



Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes

Toshiaki Fukui, Haruyuki Atomi, Tamotsu Kanai, et al.

Genome Res. 2005 15: 352-363

Access the most recent version at doi:[10.1101/gr.3003105](https://doi.org/10.1101/gr.3003105)

References This article cites 78 articles, 37 of which can be accessed free at:
<http://genome.cshlp.org/content/15/3/352.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes

Toshiaki Fukui,^{1,2,4} Haruyuki Atomi,^{1,2} Tamotsu Kanai,^{1,2} Rie Matsumi,¹ Shinsuke Fujiwara,³ and Tadayuki Imanaka^{1,2,5}

¹Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, and ²Katsura Int'tech Center, Kyoto University, Katsura, Nishikyo-ku, Kyoto 615-8510, Japan; ³Department of Bioscience, Nanobiotechnology Research Center, School of Science and Technology, Kwansai Gakuin University, 2-1 Gakuen, Sanda 669-1337, Japan

The genus *Thermococcus*, comprised of sulfur-reducing hyperthermophilic archaea, belongs to the order Thermococcales in Euryarchaeota along with the closely related genus *Pyrococcus*. The members of *Thermococcus* are ubiquitously present in natural high-temperature environments, and are therefore considered to play a major role in the ecology and metabolic activity of microbial consortia within hot-water ecosystems. To obtain insight into this important genus, we have determined and annotated the complete 2,088,737-base genome of *Thermococcus kodakaraensis* strain KOD1, followed by a comparison with the three complete genomes of *Pyrococcus* spp. A total of 2306 coding DNA sequences (CDSs) have been identified, among which half (1165 CDSs) are annotatable, whereas the functions of 41% (936 CDSs) cannot be predicted from the primary structures. The genome contains seven genes for probable transposases and four virus-related regions. Several proteins within these genetic elements show high similarities to those in *Pyrococcus* spp., implying the natural occurrence of horizontal gene transfer of such mobile elements among the order Thermococcales. Comparative genomics clarified that 1204 proteins, including those for information processing and basic metabolisms, are shared among *T. kodakaraensis* and the three *Pyrococcus* spp. On the other hand, among the set of 689 proteins unique to *T. kodakaraensis*, there are several intriguing proteins that might be responsible for the specific trait of the genus *Thermococcus*, such as proteins involved in additional pyruvate oxidation, nucleotide metabolisms, unique or additional metal ion transporters, improved stress response system, and a distinct restriction system.

[Supplemental material is available online at www.genome.org. The whole genome of *Thermococcus kodakaraensis* KOD1 has been deposited in the DDBJ/EMBL/GenBank databases under the project accession no. APO06878.]

The euryarchaeal order Thermococcales, composed of two major genera *Thermococcus* and *Pyrococcus* (Huber and Stetter 2001; Itoh 2003), may be the best-studied hyperthermophiles. They are strictly anaerobic obligate heterotrophs growing on complex proteinaceous substrates, and their growth is strongly associated with the reduction of elemental sulfur. Alternatively, with a few exceptions, they are capable of gaining energy by fermentation of peptides, amino acids, and sugars, forming acids, CO₂, and H₂ in the absence of elemental sulfur (Amend and Shock 2001; Huber and Stetter 2001).

The genus *Pyrococcus*, with a higher optimum growth temperature (95°C–103°C) than *Thermococcus* (75°C–93°C), has fascinated many microbiological researchers and has often been used as the source organism for both fundamental and application-based aspects of research. Therefore, although within the same genus, the complete genome analyses of three species—*Pyrococcus horikoshii* (Kawarabayashi et al. 1998), *Pyrococcus furiosus*

(Robb et al. 2001), and *Pyrococcus abyssi* (Cohen et al. 2003)—have been performed. In contrast to *Pyrococcus*, the genus *Thermococcus* contains the highest number of characterized isolates (Itoh 2003). Recent culture-dependent and culture-independent studies have indicated that the members of *Thermococcus* are more ubiquitously present in various deep-sea hydrothermal vent systems than those of *Pyrococcus* (Orphan et al. 2000; Holden et al. 2001). Consequently, *Thermococcus* strains, with their larger population, are considered to play a major role in the ecology and metabolic activity of microbial consortia within marine hot-water ecosystems. However, despite the importance of this genus, no complete genome sequence has been determined for *Thermococcus*. The *Thermococcus* genome can be expected to encode genes responsible for various cellular functions that provide an advantage for these strains in natural high-temperature habitats.

Thermococcus kodakaraensis KOD1 was isolated from a solfatar on the shore of Kodakara Island, Kagoshima, Japan (Mori-kawa et al. 1994; Atomi et al. 2004). Since the isolation, an abundant number of genes and their protein products from this archaeon have been examined (Imanaka and Atomi 2002), such as DNA polymerase, commercially available as an excellent enzyme for PCR amplification (Nishioka et al. 2001). Moreover, as we recently developed the first gene disruption system for a hyper-

⁴Present address: Department of Bioengineering, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama 226-8501, Japan.

⁵Corresponding author.

E-mail imanaka@sbchem.kyoto-u.ac.jp; fax 81-75-383-2778.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.3003105>. Article published online before print in February 2005.

thermophile with this archaeon (Sato et al. 2003), *T. kodakaraensis* can be regarded as one of the most useful model organisms in the research on hyperthermophiles. Here we describe the complete genome analysis of *T. kodakaraensis* KOD1 as well as a comparison with *Pyrococcus* genomes, to gain further insight into the intriguing order Thermococcales.

Results and Discussion

General features

T. kodakaraensis has a single circular chromosome of 2,088,737 bp with no exochromosomal element, and the general features are listed in Table 1. There are a single 16S–23S rRNA cluster including tRNA^{Ala} (TGC) within the intergenic region, two copies of 5S rRNA, and a 7S rRNA for signal recognition particle. In all, 46 tRNA genes are scattered over the genome, among which tRNA^{Met} and tRNA^{Trp} are predicted to harbor an intron.

By a combination of coding potential prediction and homology search, 2306 coding DNA sequences (CDSs) with an average length of 833 bp, covering 92% of the overall genome, can be identified on the genome (Fig. 1; all genes along with the annotations are listed in Supplemental Table 1). These are in good accordance with the general correlation between the microbial genome size and the predicted CDS numbers. When protein similarity was searched at a genome-wide scale, 77% of *T. kodakaraensis* proteins show the highest similarities toward proteins from the class Thermococci, including 240 proteins that are identified to be Thermococcales-specific.

Through homology and domain searches of the translated products of CDSs, functions of half (1165 CDSs) are annotatable, and 9% of CDSs (205 CDSs) show homology against proteins with general function prediction only. The functions of the remains (41%, 936 CDSs) cannot be predicted from the deduced primary structures, among which 261 CDSs are unique to *T. ko-*

dakaraensis with no significant homology to any known sequences to date. We identified 186 multigene families, and the largest family is the ATP-binding subunit of ABC transporters, containing 36 members, as is the case in other microorganisms. The general metabolisms and solute transport in *T. kodakaraensis*, deduced from the annotatable CDSs, are illustrated in Figure 2. The detailed features of gains and losses of genes in this archaeon are described below in the sections dealing on specific cellular functions.

Probable mobile elements

The *T. kodakaraensis* genome harbors seven genes encoding putative transposases (TK0298, TK0495, TK0654, TK0850, TK0931, TK0932, and TK1842). Six of them, except for TK0932, are related to orfB of the IS605 family (Mahillon and Chandler 1998), suggesting that these transposases share a common phylogenetic origin. IS605 was originally identified in *Helicobacter pylori*, and the proteins homologous to orfB (assigned as COG0675 in COGs) comprise a ubiquitous transposase group in the archaeal genomes. TK0932 is homologous to orfA of IS605, constituting an entire composite of the IS605-like element along with the adjacent TK0931. TK0654 and TK1842 genes are associated with a gene encoding a protein weakly related to site-specific integrase (previously annotated as resolvase in *Mycobacterium tuberculosis*), as seen in the IS607 element from *H. pylori*. Among the transposases in *T. kodakaraensis*, TK0298, TK0495, and TK0850 were 87%–90% identical (at the amino acid level) to one another. These may be the result of a rather ancient event where duplicated translocation or multiple integration of one mobile element occurred within this species. Although no inverted repeats can be found in the terminal regions of these transposase genes, it has been reported that the IS200/IS605 family generally does not carry the terminal inverted repeats, unlike the general insertion sequence (IS) elements (Mahillon and Chandler 1998). A crenarchaeon, *Sulfolobus solfataricus*, has been demonstrated to

show a high spontaneous mutation rate mediated by active IS elements belonging to the IS4–IS5 family (Martusewitsch et al. 2000). However, through our isolation of 5-fluoroorotic acid-resistant mutants of *T. kodakaraensis* (Sato et al. 2003), we could not observe any inactivation of the *pyrE* or *pyrF* gene by IS elements. The transposases in *T. kodakaraensis* are thought to have already lost their transposition activity during evolution after the final transposition.

In the crenarchaea *Sulfolobus*, site-specific integration of viruses or virus-related plasmids occur at tRNA genes with the same integrase-mediated mechanism (Peng et al. 2000). Similar regions bordered by a partitioned integrase gene have also been found in euryarchaea *P. horikoshii* and *Methanococcus jannaschii* (Makino et al. 1999; She et al. 2001). We can identify four potential regions, TKV1–TKV4, where genetic elements have been integrated into the *T. kodakaraensis* genome (Fig. 1). In the regions of TKV1 (23,592 bp), TKV2 (27,203 bp), and TKV3 (27,886 bp),

Table 1. General features of *Thermococcus kodakaraensis* genome along with three *Pyrococcus* genomes

	<i>T. kodakaraensis</i>	<i>P. abyssi</i>	<i>P. horikoshii</i>	<i>P. furiosus</i>
Genome features				
Genome size (bp)	2,088,737	1,765,118	1,738,505	1,908,256
Protein-coding regions	92.1%	91.1% ^a	91.2% ^a	92.5% ^a
GC content	52.0%	44.7%	41.9%	40.8%
Stable RNAs				
tRNAs	46	46	46	46
tRNAs with introns	Trp, Met	Trp, Met	Trp, Met	Trp, Met
rRNAs	16S-23S	16S-23S	16S-23S	16S-23S
Others	5S (×2), 7S RNaseP	5S (×2), 7S RNaseP	5S (×2), 7S RNaseP	5S (×2), 7S RNaseP
CDSs	2306	1784 ^a	2064 ^a	2065 ^a
Mobile elements				
Inteins	16	14 ^b	14 ^b	10 ^b
Transposases	7	5	3	>30
Virus-related regions	4	0	2	0
Long clusters of tandem repeats^c				
R1 family	0	1 ^c	3 ^c	0 ^c
R2 family	3	3 ^c	3 ^c	7 ^c

^aBased on AL09836 for *P. abyssi*, BA000001 for *P. horikoshii*, and AL009950 for *P. furiosus* from European Bioinformatics Institute (EBI).

^bInBASE (Perler 2002).

^cZivanovic et al. (2002).

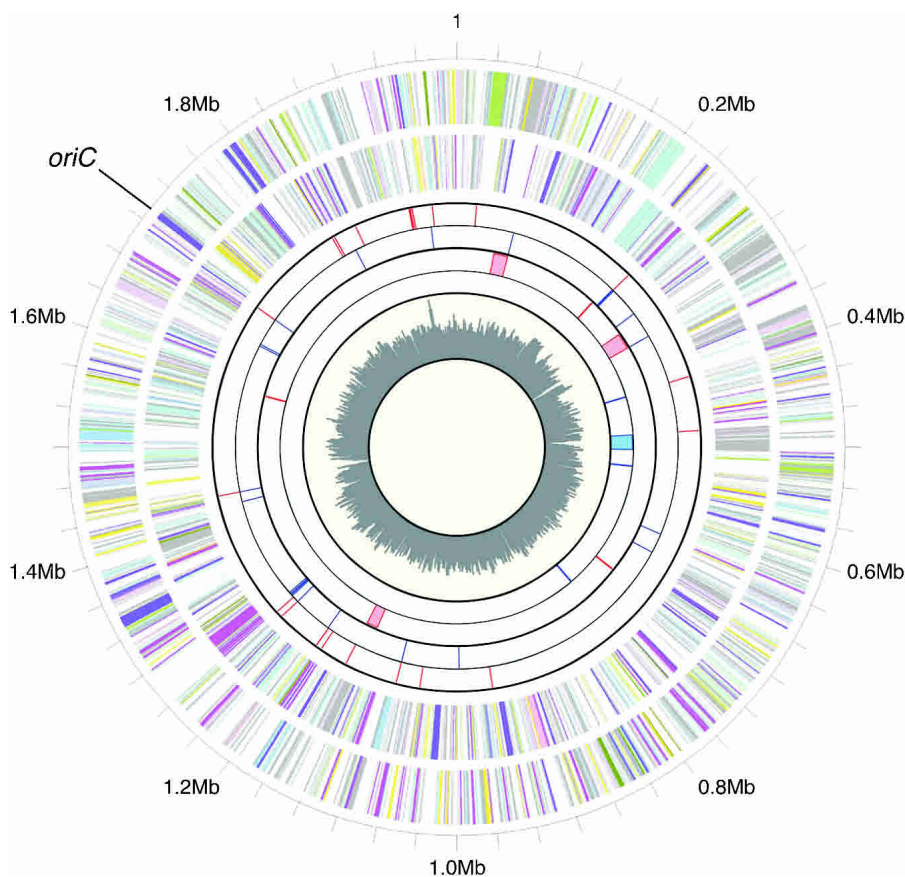


Figure 1. Circular representation of the *T. kodakaraensis* genome. Circles, from outer to inner, display (1) Physical map scaled in megabases from base 1 (outermost circle). The putative replication origin (*oriC*) was predicted from comparison with the *P. abyssi* genome. (2) Predicted protein-coding regions in the clockwise direction (the second circle) and in the counterclockwise direction (the third circle), respectively. The displayed genes are color-coded according to the functional classification: (magenta) translation/ribosomal structure/biogenesis; (pink) transcription; (pale pink) DNA replication/recombination/repair; (forest green) cell division/chromosome partitioning; (yellow) post-translational modification/protein turnover/chaperones; (light yellow) cell envelope biogenesis/outer membrane; (light green) cell motility/secretion; (pale green) inorganic ion transport/metabolism; (medium turquoise) signal transduction mechanisms; (purple) energy production/conversion; (light blue) carbohydrate transport/metabolism; (cyan) amino acid transport/metabolism; (violet) nucleotide transport/metabolism; (pale turquoise) coenzyme metabolism; (medium purple) lipid metabolism; (light sky blue) secondary metabolites biosynthesis/transport/catabolism; (light gray) general function prediction only; (dark gray) function unknown. (3) Predicted tRNA coding regions in the clockwise direction (red in the fourth circle) and in the counterclockwise direction (blue in the fifth circle), respectively. (4) Predicted mobile elements in the clockwise direction (red in the sixth circle) and in the counterclockwise direction (blue in the seventh circle), respectively. The vertical lines indicate predicted transposase genes, and the boxes indicate virus-related regions. (4) G+C percent content in a 10-kb window and 1-kb incremental shift, where the values larger than the minimum value (38%) are plotted (innermost circle).

genes corresponding to the N-terminal region of the integrase (homologous to integrase from SSV1 virus for *Sulfolobus shibatae*) are overlapped with downstream halves of tRNA^{Val}, tRNA^{Glu}, and tRNA^{Arg}, respectively, with an opposite orientation. These overlapped regions (48 bp) are predicted to be attachment (*att*) sites, since the genes for the C-terminal region of the integrase also contain the respective overlapped sequences with perfect match (TKV1) or high identities (71% in TKV2 and 69% in TKV3). These *att* sequences were highly homologous to those seen in *P. horikoshii*, *M. jamaeschii*, and SSV1 and pRN family plasmids for *Sulfolobus*. In contrast, orientation of the *att* site in TKV4 (18,824 bp) is the same as that of tRNA^{Leu}, and the sequence is distinct from

those for TKV1–TKV3 or for related elements in other archaea, suggesting a phylogenetic difference of TKV4 from the others. These TKV regions contain many small CDSs, but a large fraction of them show no homology to known genes, even to the genes within the elements from *Sulfolobus* spp. TKV2 and TKV3 are related to one another, as many genes shared significant similarities with conserved organization. Especially, 8.1-kb regions in TKV2 and TKV3 are nearly identical at the nucleotide level. It is interesting that a few orthologs involved in DNA replication are encoded in the TKV regions, the replication licensing factor MCM homologs in TKV1 (TK0096) and TKV4 (TK1361), and the DNA polymerase sliding clamp (PCNA) homolog in TKV3 (TK0582). These orthologs might be functional in the selfish utilization of the replication system in the host cell for these elements.

It has been known that archaeal genomes often contain inteins, in-frame intervening sequences excised by a self-catalytic mechanism after translation (Perler 2002). In the *T. kodakaraensis* genome, 15 inteins can be identified, among which 13 inteins contain a DOD class homing endonuclease domain. While all the ex-teins belong to previously identified ex-tein categories, endonuclease domain-coding inteins in genes for RadA and DP2 subunit of DNA polymerase II are unique in *T. kodakaraensis* when compared to previously known archaeal inteins.

Comparison with *Pyrococcus* genomes

Table 1 also lists the general features of previously determined genomes of three *Pyrococcus* species. As a well-known difference between the genera *Thermococcus* and *Pyrococcus*, the average GC content of the *T. kodakaraensis* genome (52.0%) is higher than the 40%–45% observed for *Pyrococcus* genomes. The CDS identification and the following gene

family detection clarified the absence of large noncoding regions and the low extent of gene duplications in the *T. kodakaraensis* genome, as described above. These facts suggested that the slightly larger genome of *T. kodakaraensis* with more CDSs than those of *Pyrococcus* could be attributed to the differential gain of genes during the evolution after branching off from the genus *Pyrococcus*. The gains of genes are generally assumed to occur through horizontal gene transfer (Ochman et al. 2000; Mira et al. 2002). Although a possibility of lateral gene transfer of a 16-kb composite transposon between two closely related hyperthermophilic archaea, *P. furiosus* and *Thermococcus litoralis*, has been reported (Diruggiero et al. 2000), such a composite transposon

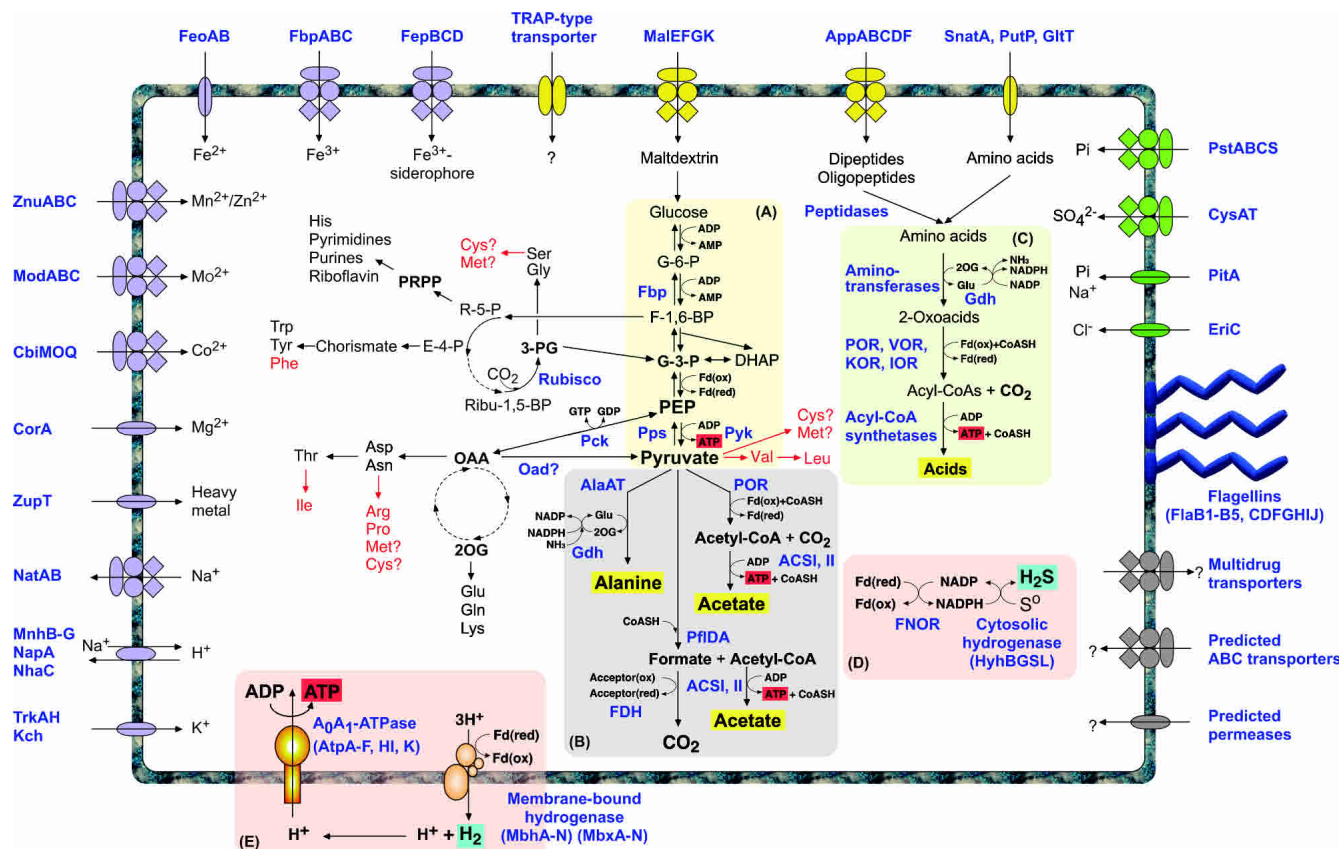


Figure 2. An overview of central metabolism and solute transport in *T. kodakaraensis*. Components or pathways for which no predictable enzymes were found are represented in red. Each gene product with a predicted function in ion or solute transport is illustrated on the membrane. The transporters and permeases are grouped by substrate specificity, as (violet) cations, (green) anions, (yellow) carbohydrates/carboxylates/amino acids, and (gray) unknown. (A) Glycolysis (modified Embden-Meyerhof pathway). (B) Pyruvate degradation. (C) Amino acid degradation. (D) Sulfur reduction. (E) Hydrogen evolution and formation of proton-motive force, coupled with ATP generation. (DHAP) Dihydroxyacetone phosphate; (E-4-P) erythrose 4-phosphate; (F-1,6-BP) fructose 1,6-bisphosphate; (G-3-P) glyceraldehyde 3-phosphate; (G-6-P) glucose 6-phosphate; (OAA) oxaloacetate; (2OG) 2-oxoglutarate; (PEP) phosphoenolpyruvate; (3-PG) 3-phosphoglycerate; (PRPP) 5-phosphoribosyl 1-pyrophosphate; (R-5-P) ribose 5-phosphate; (Ribul-1,5-BP) ribulose 1,5-bisphosphate; (ACS) acetyl-CoA synthetase (ADP-forming); (AlaAT) alanine aminotransferase; (Fbp) fructose 1,6-bisphosphatase; (FDH) formate dehydrogenase; (FNOR) ferredoxin:NADP oxidoreductase; (Gdh) glutamate dehydrogenase; (IOR) indolepyruvate:ferredoxin oxidoreductase; (KOR) 2-oxoacid:ferredoxin oxidoreductase; (PflDA) pyruvate formate lyase and its activating enzyme; (Oad) oxaloacetate decarboxylase; (Pck) phosphoenolpyruvate carboxykinase; (POR) pyruvate:ferredoxin oxidoreductase; (Pps) phosphoenolpyruvate synthase; (Pyk) pyruvate kinase; (VOR) 2-oxoisovalerate:ferredoxin oxidoreductase; (AppABCD) ABC-type dipeptide/oligopeptide transporter; (CbiMOQ) ABC-type Co^{2+} transporter; (CorA) $\text{Mg}^{2+}/\text{Co}^{2+}$ transporter; (CysAT) ABC-type sulfate transporter; (EriC) voltage-gated Cl^- channel protein; (FbpABC) ABC-type Fe^{3+} transporter; (FeoAB) Fe^{2+} transporter; (FepBCD) ABC-type Fe^{3+} -siderophore transporter; (Glt) H^+ /glutamate symporter; (Kch) Ca^{2+} -gated K^+ channel protein; (MalEFGK) ABC-type maltodextrin transporter; (MnhB-G) multisubunit Na^+/H^+ antiporter; (ModABC) ABC-type Mo^{2+} transporter; (NapA and NhaC) Na^+/H^+ antiporter; (NatAB) ABC-type Na^+ efflux pump; (PitA) Na^+ /phosphate symporter; (PstABCS) ABC-type phosphate transporter; (PutP) Na^+ /proline symporter; (SnatA) small neutral amino acid transporter; (TrkAH) Trk-type K^+ transporter; (ZnuABC) ABC-type $\text{Mn}^{2+}/\text{Zn}^{2+}$ transporter; (ZupT) heavy-metal cation transporter.

flanked by identical IS elements is not present in the *T. kodakaraensis* genome. However, comparative genomics among Thermococcales reveal several mobile elements exhibiting high similarities over the genus, as described in the Supplemental material.

We carried out clustering analysis of proteins from the predicted CDSs in the four Thermococcales species, as described in Methods. Figure 3 visualizes locations of the homologous genes on the chromosomes of Thermococcales. It is obvious that there is no large vacant region unrelated to *Pyrococcus* genomes in the *T. kodakaraensis* genome, indicating the absence of a recent transfer of a large segment derived from a distantly related lineage. The result also represents chromosome shuffling-driven evolution among *Pyrococcus* spp., as described previously (Lecomte et al. 2001; Zivanovic et al. 2002). In contrast to the well-conserved gene context among *Pyrococcus* genomes, the counterparts are

highly translocated within the *T. kodakaraensis* genome probably due to progressed rearrangement, so there is no longer segmental synteny between *T. kodakaraensis* and *Pyrococcus* genomes. Generally, genome rearrangement is faster than evolution of protein sequences, and the frequency of single-gene translocations has been pointed out to correlate with the number of IS elements and repeat sequences in the genome (Mira et al. 2002). The highly rearranged genome of *T. kodakaraensis* from *Pyrococcus* genomes with the relatively low fraction of IS elements and repeats may simply be a reflection of evolutionary distance, or alternatively, might suggest an acceleration of rearrangement under extremely high-temperature environments. Comparison to a genome of another *Thermococcus* with a lower growth temperature will help us understand this matter. As shown in Figure 4, since more than half of the proteins (1204 proteins) were shared by *T. kodaka-*

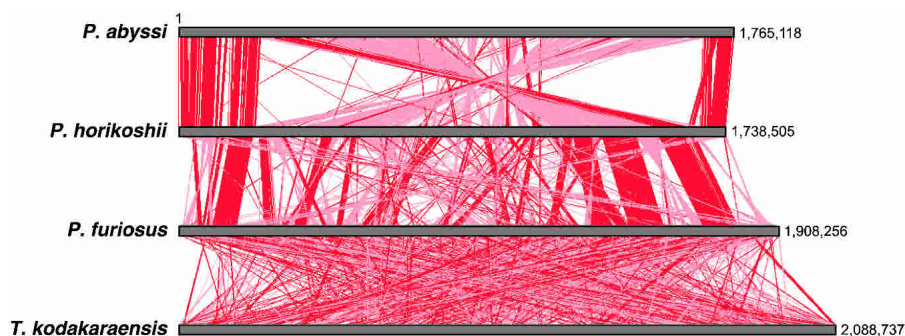


Figure 3. An integrated view of the genomic location of homologous CDSs in *P. abyssi* versus *P. horikoshii*, *P. horikoshii* versus *P. furiosus*, and *P. furiosus* versus *T. kodakaraensis*. Pairs of homologous CDSs were generated by clustering analysis of the proteins from four Thermococcales strains. The homologous CDSs between two organisms are shown by connected red lines for pairs in the same direction and pale pink lines for those in the opposite direction.

raensis and all three *Pyrococcus* spp., this shared set might relate to the common ancestor of Thermococcales. The remaining 700–1100 proteins are distributed among the order, and 689 proteins are unique in *T. kodakaraensis* and not present in *Pyrococcus* spp., as listed in Supplemental Table 2. Although many of them encode proteins with unknown function, there are several intriguing genes that might be responsible for the specific trait of the genus *Thermococcus*. On the other hand, 130 proteins are commonly present in three *Pyrococcus* spp. but not in *T. kodakaraensis*, as listed in Supplemental Table 3. It has been reported that *M. jannaschii* cells contain highly multiple copies of the chromosome especially in the exponential phase (up to 15) (Malandrin et al. 1999). *T. kodakaraensis* harbors the largest chromosome that has been sequenced among the Thermococcales and exhibits a relatively low growth temperature. It will be of interest in the future to determine the chromosome copy number of this archaeon and examine if there are any relationships with genome size and growth temperature.

Recent comparative genomics has allowed us to detect and consider genome-wide amino acid replacements among proteins from different sources of interest. Several studies have focused on the comparison between organisms with different growth temperatures to obtain insight on protein thermostability (McDonald et al. 1999; Chakravarty and Varadarajan 2000). In particular, recent studies have focused on comparison within related lineages, mesophilic and thermophilic *Methanococcus* species (Haney et al. 1999), and two mesophilic *Corynebacterium* species (*Corynebacterium efficiens* and *Corynebacterium glutamicum*) with slightly different optimum temperatures for growth (Nishio et al. 2003). We examined amino acid replacements between proteins from *T. kodakaraensis* and *Pyrococcus* species, since the optimum temperature for growth of the former (85°C) is considerably lower than those of the latter (98°C–103°C) despite their similar physiological properties. Such an analysis among closely related hyperthermophiles has not been reported. We selected 917 sets of proteins commonly existing among Thermococcales with high homology (>60% identity as well as >80% overlap), aligned them by using CLUSTALW, and extracted specific amino acid replacements along with the attributed nucleotide changes. Table 2 summarizes the frequently observed replacements. The five most frequent replacements in proteins from *T. kodakaraensis* to *Pyrococcus* spp. are Val–Ile, Arg–Lys, Leu–Ile, Asp–Glu, and Glu–Lys. However, although the frequency and direction varied depending on the compared pair, these replacements are com-

monly observed among *Pyrococcus*, indicating that they do not correlate with the distinct properties of *Thermococcus* from those of *Pyrococcus*. It should be noted that a comparison between mesophilic and thermophilic *Methanococcus* species also revealed the high frequency of these replacements, including the replacement between Glu and Lys accompanied by an inverse of electric charge. These results may suggest frequent replacements of these amino acid pairs among archaeal proteins in general. We further found that Ala in *T. kodakaraensis* proteins is replaced by various residues, especially by Val, Ser, Glu, and Lys, in the *Pyrococcus* counterparts, with significant directional bias. The replace-

ments of Ala to others occupies 36.3% of total biased replacements between *T. kodakaraensis* and *Pyrococcus* spp., where the many replacements are attributed to changes in one to all bases of the GCG or GCC codons. Although the effects of these replacements on the protein stability is still unclear, the replacement of various residues to Ala might be the result of adaptation of *Thermococcus* to a lower temperature range, or vice versa. A detailed analysis of protein structure between the specific counterparts in *T. kodakaraensis* and the *Pyrococcus* spp. should provide further clues to the principles of protein thermostability.

Information processing

The information processing machinery in Archaea is similar but simplified when compared to that in Eucarya (Grabowski and Kelman 2003). Among the archaea, information processing has been studied most in *Pyrococcus* spp. Almost all of the genes for DNA replication, DNA repair and recombination, transcription,

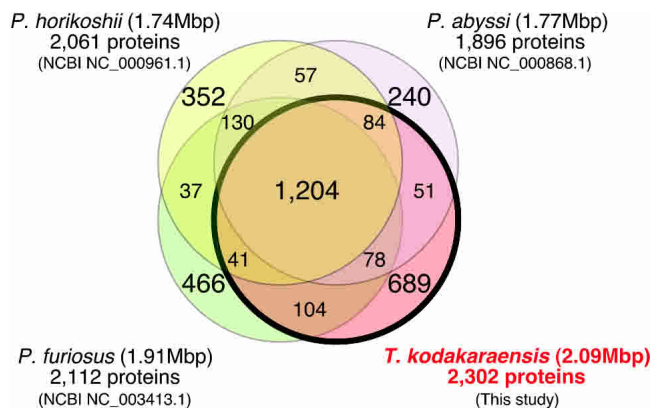


Figure 4. A Venn diagram of the shared and unique portions of the proteomes of four Thermococcales strains. The protein sets for *Pyrococcus* spp. were obtained from the RefSeq collection in NCBI (<http://www.ncbi.nlm.nih.gov/RefSeq/>) as NC_000868.1 for *P. abyssi*, NC_003413.1 for *P. furiosus*, and NC_000961.1 for *P. horikoshii*. Prior to the clustering analysis, all intein sequences were subtracted from the corresponding sequences. Thirteen proteins in *P. furiosus* (putative transposases) and four proteins in *T. kodakaraensis* (hypothetical proteins within TKV3 region) had been excluded from the protein sets because of complete duplication within the respective genomes.

Table 2. Frequent amino acid substitutions in the orthologous proteins of *T. kodakaraensis* and *Pyrococcus* spp.

			Forward	Reverse	Biased	Frequent nucleotide changes
<i>T. kodakaraensis</i> → <i>Pyrococcus</i> spp.			Total	65,053 (Average)	5664 (Average)	
Val	Ile		3745	2174	1571	GT(TC)→AT(AT)
Arg	Lys		3233	2149	1084	AG(AG)→AA(AG)
Leu	Ile		2358	1399	959	CT(TGC)→AT(AT)
Asp	Glu		1554	1382	172	GAC→GA(AG)
Glu	Lys		1494	1002	492	GAG→AA(AG)
Ala	XXX		5220	3164	2056	GC(GC)→XXX
<i>P. furiosus</i> → <i>P. abyssi</i>			Total	45,077	4401	
Lys	Arg		2726	1989	737	AA(AG)→AG(AG)
Ile	Val		2704	2283	421	AT(AT)→GT(ATG)
Ile	Leu		1435	1366	69	AT(AT)→CT(ACT)
Glu	Asp		1397	1090	307	GA(AG)→GAT
Lys	Glu		984	902	82	AA(AG)→GAG
Ala	XXX		2476	2472	4	GC(ATC)→XXX
<i>P. abyssi</i> → <i>P. horikoshii</i>			Total	33,739	3013	
Arg	Lys		2489	2040	449	AG(AG)→AA(AG)
Val	Ile		2482	1972	510	GT(AT)→AT(ATC)
Asp	Glu		1045	1028	17	GAT→GA(AG)
Leu	Ile		1031	963	68	(TC)TX→AT(AT)
Glu	Lys		829	617	212	GAG→AA(AG)
Ala	XXX		1950	1569	381	GC(ATC)→XXX
<i>P. horikoshii</i> → <i>P. furiosus</i>			Total	44,411	3117	
Ile	Val		2669	2551	118	AT(ATC)→GT(AT)
Arg	Lys		2592	2290	302	AG(AG)→AA(AG)
Ile	Leu		1336	1325	11	AT(AT)→(TC)T(AT)
Asp	Glu		1297	1059	238	GAT→GA(AG)
Lys	Glu		989	851	138	AA(AG)→GAA
Ala	XXX		2225	2610	-385	GC(ATC)→XXX

A sequence set of orthologous proteins highly conserved in all Thermococcales strains (identity $\geq 60\%$ and overlap $\geq 80\%$, total 917 sets) were aligned by CLUSTALW, and amino acid substitutions in each alignment were extracted.

and translation including RNA modification, previously identified in *Pyrococcus* genomes (Cohen et al. 2003), are also present in *T. kodakaraensis* with high similarities, indicating a common machinery among Thermococcales. The archaeal chromosomal replication origin (*oriC*) has been first identified in *P. abyssi* within the noncoding region located upstream of a gene for a homolog of eucaryal initiator Orc1/Cdc6 (Mylykallio et al. 2000). In *T. kodakaraensis*, a similar AT-rich region located upstream of a gene for the Orc1/Cdc6 homolog (TK1901) is also expected to include the *oriC* region. In contrast to *P. abyssi* and *P. horikoshii*, the "replication island" near *oriC* is separated into two regions, where genes for the Orc1/Cdc6 homolog and DNA polymerase subunits DP1 (TK1902) and DP2 (TK1903) are 290 kb away from genes for RadB (TK2231) and RF-C subunits RfcS (TK2218) and RfcL (TK2219). It is surprising that we cannot identify a distant homolog of Dna2 helicase (Higashibata et al. 2003) in *T. kodakaraensis* despite the high similarities of the other proteins.

Restriction system

Recent comparison between genomes of *P. horikoshii* and *P. abyssi* has detected a linkage between restriction-modification genes and genome rearrangements (Chinen et al. 2000). Interestingly, *T. kodakaraensis* is implied to harbor a quite distinct restriction system when compared to those in *Pyrococcus* spp. First, there are two unique fusions of endonuclease and methyltransferase (TK1158 and TK1460) identified only in *T. kodakaraensis*. Second,

in contrast to the presence of a homolog of cytosine-specific methylase Dcm (COG0270) in *Pyrococcus* spp., the corresponding gene is lacking in *T. kodakaraensis*. On the contrary, two sets of gene clusters (TK0795–0794 and TK1009–1010) are uniquely identified in *T. kodakaraensis*, which are weakly homologous to methylcytosine restriction system components McrB and McrC classified as COG1401 and COG4268, respectively. This suggests that, unlike *Pyrococcus* spp., *T. kodakaraensis* may have chromosomal DNA without cytosine methylation, and restricts cytosine-methylated DNA. Nevertheless, preliminary results suggested that the methylation of DNA added is essential for transformation of *T. kodakaraensis* cells (K. Manabe, T. Fukui, H. Atomi, and T. Imanaka, unpubl.). Although further examination is necessary, the presence of a distinct restriction system in *T. kodakaraensis* might correlate with the higher natural competency of this archaeon than that of *P. abyssi* (Lucas et al. 2002), which possesses a type I restriction system.

Sugar metabolism

T. kodakaraensis harbors a variety of sugar degradation enzymes as expected from its growth properties. α -Amylase (TK1884) (Tachibana et al. 1996), pullulanase (TK0977), and amylopullulanase (TK1774) contain putative signal sequences at the N-terminal end, suggesting that these proteins are secreted and cleave α -1,4 and α -1,6 bonds of α -linked glucans at the outside of cells. Pullulanase, originally identified and characterized in *Thermococcus aggregans* (Niehaus et al. 2000), might be a *Thermococcus*-specific enzyme as the gene is absent in *Pyrococcus* genomes. The resulting oligosaccharides can be imported into the cells by the function of the MalEFGK transporter (TK1771–1773, TK1775). It has been reported that *P. furiosus* has at least two clusters of MalEFGK transporter with different substrate specificities. One could uptake maltooligosaccharides longer than a trimer, while the other within the 16-kb composite transposon described above was specific to maltose as well as trehalose (Koning et al. 2002). The MalEFGK transporter in *T. kodakaraensis* likely corresponds to the former taking into account the higher sequence similarities. This result is consistent with the fact that maltotriose and longer oligomers, including starch, can support the sulfur-independent growth of *T. kodakaraensis*, whereas maltose cannot, unlike the case of *P. furiosus*. The imported oligosaccharides are further degraded to monosaccharides by intracellular 4- α -glucanotransferase (TK1809) (Tachibana et al. 1997) and probable isozymes (TK1436, TK1830), along with maltodextrin phosphorylase (TK1406) and possibly a probable α -glucosidase (TK1434) identified in *Thermococcus hydrothermalis* (Galichet and Belarbi 1999). Furthermore,

there is a set of extracellular cyclodextrin glucanotransferase (TK2172) (Rashid et al. 2002a) and cyclodextrinase (TK1770), suggesting the presence of an alternative starch utilization pathway via cyclic oligomers proposed in *Thermococcus* sp. B1001 (Hashimoto et al. 2001) in addition to the authentic one via linear oligomers. The genes involved in the pathway via cyclic oligomers are included in the genes gained by *T. kodakaraensis* and *P. furiosus*, but not in *P. abyssi* or *P. horikoshii*.

P. furiosus possesses proteins for degradation of cellulose, such as β -glucosidase, endo- β -1,3-glucanase, and cellobiose-uptake ABC transporters, whereas genes corresponding to these proteins are absent in *T. kodakaraensis*. This agrees with the inability of this archaeon to grow on cellobiose or longer β -linked glucans. On the other hand, we have found that *T. kodakaraensis* can assimilate chitin (poly- β -1,4-*N*-acetylglucosamine) and related oligomers through a novel chitinolytic pathway composed of extracellular chitinase with dual catalytic domains (TK1765) (Tanaka et al. 1999, 2001), *N*-acetylchitobiose deacetylase (TK1764) (Tanaka et al. 2004), and exo- β -D-glucosaminidase (TK1754) (Tanaka et al. 2003). Although the presence of chitinase gene is limited to *T. kodakaraensis* and *P. furiosus* (split to two CDSs), a gene cluster from exo- β -D-glucosaminidase to *N*-acetylchitobiose deacetylase, including an ABC transporter supposed to be for incorporation of *N*-acetylchitobiose, is highly conserved among Thermococcales with the same gene order. These facts suggest that the cluster may have a function not only in chitin utilization but may also have additional roles, such as in turnover of chitin-related sugars in the outer layer of the cells.

Central metabolism

The modified Embden-Meyerhof pathway proposed in *Pyrococcus* spp. is also present in *T. kodakaraensis* with two ADP-dependent kinases, glucokinase (TK1110) and phosphofructokinase (TK0376), glyceraldehyde-3-phosphate (GAP):ferredoxin oxidoreductase (van der Oost et al. 1998) (TK2163), and other enzymes (Fig. 2A). Although GAP dehydrogenase (phosphorylating) (TK0765) and phosphoglycerate kinase (TK1146) are responsible for the conventional two-step oxidation of GAP, these enzymes in Thermococcales have been proposed to act in the gluconeogenic direction. *T. kodakaraensis* and a limited number of other archaea (*P. furiosus*, *Sulfolobus* spp., *Aeropyrum pernix*, and *Halo-bacterium* sp. NRC-1) possess orthologs for GAP dehydrogenase (nonphosphorylating) (TK0705), although the actual substrate for this enzyme has not been clarified. The enzyme that is mainly responsible for the final step of glycolysis is still a matter to be solved (Hutchins et al. 2001; Sakuraba and Ohshima 2002). The two candidates are pyruvate kinase (TK0511), which generates pyruvate and ATP from phosphoenolpyruvate (PEP), and the reverse reaction of PEP synthase (TK1292) accompanied with ATP formation from AMP. Another key enzyme in gluconeogenesis is fructose-1,6-bisphosphatase (FBPase), which had been missing on the genomes of (hyper)thermophiles. We have recently demonstrated the power of the genetic tool for *T. kodakaraensis* by proving that a divergent FBPase (TK2164) (Rashid et al. 2002b), but not IMPase/FBPase (TK0787) (Stec et al. 2000), is the bona fide FBPase in gluconeogenesis using the gene disruption technique (Sato et al. 2004). The genomic properties for pentose-phosphate metabolism in *T. kodakaraensis*, including a unique pentagonal Rubisco (TK2290) (Ezaki et al. 1999; Maeda et al. 1999; Kitano et al. 2001), are described in the Supplemental material.

In *T. kodakaraensis*, potential members of the TCA cycle are limited to putative fumarase (TK1965–1964) and 2-oxoacid:ferredoxin oxidoreductases (KORs; TK1126–1123, and TK1131–1129). Interestingly, among Thermococcales, only *P. furiosus* possesses orthologs for C6 acids-processing enzymes (citrate synthase, aconitase, and isocitrate dehydrogenase). Most recently, an archaeal-type phosphoenolpyruvate carboxylase has been identified from *S. solfataricus* (Ettema et al. 2004), but it is interesting that *T. kodakaraensis* does not possess the ortholog that is present in *Pyrococcus* spp. The enzymes for interconversion of C3- and C4-acids in *T. kodakaraensis* are phosphoenolpyruvate carboxykinase (TK1405) (Fukuda et al. 2004) and malic enzyme (TK1963). There is no candidate for malate dehydrogenase in *T. kodakaraensis* and *P. furiosus*, despite their presence in *P. abyssi* and *P. horikoshii* clustered with a gene for malic enzyme orthologs.

Pyruvate generated via glycolysis or imported into the cells is further oxidized to form acetyl-CoA together with a reducing equivalent (reduced ferredoxin) by pyruvate:ferredoxin oxidoreductase (POR; TK1978, TK1982–1984), followed by conversion to acetate as an end product by the reverse reaction of acetyl-CoA synthetases (ACS I; TK0665–TK0465, ACS II; TK0139–TK0943) accompanied by ATP formation (Adams et al. 2001; Fig. 2B). On the contrary, pyruvate is also converted to alanine by alanine aminotransferase (TK1094) using glutamate as an amino donor, which is reductively regenerated from 2-oxoglutarate by glutamate dehydrogenase (TK1431). We have confirmed that acetate and alanine actually accumulated in the medium as major end products during cultivation of *T. kodakaraensis* on pyruvate or starch (Kanai et al. 2005). In addition to these common pathways, we found a unique gene for pyruvate-formate lyase (TK0289) that generates acetyl-CoA and formate from pyruvate. This gene is clustered with a gene for a possible specific activating enzyme (TK0290) with a role in the formation of active glycol radical in the protein of pyruvate-formate lyase. The orthologs for pyruvate-formate lyase are widely distributed in bacteria, but present only in *T. kodakaraensis*, *Archaeoglobus fulgidus*, and *Methanobacterium thermoautotrophicum* among known archaea. In bacteria, the resulting formate is generally oxidized to CO₂ by formate dehydrogenase (FDH) or that coupled with hydrogen-producing hydrogenase (formate-hydrogen lyase). Although many archaea possess proteins related to the α -subunit of *Escherichia coli* FDH-H (TK0214 in *T. kodakaraensis*), it has been reported that benzylviologen-dependent FDH activity was not detected in *P. furiosus* grown on various carbon sources (Adams et al. 2001). In contrast, we can observe high FDH activity with methylviologen in the extract of *T. kodakaraensis* (K. Obuchi, T. Kanai, H. Atomi, and T. Imanaka, unpubl.). This activity is assumed to be derived from a gene encoding an additional FDH α -subunit (TK2076) organized with genes for two 4Fe–4S cluster-binding proteins (TK2077 and TK2078) and a probable formate transporter (TK2079). The concerted action of pyruvate-formate lyase and formate dehydrogenase may play a key role in acetate production as well as in generation of reducing equivalents from pyruvate as a bypass of the pathway by POR and ACSs. This alternative pathway for pyruvate oxidation might contribute to efficient carbon and energy metabolisms responding to changes in extracellular environments.

Amino acid metabolism

As well as most species belonging to the order Thermococcales, *T. kodakaraensis* prefers proteins/peptides and amino acids as car-

bon and energy sources. The genomic properties of *T. kodakaraensis* for protein/peptide degradation and the following oxidation are described in the Supplemental material.

The distribution of amino acid biosynthesis pathways is highly diverse even among the closely related Thermococcales. For example, *P. horikoshii* apparently lacks entire sets of genes for synthesis of Arg, His, Ile, Val, Leu, Phe, Tyr, and Trp, whereas *P. abyssi* possesses Ile, Val, Leu, and Trp biosynthesis pathways (Cohen et al. 2003). However, some discrepancy has been indicated between the experimental amino acid requirement (Watrin et al. 1995) and prediction from the genome analysis of *P. abyssi* (Cohen et al. 2003). With respect to *T. kodakaraensis*, the genes for biosynthesis of Ile, Val, Leu, and Arg are entirely lacking, and *pheA* encoding prephenate dehydratase for Phe biosynthesis is absent. Most enzymes necessary for Lys biosynthesis via α -amino adipic acid, the prokaryotic AAA pathway recently identified in the thermophilic bacterium *Thermus thermophilus* (Kobashi et al. 1999), are highly conserved in Thermococcales with the same gene order. *T. kodakaraensis* and *P. furiosus* harbor a His biosynthesis pathway, of which corresponding genes (*his* operon) are clustered with *trp* and *aro* operons for aromatic amino acid biosynthesis. Our recent investigation demonstrated that *T. kodakaraensis* requires at least Ile, Val, Leu, and Arg for growth (data not shown), which are consistent with the lack of biosynthesis pathways deduced from the genome analysis.

Energy metabolism

The simple respiration system composed of membrane-bound hydrogenase and A_0A_1 -type ATP synthase (Fig. 2E), recently proposed in *P. furiosus* (Sapra et al. 2003), appears to be functional in all Thermococcales, since the genes corresponding to two membrane-bound hydrogenases (Mbh, TK2080–2093; and Mbx, TK1226–1214) and A_0A_1 -ATPase (TK1596–1604) are highly conserved with the same gene order. In the presence of elemental sulfur or polysulfide, *Thermococcus* and *Pyrococcus* prefer to use it as a terminal electron acceptor, forming H_2S during growth. A potential candidate responsible for the sulfur reduction is an NADPH-dependent cytoplasmic NiFe-hydrogenase (Ma et al. 2000) coupled with ferredoxin:NADP oxidoreductases (Ma and Adams 2001) (Fig. 2D). While *P. furiosus* and *P. abyssi* possess two NiFe-hydrogenases, only one enzyme corresponding to hydrogenase I can be identified in *T. kodakaraensis* (TK2072–2069) (Kanai et al. 2003) and *P. horikoshii*. Two ferredoxin:NADP oxidoreductases, of which two subunits are clustered, respectively (TK1326–1325 and TK1685–1684), are also shared by Thermococcales with high identities. Unlike in *Pyrococcus*, a unique gene (TK1481) is identified in *T. kodakaraensis* that shows significant homology to NADH:polysulfide oxidoreductase from *Thermotoga neapolitana* in the latest database, although the characteristics of this bacterial enzyme have not yet been described. It should be noted that the primary structure of TK1481 is related to those of H_2O -forming NADH oxidases. The related NADH oxidases, including paralogs in *T. kodakaraensis* (TK0304, TK1299, and TK1392), might also function in reduction of sulfur rather than oxygen.

Nucleotide metabolism

T. kodakaraensis and *Pyrococcus* spp. harbor almost complete pathways for purine and pyrimidine biosynthesis, whereas a few notable differences are described in the Supplemental material. One of the interesting gains of genes for nucleotide metabolism

in *T. kodakaraensis* is *deoC* for deoxyribose-5-phosphate aldolase (TK2104), *deoB* for a novel phosphopentomutase (TK1777) (Rashid et al. 2004), and *udp* for uridine phosphorylase (TK1479). Along with thymidine phosphorylase DeoA (TK0352) and probable purine nucleoside phosphorylases (TK1482 and TK1895), these enzymes are expected to complete a metabolic link between nucleoside metabolism and central metabolism to recycle the pentose moiety derived from nucleotides. This linkage may also be functional in purine and pyrimidine salvage, or biosynthesis of deoxyribose-1-phosphate. Unlike many archaea including *Pyrococcus*, *T. kodakaraensis* further possesses a clear-cut homolog of thymidine kinase (TK1318) for providing thymidylate via salvage of thymidine, in addition to de novo synthesis of thymidylate catalyzed by a newly identified class of thymidylate synthase, ThyX (TK0870) (Myllykallio et al. 2002). The gains of these genes may allow efficient utilization of nucleotides in *T. kodakaraensis*.

Coenzyme metabolism

T. kodakaraensis and *Pyrococcus* spp. contain a nearly common gene set for biosynthesis of coenzyme A, pyridine nucleotide, thiamin, pyridoxine, and molybdopterin. With regard to thiamine biosynthesis, proteins involved in sulfur activation and transfer (ThiS and ThiI) are shared by Thermococcales (TK1093 and TK0368, respectively, in *T. kodakaraensis*), although they commonly lack 1-deoxy-D-xylulose 5-phosphate (DXP) synthase, ThiG, and ThiH for thiazole ring formation. The pathway for DXP-independent thiazole ring formation in archaea still remains to be clarified. *T. kodakaraensis* further lacks genes for hydroxyethylthiazole kinase ThiM and thiamine-phosphate synthase ThiE that are clustered in *Pyrococcus* genomes. Especially, ThiE is essential for linking the pyrimidine moiety with the thiazole ring. This fact might suggest the presence of an alternative enzyme responsible for this step, or the inability of *T. kodakaraensis* to generate active thiamine cofactor.

Thermococcales also share the archaeal salvage pathway for cobinamide, a coenzyme B12 precursor, recently revealed in *Halobacterium* sp. NRC-1 (Woodson et al. 2003; Woodson and Escalante-Semerena 2004). *T. kodakaraensis* and *P. furiosus* further encode *cobQ* for cobyrinic acid synthase (TK0854), suggesting the ability to salvage cob(I)rinic acid, a,c-diamide in addition to cobinamide as B12 precursors. They also contain complete pathways for riboflavin biosynthesis, unlike *P. abyssi* and *P. horikoshii*. However, an obvious ortholog for FAD synthetase, catalyzing the formation of FMN and FAD from riboflavin, is missing in not only these Thermococcales but also in all known archaeal genomes. Other uncharacterized kinases and nucleotidyltransferases may function in FMN and FAD generation.

Transporters

While several transporters are shared among Thermococcales, the clustering analysis identified unique or additional genes for various metal cation transporters in *T. kodakaraensis*, such as ABC-type iron(III) transporter (TK0570–0572), ABC-type iron(III)-siderophore transporter (TK0706–0708), ABC-type molybdate transporter (TK0717–0719), and iron(II) transporter (TK0716–0714). The genes for the last three transporters are highly clustered, suggesting a recent gain via horizontal gene transfer. Moreover, *T. kodakaraensis* encodes additional copies of genes for monovalent cation transporters, sodium/phosphate symporter (TK0761), and NapA-type sodium/hydrogen anti-

porter (TK0627), but lacks sodium/alanine symporter. Calcium-gated potassium channel protein, whose crystal structure and transport mechanism has recently been determined on an ortholog from *M. thermoautotrophicum* (Jiang et al. 2002), is only present in *T. kodakaraensis* (TK1900) among Thermococcales. The presence of these additional transporters is expected to contribute to efficient cation uptake required to support cell survival and growth even under natural metal-limited environments.

T. kodakaraensis further harbors two components for a unique TRAP-type transporter (TK0935 and TK0937), although the transport specificity is unknown. All known microbial genomes contain many orthologs belonging to COG1131 annotated as an ATPase component of an ABC-type multidrug transporter. Among Thermococcales, *T. kodakaraensis* and *P. furiosus* possess more orthologs of COG1131 (16–17 orthologs) than *P. abyssi* and *P. horikoshii* (seven orthologs). On the other hand, only one ABC-type sugar transporter (predicted to be malt-dextrin-specific, as described above) is identified in *T. kodakaraensis*, in contrast to three to four transporters in *Pyrococcus* spp.

Stress response

In hyperthermophilic archaea, it has been proposed that the small heat-shock protein, prefoldin, and FK506-binding protein (FKBP) function as chaperones in repressing aggregation of unfolded proteins, followed by refolding with group II chaperonin (Hsp60 homolog). *Pyrococcus* genomes share Hsp20 (small heat-shock protein), α - and β -subunits of prefoldin, chaperonin with a single subunit, and long-type FKBP. Interestingly, *T. kodakaraensis* encodes another copy of genes for prefoldin (TK1121–1122 for α - and β -subunits, respectively), chaperonin α -subunit (TK0678) (Izumi et al. 1999), and short-type FKBP (TK1850) in addition to genes corresponding to the *Pyrococcus* chaperones (TK1005 and TK0643 for prefoldin α - and β -subunits, respectively; TK2303 for chaperonin β -subunit; TK0441 for long-type FKBP fused to uncharacterized C-terminal region). It seems that this archaeon has gained an improved chaperone system that may provide an advantage to grow within a wider temperature range. We have previously demonstrated that the unique α -subunit of chaperonin TK0678 (Cpka) exists as the major subunit in the functional complex at low cultivation temperature (70°C), whereas β -subunit TK2303 (Cpkb) was induced and increased in the complex at elevated temperatures (Izumi et al. 2001). Likewise, a short-type FKBP from *M. jannaschii* and from a closely related *Thermococcus* sp. KS-1 were induced at temperatures lower than the growth optimum (Ideno et al. 2001; Ideno and Maruyama 2002). These facts raise a possibility that the additional copies of prefoldin subunits might also act at a lower growth temperature. When low temperatures cause unfavorable folding and the resulting loss of catalytic abilities of extremely thermostable proteins from hyperthermophiles, these improved chaperone proteins might allow maintenance of protein functions at lower temperature conditions (namely, “cold” adaptation).

Against oxygen stress, anaerobic microorganisms can detoxify the superoxide anion by superoxide reductase Sor in place of superoxide dismutase in aerobic organisms (Jenney Jr. et al. 1999). The organization of *sor* (TK0525) with genes for rubredoxin (TK0524) and ruberythrin (TK0523) is highly conserved among Thermococcales. Multiple NADH oxidases and peroxidases are also assumed to act on the oxygen detoxification. Protein disulfide oxidoreductase (TK1085) can repair the oxidized disulfide bond in a polypeptide in cooperation with thiore-

doxin reductase (TK2100), although the exact functions remain unclear. Thermococcales conserve these two redox proteins with very high homology (>80% identities). *T. kodakaraensis* further possesses a distinct repair enzyme, protein-methionine sulfoxide reductase, Msr (TK0819). There are two kinds of Msr, (*S*)-specific MsrA and (*R*)-specific MsrB, depending on the recognition specificity toward the asymmetric sulfur in the methionine sulfoxide. A pair of MsrA and MsrB are widely distributed in bacteria and eucarya, while the archaeal distribution is rather limited; the orthologs are identified in thermophilic *M. thermoautotrophicum* and mesophilic *Methanosarcina acetivorans*, *Methanosarcina mazei*, and *Halobacterium* sp. NRC-1. *S. solfataricus* encodes only the *msrA* gene. Since TK0819 is a fusion protein of MsrA and MsrB as seen in some bacteria, the expected ability of *T. kodakaraensis* to reduce both epimeric forms of methionine sulfoxide is a unique property among hyperthermophiles.

Conclusions

The complete genome analysis of *T. kodakaraensis* provides us with a wealth of new information about differences of *Thermococcus* from closely related *Pyrococcus*, both well-studied genera among hyperthermophilic archaea. They share common information processing machinery, but *T. kodakaraensis* has gained several cellular functions in comparison with *Pyrococcus* species, as summarized in Table 3. The improved degradation system for polysaccharides, additional pyruvate oxidation pathway, and other additional metabolic pathways are expected to allow efficient utilization and recycling of carbons. Another considerable gain is the more abundant metal cation transporters, which may be an advantage to survive and grow under natural environments. Additional proteins in the chaperone system and a unique enzyme for repair of oxidized proteins will serve to maintain the protein functions in broader growth environments, such as relatively low temperature and loose anaerobic conditions. Although it is not clear at present whether these functional gains are common properties among *Thermococcus* strains in general, there is a possibility that such gained functions lead to the more ubiquitous distribution and larger population of *Thermococcus*

Table 3. Gains and losses of cellular functions in *T. kodakaraensis* when compared to *Pyrococcus* spp.

Gains	
Improved α -polysaccharide degradation ^a	
Chitin degradation ^a	
Additional pyruvate oxidation	
Tyr biosynthesis	
His biosynthesis ^a	
Riboflavin biosynthesis ^a	
Deoxyribose metabolism	
Thymidine salvage	
Additional cation transporters (Fe ²⁺ , Fe ³⁺ , Mo ²⁺ , K ⁺ , Na ⁺)	
Improved stress response system (especially for cold adaptation?)	
Losses	
Ile, Val, and Leu biosynthesis ^b	
Thiamine biosynthesis (partially)	
Others	
Distinct restriction system	

^aPresent in *T. kodakaraensis* and *P. furiosus*, and not in *P. abyssi* and *P. horikoshii*.

^bPresent in *P. abyssi* and *P. furiosus*, and not in *T. kodakaraensis* and *P. horikoshii*.

strains in natural environments when compared to *Pyrococcus* strains.

Methods

Sequencing and assembly of the whole-genome shotgun library

The sequencing of the *T. kodakaraensis* genome was performed by Dragon Genomics (Mie, Japan) using whole-genome random shotgun methods with sheared chromosomal DNA. Briefly, pUC18-based libraries containing 2-kb and 5-kb inserts were constructed, and both strands of 15,552 and 4800 clones from the 2-kb and 5-kb libraries, respectively, were sequenced (7.8-fold coverage) with a MegaBACE 1000 sequencer (Amersham Biosciences). The sequence was assembled by using Paracel Clustering Package with CAP4 algorithm (Paracel), resulting in 51 major contigs on 16 major scaffolds. The obtained sequences were estimated to have an error rate of 1 per 10,000 bases with an average Phrap score of ~40. Gaps in the sequence were filled by direct PCR sequencing with primers designed to anneal to each end of neighboring contigs.

Gene prediction and annotation

The coding potential analysis of the entire genome was performed by a combination of the Critica program (Badger and Olsen 1999) with a set of CDSs from three *Pyrococcus* spp. and the GeneHacker Plus program (Yada et al. 2001) using a hidden Markov model initially trained with CDSs predicted by Critica and gradually optimized with its own predicted CDSs. The unique set of deduced protein sequences was subjected to BLASTp searches (Altschul et al. 1997) against the nonredundant (nr) protein database from NCBI and the SWISS-PROT protein database. Additional genes were identified through detection of open reading frames longer than 90 bp within the intergenic regions, followed by similarity and motif searches of the deduced amino acid sequences. The protein sequences were further analyzed by InterPro (Mulder et al. 2003), and searched against the COGs (Clusters of Orthologous Groups of proteins) database (Tatusov et al. 2003). Transmembrane regions were predicted by the SOSUI program (Mitaku et al. 1999). tRNA-encoding regions were predicted by use of the tRNA scan-SE (Lowe and Eddy 1997). These analyses were assisted by Genome Gambler software (Sakiyama et al. 2000) and the XanaGenome microbial genome database from Xanagen (Tokyo, Japan). Intein regions were identified by searching the InBase database (Perler 2002; <http://www.neb.com/inteins/html>), and putative transposase genes were analyzed by IS Finder (<http://www-is.biotoul.fr/is.html>). A search for gene families in the genome was performed by BLASTCLUST.

Comparative analyses

For clustering of proteins in Thermococcales, three *Pyrococcus* genomes from the NCBI RefSeq collection (<http://www.ncbi.nlm.nih.gov/RefSeq/>) at January 2004 were used to avoid discrepancies of CDS prediction among the genomes. In addition, the intein sequences and perfectly duplicated proteins were subtracted and excluded from a set of protein sequences from *T. kodakaraensis* and the three *Pyrococcus* spp., respectively, prior to the analysis. An all-vs-all BLASTp search was performed with the protein set to identify pairwise matches above $E \leq 10^{-10}$ over $\geq 80\%$ coverage. Subsequently, single-linkage clustering of the matches was carried out. In order to remove paralogs in each cluster, complete-linkage clustering was further adopted against clusters containing multiple entries from one genome.

Data deposition

The whole genome of *T. kodakaraensis* KOD1 has been deposited in the DDBJ/EMBL/GenBank databases under the project accession no. AP006878.

Acknowledgments

We are grateful to T. Kozuki and K. Doga of Xanagen Inc. for clustering analysis of proteins and computing assistance. This work was supported by a Core Research for Evolutional Science and Technology from the Japan Science and Technology Corporation (CREST-JST) to T.I. and by a Grant-in-Aid for Scientific Research (S) from the Ministry of Education, Science, Sports and Culture (no. 14103011) to T.I.

References

- Adams, M.W., Holden, J.F., Menon, A.L., Schut, G.J., Grunden, A.M., Hou, C., Hutchins, A.M., Jenney Jr., F.E., Kim, C., Ma, K., et al. 2001. Key role for sulfur in peptide metabolism and in regulation of three hydrogenases in the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Bacteriol.* **183**: 716–724.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.
- Amend, J.P. and Shock, E.L. 2001. Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and Bacteria. *FEMS Microbiol. Rev.* **25**: 175–243.
- Atomi, H., Fukui, T., Kanai, T., Morikawa, M., and Imanaka, T. 2004. Description of *Thermococcus kodakaraensis* sp. nov., a well studied hyperthermophilic archaeon previously reported as *Pyrococcus* sp. KOD1. *Archaea* **1**: 263–267.
- Badger, J.H. and Olsen, G.J. 1999. CRITICA: Coding region identification tool invoking comparative analysis. *Mol. Biol. Evol.* **16**: 512–524.
- Chakravarty, S. and Varadarajan, R. 2000. Elucidation of determinants of protein stability through genome sequence analysis. *FEBS Lett.* **470**: 65–69.
- Chinen, A., Uchiyama, I., and Kobayashi, I. 2000. Comparison between *Pyrococcus horikoshii* and *Pyrococcus abyssi* genome sequences reveals linkage of restriction-modification genes with large genome polymorphisms. *Gene* **259**: 109–121.
- Cohen, G.N., Barbe, V., Flament, D., Galperin, M., Heilig, R., Lecompte, O., Poch, O., Prieur, D., Querellou, J., Ripp, R., et al. 2003. An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*. *Mol. Microbiol.* **47**: 1495–1512.
- Diruggiero, J., Dunn, D., Maeder, D.L., Holley-Shanks, R., Chatard, J., Horlacher, R., Robb, F.T., Boos, W., and Weiss, R.B. 2000. Evidence of recent lateral gene transfer among hyperthermophilic archaea. *Mol. Microbiol.* **38**: 684–693.
- Ettema, T.J., Makarova, K.S., Jellema, G.L., Gierman, H.J., Koonin, E.V., Huynen, M.A., de Vos, W.M., and van der Oost, J. 2004. Identification and functional verification of archaeal-type phosphoenolpyruvate carboxylase, a missing link in archaeal central carbohydrate metabolism. *J. Bacteriol.* **186**: 7754–7762.
- Ezaki, S., Maeda, N., Kishimoto, T., Atomi, H., and Imanaka, T. 1999. Presence of a structurally novel type ribulose-bisphosphate carboxylase/oxygenase in the hyperthermophilic archaeon, *Pyrococcus kodakaraensis* KOD1. *J. Biol. Chem.* **274**: 5078–5082.
- Fukuda, W., Fukui, T., Atomi, H., and Imanaka, T. 2004. First characterization of an archaeal GTP-dependent phosphoenolpyruvate carboxykinase from the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J. Bacteriol.* **186**: 4620–4627.
- Galichet, A. and Belarbi, A. 1999. Cloning of an α -glucosidase gene from *Thermococcus hydrothermalis* by functional complementation of a *Saccharomyces cerevisiae* mal11 mutant strain. *FEBS Lett.* **458**: 188–192.
- Grabowski, B. and Kelman, Z. 2003. Archeal DNA replication: Eukaryal proteins in a bacterial context. *Annu. Rev. Microbiol.* **57**: 487–516.
- Haney, P.J., Badger, J.H., Buldak, G.L., Reich, C.I., Woese, C.R., and Olsen, G.J. 1999. Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. *Proc. Natl. Acad. Sci.* **96**: 3578–3583.

- Hashimoto, Y., Yamamoto, T., Fujiwara, S., Takagi, M., and Imanaka, T. 2001. Extracellular synthesis, specific recognition, and intracellular degradation of cyclomaltoextrins by the hyperthermophilic archaeon *Thermococcus* sp. strain B1001. *J. Bacteriol.* **183**: 5050–5057.
- Higashibata, H., Kikuchi, H., Kawarabayasi, Y., and Matsui, I. 2003. Helicase and nuclease activities of hyperthermophile *Pyrococcus horikoshii* Dna2 inhibited by substrates with RNA segments at 5'-end. *J. Biol. Chem.* **278**: 15983–15990.
- Holden, J.F., Takai, K., Summit, M., Bolton, S., Zyskowski, J., and Baross, J.A. 2001. Diversity among three novel groups of hyperthermophilic deep-sea *Thermococcus* species from three sites in the northeastern Pacific Ocean. *FEMS Microbiol. Ecol.* **36**: 51–60.
- Huber, R. and Stetter, K.O. 2001. Discovery of hyperthermophilic microorganisms. *Methods Enzymol.* **330**: 11–24.
- Hutchins, A.M., Holden, J.F., and Adams, M.W. 2001. Phosphoenolpyruvate synthetase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Bacteriol.* **183**: 709–715.
- Ideno, A. and Maruyama, T. 2002. Expression of long- and short-type FK506 binding proteins in hyperthermophilic archaea. *Gene* **292**: 57–63.
- Ideno, A., Yoshida, T., Iida, T., Furutani, M., and Maruyama, T. 2001. FK506-binding protein of the hyperthermophilic archaeum, *Thermococcus* sp. KS-1, a cold-shock-inducible peptidyl-prolyl *cis-trans* isomerase with activities to trap and refold denatured proteins. *Biochem. J.* **357**: 465–471.
- Imanaka, T. and Atomi, H. 2002. Catalyzing “hot” reactions: Enzymes from hyperthermophilic Archaea. *Chem. Rec.* **2**: 149–163.
- Itoh, T. 2003. Taxonomy of nonmethanogenic hyperthermophilic and related thermophilic archaea. *J. Biosci. Bioeng.* **96**: 203–212.
- Izumi, M., Fujiwara, S., Takagi, M., Kanaya, S., and Imanaka, T. 1999. Isolation and characterization of a second subunit of molecular chaperonin from *Pyrococcus kodakaraensis* KOD1: Analysis of an ATPase-deficient mutant enzyme. *Appl. Environ. Microbiol.* **65**: 1801–1805.
- Izumi, M., Fujiwara, S., Takagi, M., Fukui, K., and Imanaka, T. 2001. Two kinds of archaeal chaperonin with different temperature dependency from a hyperthermophile. *Biochem. Biophys. Res. Commun.* **280**: 581–587.
- Jenney Jr., F.E., Verhagen, M.F., Cui, X., and Adams, M.W. 1999. Anaerobic microbes: Oxygen detoxification without superoxide dismutase. *Science* **286**: 306–309.
- Jiang, Y., Lee, A., Chen, J., Cadene, M., Chait, B.T., and MacKinnon, R. 2002. Crystal structure and mechanism of a calcium-gated potassium channel. *Nature* **417**: 515–522.
- Kanai, T., Ito, S., and Imanaka, T. 2003. Characterization of a cytosolic NiFe-hydrogenase from the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J. Bacteriol.* **185**: 1705–1711.
- Kanai, T., Imanaka, H., Nakajima, A., Uwamori, K., Omori, Y., Fukui, T., Atomi, H., and Imanaka, T. 2005. Continuous hydrogen production by the hyperthermophilic archaeon, *Thermococcus kodakaraensis* KOD1. *J. Biotechnol.* (in press).
- Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., et al. 1998. Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3. *DNA Res.* **5**: 55–76.
- Kitano, K., Maeda, N., Fukui, T., Atomi, H., Imanaka, T., and Miki, K. 2001. Crystal structure of a novel-type archaeal rubisco with pentagonal symmetry. *Structure (Camb)* **9**: 473–481.
- Kobashi, N., Nishiyama, M., and Tanokura, M. 1999. Aspartate kinase-independent lysine synthesis in an extremely thermophilic bacterium, *Thermus thermophilus*: Lysine is synthesized via α -aminoadipic acid not via diaminopimelic acid. *J. Bacteriol.* **181**: 1713–1718.
- Koning, S.M., Konings, W.N., and Driessen, A.J.M. 2002. Biochemical evidence for the presence of two α -glucoside ABC-transport systems in the hyperthermophilic archaeon *Pyrococcus furiosus*. *Archaea* **1**: 19–25.
- Lecompte, O., Ripp, R., Puzos-Barbe, V., Duprat, S., Heilig, R., Dietrich, J., Thierry, J.C., and Poch, O. 2001. Genome evolution at the genus level: Comparison of three complete genomes of hyperthermophilic archaea. *Genome Res.* **11**: 981–993.
- Lowe, T.M. and Eddy, S.R. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**: 955–964.
- Lucas, S., Toffin, L., Zivanovic, Y., Charlier, D., Moussard, H., Forterre, P., Prieur, D., and Erauso, G. 2002. Construction of a shuttle vector for, and spheroplast transformation of, the hyperthermophilic archaeon *Pyrococcus abyssi*. *Appl. Environ. Microbiol.* **68**: 5528–5536.
- Ma, K. and Adams, M.W. 2001. Ferredoxin:NADP oxidoreductase from *Pyrococcus furiosus*. *Methods Enzymol.* **334**: 40–45.
- Ma, K., Weiss, R., and Adams, M.W. 2000. Characterization of hydrogenase II from the hyperthermophilic archaeon *Pyrococcus furiosus* and assessment of its role in sulfur reduction. *J. Bacteriol.* **182**: 1864–1871.
- Maeda, N., Kitano, K., Fukui, T., Ezaki, S., Atomi, H., Miki, K., and Imanaka, T. 1999. Ribulose biphosphate carboxylase/oxygenase from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1 is composed solely of large subunits and forms a pentagonal structure. *J. Mol. Biol.* **293**: 57–66.
- Mahillon, J. and Chandler, M. 1998. Insertion sequences. *Microbiol. Mol. Biol. Rev.* **62**: 725–774.
- Makino, S., Amano, N., Koike, H., and Suzuki, M. 1999. Prophages inserted in archaeobacterial genomes. *Proc. Japan Acad. Ser. B* **75**: 166–171.
- Malandrin, L., Huber, H., and Bernander, R. 1999. Nucleoid structure and partition in *Methanococcus jannaschii*: An archaeon with multiple copies of the chromosome. *Genetics* **152**: 1315–1323.
- Martusewitsch, E., Sensen, C.W., and Schleper, C. 2000. High spontaneous mutation rate in the hyperthermophilic archaeon *Sulfolobus solfataricus* is mediated by transposable elements. *J. Bacteriol.* **182**: 2574–2581.
- McDonald, J.H., Grasso, A.M., and Rejto, L.K. 1999. Patterns of temperature adaptation in proteins from *Methanococcus* and *Bacillus*. *Mol. Biol. Evol.* **16**: 1785–1790.
- Mira, A., Klason, L., and Andersson, S.G. 2002. Microbial genome evolution: Sources of variability. *Curr. Opin. Microbiol.* **5**: 506–512.
- Mitaku, S., Ono, M., Hirokawa, T., Boon-Chieng, S., and Sonoyama, M. 1999. Proportion of membrane proteins in proteomes of 15 single-cell organisms analyzed by the SOSUI prediction system. *Biophys. Chem.* **82**: 165–171.
- Morikawa, M., Izawa, Y., Rashid, N., Hoaki, T., and Imanaka, T. 1994. Purification and characterization of a thermostable thiol protease from a newly isolated hyperthermophilic *Pyrococcus* sp. *Appl. Environ. Microbiol.* **60**: 4559–4566.
- Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Barrell, D., Bateman, A., Binns, D., Biswas, M., Bradley, P., Bork, P., et al. 2003. The InterPro Database, 2003 brings increased coverage and new features. *Nucleic Acids Res.* **31**: 315–318.
- Myllykallio, H., Lopez, P., Lopez-Garcia, P., Heilig, R., Saurin, W., Zivanovic, Y., Philippe, H., and Forterre, P. 2000. Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science* **288**: 2212–2215.
- Myllykallio, H., Lipowski, G., Leduc, D., Filee, J., Forterre, P., and Liebl, U. 2002. An alternative flavin-dependent mechanism for thymidylate synthesis. *Science* **297**: 105–107.
- Niehaus, F., Peters, A., Groudieva, T., and Antranikian, G. 2000. Cloning, expression and biochemical characterization of a unique thermostable pullulan-hydrolysing enzyme from the hyperthermophilic archaeon *Thermococcus aggregans*. *FEMS Microbiol. Lett.* **190**: 223–229.
- Nishio, Y., Nakamura, Y., Kawarabayasi, Y., Usuda, Y., Kimura, E., Sugimoto, S., Matsui, K., Yamagishi, A., Kikuchi, H., Ikeo, K., et al. 2003. Comparative complete genome sequence analysis of the amino acid replacements responsible for the thermostability of *Corynebacterium efficiens*. *Genome Res.* **13**: 1572–1579.
- Nishioka, M., Mizuguchi, H., Fujiwara, S., Komatsubara, S., Kitabayashi, M., Uemura, H., Takagi, M., and Imanaka, T. 2001. Long and accurate PCR with a mixture of KOD DNA polymerase and its exonuclease deficient mutant enzyme. *J. Biotechnol.* **88**: 141–149.
- Ochman, H., Lawrence, J.G., and Groisman, E.A. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299–304.
- Orphan, V.J., Taylor, L.T., Hafenbradl, D., and Delong, E.F. 2000. Culture-dependent and culture-independent characterization of microbial assemblages associated with high-temperature petroleum reservoirs. *Appl. Environ. Microbiol.* **66**: 700–711.
- Peng, X., Holz, I., Zillig, W., Garrett, R.A., and She, Q. 2000. Evolution of the family of pRN plasmids and their integrase-mediated insertion into the chromosome of the crenarchaeon *Sulfolobus solfataricus*. *J. Mol. Biol.* **303**: 449–454.
- Perler, F.B. 2002. InBase: The Intein database. *Nucleic Acids Res.* **30**: 383–384.
- Rashid, N., Cornista, J., Ezaki, S., Fukui, T., Atomi, H., and Imanaka, T. 2002a. Characterization of an archaeal cyclodextrin glucanotransferase with a novel C-terminal domain. *J. Bacteriol.* **184**: 777–784.
- Rashid, N., Imanaka, H., Kanai, T., Fukui, T., Atomi, H., and Imanaka, T. 2002b. A novel candidate for the true fructose-1,6-bisphosphatase in archaea. *J. Biol. Chem.* **277**: 30649–30655.
- Rashid, N., Imanaka, H., Fukui, T., Atomi, H., and Imanaka, T. 2004. Presence of a novel phosphopentomutase and a 2-deoxyribose

- 5-phosphate aldolase reveals a metabolic link between pentoses and central carbon metabolism in the hyperthermophilic archaeon *Thermococcus kodakaraensis*. *J. Bacteriol.* **186**: 4185–4191.
- 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. Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: Implications for physiology and enzymology. *Methods Enzymol.* **330**: 134–157.
- Sakiyama, T., Takami, H., Ogasawara, N., Kuhara, S., Kozuki, T., Doga, K., Ohyama, A., and Horikoshi, K. 2000. AIN automated system for genome analysis to support microbial whole-genome shotgun sequencing. *Biosci. Biotechnol. Biochem.* **64**: 670–673.
- Sakuraba, H. and Ohshima, T. 2002. Novel energy metabolism in anaerobic hyperthermophilic archaea: A modified Embden-Meyerhof pathway. *J. Biosci. Bioeng.* **93**: 441–448.
- Sapra, R., Bagramyan, K., and Adams, M.W. 2003. A simple energy-conserving system: Proton reduction coupled to proton translocation. *Proc. Natl. Acad. Sci.* **100**: 7545–7550.
- Sato, T., Fukui, T., Atomi, H., and Imanaka, T. 2003. Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J. Bacteriol.* **185**: 210–220.
- Sato, T., Imanaka, H., Rashid, N., Fukui, T., Atomi, H., and Imanaka, T. 2004. Genetic evidence identifying the true gluconeogenic fructose-1,6-bisphosphatase in *Thermococcus kodakaraensis* and other hyperthermophiles. *J. Bacteriol.* **186**: 5799–5807.
- She, Q., Peng, X., Zillig, W., and Garrett, R.A. 2001. Gene capture in archaeal chromosomes. *Nature* **409**: 478.
- Stec, B., Yang, H., Johnson, K.A., Chen, L., and Roberts, M.F. 2000. MJ0109 is an enzyme that is both an inositol monophosphatase and the 'missing' archaeal fructose-1,6-bisphosphatase. *Nat. Struct. Biol.* **7**: 1046–1050.
- Tachibana, Y., Leclere, M.M., Fujiwara, S., Takagi, M., and Imanaka, T. 1996. Cloning and expression of the α -amylase gene from the hyperthermophilic archaeon *Pyrococcus* sp. KOD1, and characterization of the enzyme. *J. Ferment. Bioeng.* **82**: 224–232.
- Tachibana, Y., Fujiwara, S., Takagi, M., and Imanaka, T. 1997. Cloning and expression of the 4- α -glucanotransferase gene from the hyperthermophilic archaeon *Pyrococcus* sp. KOD1, and characterization of the enzyme. *J. Ferment. Bioeng.* **83**: 540–548.
- Tanaka, T., Fujiwara, S., Nishikori, S., Fukui, T., Takagi, M., and Imanaka, T. 1999. A unique chitinase with dual active sites and triple substrate binding sites from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. *Appl. Environ. Microbiol.* **65**: 5338–5344.
- Tanaka, T., Fukui, T., and Imanaka, T. 2001. Different cleavage specificities of the dual catalytic domains in chitinase from the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J. Biol. Chem.* **276**: 35629–35635.
- Tanaka, T., Fukui, T., Atomi, H., and Imanaka, T. 2003. Characterization of an exo- β -D-glucosaminidase involved in a novel chitinolytic pathway from the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J. Bacteriol.* **185**: 5175–5181.
- Tanaka, T., Fukui, T., Fujiwara, S., Atomi, H., and Imanaka, T. 2004. Concerted action of diacetylchitobiose deacetylase and exo- β -D-glucosaminidase in a novel chitinolytic pathway in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J. Biol. Chem.* **279**: 30021–30027.
- Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., et al. 2003. The COG database: An updated version includes eukaryotes. *BMC Bioinformatics* **4**: 41.
- van der Oost, J., Schut, G., Kengen, S.W., Hagen, W.R., Thomm, M., and de Vos, W.M. 1998. The ferredoxin-dependent conversion of glyceraldehyde-3-phosphate in the hyperthermophilic archaeon *Pyrococcus furiosus* represents a novel site of glycolytic regulation. *J. Biol. Chem.* **273**: 28149–28154.
- Watrin, L., Martinjezequel, V., and Prieur, D. 1995. Minimal amino-acid-requirements of the hyperthermophilic archaeon *Pyrococcus abyssi*, isolated from deep-sea hydrothermal vents. *Appl. Environ. Microbiol.* **61**: 1138–1140.
- Woodson, J.D. and Escalante-Semerena, J.C. 2004. CbiZ, an amidohydrolase enzyme required for salvaging the coenzyme B12 precursor cobinamide in archaea. *Proc. Natl. Acad. Sci.* **101**: 3591–3596.
- Woodson, J.D., Zayas, C.L., and Escalante-Semerena, J.C. 2003. A new pathway for salvaging the coenzyme B12 precursor cobinamide in archaea requires cobinamide-phosphate synthase (CbiB) enzyme activity. *J. Bacteriol.* **185**: 7193–7201.
- Yada, T., Totoki, Y., Takagi, T., and Nakai, K. 2001. A novel bacterial gene-finding system with improved accuracy in locating start codons. *DNA Res.* **8**: 97–106.
- Zivanovic, Y., Lopez, P., Philippe, H., and Forterre, P. 2002. *Pyrococcus* genome comparison evidences chromosome shuffling-driven evolution. *Nucleic Acids Res.* **30**: 1902–1910.

Web site references

- <http://www-is.biotoul.fr/is.html>; IS Finder.
<http://www.ncbi.nlm.nih.gov/RefSeq/>; NCBI RefSeq collection.

Received July 15, 2004; accepted in revised form December 21, 2004.