ISSN 0006-2979, Biochemistry (Moscow), 2012, Vol. 77, No. 8, pp. 799-812. © Pleiades Publishing, Ltd., 2012.
Original Russian Text © A. V. Mardanov, N. V. Ravin, 2012, published in Biokhimiya, 2012, Vol. 77, No. 8, pp. 965-980.

REVIEW

The Impact of Genomics on Research in Diversity and Evolution of Archaea

A. V. Mardanov and N. V. Ravin*

Laboratory of Molecular Cloning, Centre "Bioengineering", Russian Academy of Sciences, pr. 60-let Oktyabrya 7-1, 117312 Moscow, Russia; fax: (499) 135-0571; E-mail: nravin@biengi.ac.ru; nravin@mail.ru

Received March 12, 2012

Abstract—Since the definition of archaea as a separate domain of life along with bacteria and eukaryotes, they have become one of the most interesting objects of modern microbiology, molecular biology, and biochemistry. Sequencing and analysis of archaeal genomes were especially important for studies on archaea because of a limited availability of genetic tools for the majority of these microorganisms and problems associated with their cultivation. Fifteen years since the publication of the first genome of an archaeon, more than one hundred complete genome sequences of representatives of different phylogenetic groups have been determined. Analysis of these genomes has expanded our knowledge of biology of archaea, their diversity and evolution, and allowed identification and characterization of new deep phylogenetic lineages of archaea. The development of genome technologies has allowed sequencing the genomes of uncultivated archaea directly from enrichment cultures, metagenomic samples, and even from single cells. Insights have been gained into the evolution of key biochemical processes in archaea, such as cell division and DNA replication, the role of horizontal gene transfer in the evolution of archaea, and new relationships between archaea and eukaryotes have been revealed.

DOI: 10.1134/S0006297912080019

Key words: archaea, genome, evolution, sequencing, phylogeny, gene systematics, DNA replication, horizontal gene transfer

Archaea were described by Carl Woese as a separate domain of life, along with bacteria and eukaryotes [1], and since then they have become one of the most interesting objects of modern microbiology, molecular biology, and biochemistry. For a long time methanogens, widely distributed in various anaerobic habitats, represented the best known group of archaea. Other groups of archaea were considered as extremophiles because they inhabited specific ecological niches where other organisms could not exist – under conditions of high temperature (thermophiles), in acidic or alkaline environments (acidophiles or alkaliphiles), and under high salinity (halophiles). Thermophilic archaea have been found in hot springs of Yellowstone Park in the USA, Iceland, Kamchatka, New Zealand, etc. Among archaea, microorganisms living under extreme conditions of both temperature and pH (thermoacidophiles and thermoalkaliphiles) have been found [2, 3]. Archaea have been found in deep sea hydrothermal vents at depths more than 1000 m where water remains liquid even at temperatures significantly higher than 100°C. These studies have resulted in isolation of microorganisms capable of growing at 115°C [4] and even at 121°C [5]. Works performed during the two last decades, and especially works associated with analysis of gene sequences of ribosomal 16S RNAs isolated directly from natural sources, have revealed the global distribution of archaea: they have been found in soils [6-8], sea water [9, 10] and sediments [11-13], fresh waters [14, 15], deep subterranean habitats [16], and the intestines of humans and animals [17]. Archaea are important components of microbial consortia in various habitats, e.g. in the ocean their fraction is estimated as 10-40% of all microorganisms [18]. Wide distribution and multiplicity of archaea in different marine and terrestrial ecological niches suggest that not only methanogens but also other groups of archaea can play important roles in global biogeochemical processes. In particular, this is exemplified by the key role of archaea in nitrification processes in the ocean [19] and soil [20].

Despite the great interest in archaea, there was no pronounced success in studies on their molecular biology, evolution, and metabolism until the mid-1990s. This was mainly associated with difficulties in cultivation of the

Abbreviations: FISH, fluorescent *in situ* hybridization; kb, thousand nucleotide pairs; Mb, million nucleotide pairs; rRNA, ribosomal RNA.

* To whom correspondence should be addressed.

majority of these microorganisms and with a virtually complete absence of genetic "tools" (genetic transformation, expression vectors, gene knockouts, etc.) that would be similar to those developed for such models as the bacterium *Escherichia coli*. Many evolutionarily ancient lineages of archaea that are known due to the 16S rRNA sequences from natural sequences until now have no cultured representatives and remain practically unstudied. Therefore, the development of genomics resulting in "the genome revolution" in microbiology in the mid-1990s also significantly contributed to studies on archaea. Initially "medical" objects, i.e. agents of the most important human diseases, were the main targets of microbial genomic projects, but soon organisms interesting for biotechnology, ecology, and evolution studies also became objects of genomics. Thus, the archaeon *Methanococcus jannaschii* became the third organism for which the complete nucleotide sequence of the genome was determined [21]. The role of genome sequencing in the study on *M. jannaschii* is well illustrated by the number of corresponding publications: 13 papers were published during the period from the discovery of this archaeon in 1983 until the sequencing of its genome in 1996, and more than five hundred works were published during the subsequent 15 years. The development of genome techniques allowed sequencing of genomes of uncultivated archaea directly from enrichment cultures [22, 23], metagenomic samples [24-26], and even from single cells [27]. The rapid development of genome technologies has promoted an increase in the number of complete genome sequences and enlargement of our knowledge about the biology of archaea, their diversity and evolution [28]. By 2011, about one hundred complete genomes of archaea have been sequenced, and 10 of them have been determined by us in the Centre "Bioengineering" of the Russian Academy of Sciences in collaboration with the group of Dr. E. A. Bonch-Osmolovskaya (Institute of Microbiology, Russian Academy of Sciences) [29]. In this review, the most important recent results in archaeal biology, diversity, and evolution obtained due to genome approaches will be considered.

MAIN FEATURES OF ARCHAEAN GENOMES

The size of archaeal genomes varies from 0.49 Mb in the obligate symbiont *Nanoarchaeum equitans* to 5.75 Mb in the methanogenic archaeon *Methanosarcina acetivorans* that possesses diverse metabolic pathways. As in the case of bacteria, the size of the archaeal genome is determined by the combination of genetic processes, which comprise gene duplication with their subsequent diversification, gene acquisition as a result of horizontal transfer events, gene loss in some lineages, etc. In the overwhelming majority of cases, the genome size is directly proportional to the number of encoded genes at the ratio of about one gene per thousand nucleotide pairs. The genome size is the smallest in symbionts, such as *N. equitans*. Among free-living species, small genomes 1.2-1.3 Mb in size are characteristic for organisms inhabiting stable ecological niches and having specialized metabolism exemplified by some archaea of the order Desulfurococcales and by nanohaloarchaea. Organisms with large genomes such as the archaeon *M. acetivorans* isolated from marine sediments inhabit structurally complicated ecological niches under variable conditions.

All genomes of archaea are circular DNA molecules and as a rule has one chromosome, except *Haloarcula marismortui*, which has two circular chromosomes. All plasmids of archaea also are circular, and no linear replicons have been found to date. As in the case of bacteria, chromosomes of the majority of archaea have only one replication initiation site [30, 31], although some archaea are found to have two or three concurrently active *ori*-sites [32, 33]. However, the enzymatic apparatus of DNA replication in archaea is similar to that in eukaryotes rather than bacteria. The G + C content of archaeal genomes varies from 27.63% (*Methanosphaera stadtmanae*) to 68.01% (*Halobacterium salinarum*). Note that the G + C content in archaea does not correlate with the temperature of growth. Genomes of the majority of hyperthermophilic species have a low G + C content and the chromosome stability depends on the complex of DNA-binding proteins. The reverse gyrase, an enzyme responsible for positive supercoiling of DNA, is virtually the only protein specific for hyperthermophiles. The *in vivo* role of this enzyme is not quite clear, but it was found not only in archaea but also in hyperthermophilic bacteria, which may have gained it as a result of horizontal transfer of the gene from archaea [34].

In archaea, introns are often found in genes of ribosomal and transfer RNAs [35, 36] but not in the protein-encoding genes, where an intron has only been detected in the pseudouridine synthase gene in some crenarchaea [37]. In genomes of archaea different mobile elements are found, in particular, transposons encoding enzymes required for their translocation and also miniature inverted repeat transposable elements (MITE). The latter do not contain genes of transposases, and their translocation depends on the "parental" transposon activity. The number of mobile elements in the genomes of archaea varies over wide limits, some genomes do not have mobile elements at all, whereas other genomes can have dozens of mobile elements. For example, the genome of *Sulfolobus solfataricus* contains 201 mobile elements representing 11% of the genome [38], whereas no active mobile elements have been detected in the genome of another representative of the same genus, *Sulfolobus acidocaldarius* [39]. Transposons of the same families are found in both archaea and bacteria, which suggests a possibility of translocation of mobile elements among bacteria and archaea inhabiting the same ecological niches.

The majority of archaeal genomes have been found to contain clusters of short repeated sequences (20-40 bp) separated by spacers with unique sequences. These clusters found also in genomes of many bacteria were termed CRISPR (Clustered Regularly Interspaced Palindromic Short Repeats) [40]. The spacer sequences are supposed to be "selected" from viruses, plasmids, and transposons [41], whereas the CRISPR systems inhibit the expansion of mobile elements with sequences identical to those of the spacers due to a mechanism similar to RNA interference in eukaryotes [42-44].

MAIN PHYLOGENETIC LINEAGES OF ARCHAEA

From the early 1980s the systematics of archaea was developed on the base of the 16S rRNA gene sequences used by Woese for constructing a universal phylogenetic system of prokaryotes [45]. This approach was later used in the phylogenetic analysis of both cultured archaea and "uncultivated" lineages represented only by nucleotide sequences of 16S rRNA genes isolated from natural sources [46, 47]. However, the subsequent studies revealed that using only the 16S rRNA gene sequences was insufficient to determine positions of basic branches of archaea at the base of the phylogenetic tree because of either a lack of phylogenetically important sequence differences or nucleotide composition biases leading to artifacts in constructing the phylogenetic trees [48]. At present conservative protein markers, in particular ribosomal proteins, are used for this purpose [22, 24, 49, 50]. Although a significant number of archaeal genomes have been sequenced only during recent years, some significant changes were introduced into the phylogeny of archaea immediately after the first phylogenetic trees based on ribosomal proteins were published [28].

As early as archaea were described as a separate domain of living organisms, it was established that three known groups of archaea, the halobacteria, methanogens, and members of the genus *Thermoplasma*, were evolutionary related, whereas a fourth group, the *Sulfolobus* genus, was shown to present a separate evolutionary branch. Thus, in the domain Archaea the two main phyla were established [51]: Crenarchaeota (crenarchaea) and Euryarchaeota (euryarchaea). Crenarchaea and euryarchaea not only form separate branches on the phylogenetic tree (Fig. 1), but they are significantly different in organization of the apparatus of genome replication and expression, cell division, and many other basic genetic processes (see below). Although during the last decade phylogenetically remote archaeal lineages have been found that are proposed to be described as new phyla, the majority of cultured archaea belong to just these two main phyla.

All cultured crenarchaea belong to only the class Thermoprotei, which is virtually totally represented by thermophiles and includes three orders: Thermoproteales, Sulfolobales, and Desulfurococcales. Members of the genus *Sulfolobus* were isolated by Thomas Brock in the mid-1970s from thermal springs of Yellowstone Park in the USA [52]. Sulfolobales are mainly aerobic lithoautotrophic microorganisms capable of oxidizing sulfur and its reduced compounds. The majority of Sulfolobales are thermoacidophiles, many of which were isolated from such acidic thermal habitats as solfataras; their low pH can be due to activity of Sulfolobales, which aerobically oxidize sulfur compounds with production of sulfuric acid. Some members of this order (*Acidianus*, *Stygiolobus*, etc.) are anaerobes that can oxidize hydrogen or organic compounds using sulfur as an electron acceptor [53].

During the 1980s and 1990s, many genera of crenarchaea were isolated, which are mainly neutrophilic or moderately acidophilic hyperthermophilic anaerobes [54]. They are members of the orders Thermoproteales and Desulfurococcales, which comprise, respectively, rod-shaped and coccus-like hyperthermophilic archaea. These two orders include both lithoautotrophic and organotrophic members. Sulfur plays an important role in the metabolism of many, acting as an electron acceptor [55]. Many representatives of Thermoproteales can anaerobically oxidize organic substances and are also capable of chemoautolithotrophic growth with fixation of carbon dioxide [53]. Analysis of genomes of Thermoproteales has revealed genes of membrane-bound hydrogenase and sulfur reductase [56, 57], the combined activity of which is responsible for chemolithoautotrophic growth due to oxidation of hydrogen coupled to reduction of sulfur. During heterotrophic growth, these archaea can completely oxidize organic substances to carbon dioxide and water in the tricarboxylic acid cycle [58]. Some Thermoproteales, such as representatives of the genus *Pyrobaculum*, can use a wide spectrum of electron acceptors including nitrate, oxidized iron compounds, and arsenate, and their genomes encode a large number of membrane-bound oxidoreductases [59-61].

Metabolic possibilities determined by the genomes of Desulfurococcales [62-64] are more limited, most performing anaerobic fermentation of organic substances with reduction of sulfur, but chemolithoautotrophs are also known, e.g. members of the genus *Ignicoccus* [65, 66]. Phylogenetic trees constructed on the base of both 16S rRNA and ribosomal protein sequences suggest that Thermoproteales are evolutionarily the most ancient group, whereas Sulfolobales and Desulfurococcales branched out later.

Recently, some crenarchaeal genera have been described and separated as the new orders Acidilobales and Fervidicoccales. Representatives of the order Acidilobales [67] comprising the genera *Acidilobus* and *Caldisphaera* are anaerobic thermoacidophiles. These microorganisms isolated in thermal springs of Kamchatka can completely oxidize organic substances to carbon

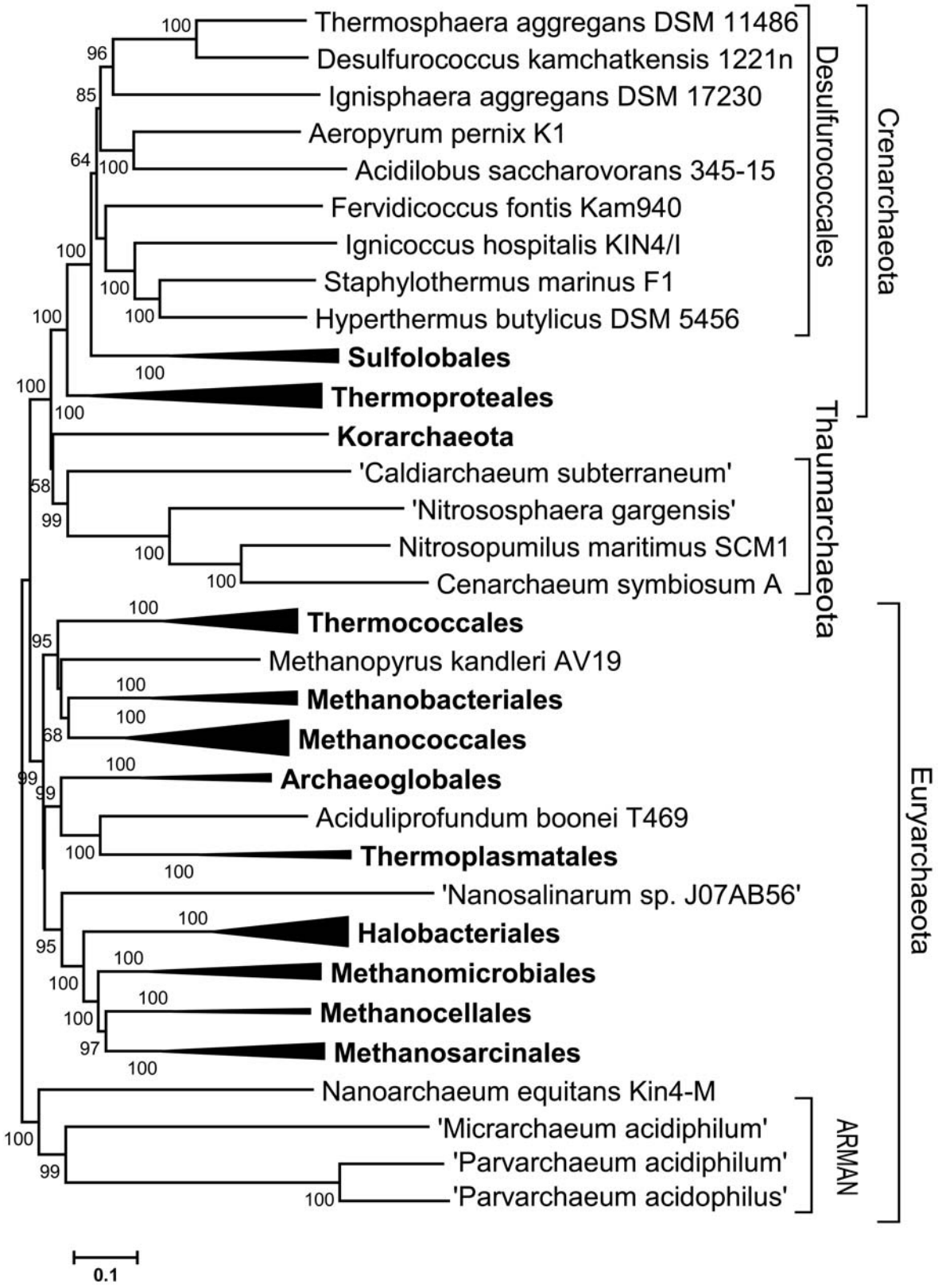

Fig. 1. Phylogenetic tree of the main groups of archaea constructed using the Minimum Evolution approach on the basis of comparing concatenated sequences of 41 ribosomal proteins. Other approaches (Maximum Likelihood, Neighbor-Joining) resulted in a similar tree structure. The MUSCLE v3.7 program was used to align 93 sequences, and the trees were constructed using the MEGA 5 Beta program. Numbers at the branches indicate bootstrap support, only values higher than 50% being shown. The scale corresponds to 0.1 substitution per site.

dioxide and water in the tricarboxylic acid cycle under anaerobic conditions using sulfur as an electron acceptor [68]. At present, the classification of Acidilobales as a new order is under discussion [69]. About half of the genes of *Acidilobus saccharovorans* are most alike their analogs from the crenarchaeon *Aeropyrum pernix* [62] of the order Desulfurococcales, but the resemblance to other archaea of the order Desulfurococcales is less than to members of the orders Thermoproteales and Sulfolobales [68]. Another newly proposed order of thermophilic archaea, Fervidicoccales, includes the cultured species *Fervidicoccus fontis* and also different "uncultivated" members, which form a separate branch on the 16S rRNA tree [70]. *Fervidicoccus fontis* is a moderate thermoacidophile capable of anaerobic fermentation of proteinaceous substrates with production of acetate and hydrogen [70]. Representatives of this order were found in thermal springs of Kamchatka, Yellowstone Park, and Iceland.

Euryarchaeota, the second phylum of the archaeal domain, includes ten orders: Methanopyrales, Thermococcales, Archaeoglobales, Methanobacteriales, Methanococcales, Methanomicrobiales, Methanosarcinales, Methanocellales, Thermoplasmatales, and Halobacteriales. All these orders have cultured representatives; Methanopyrales, Thermococcales, and Archaeoglobales include only thermophiles.

The order Thermococcales mainly comprises anaerobic heterotrophic archaea fermenting carbohydrates and/or protein substrates [53]. Sulfur reduced to hydrogen sulfide can be used as an electron acceptor, and its presence stimulates the growth and is necessary for some species. Representatives of Thermococcales are widely distributed in marine hydrothermal ecosystems and also in subterranean thermal habitats (oil reservoirs, etc.), but species isolated from terrestrial hydrothermal vents are also known [71]. As a rule, archaea of this order can be easily grown under laboratory conditions, and many of them are used as models in studies on biology of archaea and as sources of new thermostable enzymes (such as the thermostable Pfu polymerase).

The order Archaeoglobales includes genera *Archaeoglobus*, *Ferroglobus*, and *Geoglobus*. *Archaeoglobus* is specified by its ability to reduce sulfate, and this property is very rare in archaea. In addition to the *Archaeoglobus* [72], sulfate reduction pathway has been found only in genomes of some crenarchaea of the order Thermoproteales [57, 73]. The *Archaeoglobus* representatives can use as electron donors different organic substances and hydrogen; they are isolated from marine hydrothermal vents and oil holes. Other members of the order Archaeoglobales can use as oxidizers nitrate or Fe^{3+}, and they oxidize anaerobically organic acids including acetate [53].

The phylum Euryarchaeota includes six orders that join methanogenic archaea – Methanopyrales, Methanobacteriales, Methanococcales, Methanomicrobiales, Methanocellales, and Methanosarcinales [53]. The order Methanopyrales has only one species, *Methanopyrus kandleri*, which inhabits deep-water hydrothermal vents, and the optimal temperature for its growth is about 105°C [74]. Other orders include both thermophilic and mesophilic members, many of which were known long ago before archaea were defined as a special domain of living organisms.

The order Halobacteriales includes halophilic archaea inhabiting salt lakes and other ecological niches with high concentrations of NaCl, which can be above 150-200 g/liter [75]. Cultured haloarchaea are mesophilic organisms, and the optimal temperature for their growth is not higher than 50°C [53]. However, 16S rRNA sequences assigned to Halobacteriales were detected in hot springs with temperature of 50-70°C and relatively low salinity [76].

Euryarchaea of the order Thermoplasmatales are known from the 1970s. They are moderately thermophilic or mesophilic microorganisms growing at low pH values (*Picrophilus torridus* at $pH \geq 0$) in acidic thermal springs, solfataras, and volcanic soils [53]. Members of genera *Thermoplasma* and *Picrophilus* are aerobic organisms (facultative and obligate, respectively) oxidizing different organic substances, mainly proteinaceous substrates. Under anaerobic conditions *Thermoplasma* sp. can use sulfur as an electron acceptor [53]. The genus *Ferroplasma* includes slowly growing mesophilic archaea oxidizing iron under aerobic conditions [77]. Some species can grow anaerobically on organic substrates using Fe^{3+} as an oxidant. Members of this genus are resistant to high concentrations of metals (Fe, Zn, Cu, Cd, As, etc.) that are toxic for the majority of other organisms. Archaea of the genera *Thermoplasma* and *Ferroplasma* are characterized by the absence of a cell wall.

ROLE OF GENOMICS IN IDENTIFICATION AND CHARACTERIZATION OF NEW LINEAGES OF ARCHAEA

The main lineages of cren- and euryarchaea presented above have been described by isolation of pure cultures of the microorganisms and their characterization. Many of these lineages were known before the appearance of molecular biology and genomic techniques, and methanogenic archaea have been under study for more than hundred years. The sequencing of full genomes of these organisms has significantly contributed to studies on their biology, but the role of genomics is far greater than characterization of earlier known organisms: many new lineages of archaea became known only following the application of molecular methods for studies on natural diversity.

Thus, analyzing natural microbial consortia by 16S rRNA molecular approaches has shown that as a rule no more than 0.1-1% of microorganisms can be cultured under standard laboratory conditions. In particular, analysis of sequences of 16S rRNA genes isolated from samples taken in terrestrial and marine thermal springs, aquatic environments, marine and oceanic sediments, and soils has resulted in identification of some dozens of new phylogenetic lineages of archaea of high taxonomic level, and many of them occupy basic positions on the evolutionary tree [78-80]. Discovery of these new groups of archaea stimulated works aimed at obtaining their members in pure cultures suitable for subsequent microbiological, biochemical, and genomic analysis. However, attempts to isolate pure cultures for some new lineages have been unsuccessful, and in such cases genomic approaches can be used as the main tools because they allow reconstruction of genomes of the organisms under study as a result of sequencing of metagenomic samples isolated from natural sources or by sequencing of genomes of individual cells.

Many new groups of thermophilic archaea have been found in the hot spring Obsidian Pool (75-90°C) in Yellowstone National Park. The majority of them are new archaeal lineages of the Crenarchaeota phylum [46]. However, some sequences of 16S rRNA genes could not be assigned to the known archaeal phyla based on the level of homology, and they were described as a new phylum, Korarchaeota. Using specific PCR primers, members of Korarchaeota have been detected in hot springs of Iceland and Kamchatka [81, 82]. The appropriateness of the separation of Korarchaeota as a new phylum was confirmed by sequencing and analysis of the full genome of *Korarchaeum cryptofilum*, which was the first member of this group obtained in a stable enrichment culture [22]. Analysis of the genomic data revealed that the genome lacked genes of many biosynthetic pathways, and just this could prevent the isolation of this microorganism as a pure culture [22]. The genome was shown to contain crenarchaeal-specific genes of ribosomal proteins and subunits of RNA polymerase [22] and also euryarchaeal-specific genes encoding the maturation of tRNA, replication and repair of DNA, and cell division [22]. All these findings suggest that Korarchaeota represent a separate phylum different from cren- and euryarchaea (Fig. 1).

The description of a new phylum of archaea, Thaumarchaeota, based on results of the full genome analysis seems to be one of the most important recent results [48]. Initially this lineage was known only by sequences of 16S rRNA isolated from sea water and called Marine group I. Representatives of this group found in marine ecosystems were described as mesophilic crenarchaea, but during recent years thermophilic species were also found, e.g. *Nitrosocaldus yellowstonii* [83]. Analysis of the genome of the Marine group I member *Cenarchaeum symbiosum*, which is a marine sponge symbiont, revealed in it the presence of specific features of both eury- and crenarchaea [23]; therefore the authors, based on results of the phylogenetic analysis of 16S rRNA and ribosomal protein sequences, classified this lineage as a new phylum of archaea, Thaumarchaeota [48]. The appropriateness of definition of this lineage as a new phylum was confirmed by results of sequencing the genomes of other members of this group, the marine archaeon *Nitrosopumilus maritimus* [49] and two archaeons isolated from soil, *Nitrosophaera gargensis* [50] and *Nitrosoarchaeum koreensis* [84]. The reconstruction of a virtually full genome (1.8 Mb) of a fresh-water thaumarchaeon by sequencing DNA from several individual cells is also an example of description of a new microorganism by only genomic methods [27]. Archaea of the phylum Thaumarchaeota are characterized by their ability for chemolithoautotrophic growth due to aerobic oxidation of ammonium, and this explains the great interest in this group [85]. Nitrification is a key element of the global nitrogen cycle, and it was described earlier only for proteobacteria [86]. Considering the multiplicity of the Thaumarchaeota in marine and soil ecosystems, they probably play the leading role in nitrification processes [20, 87].

The genome of *Caldiarchaeum subterraneum*, which is a member of another uncultivated lineage HWCG I (Hot Water Crenarchaeotic group I) related to Thaumarchaeota, has been reconstructed from sequences obtained from a metagenomic sample isolated from thermal waters in a gold mine [24]. Based on the presence of homologs of cren-, eury-, and thaumarchael genes in the genome and on the presence of a specific eukaryotic ubiquitin-like system of protein modification, the authors initially supposed to separate HWCG I as a new phylum different from Thaumarchaeota. However, the appropriateness of this separation is still under discussion [69] because on phylogenetic trees based on the ribosomal protein sequences, *C. subterraneum* is clustered with Thaumarchaeota resulting in a "basal" phylogenetic branch of this type, and differences between *C. subterraneum* and other thaumarchaea are similar to differences between different orders of Euryarchaeota [69]. Note that analysis of the *C. subterraneum* genome did not reveal pathways of ammonium oxidation, and, on taking into account the "basal" position of this organism, this indicates that it either has lost the ability for nitrification that was present in the common ancestor of Thaumarchaeota or that the corresponding genes were gained by mesophilic and some thermophilic members of Thaumarchaeota at a later stage of evolution.

An important result of recent years is determination of the greater part (82-90%) of the genome sequence of ANME-1, which represents one of three uncultivated euryarchaeal lineages capable of anaerobic methane oxidation in consortium with sulfate-reducing bacteria. The genome was determined by sequencing a metagenome of methane-oxidizing microbial mat taken from the bottom of the Black Sea [26]. The data obtained on metabolism

of these archaea confirmed the hypothesis of a reverse methanogenesis pathway for the anaerobic oxidation of methane performed with the same enzymes [88].

Although there are many genome sequences available, the real diversity of cren- and euryarchaeal lineages, even those known for a long time, remains unknown. Thus, a number of different archaeal lineages distantly related to Thermoplasmatales has been detected on analyzing sequences of 16S rRNA genes isolated from marine and terrestrial hydrothermal vents [47], acidic thermal groundwater of Kamchatka [56], etc. A member of one such lineages, DHVE2 (Deep-Sea Hydrothermal Vent Euryarchaeota), has been successfully isolated in a pure culture [89]. Analysis of the full genome sequence of this anaerobic thermoacidophilic archaeon, *Aciduliprofundum boonei*, has shown it to be a representative of the order Thermoplasmatales, in which it forms a separate phylogenetic branch [90] (Fig. 1). This organism is an organotroph capable of fermenting various proteinaceous substrates.

Nanoarchaea seem to be the most unusual group of archaea; they were initially identified by analysis of sequences of 16S rRNA genes. The size of cells of the only nanoarchaeon cultured so far, *Nanoarchaeum equitans* [91], is only about 500 nm. This archaeon grows in deep water hydrothermal vents being attached to living cells of another archaeon, the chemolithoautotrophic crenarchaeon *Ignicoccus hospitalis*. Because of the low homology of 16S rRNA genes of the *Nanoarchaeum* with members of other recognized phyla, the nanoarchaea were supposed to represent a new phylum, the Nanoarchaeota [91]. Sequences of 16S rRNA belonging to nanoarchaea have been identified in various marine and terrestrial hydrothermal vents [92-94] and also in mesophilic ecological niches with high salinity [94]. The *N. equitans* genome has the smallest size among archaea (about 490 kb, only 552 genes); it lacks many genes encoding different biosynthetic pathways including lipid biosynthesis, and this suggests a specialized symbiosis or parasitism by this organism [95]. The gene loss in nanoarchaea was accompanied by disruption of operons and "splitting" of some genes, e.g. of the tRNA genes. Different approaches for comparative genomic analysis result in different conclusions about the antiquity of origin of this phylogenetic lineage of archaea, which can be either the deepest branch of archaea [95, 96] or the rapidly evolving lineage of euryarchaea close to the order Thermococcales [97-99]. Such presumably evolutionarily ancient characteristics as "split" genes of tRNA produced as a result of trans-splicing [96] were later found in the crenarchaeon *Caldivirga maquilingensis* [100].

Another group of archaea with small cell size (with diameter lower than 500 nm) has been identified as a minor component in biofilms that develop in mine waters characterized by a high acidity and the presence of metals (Richmond Mine, California, USA). These archaea live in association with euryarchaea of the order Thermoplasmatales. Soon after this group named ARMAN had been described [101], the sequencing of metagenomes of microbial mats provided nearly complete genomes of three representatives of this group, which were named *Parvarchaeum acidiphilum*, *Parvarchaeum acidophilus*, and *Micrarchaeum acidiphilum* [101]. The size of the genomes is about 1 Mb, and they are characterized by a high AT content and the presence of "split" genes, typical of symbiotic and parasitic microorganisms [101]. The genomes of the ARMAN lineage archaea have some specific features reflecting the evolutionary pressure directed to decreasing the genome size (compared to other archaea) – higher fraction of coding sequences in the genome, lower average length of the gene, and presence of a significant number of overlapping genes. Phylogenetic analysis of sequences of ribosomal proteins revealed that ARMAN are a separate lineage in the Euryarchaeota [69].

Archaea with small cells and genomes were recently found also among halophiles of the order Halobacteriales. The genome sequences of two such microorganisms from the salt lake Tyrrell in Australia were determined by sequencing a metagenome of a water sample from the lake, which contained cells with 0.1-0.8-μm size [102]. The phylogenetic analysis of sequences of ribosomal proteins showed that these "nanohaloarchaea", called "Candidatus *Nanosalina* sp. J07AB43" and "Candidatus *Nanosalinarum* sp. J07AB56", form a separate phylogenetic branch at the base of the order Halobacteriales. The novelty of this lineage is also shown by the absence in GenBank of homologs for about 60% of all proteins predicted for nanohaloarchaea. Analysis of the genome sequences suggests that these organisms, similarly to common haloarchaea, are aerobic heterotrophs incapable of anaerobic respiration. They are characterized by the small size of their genome (about 1.2 Mb) with low GC content, the absence of gas vacuoles characteristic of haloarchaea, and the presence of unusual pathways of carbohydrate metabolism [102]. Analysis of microorganisms from Lake Tyrrell by the FISH method using specific DNA probes showed that the cells of these nanohaloarchaea are about 0.6 μm in diameter and their fraction in the consortium is about 14%. The small volume of the cells elevates the surface to volume ratio, and this might be an adaptive feature optimizing nutrient uptake capacity. It seems that the small volume allows the nanohaloarchaea to remain suspended in the surface aerobic layers of water even in the absence of gas vesicles that are responsible for the floatation of haloarchaea.

EVOLUTION OF GENOMES – ROLES OF GENOME REARRANGEMENTS AND OF HORIZONTAL GENE TRANSFER

The comparison of genomes of closely related organisms posed some hypotheses about mechanisms of genome evolution. Thus, at the end of the 1990s the

genome sequences were determined for two species of the genus *Pyrococcus*: *Pyrococcus abyssi* [103] and *Pyrococcus horikoshii* [104]. When comparing the genomes, seven long homologous regions with conserved gene order (syntenic regions) were revealed (I-VII in Fig. 2a), and genome rearrangements were also identified that occurred after the separation of these two species: inversions of fragments I and III and transposition of the fragments V and VI [105]. Note the inversion of fragment I containing the *ori*-site. Such genome rearrangements are more frequent than other types of inversions, possibly because they do not change the direction of gene transcription relative to the direction of the movement of the replicative fork (they are coincident for the majority of highly expressed genes). A possible model explaining the inversion of fragments containing *ori*-sites is the existence of a single "replication factory" combining two replicative forks (Fig. 2b) [105].

Inversions, translocations, insertions, and deletions of relatively short fragments are other types of genome rearrangements caused by homologous or site-specific recombination associated with tRNA genes and different repeated sequences. Studies on genomes of three species of *Pyrococcus*, *Thermococcus kodakaraensis*, and also of Sulfolobales were informative for understanding the mechanisms of such events [38, 39, 106]. Thus, in

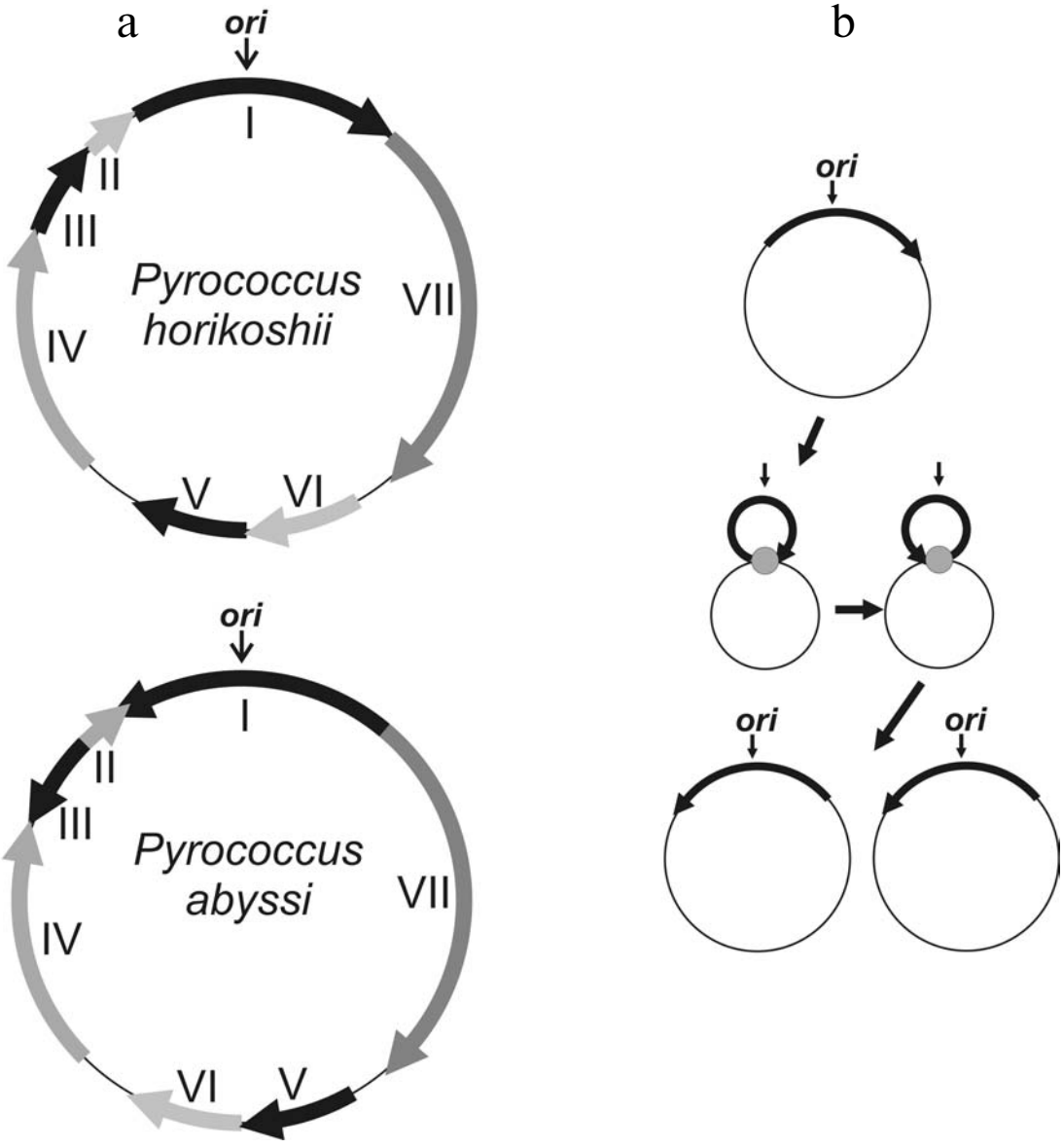

Fig. 2. Scheme of localization of syntenic fragments in the genomes of *Pyrococcus abyssi* and *Pyrococcus horikoshii* (a) and a possible mechanism of inversion of the fragment containing the *ori*-site (b). The localization and relative orientation of fragments I-VII are shown by arrows. The position of the replication initiation site is indicated with the arrow.

Pyrococcus and *Sulfolobus* the remnants of tRNA genes were found at borders of the recombination fragments. A large number of mobile elements in the *S. solfataricus* and *Sulfolobus tokodaii* genomes explain the high frequency of genome rearrangements in these species caused by recombination events between the repeated sequences. The high number of rearrangements in the *Pyrococcus furiosus* [107] genome relative to the *P. abyssi* and *P. horikoshii* genomes can be explained in a similar way, namely, by the presence of 23 insertion elements. A correlation was found between the distribution of mobile elements in the genome and borders of syntenic regions, and this suggests the key role of recombination events in the genome rearrangements during the evolution of *P. furiosus* [105].

Losses of genes, their duplications, and integration of foreign DNA are main factors influencing genome content. Although it is rather difficult to evaluate the frequency of gene losses, this process can be active in some groups of archaea. Thus, in some lineages of archaea about 15% of ribosomal protein genes can be lost [108]. Frequent gene losses and non-orthologous substitutions have been described for components of metabolic pathways [109]. In some cases, expansion of certain protein families occurs [109], e.g. about half of 4540 genes of the methanogen *M. acetivorans* possessing different metabolic pathways belong to 539 multigenic protein families [110].

Horizontal transfer or gene loss events can be predicted by comparing genomes of closely related organisms. For example, in the genome of *P. furiosus* a region was identified with the length of about 16 kb that encodes a complex of enzymes responsible for degradation and transport of polysaccharides [111]. This region also exists in the genomes of *Thermococcus sibiricus* and *Thermococcus litoralis*, but is absent in other closely related species (*T. kodakaraensis*, *Thermococcus onnurineus*, *P. abyssi*, and *P. horikoshii*), which favors the hypothesis of its horizontal transfer between archaea of the order Thermococcales. Such "gene islands" encoding enzyme complexes, which provide a selective advantage to a recipient in a certain ecological niche, can be transferred also between organisms of different domains. A genome fragment of about 20 kb length encoding the enzymes responsible for hydrolysis of beta-linked polysaccharides (cellulose, laminarin, etc.) and transport of sugars into the cell was found in the genome of *T. sibiricus* as an insertion in the cluster of ribosomal protein genes [112]. Such an insertion was absent in the genomes of closely related archaeons *T. kodakaraensis* and *T. onnurineus*; moreover, the closest homologs of enzymes encoded by this gene island were found in the genomes of thermophilic bacteria of the *Thermotoga* genus, which suggests the possibility of the horizontal transfer of this fragment.

In some specialized ecological niches the efficiency of horizontal gene transfer can be rather high, even between evolutionarily distant groups of archaea. The genomes of representatives of the orders *Thermoplasmatales* (*Picrophilus torridus* and *Thermoplasma acidophilum*) and *Sulfolobales* inhabiting hot acidic springs contain a rather large fraction of genes (6-11%) that are absent in organisms living under other conditions [113]. As a result of the frequent horizontal transfer of genes between these evolutionarily distant cren- and euryarchaea, the proteomes of these organisms are more like one another than to the proteomes of evolutionarily close organisms of the corresponding phyla [113]. Thus, 58% of the *P. torridus* genes have homologs in *S. solfataricus*, but only 35% of the genes were found to have homologs in the genome of the euryarchaeon *P. furiosus*, which is closer evolutionarily but inhabiting a different ecological niche (a submarine hydrothermal vent). The set of horizontally transferred genes mainly includes different proteolytic enzymes and transport-related proteins but not vertically inherited components of DNA replication, transcription, and translation machineries [114]. Genes similar to the *Sulfolobus* genes are distributed in the *T. acidophilum* genome nonrandomly, and therefore it was supposed that the genetic exchange included a few events of horizontal transfer of large genome fragments [114]. About 13% of the *P. torridus* genes have homologs among the *S. solfataricus* genes but not among the *T. acidophilum* genes, and this indicates that these genes have been transferred between *P. torridus* and *S. solfataricus* relatively recently, already after the divergence of the genera *Picrophilus* and *Thermoplasma* occurred. A high number of homologs of the *Sulfolobales* proteins in the proteome of *A. saccharovorans* from the order *Acidilobales* is another example of intense horizontal transfer of genes between thermoacidophiles [68]. However, no evidence of genetic exchange between *A. saccharovorans* and *Thermoplasmatales* has been found.

Analysis of full genomes of archaea has shown that, notwithstanding a constantly increasing number of sequenced genomes, a significant fraction of genes of every organism (usually 10-30%) does not have homologs in databases or is homologous only to genes of closely related organisms, and their functions are also unknown. Such genes are often found in the chromosome as clusters containing up to several dozens of genes. Such clusters can also include genes encoding archaeal proteins with known functions. One of the possible explanations is a hypothesis about plasmid or viral origin of these genes as a result of integration of the corresponding extrachromosomal elements into the genome. This correlates with the existence of a great diversity of such genetic elements in nature, and the majority of the proteins of archaeal viruses have no homologs in the databases [115]. For example, the genome of the Pyrobaculum Spherical Virus, which is about 40 kb in size, contains 49 open reading frames with unknown functions [116]. The majority of such viral genes integrated into the host genome do not provide a

selective advantage for the cell and became rapidly eliminated as is shown by the uniqueness of such "viral islands" concurrently detected only in the genomes of very close species newly separated from the common ancestor into the genome of which the integration has occurred. Thus, various viruses and plasmids can act not only as transporters of genetic material between different species of archaea, but they also remain the largest and virtually uncharacterized reservoir of genetic information entering the archaeal genomes from the outside [117].

DIFFERENCES BETWEEN THE MAIN ARCHAEAL LINEAGES ON THE GENOMIC LEVEL

Initially, archaea were subdivided into cren- and euryarchaea based on the 16S rRNA phylogeny. Comparative analysis of genomes confirmed such division because a number of key proteins involved in the replication of DNA, cell division, and translation were specific for cren- and euryarchaea (table). Thus, cell division in euryarchaea is provided for by a bacteria-specific system with the protein FtsZ as a key element. In crenarchaea this protein is absent, and cell division is provided for by the system CdvABC [118], which is related to the eukaryotic ESCRT-III (in Desulfurococcales and Sulfolobales) or by an uncharacterized actin-based system (Thermopro-teales). Crenarchaea are characterized by the absence of DNA polymerase of the D family and histones, whereas no genes of the ribosomal proteins S30, S26, S25, and L13 were found in the genomes of Euryarchaeota. Korarchaeota and Thaumarchaeota are characterized by different combinations of cren- and euryarchaeal features and also by some unique traits (table). Thus, among components of cell division both FtsZ and actins were found in Korarchaeota, whereas thaumarchaea were shown to have FtsZ and CdvABC. The phylum Thaumarchaeota is characterized by the presence of DNA topoisomerase type IB, which seems to substitute the type IA enzyme found in all other of archaeal phyla (table).

Such a mosaic distribution of the characters determining the most important biological processes can be a result of evolutionarily ancient events of horizontal gene transfer between different lineages and also a consequence of the loss of genes initially present in the common ancestor of archaea in different lineages as they diversified [69]. The hypothesis of gene loss is favored by results of the comparative study on 28 full genomes of archaea, which allowed construction of a model of genome evolution that suggests the gaining, duplication, and loss of genes in the course of evolution [119]. In the framework of this hypothesis, the common ancestor of archaea is a not less complex organism than its modern specialized descendants. The sequencing and analysis of genomes of other evolutionarily ancient "uncultivated"

Components of systems responsible for DNA replication and maintenance, cell division, transcription, and translation in different lineages of archaea

	Crenarchaeota	Thaumarchaeota	Korarchaeota	Euryarchaeota
Families of DNA polymerases	BI, BII	BII, D	BI, BII, D	BI, D
Topoisomerase IA	+	+/–	+	+
Topoisomerase IB	–	+	–	–
Histones	– (D, S) + (T)	+	+	+*
FtsZ	–	+	+	+
CdvABC	+ (D, S) – (T)	+	–	–
Actin	– (D, S) + (T)	–	+	–**
ATPase of the SMC family involved in chromosome segregation	–	+	+	+
RNA polymerase subunit Rpb8	+	–	+	–
Ribosomal proteins S25, S26, S30	+	+	+	–

Notes: +, is present; –, is absent; +/–, is present in some members; D, Desulfurococcales; S, Sulfolobales; T, Thermoproteales; *, is absent in Thermoplasmatales; **, is present in Thermoplasmatales.

lineages of archaea are promising for the better understanding of early stages of the evolution of life, especially of archaea and eukaryotes.

During the last 15 years more than 100 full genomes of archaea have been determined, and now there are full genome sequences for the majority of the cultured lineages of archaea on the genus/family level. As the genome sequencing methods became more elaborated, especially after introducing the method of parallel pyrosequencing into practice in the end of the 2000s, the sequencing of genomes of microorganisms became a standard procedure. Along with successfully developing genetic tools [120], the genome data made a key contribution to studies on archaea. Interest in genomics of archaea continues and seems likely to increase in the future. The works of the last decade have shown that the known cultured lineages represent only a small part of the true diversity of archaea, which are widely distributed and play important ecological roles in "non-extreme" ecological niches. The majority of these lineages are still not cultured under laboratory conditions, and this makes genomics the main approach for their study. It seems that in the nearest future the main achievements in the genomics of archaea will be associated with sequencing of the genomes of uncultivated lineages of archaea directly from enrichment cultures, metagenomic specimens, and from single cells. Not less important will be sequencing of the metagenomes of natural microbial communities, including their viral component that seems to be a virtually still unexplored reservoir of genetic information.

We are grateful to Dr. E. A. Bonch-Osmolovskaya (Winogradsky Institute of Microbiology, Russian Academy of Sciences) for fruitful discussions concerning problems of systematics of archaea.

The work of the Laboratory of Molecular Cloning of the Centre "Bioengineering" of the Russian Academy of Sciences in the field of archaeal genomics was supported by the Ministry of Education and Science of Russia (state contracts 02.512.11.2201 and P1049) and by the Russian Foundation for Basic Research (project 11-04-00671).

REFERENCES

1. Woese, C. R., and Fox, G. E. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5088-5090.
2. Schleper, C., Puehler, G., Holz, I., Gambacorta, A., Janekovic, D., Santarius, U., Klenk, H. P., and Zillig, W. (1995) *J. Bacteriol.*, **177**, 7050-7059.
3. Li, Y., Mandelco, L., and Wiegel, J. (1993) *Int. J. Syst. Bacteriol.*, **43**, 450-460.
4. Blochl, E., Rachel, R., Burggraf, S., Hafenbradl, D., Jannasch, H. W., and Stetter, K. O. (1997) *Extremophiles*, **1**, 14-21.
5. Kashefi, K., and Lovely, D. R. (2003) *Science*, **301**, 934.
6. Bintrim, S. B., Donohue, T. J., Handelsman, J., Roberts, G. P., and Goodman, R. M. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 277-282.
7. Ochsenreiter, T., Selezi, D., Quaiser, A., Bonch-Osmolovskaya, L., and Schleper, C. (2003) *Environ. Microbiol.*, **5**, 787-797.
8. Galand, P. E., Fritze, H., Conrad, R., and Yrjala, K. (2005) *Appl. Environ. Microbiol.*, **71**, 2195-2198.
9. DeLong, E. F. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 5685-5689.
10. Fuhrman, J. A., McCallum, K., and Davis, A. A. (1992) *Nature*, **356**, 148-149.
11. Boetius, A., Ravenschlag, K., Schubert, C. J., Rickert, D., Widdel, F., Gieseke, A., Amann, R., Jorgensen, B. B., Witte, U., and Pfannkuche, O. (2000) *Nature*, **407**, 623-626.
12. Orphan, V. J., House, C. H., Hinrichs, K. U., McKeegan, K. D., and DeLong, E. F. (2001) *Science*, **293**, 484-487.
13. Mills, H. J., Martinez, R. J., Story, S., and Sobecky, P. A. (2005) *Appl. Environ. Microbiol.*, **71**, 3235-3247.
14. MacGregor, B. J., Moser, D. P., Alm, E. W., Nealson, K. H., and Stahl, D. A. (1997) *Appl. Environ. Microbiol.*, **63**, 1178-1181.
15. Keough, B. P., Schmidt, T. M., and Hicks, R. E. (2003) *Microb. Ecol.*, **46**, 238-248.
16. Takai, K., Moser, D. P., DeFlaun, M., Onstott, T. C., and Fredrickson, J. K. (2001) *Appl. Environ. Microbiol.*, **67**, 5750-5760.
17. Lepp, P. W., Brinig, M. M., Ouverney, C. C., Palm, K., Armitage, G. C., and Relman, D. A. (2004) *Proc. Natl. Acad. Sci. USA*, **101**, 6176-6181.
18. Karner, M. B., DeLong, E. F., and Karl, D. M. (2001) *Nature*, **409**, 507-510.
19. Konneke, M., Bernhard, A. E., de la Torre, J. R., Walker, C. B., Waterbury, J. B., and Stahl, D. A. (2005) *Nature*, **437**, 543-546.
20. Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G. W., Prosser, J. I., Schuster, S. C., and Schleper, C. (2006) *Nature*, **442**, 806-809.
21. Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S., and Venter, J. C. (1996) *Science*, **273**, 1058-1073.
22. Elkins, J. G., Podar, M., Graham, D. E., Makarova, K. S., Wolf, Y., Randau, L., Hedlund, B. P., Brochier-Armanet, C., Kunin, V., Anderson, I., Lapidus, A., Goltsman, E., Barry, K., Koonin, E. V., Hugenholtz, P., Kyrpides, N., Wanner, G., Richardson, P., Keller, M., and Stetter, K. O. (2008) *Proc. Natl. Acad. Sci. USA*, **105**, 8102-8107.
23. Hallam, S. J., Konstantinidis, K. T., Putnam, N., Schleper, C., Watanabe, Y., Sugahara, J., Preston, C., de la Torre, J., Richardson, P. M., and DeLong, E. F. (2006) *Proc. Natl. Acad. Sci. USA*, **103**, 18296-18301.
24. Nunoura, T., Takaki, Y., Kakuta, J., Nishi, S., Sugahara, J., Kazama, H., Chee, G. J., Hattori, M., Kanai, A., Atomi, H., Takai, K., and Takami, H. (2011) *Nucleic Acids Res.*, **39**, 3204-3223.
25. Baker, B. J., Comolli, L. R., Dick, G. J., Hauser, L. J., Hyatt, D., Dill, B. D., Land, M. L., Verberkmoes, N. C.,

Hettich, R. L., and Banfield, J. F. (2010) *Proc. Natl. Acad. Sci. USA*, **107**, 8806-8811.
26. Meyerdierks, A., Kube, M., Kostadinov, I., Teeling, H., Glockner, F. O., Reinhardt, R., and Amann, R. (2010) *Environ. Microbiol.*, **12**, 422-439.
27. Blainey, P. C., Mosier, A. C., Potanina, A., Francis, C. A., and Quake, S. R. (2011) *PLoS One*, **6**, e16626.
28. Gribaldo, S., and Brochier-Armanet, C. (2006) *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, **361**, 1007-1022.
29. Bonch-Osmolovskaya, E. A., and Ravin, N. V. (2010) *Vestn. Ros. Akad. Nauk*, **80**, 977-984.
30. Matsunaga, F., Forterre, P., Ishmo, Y., and Myllykallio, H. (2001) *Proc. Natl. Acad. Sci. USA*, **98**, 11152-11157.
31. Myllykallio, H., Lopez, P., Lopez-Garcia, P., Heilig, R., Saurin, W., Zivanovic, Y., Philippe, H., and Forterre, P. (2000) *Science*, **288**, 2212-2215.
32. Robinson, N. P., Dionne, I., Lundgren, M., Marsh, V. L., Bernander, R., and Bell, S. D. (2004) *Cell*, **116**, 25-38.
33. Robinson, N. P., and Bell, S. D. (2007) *Proc. Natl. Acad. Sci. USA*, **104**, 5806-5811.
34. Forterre, P. (2002) *Curr. Opin. Microbiol.*, **5**, 525-532.
35. Kjems, J., and Garrett, R. A. (1988) *Cell*, **54**, 693-703.
36. Kjems, J., Leffers, H., Olesen, T., and Garrett, R. A. (1989) *J. Biol. Chem.*, **264**, 17834-17837.
37. Watanabe, Y., Yokobori, S., and Kawarabayasi, Y. (2002) *Tanpakushitsu Kakusan Koso*, **47**, 833-836.
38. She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Wei her, C. C., Clausen, I. G., Curtis, B. A., De Moors, A., Erauso, G., Fletcher, C., Gordon, P. M., Heikamp-de Jong, I., Jeffries, A. C., Kozera, C. J., Medina, N., Peng, X., Thi-Ngoc, H. P., Redder, P., Schenk, M. E., Thenault, C., Tolstrup, N., Charlebois, R. L., Doolittle, W. F., Duguet, M. T., Gaasterland, R. A., Garrett, M. A., Ragan, C. W., Sensen, C. W., and van der Oost, J. (2001) *Proc. Natl. Acad. Sci. USA*, **98**, 7835-7840.
39. Chen, L., Brugger, K., Skovgaard, M., Redder, P., She, Q., Torarinsson, E., Greve, B., Awayez, M., Zibat, A., Klenk, H. P., and Garrett, R. A. (2005) *J. Bacteriol.*, **187**, 4992-4999.
40. Jansen, R., Embden, J. D., Gaastra, W., and Schouls, L. M. (2002) *Mol. Microbiol.*, **43**, 1565-1575.
41. Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005) *Mol. Evol.*, **60**, 174-182.
42. Lillestol, R. K., Redder, P., Garrett, R. A., and Brugger, K. (2006) *Archaea*, **2**, 59-72.
43. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., and Horvath, P. (2007) *Science*, **315**, 1709-1712.
44. Marraffini, L. A., and Sontheimer, E. J. (2010) *Nat. Rev. Genet.*, **11**, 181-190.
45. Winker, S., and Woese, C. R. (1991) *Syst. Appl. Microbiol.*, **14**, 305-310.
46. Barns, S. M., Delwiche, C. F., Palmer, J. D., and Pace, N. R. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 9188-9193.
47. Takai, K., and Horikoshi, K. (1999) *Genetics*, **152**, 1285-1297.
48. Brochier-Armanet, C., Boussau, B., Gribaldo, S., and Forterre, P. (2008) *Nat. Rev. Microbiol.*, **6**, 245-252.
49. Walker, C. B., de la Torre, J. R., Klotz, M. G., Urakawa, H., Pinel, N., Arp, D. J., Brochier-Armanet, C., Chain, P. S., Chan, P. P., Gollabgir, A., Hemp, J., Hugler, M., Karr, E. A., Konneke, M., Shin, M., Lawton, T. J., Lowe, T., Martens-Habbena, W., Sayavedra-Soto, L. A., Lang, D., Sievert, S. M., Rosenzweig, A. C., Manning, G., and Stahl, D.A. (2010) *Proc. Natl. Acad. Sci. USA*, **107**, 8818-8823.
50. Spang, A., Hatzenpichler, R., Brochier-Armanet, C., Rattei, T., Tischler, P., Spieck, E., Streit, W., Stahl, D. A., Wagner, M., and Schleper, C. (2010) *Trends Microbiol.*, **18**, 331-340.
51. Woese, C. R., Kandler, O., and Wheelis, M. L. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 4576-4579.
52. Brock, T. D., Brock, K. M., Belly, R. T., and Weiss, R. L. (1972) *Arch. Microbiol.*, **84**, 54-68.
53. Kletzin, A. (2007) in *Archaea: Molecular and Cellular Biology* (Cavicchioli R., ed.) ASM Press Washington, USA, pp. 14-93.
54. Stetter, K. O. (1996) *FEMS Microbiol. Rev.*, **18**, 149-158.
55. Bonch-Osmolovskaya, E. A. (1994) *FEMS Microbiol. Rev.*, **15**, 65-77.
56. Mardanov, A. V., Gumerov, V. M., Beletsky, A. V., Prokofeva, M. I., Bonch-Osmolovskaya, E. A., Ravin, N. V., and Skryabin, K. G. (2011) *J. Bacteriol.*, **193**, 3156-3157.
57. Siebers, B., Zaparty, M., Raddatz, G., Tjaden, B., Albers, S. V., Bell, S. D., Blombach, F., Kletzin, A., Kyrpides, N., Lanz, C., Plagens, A., Rampp, M., Rosinus, A., von Jan, M., Makarova, K. S., Klenk, H. P., Schuster, S. C., and Hensel, R. (2011) *PLoS One*, **6**, e24222.
58. Selig, M., and Schonheit, P. (1994) *Arch. Microbiol.*, **162**, 286-294.
59. Fitz-Gibbon, S. T., Ladner, H., Kim, U. J., Stetter, K. O., Simon, M. I., and Miller, J. H. (2002) *Proc. Natl. Acad. Sci. USA*, **99**, 984-989.
60. Mardanov, A. V., Gumerov, V. M., Slobodkina, G. B., Beletsky, A. V., Bonch-Osmolovskaya, E. A., Ravin, N. V., and Skryabin, K. G. (2012) *J. Bacteriol.*, **194**, 727-728.
61. Cozen, A. E., Weirauch, M. T., Pollard, K. S., Bernick, D. L., Stuart, J. M., and Lowe, T. M. (2009) *J. Bact.*, **191**, 782-794.
62. Kawarabayasi, Y., Hino, Y., Horikawa, H., Yamazaki, S., Haikawa, Y., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A., Kosugi, H., Hosoyama, A., Fukui, S., Nagai, Y., Nishijima, K., Nakazawa, H., Takamiya, M., Masuda, S., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., Kubota, K., Nakamura, Y., Nomura, N., Sako, Y., and Kikuchi, H. (1999) *DNA Res.*, **6**, 83-101.
63. Ravin, N. V., Mardanov, A. V., Beletsky, A. V., Kublanov, I. V., Kolganova, T. V., Lebedinsky, A. V., Chernyh, N. A., Bonch-Osmolovskaya, E. A., and Skryabin, K. G. (2009) *J. Bacteriol.*, **191**, 2371-2379.
64. Anderson, I. J., Dharmarajan, L., Rodriguez, J., Hooper, S., Porat, I., Ulrich, L. E., Elkins, J. G., Mavromatis, K., Sun, H., Land, M., Lapidus, A., Lucas, S., Barry, K., Huber, H., Zhulin, I. B., Whitman, W. B., Mukhopadhyay, B., Woese, C., Bristow, J., and Kyrpides, N. (2009) *BMC Genomics*, **10**, 145.
65. Paper, W., Jahn, U., Hohn, M. J., Kronner, M., Nather, D. J., Burghardt, T., Rachel, R., Stetter, K. O., and Huber, H. (2007) *Int. J. Syst. Evol. Microbiol.*, **57**, 803-808.
66. Podar, M., Anderson, I., Makarova, K. S., Elkins, J. G., Ivanova, N., Wall, M. A., Lykidis, A., Mavromatis, K.,

Sun, H., Hudson, M. E., Chen, W., Deciu, C., Hutchison, D., Eads, J. R., Anderson, A., Fernandes, F., Szeto, E., Lapidus, A., Kyrpides, N. C., Saier, M. H., Jr., Richardson, P. M., Rachel, R., Huber, H., Eisen, J. A., Koonin, E. V., Keller, M., and Stetter, K. O. (2008) *Genome Biol.*, **9**, R158.

67. Prokofeva, M. I., Kostrikina, N. A., Kolganova, T. V., Tourova, T. P., Lysenko, A. M., Lebedinsky, A. V., and Bonch-Osmolovskaya, E. A. (2009) *Int. J. Syst. Evol. Microbiol.*, **59**, 3116-3122.
68. Mardanov, A. V., Svetlitchnyi, V. A., Beletsky, A. V., Prokofeva, M. I., Bonch-Osmolovskaya, E. A., Ravin, N. V., and Skryabin, K. G. (2010) *Appl. Environ. Microbiol.*, **76**, 5652-5657.
69. Brochier-Armanet, C., Forterre, P., and Gribaldo, S. (2011) *Curr. Opin. Microbiol.*, **14**, 274-281.
70. Perevalova, A. A., Bidzhieva, S., Kublanov, I. V., Hinrichs, K. U., Liu, X. L., Mardanov, A. V., Lebedinsky, A. V., and Bonch-Osmolovskaya, E. A. (2010) *Int. J. Syst. Evol. Microbiol.*, **60**, 2082-2088.
71. Bertoldo, C., and Antranikian, G. (2006) in *The Prokaryotes*, 3rd Edn (M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, and E. Stackebrandt, eds.) Springer-Verlag, New York, Vol. 3, pp. 69-81.
72. Klenk, H. P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. E., Ketchum, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D., Richardson, D. L., Kerlavage, A. R., Graham, D. E., Kyrpides, N. C., Fleischmann, R. D., Quackenbush, J., Lee, N. H., Sutton, G. G., Gill, S., Kirkness, E. F., Dougherty, B. A., McKenney, K., Adams, M. D., Loftus, B., Peterson, S., Reich, C. I., McNeil, L. K., Badger, J. H., Glodek, A., Zhou, L., Overbeek, R., Gocayne, J. D., Weidman, J. F., McDonald, L., Utterback, T., Cotton, M. D., Spriggs, T., Artiach, P., Kaine, B. P., Sykes, S. M., Sadow, P. W., D'Andrea, K. P., Bowman, C., Fujii, C., Garland, S. A., Mason, T. M., Olsen, G. J., Fraser, C. M., Smith, H. O., Woese, C. R., and Venter, J. C. (1997) *Nature*, **390**, 364-370.
73. Gumerov, V. M., Mardanov, A. V., Beletsky, A. V., Prokofeva, M. I., Bonch-Osmolovskaya, E. A., Ravin, N. V., and Skryabin, K. G. (2011) *J. Bacteriol.*, **193**, 2355-2356.
74. Kurr, M., Huber, R., Konig, H., Jannasch, H. W., Fricke, H., Trincone, A., Krstjansson, J. K., and Stetter, K. O. (1991) *Arch. Microbiol.*, **156**, 239-247.
75. Oren, A. (2006) in *The Prokaryotes* (Dworkin, M., ed.) *Springer*, **3**, 113-164.
76. Sayeh, R., Birrien, J. L., Alain, K., Barbier, G., Hamdi, M., and Prieur, D. (2010) *Extremophiles*, **14**, 501-514.
77. Golyshina, O. V., Pivovarova, T. A., Karavaiko, G. I., Kondrat'eva, T. F., Moore, E. R. B., Abraham, W. R., Lunsdorf, H., Timmis, K. N., Yakimov, M. M., and Golyshin, P. N. (2000) *Int. J. Syst. Evol. Microbiol.*, **50**, 997-1006.
78. Sorensen, K. B., and Teske, A. (2006) *Appl. Environ. Microbiol.*, **72**, 4596-4603.
79. Teske, A., and Sorensen, K. B. (2008) *ISME J.*, **2**, 3-18.
80. Schleper, C. (2007) *Archaea: Evolution, Physiology, and Molecular Biology* (Garrett, R., and Hans-Peter Klenk, H.-P., eds.) Blackwell Publishing Ltd.
81. Reigstad, L. J., Jorgensen, S. L., and Schleper, C. (2010) *ISME J.*, **4**, 346-356.
82. Gumerov, V. M., Mardanov, A. V., Beletsky, A. V., Bonch-Osmolovskaya, E. A., and Ravin, N. V. (2011) *Mikrobiologiya*, **80**, 258-265.
83. Nunoura, T., Hirayama, H., Takami, H., Oida, H., Nishi, S., Shimamura, S., Suzuki, Y., Inagaki, F., Takai, K., Nealson, K. H., and Horikoshi, K. (2005) *Environ. Microbiol.*, **7**, 1967-1984.
84. Kim, B. K., Jung, M. Y., Yu, D. S., Park, S. J., Oh, T. K., Rhee, S. K., and Kim, J. F. (2011) *J. Bacteriol.*, **193**, 5539-5540.
85. Konneke, M., Bernhard, A. E., de la Torre, J. R., Walker, C. B., Waterbury, J. B., and Stahl, D. A. (2005) *Nature*, **437**, 543-546.
86. Schleper, C., and Nicol, G. W. (2010) *Adv. Microb. Physiol.*, **57**, 1-41.
87. Wuchter, C., Abbas, B., Coolen, M. J., Herfort, L., van Bleijswijk, J., Timmers, P., Strous, M., Teira, E., Herndl, G. J., Middelburg, J. J., Schouten, S., and Sinninghe Damste, J. S. (2006) *Proc. Natl. Acad. Sci. USA*, **103**, 12317-12322.
88. Hallam, S. J., Putnam, N., Preston, C. M., Detter, J. C., Rokhsar, D., Richardson, P. M., and DeLong, E. F. (2004) *Science*, **305**, 1457-1462.
89. Reysenbach, A.-L., Liu, Y., Banta, A. B., Beveridge, T. J., Kirshtein, J. D., Schouten, S., Tivey, M. K., von Damm, K. L., and Voytek, M. A. (2006) *Nature*, **442**, 444-447.
90. Reysenbach, A. L., and Flores, G. E. (2008) *Geobiology*, **6**, 331-336.
91. Huber, H., Hohn, M. J., Rachel, R., Fuchs, T., Wimmer, V. C., and Stetter, K. O. (2002) *Nature*, **417**, 63-67.
92. Hohn, M. J., Hedlund, B. P., and Huber, H. (2002) *Syst. Appl. Microbiol.*, **25**, 551-554.
93. McCliment, E. A., Voglesonger, K. M., O'Day, P. A., Dunn, E. E., Holloway, J. R., and Cary, S. C. (2006) *Environ. Microbiol.*, **8**, 114-125.
94. Casanueva, A., Galada, N., Baker, G. C., Grant, W. D., Heaphy, S., Jones, B., Yanhe, M., Ventosa, A., Blamey, J., and Cowan, D. A. (2008) *Extremophiles*, **12**, 651-656.
95. Waters, E., Hohn, M. J., Ahel, I., Graham, D. E., Adams, M. D., Barnstead, M., Beeson, K. Y., Bibbs, L., Bolanos, R., Keller, M., Kretz, K., Lin, X., Mathur, E., Ni, J., Podar, M., Richardson, T., Sutton, G. G., Simon, M., Soil, D., Stetter, K. O., Short, J. M., and Noordewier, M. (2003) *Proc. Natl. Acad. Sci. USA*, **100**, 12984-12988.
96. Di Giulio, M. (2007) *Gene*, **401**, 108-113.
97. Makarova, K. S., and Koonin, E. V. (2005) *Curr. Opin. Microbiol.*, **117**, 52-67.
98. Brochier, C., Forterre, P., and Gribaldo, S. (2005) *BMC Evol. Biol.*, **5**, 36.
99. Forterre, P., Gribaldo, S., and Brochier-Armanet, C. (2009) *J. Biol.*, **8**, 7.
100. Fujishima, K., Sugahara, J., Kikuta, K., Hirano, R., Sato, A., Tomita, M., and Kanai, A. (2009) *Proc. Natl. Acad. Sci. USA*, **106**, 2683-2687.
101. Baker, B. J., Tyson, G. W., Webb, R. I., Flanagan, J., Hugenholtz, P., Allen, E. E., and Banfield, J. F. (2006) *Science*, **314**, 1933-1935.
102. Narasingarao, P., Podell, S., Ugalde, J. A., Brochier-Armanet, C., Emerson, J. B., Brocks, J. J., Heidelberg, K. B., Banfield, J. F., and Allen, E. E. (2012) *ISME J.*, **6**, 81-93.

103. Cohen, G. N., Barbe, V., Flament, D., Galperin, M., Heilig, R., Lecompte, O., Poch, O., Prieur, D., Querellou, J., Ripp, R., Thierry, J. C., van der Oost, J., Weissenbach, J., Zivanovic, Y., and Forterre, P. (2003) *Mol. Microbiol.*, **47**, 1495-1512.
104. Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y., Sakai, M., Ogura, K., Otsuka, R., Nakazawa, H., Takamiya, M., Ohfuku, Y., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., and Kikuchi, H. (1998) *DNA Res.*, **5**, 147-155.
105. Zivanovic, Y., Lopez, P., Philippe, H., and Forterre, P. (2002) *Nucleic Acids Res.*, **30**, 1902-1910.
106. Brugger, K., Torarinsson, E., Redder, P., Chen, L., and Garrett, R. A. (2004) *Biochem. Soc. Trans.*, **32**, 179-183.
107. Robb, F. T., Maeder, D. L., Brown, J. R., DiRuggiero, J., Stump, M. D., Yeh, R. K., Weiss, R. B., and Dunn, D. M. (2001) *Methods Enzymol.*, **330**, 134-157.
108. Lake, J. A., Henderson, E., Oakes, M., and Clark, M. W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 3786-3790.
109. Makarova, K. S., Aravind, L., Galperin, M. Y., Grishin, N. V., Tatusov, R. L., Wolf, Y. I., and Koonin, E. V. (1999) *Genome Res.*, **9**, 608-628.
110. Galagan, J. E., Nusbaum, C., Roy, A., Endrizzi, M. G., Macdonald, P., FitzHugh, W., Calvo, S., Engels, R., Smirnov, S., Atnoor, D., Brown, A., Allen, N., Naylor, J., Stange-Thomann, N., DeArellano, K., Johnson, R., Linton, L., McEwan, P., McKernan, K., Talamas, J., Tirrell, A., Ye, W., Zimmer, A., Barber, R. D., Cann, I., Graham, D. E., Grahame, D. A., Guss, A. M., Hedderich, R., Ingram-Smith, C., Kuettner, H. C., Krzycki, J. A., Leigh, J. A., Li, W., Liu, J., Mukhopadhyay, B., Reeve, J. N., Smith, K., Springer, T. A., Umayam, L. A., White, O., White, R. H., Conway de Macario, E., Ferry, J. G., Jarrell, K. F., Jing, H., Macario, A. J., Paulsen, I., Pritchett, M., Sowers, K. R., Swanson, R. V., Zinder, S. H., Lander, E., Metcalf, W. W., and Birren, B. (2002) *Genome Res.*, **12**, 532-542.
111. Diruggiero, J., Dunn, D., Maeder, D. L., Holley-Shanks, R., Chatard, J., Horlacher, R., Robb, F. T., Boos, W., and Weiss, R. B. (2000) *Mol. Microbiol.*, **38**, 684-693.
112. Mardanov, A. V., Ravin, N. V., Svetlitchnyi, V. A., Beletsky, A. V., Miroshnichenko, M. L., Bonch-Osmolovskaya, E. A., and Skryabin, K. G. (2009) *Appl. Environ. Microbiol.*, **75**, 4580-4588.
113. Futterer, O., Angelov, A., Liesegang, H., Gottschalk, G., Schleper, C., Schepers, B., Dock, C., Antranikian, G., and Liebl, W. (2004) *Proc. Natl. Acad. Sci. USA*, **101**, 9091-9096.
114. Ruepp, A., Graml, W., Santos-Martinez, M. L., Koretke, K. K., Volker, C., Mewes, H. W., Frishman, D., Stocker, S., Lupas, A. N., and Baumeister, W. (2000) *Nature*, **407**, 508-513.
115. Pina, M., Bize, A., Forterre, P., and Prangishvili, D. (2011) *FEMS Microbiol. Rev.*, **35**, 1035-1054.
116. Haring, M., Peng, X., Brugger, K., Rachel, R., Stetter, K. O., Garrett, R. A., and Prangishvili, D. (2004) *Virology*, **323**, 233-242.
117. Cortez, D., Forterre, P., and Gribaldo, S. (2009) *Genome Biol.*, **10**, 65.
118. Lindas, A. C., Karlsson, E. A., Lindgren, M. T., Ettema, T. J., and Bernander, R. (2008) *Proc. Natl. Acad. Sci. USA*, **105**, 18942-18946.
119. Csuros, M., and Miklos, I. (2009) *Mol. Biol. Evol.*, **26**, 2087-2095.
120. Leigh, J. A., Albers, S. V., Atomi, H., and Allers, T. (2011) *FEMS Microbiol. Rev.*, **35**, 577-608.