomes: moving forward industrial biotechnology. Chem Technol Biotechnol 82:421–423
54. Welbaum GE, Sturz AV, Dong ZM and Nowak J (2004) Managing soil microorganisms to improve productivity of agro-ecosystems. Crit Rev Plant Sci 23:175–193
55. Tsai SH, Liu CP and Yang SS (2007) Microbial conversion of food wastes for biofertilizer production with thermophilic lipolytic microbes. Renew Energy 32:904–915
56. Dethlefsen L, McFall-Ngai M and Relman DA (2007) An ecological and evolutionary perspective on human-microbe mutualism and disease. Nature 449:811–818
57. Zoetendal EG, Vaughan EE and de Vos WM (2006) A microbial world within us. Mol Microbiol 59:1639–1650
58. Martínez JL, Baquero F and Andersson DI (2007) Predicting antibiotic resistance. Nat Rev Microbiol 5:958–965
59. Courtois S, Cappellano CM, Ball M, Francou FX, Normand P, Helynck G, Martinez A, Kolvek SJ, Hopke J, Osburne MS, August PR, Nalin R, Guérineau M, Jeannin P, Simonet P and Pernodet JL (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. Appl Environ Microbiol 69:49–55
60. Li X and Qin L (2005) Metagenomics-based drug discovery and marine microbial diversity. Trends Biotechnol 23: 539–543
61. D'Costa VM, Griffiths E and Wright GD (2007) Expanding the soil antibiotic resistome: exploring environmental diversity. Curr Opin Microbiol 10:481–489
62. Pontes DS, Lima-Bittencourt CI, Chartone-Souza E and Amaral Nascimento AM (2007) Molecular approaches: advantages and artifacts in assessing bacterial diversity. J Ind Microbiol Biotechnol 34:463–473
63. Maron PA, Ranjard L, Mougel C and Lemanceau P (2007) Metaproteomics: a new approach for studying functional microbial ecology. Microb Ecol 53:486–493
64. Cook KL and Sayler GS (2003) Environmental application of array technology: promise, problems and practicalities. Curr Opin Biotechnol 14:311–318
65. Mashego MR, Rumbold K, De Mey M, Vandamme E, Soetaert W and Heijnen JJ (2007) Microbial metabolomics: past, present and future methodologies. Biotechnol Lett 29: 1–16
66. Danchin A (2004) The bag or the spindle: the cell factory at the time of systems' biology. Microb Cell Fact 3:13
67. Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, Hutchison CA 3rd, Smith HO and Venter JC

(2006) Essential genes of a minimal bacterium. Proc Natl Acad Sci USA 103:425–430

68. Lartigue C, Glass JI, Alperovich N, Pieper R, Parmar PP, Hutchison CA 3rd, Smith HO and Venter JC (2007) Genome transplantation in bacteria: changing one species to another. Science 317:632–638

69. Gabor E, Liebeton K, Niehaus F, Eck J and Lorenz P (2007) Updating the metagenomics toolbox. Biotechnol J 2:201–206

70. Atsumi S, Hanai T and Liao JC (2008) Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature 451:86–89