Science in China Series C: Life Sciences

www.scichina.com
life.scichina.com
www.springerlink.com

Special Topics

Research progress in genomics of environmental and industrial microorganisms

WANG Lei[1,2†], LIU Bin[1,2] & ZHOU ZheMin[1,2]

[1] TEDA School of Biological Sciences and Biotechnology, Nankai University, Tianjin 300457, China;
[2] Tianjin Research Center for Functional Genomics and Biochip, Tianjin 300457, China

Microbes contribute to geochemical cycles in the ecosystem. They also play important roles in biodegradation and bioremediation of contaminated environments, and have great potential in energy conversion and regeneration. Up to date, at least 150 genomes of non-pathogenic microbes have been sequenced, of which, the majority are bacteria from various environments or of industrial uses. The emerging field 'metagenomics' in combination with the high-throughput sequencing technology offers opportunities to discover new functions of microbes in the environment on a large scale, and has become the 'hot spot' in the field of environmental microbiology. Seven genomes of bacteria from various extreme environments, including high temperature, high and low pressure, and extreme acidic regions, have been sequenced by researchers in China, leading to the discovery of metabolic pathways, genetic functions and new enzymes, which are related to the niches those bacteria occupy. These results were published in *Nature*, *PNAS*, *Genome Research* and other top international journals. In the meantime, several groups in China have started 'metagenomics' programs. The outcomes of these researches are expected to generate a considerable number of novel findings, taking Chinese researchers to the frontier of genomics for environmental and industrial microorganisms.

genomics, microbiology and metagenomics

The environmental microorganism is one of the key factors in maintaining the circulation of the energy and materials in the ecosphere. Not only does it play an important role in the degradation of pollutants and injurants, but it also has potential application value in the research field of energy production and reproducible utilization. Among more than 150 sequenced genomes of avirulent microorganism, most are environmental and industrial bacteria. With the use of the new generation sequencing technique and the appearance of the method of Metagenomics, the researches on microorganic genomes in the world have stepped into the stage of high-flux and high-throughput. In China, the 7 environmental bacteria of which genomes have been sequenced cover the bacteria in all kinds of extreme environments, such as thermophilic, acidophilic, highly pressurized and low-pressurized environments. From them, the researchers have found many metabolic pathways, genetic functions and important enzymes which have a close relationship with the environmental and industrial application. At present, many enterprises in China have started the metagenomic program, and it is certain that they will bring a large number of original results to the field of the genomics of environmental and industrial microorganisms.

1 Generality

There are a lot of invisible microorganisms living in our surroundings. Researches have found that there are more

Received October 3, 2008; accepted October 8, 2008
doi: 10.1007/s11427-009-0013-8
†Corresponding author (email: wanglei@nankai.edu.cn)
Supported by the National High-tech Research Development Program of China (Grant No. 2007AA02Z106, 2007AA021303 and 2007AA020703) and the National Natural Science Foundation of China (Grant No. 30530010)

than 1 billion[1] different bacteria in a gram of the soil, of which 99.8% are unknown to people[2] because they can not be cultured in the lab. We did not have the ability to carry out further researches on these lives until the development of the genomics research.

The environmental microorganism is defined to be all the microorganisms which exist in the soil, the water, the atmosphere and between the rock stratums, even those living in human-beings, animals and plants. Many microorganisms can live in all kinds of extreme environments. The environmental microorganism is the key factor in maintaining the circulation of the energy and the materials in the ecosphere. Due to the diversity of the metabolism of the environmental bacteria, the natural circulation of the basic elements of many lives all needs the role of bacteria. For example, alcoholization and nitrogen fixation are the specific metabolic process of some bacteria, so they play a key role in the circulation of carbon and nitrogen. The bacteria can also make many particular materials as the oxidizer, such as H_2S or a variety of aromatics materials as energy[3,4].

The environmental microorganisms have many uses, including biodegradation of sorts of pollutants (such as petroleum) and other hazardous material; researching and producing new products which can cure or prevent illness; the energy production and the research on the regenerative energy resource (methane and hydrogen); producing chemical catalyzers, reagents and enzymes to improve the efficiency of the process of the industrial production; governing the emission of the greenhouse gases (such as carbon dioxide) in the environment.

The traditional sequencing method costs a large amount of money and time, so people put much focus on the pathogenic bacteria which have a close relationship with human-beings. However, the genomes of more than 150 non-pathogenic microorganisms have been sequenced, and most of them are environmental and industrial bacteria, such as the sequencing of *Pyrobaculum aerophilum*[5], which was carried out jointly by the scientists from University of California at Los Angeles, California Institute of Technology and University of Regensburg. *P. aerophilum* can live in the seawater where the temperature is above 100℃. However, it is not an obligate anaerobe, and cannot bear the existence of sulfur, which is different from its close relatives in evolution. The research team finished the sequencing of its 2.2 Mb genome and recognized 2587 protein coding regions. Finally the reason for its resistance to strikingly high temperature was found to be its lack of the mismatch repair system, which can be used to repair the DNA mutation and is present in both *Escherichia coli* and human-beings as an universal system. But this mechanism is not found in the 6 *P. aerophilums* which have been sequenced. This result increases the possibility that these bacteria are living in a mutation way, that is, living by going through unrepaired mutations successively. As the author expressed in the article, these high temperature-resistant primitive organisms may serve as a good example to prove that organisms may live in the form of perpetual mutants.

Another similar case is the H_2-Oxidizing lithoautotrophic bacterium *Ralstonia eutropha* H16, which is a metabolically versatile organism capable of subsisting, in the absence of organic growth substrates, on H_2 and CO_2 as its sole sources of energy and carbon. The bacterial strain *R. eutropha* H16 has the ability to produce and store a large amount of PHB and other high molecular polymers, which can be harnessed to produce biodegradable plastics, and thus gain much attention. In 2006, the German scientists finished the genomic sequencing of *R. eutropha*, which made people further understand the potential application of engineering bacteria[6].

With the rapid development of industries, water pollution is becoming more serious, and marine pollution is drawing worldwide attention. Of all marine pollutants, oil pollution is the most universal and serious. *Alcanivorax borkumensis*, a kind of global marine microorganism, can use petroleum hydrocarbon as the sole carbon source and energy. *A. borkumensis* can be hardly detected in an unpolluted area, but it will quickly occupy the polluted area and become the advantageous bacterium once an oil pollution happens. The scientists from the University of Bielefeld studied the genome of *A. borkumensis* SK2, and found that its genome has a lot of genes related with the degradation of petroleum hydrocarbon, and a system to clean out marine nutritive salt, especially organic and inorganic nitrogen. Additionally, *A. borkumensis* can form a biofilm at the oil-water interface and produce biosurfactants on the organism surface. All these abilities in combination have made *A. borkumensis* SK2 become an advantageous bacterium in an oil-polluted area[7], and research on *A. borkumensis* SK2 will lay a foundation for alleviating the environmental pollution caused by petroleum leakage.

The application of the new sequencing technique has greatly improved the speed of genome sequencing. The emerging metagenomics method also greatly expanded the scope of microorganisms that human can study[8]. Metagenomics is a method of getting all the genomes of microorganism from the niche for a systematic analysis. At present, many researches on microorganism genomics in the world take metagenomics as a key research method. The "Sorcerer II Expedition", the plan for genomic sequencing of marine microorganism, which is led by the J. Craig Venter Institute, has chosen 17 marine surface water samples which are provided by 79 colleagues worldwide for metagenomics research. With these samples, they got the data of 6.3 billions bp. 6.12 million new protein sequences, and more than 2000 new protein families have been annotated. By March, 2008, the "Sorcerer II Expedition" has published 126 complete genomic datasets. These researches have greatly improved people's understanding of the biodiversity in marine life[9-12].

It is well known that termites have caused billions of dollars of economic loss each year for their characteristic of eating the woods. The recent research found that it may provide a high-efficient biochemical method for the production of green biofuels. The alimentary canal of this small insect may be a "gold mine" of microorganism flora and may become the source of the enzyme which is greatly needed in the process of improving the wood quality or in the process of turning the waste into the biologic fuel. The Joint Genome Institute of American Energy Department and many other research centers have jointly sequenced the genome sequences of the flora in the intestine of the termite, and the research result was released in *Nature* on Nov. 22, 2007[4].

U.S. Department of Energy has a more enormous plan. In May 2007, the Joint Genome Center (JGI) of U.S. Department of Energy brought forward the plan of "Genomics Encyclopedia of Bacteria and Archaebacteria" (GEBA). The GEBA project is aimed at systematically filling the gaps in sequencing along the bacterial and archaeal branches of the tree of life. This project represents the first systematic attempt to use the tree of life itself as a guide to sequencing target selection. To test the feasibility of the GEBA approach, JGI is undertaking a pilot project in collaboration with DSMZ to sequence 100 bacterial and archaeal genomes based on the phylogenetic positions of organisms in the tree of life.

2 The development of environmental and industrial micro-organisms genomics research in China

The bacteria Genome Project in China started very late. In 2002, Professor JIN Qi's research group[13] finished the first bacterial genome sequencing project(*Shigella* Genome Project). At the beginning, China also paid most of their attention to the pathogen genome; however, the Bacteria Genome Project in China developed so fast that many whole bacterial genome projects which had important environmental and industrial significance had been put on the agenda successively. There are vast territories and rich natural resources in the landscapes in China which can be used as unique biodiversity sources of environmental bacteria. By taking advantage of the distinctive resources of China, the scientific community has carried out researches on all kinds of microorganisms which can adapt to the extreme environment, and the 7 environmental bacteria whose genomes have been sequenced cover the bacteria in all kinds of extreme environments such as thermophilic, acidophilic, highly pressurized and low-pressurized environments. Due to the characteristics in their metabolization and genetics, the researchers can find a lot of distinctive metabolic pathways, genetic functions and enzymes, and combine them closely with their applications in industry.

The *Geobacillus thermodenitrificans* NG80-2[14] and the *Bacillus cereus* Q1 containing the Alkane-degrading enzyme play a key role in preventing petroleum pollution and in the process of oil production. NG80-2 has the distinctive ability to degrade heavy oil, which was isolated from the Dagang Oilfield. Using the advanced methods in the fields of functional genomics, genomics, bioinformatics, molecular biology, biochemistry and so on, the scientists in the Nankai University carried out in-depth research on NG80-2, and uncovered the metabolic pathway of microbial degradation of long-chain alkanes, which is the major component of heavy oil, for the first time in the world, and studied the function of the long-chain alkanes monooxygenase gene *ladA* which encode the key protein in degradation pathway of long-chain alkanes. Such achievement is a great breakthrough in the study of petroleum microorganisms, and has important scientific and application meanings to treatment of oil pollutions and improvement of biotechnology in oil production. The determination of the ge-

nomes of *Thermoanaerobacter tengcongensis*[16] and *Shewanella piezotolerans* WP3[17] have deepened people's understanding of the physiology and the genetic features of the microorganisms in the extreme environment. The genomics research on *Methylacidiphilum infernorum* also deepened people's understanding of the taxonomy, ecology, and genetic diversity of the methanotrophic bacteria, and has provided the basis[18,19] for the further control and utilization of natural methane. The genomics research on *Ketogulonigenium* sp. and the research on the key genes which have close relationship with the metabolism of *Ketogulonigenium* sp. have provided the theoretical basis for the research of the one-step production of Vitamin C.

Though the environmental microbial genome research in China is still in the development stage, it has occupied a place in the world. Many research results have been published in many first-classed internal journals, such as *Nature*, *PNAS*, and *Genome Research* and these achievements won favorable comments from the internal academic circles. With the development of the new sequencing technique, the advantages of the microorganism resources in China will be an aid to furthering research, and at the same time, many sequencing centers have started the metagenomics project. These new developments will result in a great increase in research results in the field of the environmental and industrial micro-organisms genomics.

2.1 The genomics and proteomics research of the degraded bacteria of the long-chain alkanes NG80-2

Members of *Geobacillus* have been isolated from various terrestrial and marine environments, not only in geothermal areas, but also in temperate regions and permanently cold habitats, demonstrating their great capabilities for adaptation to a wide variety of environmental niches. *Geobacillus* spp. have attracted industrial interest for their potential applications in biotechnological processes as sources of various thermostable enzymes.

Alkanes are the major components of crude oils and are commonly found in oil contaminated environments. Biotechnological applications for microbial degradation of those pollutants are of long-standing interest. Although long-chain alkanes are more persistent in the environment than shorter-chains, only genes involved in the degradation of alkanes up to C16 have been well studied, and those for longer alkanes have not yet been reported.

G. thermodenitrificans NG80-2 was isolated from a deep-subsurface oil reservoir in Dagang oilfield, Northern China. It grows between 45 ℃ and 73℃ (optimum 65℃) and can utilize long-chain (C15-C36) alkanes as a sole carbon source. NG80-2 can produce a large amount of emulsifiers and has a very good emulsification effect on crude oil.

On the basis of finishing the whole-genome sequencing of NG80-2, Professor Wang Lei, together with his group in TEDA School of Biological Sciences and Biotechnology in Nankai University, made in-depth analysis of NG80-2 and discovered the degradation pathway of the long-chain alkanes in microorganism for the first time. Their research result was published in *PNAS*[14] in March, 2007.

The genome of *G. thermodenitrificans* NG80-2 is composed of a 3550319 base pairs (bp) chromosome and a 57693 bp plasmid. There are 3499 predicted open reading frames (ORFs), 11 rRNA operons and 87 tRNA genes, covering 86% of the genome. Putative functions were assigned to 2479 ORFs. Of the remainder, 757 showed similarity to hypothetical proteins, and 263 had no detectable homologs in the public protein databases. There are 68 putative transposase genes in intact or mutated forms.

The bioinformatics analysis revealed all the potential genes that may be involved in the degradation of alkanes. Then these genes were further screened by transcriptomic and proteomic methods, such as real time RT-PCR, 2-D electrophoresis and MALDI-TOF MS analysis. Finally the degradation pathway of the long-chain alkane including the long-chain alkane hydroxylase (monooxygenase), the alcohol dehydrogenase, the aldehyde dehydrogenase, the acyl-CoA ligase and the multi-enzymes of the β-oxidation pathway, was determined. For example, RT-PCR showed a 120-fold increase in transcription of a plasmid-borne putative monooxygenase gene (GT3499) when crude oil was used as a sole carbon source instead of sucrose, and the further functional experiment confirmed that GT3499 is the long-chain alkane monooxygenase gene. GT3499 was named *ladA*. Using sucrose-grown cells as the reference state, proteins differentially expressed in hexadecane-grown cells were investigated by 2-D electrophoresis and MALDI-TOF MS analysis. Expression of LadA was found to be induced and mainly found in extracellular fraction. And

the pathway involved in hexadecane metabolism in NG80-2 was determined. LadA is a thermophilic enzyme, which offers the major biotechnological advantage over the mesophilic enzyme. LadA is a single-component with no coenzyme requirement, soluble (extracellular), and easily expressed and purified in *E. coli*. Therefore, LadA has great potential to be used in treatment of environmental oil pollutions.

As for the analysis of general metabolism of NG80-2 and its adaptation to oil reservoirs, genes required for the synthesis of purine and pyrimidine nucleotides, fatty acids, and all 20 amino acids were identified. Complete sets of genes for the synthesis of all vitamins and cofactors are present except for biotin. However, a putative biotin transporter gene was found, and the requirement of exogenous biotin for growth of NG80-2 was confirmed. All central metabolic pathways for carbohydrates except for the Entner-Doudoroff pathway are present. NG80-2 utilizes a large variety of carbohydrates such as glycerol, cellobiose, trehalose and starch.

Genome analysis further revealed the presence of a gene cluster for utilization of plant hemicellulose xylans, and the ability of NG80-2 to utilize xylans as a sole carbon source was confirmed. NG80-2 has a large number of transporter genes for nutrient absorption and detoxification, which makes it adapt well to the oil reservoir.

In oil reservoirs, oxygen is only transiently available during water flushing used for oil production. Therefore, a flexible respiration system for quick response to changed O_2 concentration is very important for the survival of bacteria in reservoirs. The analysis revealed that *G. thermodenitrificans* is a facultative aerobe, capable of oxygen and nitrate respiration. Its aerobic respiration system contains 5 terminal oxidases. Its anaerobic respiration system can deoxidize the N_2O to N_2, and contains two sets of gene clusters containing membrane-bound nitrate reductase genes.

As for thermotolerance, mesophilic bacteria respond to heat induced stresses by induction of heat shock proteins, which removes or refolds damaged proteins. NG80-2 contains a wide range of genes encoding molecular chaperones including *dnaK* operon, genes encoding GroEL-GroES, a disulfide bond chaperone of HSP33 family, genes encoding small heat shock proteins of IbpA family, and the genes encoding ATP-dependent heat shock-responsive proteases such as HslVU, Clp, and Lon. NG80-2 has three genes encoding proteins of the 58 COG families whose members are frequently detected in archaea and thermophilic bacteria and predicted to be associated with the (hyper) thermophilic phenotype. Polyamines are commonly found in hyperthermophilic bacteria. Both NG80-2 and *G. kaustophilus* have genes encoding spermine/spermidine synthase and polyamine ABC transporters. Similar to *G. kaustophilus*, asymmetric amino acid substitutions was found in NG80-2, which may be associated with its thermoadaptation.

The presence of genes involved in utilization of xylans and degradation of oligopeptides confirms that NG80-2 originated from a soil environment. It is suggested that NG80-2 gained the capacity for alkane oxidization using plasmid pLW1071 in an oil contaminated soil before invading the oil reservoir.

2.2 The Genomics Research of *Thermoanaerobacter tengcongensis*

Thermoanaerobacter tengcongensis, isolated from a hot spring in Tengchong, Yunnan, China, is a rod-shaped, gram-negative bacterium that grows anaerobically under the extreme environment. It propagates at temperatures ranging from 50℃ to 80℃ (optimally at 75℃) and at pH values ranging between 5.5 and 9 (optimally from 7 to 7.5). The analysis of 16S rDNA indicated that it belongs to genus *Thermoanaerobacter*.

T. tengcongensis, however, has several important phenotypic properties that contradict its membership to the genus. Some of the examples include the absence of spore production, negative gram-staining result, lack of motility under cultural conditions, and exclusive metabolic pathways (such as deficiencies in lactate production and xylan utilization). To obtain a global view of genes possessed by the organism and to resolve some of the controversies at the molecular levels, as well as to understand the biology of thermophilic prokaryotes in general and find interesting thermophilic enzymes, Academician YANG Huanming, together with his group[16] in the Institute of Genetics and Developmental Biology of Chinese Academy of Sciences, sequenced and analyzed the genome *T. tengcongensis*. MB4T.

Using a whole-genome-shotgun method, the 2689 445-bp genome of MB4T was obtained. The genome encodes 2588 predicted coding sequences (CDS). Among them, 1764 (68.2%) are classified according to homology to other documented proteins, and the rest, 824 CDS (31.8%), are functionally unknown. One of the

interesting features of the *T. tengcongensis* genome is that, it has the most biased gene distribution on the leading strand, in the same direction as genome replication, among all sequenced prokaryotic genomes at that time. 86.7% of its genes are encoded on the leading strand of DNA replication. Based on protein sequence similarity, the *T. tengcongensis* genome is most similar to that of *Bacillus halodurans*, a mesophilic eubacterium.

Computational analysis on the genes involved in basic metabolic pathways supports the experimental discovery that *T. tengcongensis* metabolizes sugars as principal energy and carbon source and utilizes thiosulfate and element sulfur, but not sulfate, as electron acceptors. Such an observation seems to contradict a common feature observed in most sulfur-respiratory prokaryotes. *T. tengcongensis* has a complete set of genes constituting the glycolysis and the pentose phosphate pathways. Neither the genes related to sulfate transport systems, nor the key genes involved in the sulfate reduction are present. Secondly, in the reduction process, thiosulfate is generally reduced to sulfite and further to sulfide. Thiosulfate reductase and sulfite reductase, which play crucial roles in these steps, are not found in the *T. tengcongensis* genome. Instead, a rhodanese-related sulfurtransferase, which employs thiosulfate as electron acceptor in the presence of cyanideion, is identified.

Genome sequence analysis showed that *T. tengcongensis* does appear to be well equipped with all essential genes for flagellar biogenesis and with nearly all the genes for the chemotaxis signaling pathways. However, it remains puzzling why *T. tengcongensis* does not assemble functional flagellar under the culture conditions.

The author suggested that these "silent" genes involved in flagellar structure in *T. tengcongensis* might be activated only under certain environmental conditions or they used to be active not long before the present day.

T. tengcongensis, which is a gram-negative rod by staining, shares many genes that are characteristics of gram-positive bacteria but lacks some characteristics of gram-negative bacteria. First, sporulation is generally one of the important features for certain gram-positive and rodshaped bacteria. No spore formation has been observed in *T. tengcongensis* culture. But surprisingly, there are 23 CDS, which are related to sporulation, in the *T. tengcongensis* genome. Secondly, gram-negative organisms have lipopolysaccharides (LPS), which the gram-positive lacks. The *T. tengcongensis* genome, though having a few CDS related to lipopolysaccharide biosynthesis, does not possess three of the key genes. Thirdly, none of the four CDS involved in lipid A synthesis are found in the *T. tengcongensis* genome, although they are well documented in most of the gram-negative prokaryotes. Finally, CDS for porins unique to gram-negative bacteria also appear absent in *T. tengcongensis*.

A strong correlation between the G + C content of tDNA and rDNA genes and the optimal growth temperature is found among the sequenced thermophiles. It is concluded that thermophiles are a biologically and phylogenetically divergent group of prokaryotes that have converged to sustain extreme environmental conditions over evolutionary timescale.

This research result was published in *Genome Research*[16] in 2002.

2.3 The genome research of the extreme acidophilus methane-oxidizing bacteria

Aerobic methanotrophic bacteria consume methane as it diffuses away from methanogenic zones of the soil and sediment. They act as a biofilter to reduce methane emissions to the atmosphere, and they are therefore targets in strategies to combat global climate change. No cultured methanotroph grows optimally below pH 5, but some environments with active methane cycles are very acidic. In Hell's Gate, New Zealand, which is a geothermal area rich in abiogenic methane, the researchers isolated an extremely acidophilic methanotroph that grows optimally at pH 2.0–2.5. Unlike the known methanotrophs, it does not belong to the phylum Proteobacteria but rather to the Verrucomicrobia. Similar organisms were independently isolated from geothermal systems in Italy and Russia. Verrucomicrobia is a widespread and diverse bacterial phylum that primarily comprises uncultivated species with unknown genotypes.

Although cultivation-independent approaches detect representatives of this phylum in a wide range of environments, including soils, seawater, hot springs and human gastrointestinal tract, only a few have been isolated in pure culture.

Professor WANG Lei's group in the TEDA School of Biological Sciences and Biotechnology of Nankai University, GNS Extremophiles Research Group in New Zealand and Professor Alam's group in the University of Hawaii collaborated to obtain the genome sequence of the bacteria V4, which is the first one from a representa-

tive of the Verrucomicrobia. Isolate V4, initially named "*Methylokorus infernorum*" (and recently renamed *Methylacidiphilum infernorum*) is an autotrophic bacterium with a streamlined genome of ~2.3 Mbps that encodes simple signal transduction pathways and has a limited potential for regulation of gene expression.

Aerobic methanotrophic bacteria use monooxygenase enzymes to convert methane to methanol, which is then oxidized to formaldehyde, formate and CO_2. Analysis of draft genome of V4 detected genes encoding particulate methane monooxygenase that were homologous to the genes found in methanotrophic proteobacteria. Phylogenetic analysis of its three *pmoA* genes (encoding a subunit of particulate methane monooxygenase) placed them into a distinct cluster from proteobacterial homologues. This indicates an ancient divergence of Verrucomicrobia and Proteobacteria methanotrophs rather than a recent horizontal gene transfer of methanotrophic ability. The findings show that methanotrophy in the bacteria is more genetically diverse than previously thought.

Central metabolism of *M. infernorum* was reconstructed almost completely and revealed highly interconnected pathways of autotrophic central metabolism compared to other known methylotrophs. Known genetic modules for methanol and formaldehyde oxidation were incomplete or missing, suggesting that the bacterium uses some novel methylotrophic pathways. The *M. infernorum* genome does not encode tubulin, which was previously discovered in bacteria of the genus Prosthecobacter, or close homologs of any other signature eukaryotic proteins.

Phylogenetic analysis of ribosomal proteins and RNA polymerase subunits unequivocally supports grouping Planctomycetes, Verrucomicrobia and Chlamydiae into a single clade, the PVC superphylum, despite the dramatically different gene content in members of these three groups. Comparative-genomic analysis suggests that evolution of the *M. infernorum* lineage involved extensive horizontal gene exchange with a variety of bacteria. The genome of *M. infernorum* shows apparent adaptations for existence under extremely acidic conditions including a major upward shift in the isoelectric points of proteins.

The genome sequencing of *M. infernorum* uncovers the veil of widespread Verrucomicrobia. The analysis revealed a novel methylotrophic pathway, and showed the diversity of methanotrophic bacteria in the extreme acidic environment. These findings provide a new theoretical basis for research and control of the greenhouse effect.

The result was published in *Nature* and *Biology Direct*[18,19] in 2007 and 2008, respectively.

2.4 The genome research on *Shewanella piezolerans*

Shewanella species are widespread in various environments. *Shewanella piezotolerans* WP3 is a piezotolerant and psychrotolerant iron reducing bacterium from deep-sea sediment. To study its environmental adaptation mechanisms, the Key Llaboratory of Marine Biogenetic Resources of State Oceanic Administration and Huada Genomics Research Institute sequenced and analyzed the whole genome of *S. piezotolerans* WP3, and the result was published in *PLos One*[17].

The WP3 genome consists of a single circular chromosome of 5396476 bps with 4,944 predicted genes. WP3 has the largest genome size among the sequenced *Shewanella* genomes. Of these predicted proteins, 3326 are similar to the known proteins in current databases, 877 are conserved hypothetical proteins, and 741 are hypothetical proteins which have no database match. The genome of WP3 is mostly related to *S. loihica* PV-4 genome, a psychrotolerant bacterium isolated from iron-rich microbial mats at an active, deep sea. WP3 and PV-4 have extensive regions of similar gene order, as revealed by an obvious ‘‘X’’ pattern when the two genomes are aligned; this pattern indicates that *Shewanella* genomes have undergone extensive inversions around the origin and terminus.

The research compared paralogous gene families in MR-1 and WP3 genomes. The results clearly indicated that for each pairs of corresponding gene families in both genomes, WP3 usually has a much larger gene number in these families than MR-1 has. In WP3, these expanded families are primarily involved in transport, secretion, energy metabolism and transcriptional regulation. The existence of additional duplicated genes could provide more capacity for coping with the environmental change and less selective pressure.

WP3 genome contains 55 putative c-type cytochrome genes, which could be divided into 17 groups. WP3 can also use a variety of electron acceptors such as nitrate, fumarate, trimethylamine N-oxide (TMAO), dimethyl sulfoxide (DMSO), and insoluble metals during anaerobic growth. The large number of cytochrome c genes in WP3 could be considered an adaptation to high-pressurized deep-sea environments.

A change in membrane fluidity is a well-established response of microorganism to low temperature and high pressure exposure. WP3 has a gene cluster that can synthesize eicosapentaenoic acid (EPA), and its content in the cell membrane increases as a function of the decreased temperatures. EPA plays certain roles in the environmental adaptation of WP3.

WP3 contains seven pseudouridine synthase genes, nine pseudouridylate synthase genes and three genes for RNA modification. In general, bacteria only have 1–4 copies of these genes; this is the first report of so many pseudouridine synthase genes in a single bacteria genome. The large number of pseudouridine synthase genes was also observed in several of the *Shewanella* genomes including PV4, MR-4, MR-7, CN-32, SB2B, etc. The large number of genes for structural RNA modification in so many *Shewanella* genomes may be one of the important mechanisms for the wide environmental adaptation and distribution of the genus.

Two sets of flagella genes are present in the WP3 genome. The two sets of flagellum systems were found to be differentially regulated under low temperature and high pressure; the lateral flagellum system was found essential for its motility and living at low temperature.

The research on the genome of *S. piezotolerans* WP3 deepens the understanding of *Shewanella* adaptation to environments, and the adaptation mechanism of microorganisms to the psychrosphere/piezosphere.

2.5 Genomics Research on *Pseudomonas stutzeri* A1501

The capacity to fix nitrogen is widely distributed in phyla of Bacteria and Archaea but has long been considered to be absent from the *Pseudomonas* genus. *P. stutzeri* A1501, which was isolated from rice paddy soils in 1980, is an associative nitrogen-fixing bacterium with good nitrogen-fixing capacity. Kinds of researches on its physiology, biochemistry and genetic characteristics have been carried out during the past 20 years. Professor LIN Min and his group in Chinese Academy of Agricultural Sciences analyzed the genome of *P. stutzeri* A1501, and found that its genome is composed of a single circular chromosome of 4567418 bp[20]. Comparative genomics revealed that, among 4146 protein-encoding genes, 1977 have orthologs in each of the five other *Pseudomonas* representative species sequenced to date. The genome contains genes involved in broad utilization of carbon sources, nitrogen fixation, denitrification, degradation of aromatic compounds, biosynthesis of polyhydroxybutyrate, multiple pathways of protection against environmental stress, and other functions that presumably give *P. stutzeri* A1501 an advantage in root colonization.

Genetic information on synthesis, maturation, and functioning of nitrogenase is clustered in a 49-kb island with 66.8% average content of GC, suggesting that this property was acquired by lateral gene transfer. All genes involved in nitrate metaboliztion in A1501 are organized in a super cluster containing *nor*, *nir*, and *nos* genes, including 40 genes in total. The transcription products of these genes are closely related to the proteins involving transportation, gene regulation and reductase-like proteins, suggesting that this property was acquired by lateral gene transfer. New genes required for the nitrogen fixation process have been identified within the *nif* island.

The genomics research on the *P. stutzeri* A1501 gave us a general understanding of its metabolic networks and nitrogen-fixing mechanism in the systematic view. The regulatory mechanism of the nitrogen- fixing system and the overall regulation of the nitrogen circle (including the amino acid cycle, nitrogen fixation and denitrification) can be further studied, as well as the researches on the coupling effect among the carbon metabolism and the nitrogen metabolism and relative molecular mechanism. The strains with higher halotolerancy and higher efficiency of nitrogen-fixing can be developed based on the development nowadays. New important functional genes can be cloned with patent. The interaction mechanism between the nitrogen-fixing bacteria and plants can be studied. The high-yielding, low-consuming, haloduric microbe-plants interaction farming system can be finally developed.

The result was published in *PNAS*[20] in 2008.

2.6 Genomics Research on other environmental and industrial microorganisms

Bacillus cereus Q1 is an industrial bacterium which was first isolated by the Daqing Petroleum Administration Bureau through lab screening, evaluation, and field tests. The bacterial strain can degrade the heavy carbon chain in C20~C40, with a viscosity reduction rate above 70%. The optimum growth temperature for *B. cereus* Q1 is 28℃~99℃. This strain adapts to most of the geological and stratal conditions in China and has high application value. Besides these, it also has the ability to lower the

interfacial tension, and achieves fine viscidity reduction and oil increase effect in the field test. The genomics research and the further functional gene researches and genetic engineering on this strain can be applied to not only the geological condition of Daqing, but also other different oil fields of China, and has broad prospects for international application and far-reaching scientific significance. Professor JIN Qi and his group in the State Key Laboratory for Molecular Virology and Genetic Engineering have deciphered the whole genomes of *Bacillus cereus* Q1, which has a length of 5,214,195 bps, with 85% coding region and average 830 bp for each coded sequence. By annotation and further analysis, of the 5360 open reading frame(ORF) found, 2630 are in the leading strand and 2730 in the lagging strand; 3818 genes have known functions, 1445 have unknown functions, but with significant or partial homology with the protein sequence in the database; 87 ORF have no homology with any sequence that has been published or identified. The research on the functional genes of *B. cereus* Q1 is still under way.

Vitamin C is also known as ascorbic acid, which is a necessary vitamin for humans and has important applications in the medicine and food industries. In the 1980s, the "two-step fermentation" for the production of vitamin C was invented by Chinese scientists. This technique replaced the chemical process with biological processes, which have the advantages of low cost and no pollution. But two processing steps and three types of strains are required in this technique, which has the disadvantages of complicated operation and lots of pollution. Therefore it is necessary to reform the existing production strains with the modern genetic technology, changing the two-steps fermentation into one-step fermentation. The State Key Laboratory for Molecular Virology and Genetic Engineering and the North China Pharmaceutical Company joined together to launch the research on the functional genomics on the acid- producing *Ketogulonigenium* sp. WB0104 at the second step in the two-step fermentation. The genomics analysis shows that the genomes of *Ketogulonigenium* sp. are constructed by a ring chromosome with the length of 2765030 bp and two ring plasmids (sizes are 267968 bp and 242707 bp, respectively), The chromosome, which has an average content of G+C of 61.69%, contains 2727 ORFs, of which 72.4% have defined biological functions while the other 17.7% are not clear. A series of key genes closely related to the Gulonic acid metabolism, including Sorbitol dehydrogenase gene family, the synthetic gene cluster of coenzyme PQQ, the Sorbitol dehydrogenase of multi-subunit and so on were found through the bioinformatics analysis, gene chips, protein purification, gene expression and the mass spectrometric analysis. Through *E. coli* recombinant expression and *in vitro* active experiment for these genes and gene clusters, as well as the liquid chromatogram and mass spectrum, studies on the substrate specificity and enzymology characteristics of these key enzymes were studied. These researches laid a foundation for improving the genetic engineering of the acid-producing bacteria, the furthering construction of the zymolytic engineering bacteria and the improvement of the zymotechniques.

3 Conclusions

In the past few years, the environmental microbial genomics in China showed a tendency of rapid development. The solid foundation has been established by seting up world-level research platforms and gathering a group of well-known talents.

China is going to obtain a number of original achievements with great influence in the environmental microbial genomics in the near future by integrating the new generation sequencing platforms, the excellent R&D (Research and development) team and the improvement of bioinformatics and proteomics. During the 11th Five-Year Plan, China started the National High-Tech Research Development Plan (863 Plan) ——"Resources and environment technology field", the National Key Basic Research Development Plan (973 Plan), and the National Natural Science Foundation, all of which have supported the field of environmental microbiology. It is expected that China will achieve a great breakthrough in the field of environmental microbial genomics research in the next few years.

1 Gans J, Wolinsky M, Dunbar J. Computational improvements reveal great bacterial diversity and high metal toxicity in soil. Science, 2006, 309(5739): 1387—1390

2 Streit W R, Schmitz R A. Metagenomics——the key to the uncultured microbes. Curr Opin Microbiol, 2004, 7(5): 492—498

3 Strom S L. Microbial ecology of ocean biogeochemistry: a community perspective. Science, 2008, 320(5879): 1043—1045

4 Warnecke F, Luginbuhl P, Ivanova N. Metagenomic and functional

analysis of hindgut microbiota of a wood-feeding higher termite. Nature, 2007, 450(7169): 560—565
5 Fitz-Gibbon S T, Ladner H, Kim U J. Genome sequence of the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. Proc Natl Acad Sci USA, 2002, 99(2): 984—989
6 Pohlmann A, Fricke W F, Reinecke F. Genome sequence of the bioplastic-producing "Knallgas" bacterium *Ralstonia eutropha* H16. Nat Biotechnol, 2006, 24(10): 1257—1262
7 Schneiker S, Martins dos Santos V A, Bartels D. Genome sequence of the ubiquitous hydrocarbon-degrading marine bacterium *Alcanivorax borkumensis*. Nat Biotechnol, 2006, 24(8): 997—1004
8 Edwards R A, Rodriguez-Brito B, Wegley L. Using pyrosequencing to shed light on deep mine microbial ecology. BMC Genomics, 2006, 7: 57
9 Nicholls H. Sorcerer Ⅱ: the search for microbial diversity roils the waters. PLoS Biol, 2007, 5(3): e74
10 Rusch D B, Halpern A L, Sutton G. The Sorcerer Ⅱ Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. PLoS Biol, 2007, 5(3): e77
11 Yooseph S, Sutton G, Rusch D B. The Sorcerer Ⅱ Global Ocean Sampling expedition: expanding the universe of protein families. PLoS Biol, 2007, 5(3): e16.
12 Williamson S J, Rusch D B, Yooseph S. The Sorcerer Ⅱ Global Ocean Sampling Expedition: metagenomic characterization of viruses within aquatic microbial samples. PLoS ONE, 2008, 3(1): e1456
13 Jin Q, Yuan Z, Xu J. Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157. Nucleic Acids Res, 2002, 30: 4432—4441
14 Feng L, Wang W, Cheng J. Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir. Proc Natl Acad Sci USA, 2007, 104(13): 5602—5607
15 Makarova K S, Koonin E V. Comparative genomics of Archaea: how much have we learned in six years, and what's next? Genome Biol, 2003, 4(8): 115
16 Bao Q, Tian Y, Li W. A complete sequence of the *T tengcongensis* genome. Genome Res, 2002, 12: 689—700
17 Wang F, Wang J, Jian H. Environmental adaptation: genomic analysis of the piezotolerant and psychrotolerant deep-sea iron reducing bacterium *Shewanella piezotolerans* WP3. PLoS ONE, 2008, 3(4): e1937
18 Dunfield P F, Yuryev A, Senin P. Methane oxidation by an extremely acidophilic bacterium of the phylum Verrucomicrobia. Nature, 2007, 450(7171): 879—882
19 Hou S, Makarova K S, Saw J H. Complete genome sequence of the extremely acidophilic methanotroph isolate V4, *Methylacidiphilum infernorum*, a representative of the bacterial phylum Verrucomicrobia. Biol Direct, 2008, 3: 26
20 Yan Y, Yang J, Dou Y. Nitrogen fixation island and rhizosphere competence traits in the genome of root-associated *Pseudomonas stutzeri* A1501. Proc Natl Acad Sci USA, 2008, 105(21): 7564—7569