Molecular Microbiology (2003) **47**(6), 1495–1512

An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*

Georges N. Cohen,[1] Valérie Barbe,[2] Didier Flament,[3] Michael Galperin,[4] Roland Heilig,[2] Odile Lecompte,[5] Olivier Poch,[5] Daniel Prieur,[6] Joël Quérellou,[3] Raymond Ripp,[5] Jean-Claude Thierry,[5] John Van der Oost,[7] Jean Weissenbach,[2] Yvan Zivanovic[8] and Patrick Forterre[8]*

[1]*Institut Pasteur, 25,28 rue du Docteur Roux, 75724 Paris CEDEX 15, France.*
[2]*Genoscope, CNS, 2, rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, France.*
[3]*Ifremer, Centre de Brest, DRV/VP/LMBE, BP 70, 29280 Plouzané, France.*
[4]*National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD 20894, USA.*
[5]*CNRS-IGBMC, BP 163, 1 rue Laurent Fries, 67404 Illkirch, France.*
[6]*Université de Bretagne Occidentale, IUEM, Place Nicolas Copernic, TBI, 29280 Plouzané, France.*
[7]*Wageningen University, Hesselink van Suchtelenweg 4, 6703 CT Wageningen, the Netherlands.*
[8]*Université Paris-Sud, Centre Universitaire d'Orsay, Institut de Génétique et Microbiologie, UMR, CNRS, 8621, Bat 409, 91405 Orsay Cedex, France.*

Summary

The hyperthermophilic euryarchaeon *Pyrococcus abyssi* and the related species *Pyrococcus furiosus* and *Pyrococcus horikoshii*, whose genomes have been completely sequenced, are presently used as model organisms in different laboratories to study archaeal DNA replication and gene expression and to develop genetic tools for hyperthermophiles. We have performed an extensive re-annotation of the genome of *P. abyssi* to obtain an integrated view of its phylogeny, molecular biology and physiology. Many new functions are predicted for both informational and operational proteins. Moreover, several candidate genes have been identified that might encode missing links in key metabolic pathways, some of which have unique biochemical features. The great majority of *Pyrococcus* proteins are typical archaeal proteins and their phylogenetic pattern agrees with its position near the root of the archaeal tree. However, proteins probably from bacterial origin, including some from mesophilic bacteria, are also present in the *P. abyssi* genome.

Accepted 18 November, 2002 *For correspondence. E-mail forterre@igmors.u-psud.fr; Tel. (+ 33) 1 69 15 74 89; Fax (+ 33) 1 69 15 78 08.

© 2003 Blackwell Publishing Ltd

Introduction

The discovery of deep-sea hydrothermal vents and associated 'black smokers' has opened a novel field of research for microbiologists who were fascinated by these extreme environments where, as a result of hydrostatic pressure, hydrothermal fluids may remain liquid at temperature up to 400°C (Delaney *et al*., 1984). In 1989 the French-Japanese programme 'Starmer' organized a series of oceanographic cruises to detect and study hydrothermal activity in the SW Pacific and particularly the North Fiji Basin. From samples of smoker material and fluids collected by the man-operated submersible 'Nautile' at a depth of 2000 m, several heterotrophic anaerobic hyperthermophilic Archaea were isolated and assigned to the *Thermococcales* (Marteinsson *et al*., 1995). Among 20 isolates representing about six genomic species, strain GE5 was selected because it harboured a small multicopy plasmid, pGT5 (Erauso *et al*., 1992), and fully described as a novel species of the genus *Pyrococcus*, *P. abyssi* (Erauso *et al*., 1993). This new organism is a Gram-negative highly motile coccus, growing between 67°C and 102°C under atmospheric pressure, with an optimum at 96°C (doubling time 33 min), extended by at least 3°C when cells were cultivated under *in situ* pressure (20 MPa). This obligate heterotroph ferments peptides or mixtures of amino acids, producing acetate, isovalerate, isobutyrate, propionate, H_2 and CO_2, plus hydrogen sulphide when grown in the presence of elemental sulphur or cysteine.

Because of its deep-sea origin, *P. abyssi* was used as a model organism for physiological and enzymological studies, and biotechnological applications (Purcarea *et al*., 1994; Ladrat *et al*., 1996; Marteinsson *et al*., 1997; Dib *et al*., 1998). The ability of strain GE5 to grow on solid medium (Erauso *et al*., 1995) allowed for the design of a minimal medium (Watrin *et al*., 1995), and selection of several mutants, resistant to puromycin or auxotrophic for uracil (Watrin and Prieur, 1996; 1998). One of these

mutants was successfully used for transformation experiments with the gene *pyrE* from *Sulfolobus solfataricus* as a genetic marker inserted in a pGT5 derived shuttle vector (Lucas *et al*., 2003). *Pyrococcus abyssi* and closely related strain devoid of plasmids (e.g. GE9) appear thus promising species to develop genetic tools for hyperthermophiles.

The complete genome of *P. abyssi* GE5 (deposited at the CNCM under strain Orsay) has been sequenced at Genoscope in 1998 and released before publication (http://www.genoscope.cns.fr/Pab/). This sequence has been instrumental in the first experimental identification of a replication origin in Archaea (Myllykallio *et al*., 2000). Its rapid release allowed meaningful comparison with the genome of *Pyrococcus horikoshii* (Kawarabayasi *et al*., 1998), revealing that the terminus of replication was a hot spot of recombination in Archaea as in Bacteria (Myllykallio *et al*., 2000). Inversion of a large fragment carrying the single replication origin at its midpoint, and translocation of two fragments at the terminus, testify of a major role of replication-directed translocations in *Pyrococcus* genome evolution, as previously noticed for bacterial genomes (Makino *et al*., 2001; Zivanovic *et al*., 2002). Further comparison including the third sequenced *Pyrococcus* species, *P. furiosus* (Maeder *et al*., 1999) led to identification of DNA reorganization linked to IS-like elements and DNA integration within tRNA genes (Lecompte *et al*., 2001). These DNA rearrangements do not correlate with replication but are instead confined to one replichore (Zivanovic *et al*., 2002). Other rearrangements correspond to the mobility of long clusters of repeated sequences that could play a role in chromosome segregation (Mojica *et al*., 2000; Zivanovic *et al*., 2002). The possibility to compare three closely related species was also useful to identify small genes encoding functional RNA, such as snoRNA (Gaspin *et al*., 2000; Dennis *et al*., 2001). Comparison at the proteomic level indicates a high amount of differential gain and losses of genes among the three *pyrococci* (Lecompte *et al*., 2001). Recent gene transfer and genome polymorphism have been reported to correlate with the presence of restriction-modification genes (Chinen *et al*., 2000). Proteomic comparison of the three *pyrococci* also provided new tools to assess the relative substitution rates of proteins in different lineages, allowing the test of the molecular clock hypothesis at the interspecies level and to identify false orthologues and functional diversification (Jordan *et al*., 2001; Lecompte *et al*., 2001).

The first annotation of the *Pyrococcus* genome was released together with the sequence. A preliminary comparative description in terms of number of ORFs and RNA genes was published (Lecompte *et al*., 2001). Additional annotations were performed using the COG database (Natale *et al*., 2000; Tatusov *et al*., 2001). However, there is still no comprehensive description of a *Pyrococcus* genome in terms of gene content and function prediction. Furthermore, the number of completely sequenced genomes has dramatically increased in the last two years (including many new archaeal genomes), together with the number of previous orphan proteins now identified from biochemical analyses, challenging previous annotations. We have thus completely re-annotated the *P. abyssi* genome, taking into account the most recent biochemical and structural data available. We present here the result of this analysis, together with some comparative genomic data and analyses focusing on gene transfer and adaptation to hyperthermophily. Additional information is available on a dedicated website (see *Supplementary material*). The complete re-annotation of *P. abyssi* has been submitted as an update to the current entries for this sequence to EMBL databank (accession number AL096836).

Results and discussion

The initial genome annotation predicted the existence of 1764 genes, among which nearly 50% (864) had no attributed function (labelled as 'hypothetical proteins'). Three hundred and one new functional assignments (among which 110 have 'general function prediction' only) could be made with good confidence in this particular set of genes, mostly based on sequence similarity searches in the COG database with BLAST (Altschul *et al*., 1997; Tatusov *et al*., 1997). On the other hand, about 20 genes with function assignation in the initial annotation did not fit into any COG category, eventually leaving approximately one-third (580) orphan sequences. Overall analysis results are summarized in Table 1.

The global distribution of genes within different functional categories did not change significantly, the most salient feature remaining the net increase in new function assignment to hitherto functionless genes. Many of them (279, see Table 1) remain however, on a general function prediction level.

DNA replication, chromosome segregation and cell division

Pyrococcus abyssi is the first archaeon whose chromosomal replication origin (*oriC*) has been experimentally determined (Myllykallio *et al*., 2000; Matsunaga *et al*., 2001). *Pyrococcus abyssi oriC* is included into an intergenic region of about 800 bp located upstream of a 'replication island', encoding for the homologue of the eukaryal initiator proteins Cdc6/Orc1 (PAB2265), for the two DNA polymerase II subunits, DP1 (PAB2266) and DP2 (PAB2404) and the two clamp loader RF-C subunits (PAB0068 and PAB0069). PAB2265 is most likely involved in the initiation of DNA replication as it binds specifically to *P. abyssi oriC in vivo* (Matsunaga *et al*., 2001). Despite

Table 1. *Pyrococcus abyssi* annotation revision in light of the COG database.

Annotation revision overview	Current[a]	New annotation[b]
Number of genes	1765	1788
Genes with assigned function	901	1208
Hypothetical genes (no function assigned)	864	580[c]

Functional categories overview	Current[a]	COG assignations
Translation, ribosomal structure and biogenesis	125[d]	147[e]
Transcription	22	74
DNA replication, recombination and repair	54	73
Cell division and chromosome partitioning	16	13
Posttranslational modification, protein turnover, chaperones	53	35
Cell envelope biogenesis, outer membrane	56	50
Cell motility and secretion	25	39
Inorganic ion transport and metabolism	108	70
Signal transduction mechanisms	3	13
Energy production and conversion	122	110
Carbohydrate transport and metabolism	16	63
Amino acid transport and metabolism	80	129
Nucleotide transport and metabolism	45	50
Coenzyme metabolism	49	54
Lipid metabolism	20	15
Other	107[f]	
General function prediction only	–[f]	279[g]
Function unknown	–[f]	119

a. As appearing at http://www.cns.fr/pab/. This corresponds to the annotation currently in public databases
b. This work; see *Supplementary material*. Twenty-six new genes were added, and six suppressed.
c. This number includes genes (119) which belong to the 'function unknown' category.
d. Function assignations from current annotation were approximated to fit COG categories.
e. Items in this column correspond to COG database pathways and systems categories.
f. Equivalent function classes in the current annotation are difficult to establish.
g. One hundred and ten items in this class come from previously functionless ('hypothetical protein') genes.

a bacterial-like mode of replication (single origin, high replication speed and bidirectional replication), the *P. abyssi* genome only encodes for eukaryal-like DNA replication proteins. The functional interactions of the *P. abyssi* RF-C with human PCNA (Henneke *et al*., 2002), demonstrates a high conservation of the structural properties of these proteins from *pyrococci* to human and suggests that adjacent DNA polymerase II is involved in chromosome replication in *pyrococci*. A disproportionally large number of inteins are inserted in functionally important genes involved in nucleic acid metabolism (Liu, 2000) and, the presence of an intein in one of the two Pol II genes, indicates that this polymerase should be essential for *P. abyssi* viability.

Other important eukaryal-like replication proteins can be readily identified in the *P. abyssi* proteome (see *Supplementary material*). *Pyrococcus* species are the only Archaea in which the RP-A factor contains three subunits, as in Eukarya (Komori and Ishino, 2001). Although the smallest RP-A subunit (RP-A14: PAB2164) shares clear-cut similarities with the small subunit of eukaryal RP-A, we failed to detect homologues of RP-A14 in any other archaeal genome, except in the three *Pyrococcus* species. Another putative replication gene in the *oriC* region is PAB0067 that encodes a distantly related homologue of the eukaryal Dna2 protein. In Eukarya, the Dna2 protein, which is formed by the fusion of an helicase and an endonuclease domain, is involved in the maturation of Okazaki fragments (MacNeill, 2001). We have recently shown that Okazaki fragments have similar sizes in *P. abyssi* and Eukarya (around 150 nucleotides) thus being much shorter than bacterial Okazaki fragments (1–2 kb) (Matsunaga *et al*., 2003). Hence, the priming of Okazaki fragments in Archaea is most likely performed by the eukaryal-like primase (PAB 2235, 2236). As in Eukarya, maturation of Okazaki fragments could involve the concerted action of RP-A, FEN-1 and Dna2, all present in the *P. abyssi* proteome. Major putative *Pyrococcus* replication proteins (except the Dna2-like protein) have now been purified from *P. furiosus* and/or *P. abyssi*, and their activities *in vitro* are compatible with their proposed functions in DNA replication *in vivo* (Cann *et al*., 1998; 1999; 2001; Cann and Ishino; Bocquier *et al*., 2001; Gueguen *et al*., 2001; Liu *et al*., 2001; Henneke *et al*., 2003).

In striking contrast with the situation observed in the case of DNA replication, the two putative cell division proteins that can be detected in *P. abyssi*, MinD and FtsZ, have only bacterial homologues. The *P. abyssi* genome

encodes several paralogues of these two proteins. Searches using *P. abyssi* MinD (Gerard *et al.*, 1998) and FtsZ as reciprocal queries have allowed us to identify seven proteins of the MinD/MRP superfamily (PAB0852, 0954, 2105 1983, 0400 1795 and 0578) and three FtsZ proteins (PAB1820, 2351, 0851). One of these MinD protein and FtsZ protein (PAB0851, 0852) are associated into an operon-like structure. Archaea encode a single copy of the bacterial proteins XerC and XerD (PAB0255) involved in the resolution of dimeric chromosomes at the end of DNA replication. Another protein that can be involved in cell division and/or chromosome segregation is the universal SMC protein (PAB2109). *Pyrococcus abyssi* encodes a single type II DNA topoisomerase (Topo VI family B subunits PAB 2411 and 0407), that should be essential for the segregation of the chromosome and for the removal of topological constraints (positive superturns) during replication and transcription. Indeed, the other DNA topoisomerases encoded by *P. abyssi* can only relax negative superturns (Topo I family A – PAB 1430) or produces positive superturns (reverse gyrase – PAB 2432).

DNA repair and recombination

DNA in hyperthermophiles is continuously exposed to temperatures that increase dramatically the rate of reactions such as depurination or cytosine deamination. As the rate of introduction of spontaneous mutations does not seem to be especially high in *P. abyssi* (Watrin and Prieur, 1998; Watrin *et al.*, 1999), it should possess strong ability to repair DNA damages. Indeed, *Pyrococcus* species, including *P. abyssi*, are resistant to high doses of gamma-ray irradiation (DiRuggiero *et al.*, 1997; Gerard *et al.*, 2001).

Pyrococcus abyssi, as most other Archaea, encodes many bacterial-like enzymes that can either control the dNTP pool, directly correct modified bases, or remove modified bases to produce abasic sites (base excision repair). PAB1235 encodes a xanthosine triphosphate pyrophosphatase (preventing 6-*N*-hydroxylaminopurine mutagenesis) that has been recently identified in a structural genomic programme (Hwang *et al.*, 1999). PAB1968 encodes a homologue of MutT, an enzyme that hydrolyses oxidized deoxyguanosine nucleotide, 8-oxo-dGTP, whereas PAB1164 encodes a homologue of a dUTPase that avoids incorporation of uracil into DNA. A close relative of this enzyme encoded by the *Sulfolobus* virus SIRV1 has been biochemically characterized (Prangishvili *et al.*, 1998). PAB0260 encodes a putative O(6)-methyl guanine methyl transferase, that can directly correct DNA lesions, whereas PAB1530 and PAB0459 (endonuclease III, AP-endonuclease), encode a putative DNA glycosylase (AlkA homologue) that can remove methylated, alkyl-bases or oxidase bases. PAB0474 encodes an uracil-DNA glycosylase purified in *Thermotoga maritima* and *Archaeoglobus fulgidus* (Sandigursky and Franklin, 2000). Surprisingly, all three *Pyrococcus* species lack homologues of AP-endonucleases present in other archaeal genomes, such as the eukaryotic-like AP-endonuclease 1 and 2 and bacterial-like exonuclease III (XthA). Nevertheless, *P. abyssi* encodes several AP-endonucleases [PAB0459, PAB1103 (endonuclease IV) and PAB0916 (endonuclease V/nfI)] that could process apurinic sites produced by depurination at high temperature and abasic sites produced by the above glycosylases. PAB0034 encodes an homologue of the bacterial single-stranded nuclease RecJ, which is ubiquitous in Archaea. In Bacteria, this nuclease has been recently involved in mismatch repair. *Pyrococcus abyssi*, as most Archaea, lacks homologues of the classical bacterial/eukaryal DNA mismatch repair system, MutL and MutS. Although PAB1307 encodes a MutS2 protein, whose orthologue in *P. furiosus* has been recently characterized (Vijayvargia and Biswas, 2002), this protein does not have any detectable mismatch-specific DNA binding activity *in vitro.*

The mechanism of nucleotide excision repair (NER) has not yet been studied in Archaea *in vitro*. However, *P. abyssi* encodes a clear-cut homologue (PAB2385) of the eukaryal TFIIH helicase beta subunit (other names XPD/ERCC2, Rad3/Rad15) that is involved in transcription-coupled NER. Although PAB1312 and another superfamily II helicases of *P. abyssi* (PAB 0128) have been previously annotated as Rad 25(XPB) homologues (another TFIIH subunit), our analysis shows that PAB1312 is more related to the eukaryal ERCC3 (XPD) whereas PAB0128 is the representative of a new family of DNA/RNA helicases present in Archaea, Bacteria and viruses from the three domains (J. Filée, pers. comm.). Another protein that could be involved in NER is PAB0190 that contains a nuclease motif recently detected and experimentally verified in eukaryotic helicases of the XPF-ERCC1 family (Enzlin and Scharer, 2002).

A major pathway for DNA repair in Bacteria and Eukarya is homologous recombination. Recombinational repair is especially important to repair double-stranded breaks induced by disruption of replication forks or by gamma rays irradiation. The main archaeal recombination protein is RadA (PAB0164), an homologue of Rad51/RecA, that exhibits the properties expected for a recombinase *in vitro* (Komori *et al.*, 2000a). Archaea encode a second homologue of Rad51, called RadB (PAB2270). In the three *Pyrococci* species, the *radB* gene is located in the *oriC* region, suggesting that RadB could be involved in the repair of stalled replication forks. Indeed, RadB interacts specifically with DP1 *in vitro* (Hayashi *et al.*, 1999). In addition to RadA, homologous recombination in *P. abyssi* probably involved homologues of eukaryal pro-

teins Rad50 and Mre11 and the archaeal-specific resolvase Hjc (PAB0552) (Komori *et al.*, 2000b). The Rad50 and Mre11 proteins of *P. furiosus* exhibit the same catalytic activities *in vitro* than their eukaryal counterpart (Hopfner *et al.*, 2001). The Rad50 and Mre11 proteins are encoded in the *P. abyssi* genomes by two genes (PAB 0812 and PAB0811) that are organized into an operon with a third gene (PAB0813) coding for an homologue of the recently described *Sulfolobus* 5′-3′ exonuclease NurA (Constantinesco *et al.*, 2002). NurA could be involved, together with Rad50 and Mre11, in the processing of DSB to produce the protruding 3′ tails required by the recombinase. Another putative nuclease related to Mre11 and its bacterial homologue, SbcD, is PAB0926.

The identification of putative DNA repair genes in *Pyrococcus* should provide a good starting point to study DNA repair in radioresistant hyperthermophiles. It has been shown that the DNA double helix is very stable at the boiling point of water as long as the phosphodiester backbone remains intact (Marguet and Forterre, 1994). The prevention of DNA denaturation at the site of single or double-stranded breaks thus should be a priority for these organisms. It remains challenging to determine if the DNA repair proteins described in this section are sufficient to perform this task or if unidentified additional proteins are involved in this process.

Transcription

The basal transcriptional components of Archaea share homology with the eukaryotic RNA polymerase II system at both subunit complexity and sequence levels (Langer *et al.*, 1995). In the *P. abyssi* genome, 12 RNA polymerase subunits are encoded by six loci: *rpoL* (PAB2316), *rpo*DNK (PAB2410, PAB7131, PAB7132), *rpo*HBA′A″ (PAB7151, PAB0423, PAB0424, PAB0425), *rpo*E′E″ (PAB1105, PAB7428), *rpo*P (PAB3072), *rpoF* (PAB0732). Except for rpoL, the *rpo* genes are always found in conserved clusters that also encode ribosomal proteins or lie in the immediate vicinity of ribosomal protein genes. The cluster containing *rpo*HBA′A″ also encodes the transcription termination-antitermination factor NusA (PAB0426) whereas the transcription antitermination protein NusG (PAB2352) belongs to another ribosomal protein gene cluster.

The initiation of transcription in Archaea seems to be a minimalist version of the eukaryotic system required for RNA polymerase II initiation (Thomm, 1996). Similarly to other Archaea, *P. abyssi* genome encodes two basal transcription factors: TBP, the TATA-binding protein (PAB1726) and TFB, a homologue of TFIIB (PAB1912) but in contrast to other archaeal species, including *P. horikoshii and P. furiosus*, *P. abyssi* contains a unique copy of the TFB gene. *Pyrococcus abyssi* also encodes the transcription factor S (PAB1464) similar to both the eukaryotic transcription elongation factor TFIIS and to small subunits of eukaryotic RNA polymerases I, II and III. Indeed, this protein, formerly annotated RpoM, is not a subunit of the archaeal RNA polymerase but is able to induce RNA cleavage in the RNA polymerase like eukaryotic TFIIS (Hausner *et al.*, 2000). In addition, the *P. abyssi* genome encodes a protein (PAB0950) sharing sequence similarity with the N-terminal part of the eukaryotic TFIIE alpha subunit. Recent experiments suggested that this protein, called transcription factor E in Archaea, facilitates or stabilizes interactions between TBP and the TATA box (Bell and Jackson, 2001; Hanzelka *et al.*, 2001). As indicated above, *P. abyssi* encodes an orthologue of the TFIIH helicase beta subunit (PAB2385); a role in transcription remains to be demonstrated.

Despite the similarity of the archaeal and eukaryotic transcriptional machinery, transcription regulation in Archaea seems to involve bacterial-like proteins (Bell and Jackson, 2001). *Pyrococcus abyssi*, like all archaeal genomes, encodes a plethora of putative bacterial-like regulators, in particular proteins belonging to the Lrp/AsnC family (Brinkman *et al.*, 2000). In *Pyrococcus*, the picture of the transcription regulation is further complicated by the presence of two supplementary TBP-interacting proteins. The first one (PAB2107) is homologous to the eukaryotic TIP49 protein and is absent in most euryarchaeal sequenced genomes. The second one (PAB1510) is apparently specific to *Pyrococcus* species and has been shown to negatively regulate transcription by inhibiting the interaction between TBP and TATA-DNA in *P. kodakaraensis* (Matsuda *et al.*, 1999).

In relation to promoter structure, an analysis of the frequency of trinucleotides has been performed (see *Supplementary material*).

Translation

A single-copy rRNA operon contains 16S, $tRNA^{Ala}$ and 23S genes. Two copies of the 5S rRNA gene are present: the first one is adjacent to the $tRNA^{Asp}$(GTC) gene whereas the second is isolated on the genome, just like the 7S RNA, the RNA component of the signal recognition particle. The rRNAs modification may involve a dimethyladenosine transferase (PAB0253) and two homologues of eukaryotic snoRNP common proteins implicated in the rRNA ribose methylation: a fibrillarin-like prerRNA processing protein (PAB2306) and a Nop58p-like protein (PAB2305). A homologue of the eukaryotic IMP4 protein, a U3 snoRNP component, is also present in *P. abyssi* genome (PAB2357) (Mayer *et al.*, 2001). In line with the presence of snoRNP proteins homologues, a family of 46 box C/D small RNAs homologues of eukaryotic methylation guide small nucleolar RNAs have been experimen-

tally identified in the *P. abyssi* genome (Gaspin *et al*., 2000). In addition, three genes (PAB0419, PAB0420, PAB0421) organized in an operon-like cluster are predicted to encode the archaeal counterparts of the core subunits of the eukaryotic exosome that mediates processing and degradation of a variety of RNA (Koonin *et al*., 2001).

Sixty-two ribosomal protein genes are present and generally exhibit a higher similarity to their eukaryotic counterparts than to bacterial genes although they often occur in bacterial-like clusters, including a large operon containing 22 ribosomal protein genes. Genes encoding translational initiation factors IF1 (SUI1), IF2, eIF1A, eIF2 alpha, beta and gamma subunits, eIF2BI and II subunits, eIF5A, eIF6 as well as elongation factors EF1 alpha and beta subunits and EF2 are present. Interestingly, in addition to the genes that encode the archaeal initiation factor eIF2BI and II subunits that belong to the eukaryotic eIF2B alpha, beta and delta subunit family, we found a third gene (PAB1306) belonging to the same family and specific to the three *Pyrococcus* species. This finding further complicates the distribution scheme of eIF2B subunits in archaeal genomes as one or two (Kyrpides and Woese, 1998) and now three genes can be present, which raises the question of the actual role of eIF2B in Archaea.

In *P. abyssi*, the canonical pattern for archaeal tRNA genes is found, with 46 tRNAs isoacceptors able to decode the 61 sense codons (Marck and Grosjean, 2002). Proteins involved in modification systems of tRNAs include a tRNA intron endonuclease (PAB1099), a tRNA nucleotidyltransferase (PAB0063), an archaeosine tRNA-ribosyltransferase (PAB0740), a N2, N2-dimethylguanosine tRNA methyltransferase (PAB2092) and tRNA pseudouridine synthases A and B (PAB1701 and PAB0356, respectively). Moreover, four novel box C/D small RNAs have recently been experimentally identified in *P. abyssi* and predicted to direct 2′-O-ribose methylation onto the first position of the anticodon in $tRNA^{Leu}$(CAA), $tRNA^{Leu}$(UAA), elongator tRNA and $tRNA^{Trp}$, respectively (d'Orval *et al*., 2001). We also observed in the *P. abyssi* genome the RNase P RNA and three proteins encoding genes which are similar to some eukaryotic RNase P protein components involved in tRNA processing. Two of these genes (PAB0467 and PAB1136) have been previously reported as being orthologues of the *Saccharomyces cerevisiae* RNase P protein subunit Pop5 and Rpp1, respectively (Koonin *et al*., 2001). The third gene (PAB2126) is similar to the C-terminal domain of another eukaryotic RNase P protein subunit, the RNA processing protein Pop4.

In the *P. abyssi* genome, aminoacyl-tRNA synthetases (aaRS) are present for all amino acids (except for glutamine) including a class I LysRS (PAB0139). Interestingly, two *Pyrococcus* aaRSs exhibit unusual phylogenetic relationships: the TrpRS is more closely related to its eukaryotic counterparts than to archaeal ones and the closest homologue of *P. abyssi* ArgRS is found in *Deinococcus radiodurans*. Organisms lacking a GlnRS use an alternative synthesis mechanism, the transamidation of Glu-$tRNA^{Gln}$ to Gln-$tRNA^{Gln}$ by a Glu-$tRNA^{Gln}$ amidotransferase. In *Pyrococcus*, this function is ensured by an archaea-specific heterodimeric enzyme (PAB1901 and PAB1902) (Tumbula *et al*., 2000). Remarkably, *Pyrococcus*, *Thermoplasma* and *Pyrobaculum* genera are the sole sequenced archaeal genomes that contain an AsnRS gene and lack the homologue of the bacterial heterotrimeric amidotransferase (*gat*CAB) which amidates mis-aminoacylated Asp-$tRNA^{Asn}$ to Asn-$tRNA^{Asn}$. In this line, we found an additional ORF (PAB2356) of unknown function which is highly similar to the AsnRS/AspRS conserved catalytic core. Intriguingly, this gene exhibits the same pattern of co-presence/co-absence as the archaeal AsnRS as it is conserved in *Pyrococcus*, *Thermoplasma* and *Pyrobaculum* genomes and is absent in all other archaeal genomes. At the evolutionary level, this finding suggests the PAB2356 family arises from an ancestral duplication of the AsnRS/AspRS catalytic domain predating the crenarchaeal and euryarchaeal split. At the functional level, this protein family retains the residues involved in ATP and magnesium ions binding in both AsnRS and AspRS (Poterszman *et al*., 1994; Berthet-Colominas *et al*., 1998) but holds the three discriminative residues responsible for the aspartate recognition in AspRS. Thus, this protein constitutes an original example of an aaRS catalytic domain recruitment in a function that uncouples tRNA recognition from amino acid recognition and ATP binding.

As previously observed in the *P. furiosus* genome (Robb *et al*., 2001), four additional genes appear to code for proteins homologous to aminoacyl-tRNA synthetase non-catalytic domains. Three of these proteins (PAB1190, PAB0066 and PAB1440) of variable length (405, 213 and 159 amino acids, respectively) are homologous to the C-terminal part of alanyl-tRNA synthethases whereas PAB0278 is closely related to the extreme C-terminal domain of some methionyl-tRNA synthetase and seems to be a recurrent module of several RNA-binding proteins. In *Aquifex aeolicus*, a free-standing protein homologue of PAB0278 has been shown to code for a structure-specific tRNA-binding protein but its cellular role remains unclear (Morales *et al*., 1999).

Metabolism

Generation of energy and reductant. The catabolic degradation by *P. abyssi* of both peptides and sugars to their respective metabolites (discussed below), generally is

coupled to the generation of energy (ATP) and reductant (mainly ferredoxin). Apart from ATP as energy source to drive various reactions and processes, the re-oxidation of ferredoxin and the carboxylation of certain metabolites is anticipated to give rise to the generation of an electrochemical gradient (proton or sodium potential) across the cytoplasmic membrane of *P. abyssi*, that potentially drives various secondary processes.

All *Pyrococci* lack the genes of haem biosynthesis, and they do not seem to encode any of the classical primary H^+ pumps such as the respiratory chain oxidoreductase complexes. However, a variant of the respiratory NADH dehydrogenase complex, a multisubunit hydrogenase, has recently been proposed to be a potential energy transducing system (Silva *et al.*, 2000). In addition, genes have been identified in the *P. abyssi* genome that suggest the presence of a Na^+-translocating methylmalonyl-CoA decarboxylase, a primary Na^+ pump that is typical for anaerobic bacteria. This enzyme is most likely involved in decarboxylation of methylmalonyl-CoA and malonyl-CoA, degradation products of pyrimidine and amino acids (Met, Val, Ile). The genes encoding its three subunits and the biotin carrier protein form an operon (PAB1769-PAB1772), which is similar to the one found in other *Pyrococci*. *Pyrococcus abyssi* encodes an archaeal A_0A_1-type ATPase (PAB1179-1184, PAB1186, PAB2378-2379). At present it is not known whether its coupling ion is H^+ or Na^+, and whether it functions primarily as an ATP hydrolase or an ATP synthetase. Although using a sodium-motive force for ATP synthesis in *P. abyssi* would seem an attractive possibility, such a mechanism has not been demonstrated so far in hyperthermophilic Archaea. In addition to the A_0A_1-type ATPase, *P. abyssi* encodes an ABC-type Na^+-ATPase (PAB0434-PAB0435), homologous to the *B. subtilis* NatAB system of Na^+ export. This enzyme might function in preventing Na^+ toxicity (Hase *et al.*, 2001). Ferredoxin appears to be the major metabolic electron carrier in *pyrococci* (Fig. 1). After reduction during peptide or sugar degradation, it is mainly re-oxidized by a membrane-bound hydrogenase (Silva *et al.*, 2000), potentially generating membrane potential. In addition, ferredoxin has been suggested also to be re-oxidized by ferredoxin-NADP oxidoreductase; in that case the electrons are further transferred from NADP to external electron acceptors, such as elemental sulphur or polysulphide (Silva *et al.*, 2000; Schut *et al.*, 2001). Moreover, NADPH can also be oxidized via the conversion of pyruvate to alanine, via glutamate dehydrogenase (PAB0391) and alanine aminotransferase (PAB1810) (Ward *et al.*, 2000).

Transport. The *P. abyssi* genome encodes dozens of solute import and export systems. Several ABC transporters [extracellular high-affinity binding protein, membrane-embedded permease(s) and intracellular ATPase(s)] and secondary transporters (membrane-embedded symport or antiport systems) are present (Fig. 1). As in all available archaeal genomes, no counterpart of the bacterial phosphotransferase system (PTS) has been found in *P. abyssi*. Like many anaerobic bacteria, archaea appear to apply a transmembrane gradient of H^+ or Na^+ ions as a source of energy for secondary transport (Hase *et al.*, 2001). Accordingly, *P. abyssi* encodes a membrane permease for dicarboxylates (PAB1799), a Na^+-alanine symporter (PAB1527), and three proteins of the solute sodium symporter (TC 2.A.21) family, including a likely Na^+-proline symporter. In addition, *P. abyssi* encodes a member of the sodium neurotransmitter symporter (TC 2.A.22) family (PAB1850), whose function is still unknown. Maintenance of the cytoplasmic pH and generation of the proton-motive force in *P. abyssi* is carried out by the Na^+/H^+ antiporter NhaP (PAB1518). In addition, *P. abyssi* and other *pyrococci* encode putative Na^+/H^+ antiporters, homologous to the multisubunit antiporter MnhBCDEFG of *Staphylococcus aureus* (Hiramatsu *et al.*, 1998), which have recently been reported to be hydrogenase-associated (Silva *et al.*, 2000). *Pyrococcus abyssi* and *P. furiosus* both encode two major phosphate transporters, the Pst and Pit systems, whereas *P. horikoshii* contains only the Pit system. The Pst system is a high-affinity ABC-type transport system that is encoded by a well-conserved *pstSCAB-phoU* operon. In *P. abyssi*, however, the *pstS-pstC* region of this operon (PAB2365-0702) contains three additional genes, encoding an alkaline phosphatase (Zappa *et al.*, 2001) and two copies of an uncharacterized protein, distantly related to apurinic/apyrimidinic site (AP) endonucleases. The physiological function of this alkaline phosphatase is not clear, but it has been shown to dephosphorylate linear DNA fragments. In concert with putative AP endonucleases, it could potentially participate in the degradation of damaged DNA. *Pyrococcus abyssi* genome contains two copies of the pitA gene (PAB0927 and PAB1652). The gene encoding the arsenate efflux pump has two copies in *P. horikoshii* (PH0824 and PH0888); one is present in *P. furiosus* and none in *P. abyssi*. However, *P. abyssi* encodes an ArsB-related efflux pump for arsenite, the product of reductive detoxification of arsenate.

Peptide degradation. Pyrococcus abyssi grows well in a peptide broth, and peptide fermentation comprises its principal metabolic pathway (Erauso *et al.*, 1993; Godfroy *et al.*, 2000). The genome of *P. abyssi* contains five operons encoding ABC-type dipeptide/oligopeptide transport systems that are probably used for uptake of specific oligopeptides and amino acids. About 30 different genes in the *P. abyssi* genome code for proteins with potential proteolytic or peptidolytic activity: the alpha and beta subunits of component A proteasome, a 26S proteasome

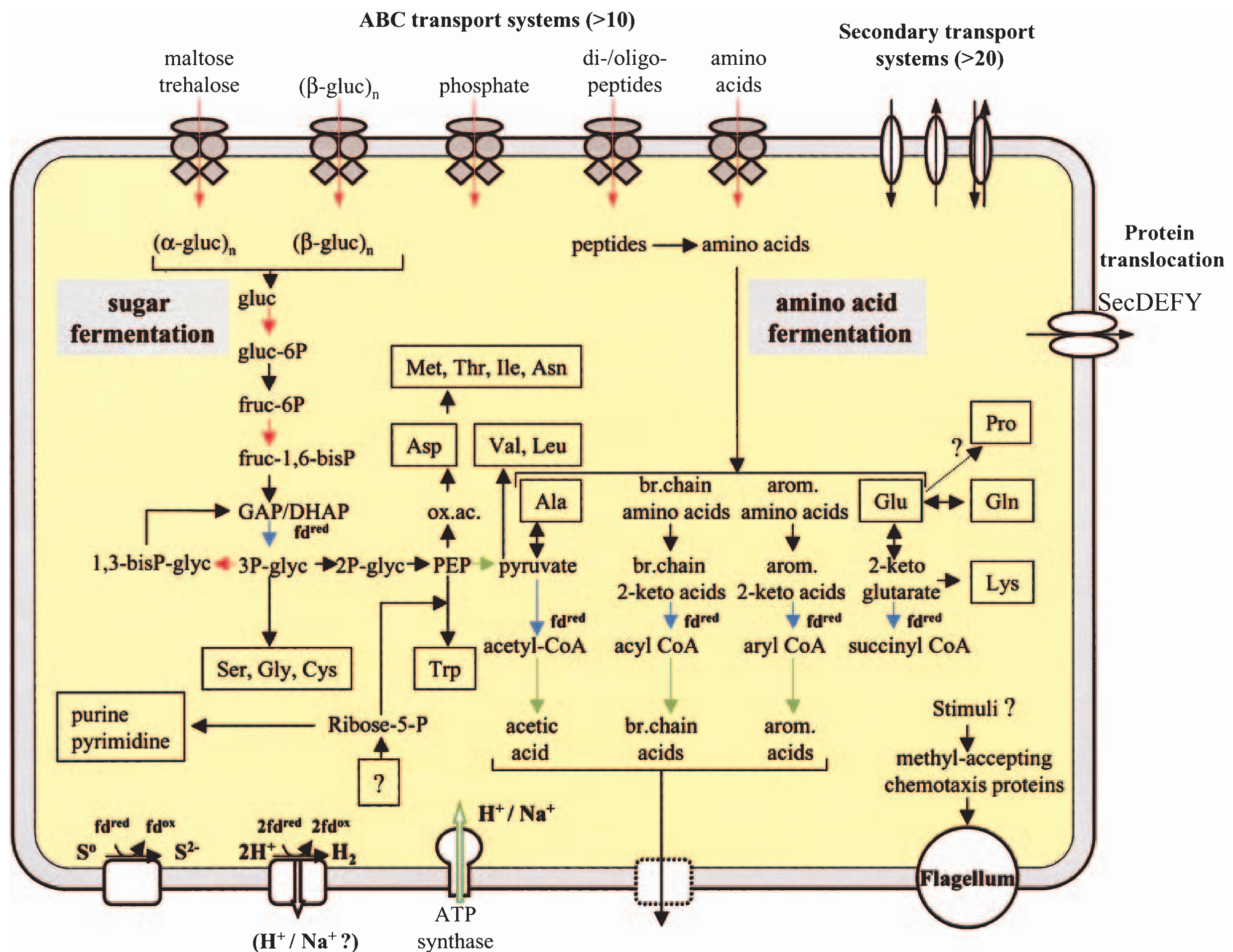

Fig. 1. Overview of fermentation and biosynthesis pathways in *P. abyssi*. Arrows denote the following reactions: chemical conversion (black), energy-consuming (red), energy yielding (green), redox (blue); conversions/intermediates that were anticipated but for which no gene/compound has yet been identified are shown as broken arrows/question mark; extracellular enzymes that hydrolyse polymers (proteases, glycosyl hydrolases) are not shown; more than 10 operons encoding ABC transport systems are present; At least 20 secondary transporters (permeases) are present (symport and antiport); all of the genes encoding enzymes of the *Pyrococcus* variant Embden–Meyerhof (EM)-type glycolytic pathway have been identified (gluc, glucose; fruc, fructose; GAP, glyceraldehyde-3P; DHAP, dihydroxyacetone-P; glyc, glycerate; PEP, phosphoenolpyruvate); only two genes that are potentially involved in the Pentose Phosphate Pathway (PPP) were identified (see text), the precursor of ribose-5P is yet unknown; many genes of the archaeal-type amino acid fermentation network have been identified; several components of ferredoxin-oxidizing complexes (possibly generating membrane potential) have been identified; 15 amino acid biosynthesis pathways have been identified (almost) completely (see text), with the exception of the Proline biosynthesis enzymes; purine and pyrimidine biosynthesis pathways are completely present; archaeal flagellar components are present, possibly responding to methyl-accepting proteins by yet unknown chemotaxis stimuli; components of a Sec/SRP-type translocation system have been identified. See text for detailed discussion.

regulatory subunit, pyrolysin-like cysteine protease (PAB1252) (De Vos *et al.*, 2001) and a serine protease of the family S9A (PAB0762), that have not been found so far in any Archaea except for *Pyrococci*. In addition, an aminopeptidase (leucine or methionine specific), a carboxypeptidase and an endopeptidase, as well as metal-dependent (zinc or cobalt) peptidases, several prolyl dipeptidases and a pyrrolidone carboxypeptidase have been detected. The resulting amino acids are then deaminated to 2-keto acids in reactions that could be catalysed by several different aminotransferases (PAB0086, PAB0501, PAB0525, PAB1523, PAB1810, PAB1921, PAB2386, PAB2440), encoded in the *P. abyssi* genome (Fig. 1). Further metabolism of these 2-keto acids includes their conversion to the corresponding CoA derivatives, catalysed by four distinct 2-ketoacid ferredoxin oxidoreductases: pyruvate oxidoreductase (PAB1470, PAB1474-1476), indolepyruvate oxidoreductase (PAB0855, PAB0857), 2-keto-isovalerate oxidoreductase (PAB1470-1473), and 2-keto-glutarate oxidoreductase (PAB0341, PAB0344-0348, PAB2359) (Fig. 1). As noted (Adams, 1999), this mechanism is clearly distinct from what has been described for mesophilic anaerobic bacteria where 2-keto acids are first decarboxylated to alde-

hydes, and then oxidized to the corresponding acids by NAD(P)-linked dehydrogenases. *Pyrococcus abyssi* encodes two acetyl-CoA synthetases (NDP-forming) that can hydrolyse these acyl-CoA derivatives into their corresponding organic acids and free CoA, saving the free energy of this reaction in the form of ATP (Mai and Adams, 1996; Sanchez *et al*., 2000) (Fig. 1). The subsequent reoxidation of ferredoxin is discussed below.

Sugar degradation. In its original description, *P. abyssi* GE5 has been reported not to grow on carbohydrates, alcohols, organic acids and individual amino acids (Erauso *et al*., 1993). In a subsequent comparison of different *pyrococci* isolates, growth of both *P. furiosus* and *P. abyssi* on starch, maltose and pyruvate has been reported; in contrast, *P. horikoshii* did not grow under these conditions (Gonzalez *et al*., 1998). Moreover, *P. abyssi* strain ST549 has been shown to be unable to use disaccharides as carbon source (Godfroy *et al*., 2000). This discrepancy could reflect variation in strain properties, but most likely reflects a lag phase required for cellular adaptation to allow the shift to an alternative carbon source. The capacity to catabolise at least some sugar substrates, would agree with the fact that the *P. abyssi* genome encodes a number of glycoside hydrolases and sugar transporters. Different endoglucanases may be involved in the extracellular degradation of both alpha-glucans (alpha-amylase, PAB0118; pullulanase, PAB0122) and beta-glucans (endo-beta-glucanase, PAB0632). Interestingly, the former hydrolase genes are clustered in an operon that encodes an ABC-type transport system that resembles the *P. furiosus* maltose/trehalose transporter, and the endo-beta-glucanase gene forms an operon with a homologue of a cellobiose transporter (Koning *et al*., 2001). Moreover, another ABC transporter operon (PAB1343-PAB1349) also contains beta-mannosidase and beta-galactosidase genes. All the genes that encode enzymes of the Embden-Meyerhof glycolytic pathway have been identified in the *P. furiosus* genome (Verhees, 2002), and orthologues are present in *P. abyssi* and *P. horikoshii* (Fig. 1). Compared to the glycolysis of bacteria and eukaryotes, the pyrococcal pathway has several unique enzymes: (i) ADP-dependent glucokinase (PAB0967); (ii) glucose-6-phosphate isomerase (PAB1199); (iii) ADP-dependent phosphofructokinase (PAB2013); (iv) fructose-bis-phosphate aldolase (DhnA-type, PAB0049); and (v) glyceraldehyde-3-phosphate oxidoreductase (PAB1315). The former non-orthologous enzymes catalyse analogous steps in the glucose degradation (Verhees, 2002). The latter enzyme, however, catalyses an unique conversion: rather than the usual ATP-yielding, NAD-dependent two-step oxidation of glyceraldehyde-3-phosphate to 3-phosphoglycerate, *Pyrococci* use a unique ferredoxin-dependent tungsten enzyme to catalyse the same conversion in a single step, without the gain of ATP (Van der Oost *et al*., 1998).

Like all anaerobic organisms, *P. abyssi* does not possess a complete citric-acid cycle (Huynen *et al*., 1999). It appears to interconvert phosphoenolpyruvate (PEP) and oxaloacetate by PEP carboxylase (PAB0016) and PEP carboxykinase (PAB1253). In addition, malate may be converted by fumarase (PAB2030/2031), an archaeal-type malate dehydrogenase (PAB1791) and malic enzyme (PAB1792). Ribose-5-phosphate isomerase (PAB0522) and potential transketolases (PAB0295-PAB0296) are the only enzymes of pentose phosphate pathway encoded in *P. abyssi*. Apparently, the enzymes involved in the synthesis of ribose-5-phosphate (precursor in biosynthetic pathways of tryptophan, purine and pyrimidine; see below) have evolved independently from their analogues in bacteria and eukaryotes. It has been proposed that the above-mentioned glycolytic aldolase, which is highly conserved in archaea, may also catalyse the formation of a pentose from glyceraldehyde-3-phosphate and acetaldehyde, a product of peptide fermentation (Galperin *et al*., 2000).

Amino acid biosynthesis. As described in detail below, the *P. abyssi* genome reveals the presence of (almost) complete pathways for the synthesis of 15 amino acids: **Trp**, **Lys**, **Ser**, **Gly**, **Cys**, **Glu**, **Gln** (partial), **Ala**, **Asp**, and **Asn**, and most likely *Met (*partial), *Val, Ile, Leu,* and *Thr*. On the contrary, *P. abyssi* appears to lack the sets of genes required for the synthesis of **Arg**, **His**, **Phe**, **Tyr**, and *Pro* (amino acids in bold indicates an agreement between genomic and experimental data, in italic *disagreement*) (Table 2). It has been shown experimentally that *P. abyssi* is indeed auxotrophic for **Arg**, **His**, **Phe** and **Tyr** (Watrin *et al*., 1995). However, in the latter study *P. abyssi* GE5 did grow in the absence of *Pro*, and it did not grow in the absence of *Met, Thr*, *Ile*, *Leu*, and *Val*, which does not agree with the presence of all genes required to express the biosynthetic routes (Table 2). Although this discrepancy may be explained by global regulatory effects, additional experiments are needed to solve this matter. A recent analysis of *P. abyssi* ST549 demonstrated auxotrophy for Arg, His, Phe, Tyr, Val, Leu, Thr, Trp, suggesting that significant differences in amino acid requirements might occur from strain to strain (A. Godfroy, per. comm.).

The entire set of enzymes involved in tryptophan and threonine biosynthesis pathways is encoded in the *P. abyssi* genome. All but two enzymes required for serine, glycine and cysteine biosynthesis are present. The enzymes that constitute the classical bacterial lysine biosynthesis pathway are not present whereas a cluster of genes encoding an alternative AAA-type route is found (Brinkman *et al*., 2002). Despite methionine auxotrophy,

Table 2. Anabolic capacity of *P. abyssi* as deduced from genome analysis.

Amino acid biosynthesis	Gene name	(predicted) PAB
Phe, Tyr biosynthesis	pheA/aroH	No
	tyrA	No
	aspC	No
Trp biosynthesis	aro-operon	297–307
	trp-operon	2043–2049
His biosynthesis	his operon	No
Ser biosynthesis	serAB	514,1207
Gly biosynthesis	glyA	2018
Thr biosynthesis*	thr-operon	1674–1678
Cys biosynthesis	cysKM	250,605
Leu biosynthesis*	leuABCD	890–894,2424
Ile, Val biosynthesis*	ilvBCD	888,889,895
Met biosynthesis*		605–608,610 1372,2094,2361
Pro biosynthesis**	novel type?	?
Lys biosynthesis (AAA-type)	lysYZJK	286–294
Arg biosynthesis	argGH	No
Ala biosynthesis	alaAT	1810
Asp biosynthesis	aspAT	Several ATs
Glu biosynthesis	gltD	1214 1738
Gln biosynthesis	glnA	1292
Asn biosynthesis	asnB	750,1605

Nucleotides and cofactors	Present
Purine biosynthesis*	Yes
Pyrimidine biosynthesis	Yes
NAD biosynthesis	Yes
Haeme biosynthesis	No
Cobalamin biosynthesis	No
Folate biosynthesis	No
Pyridoxal biosynthesis	No
Biotin biosynthesis	No
Coenzyme A biosynthesis*	Yes
Haem biosynthesis	No

Predicted genes/operons involved in amino acid biosynthesis are indicated by PAB identifier; when no gene has been identified it is indicated (No). In some cases there is a discrepancy with experimentally determined autotrophy (*) or prototrophy (**). For details, see text and *Supplementary material*.

several orthologues of bacterial enzymes involved in methionine biosynthesis can be found. In contrast, none of the proteins for proline biosynthesis has been detected in spite of a reported proline prototrophy, suggesting a novel pathway. A more detailed analysis of the different amino acids biosynthesis pathways is available (see *Supplementary material*).

Nucleotide synthesis. Like other *pyrococci*, *P. abyssi* encodes almost complete set of purine biosynthesis enzymes, including the recently described PurS subunit of the phosphoribosyl-formylglycinamidine (FGAM) synthase. The only exceptions are the *purK* gene and the *purH* gene, encoding the ATP-binding carboxylase subunit of the phosphoribosylamino-imidazole carboxylase and the bifunctional fusion protein with AICAR transformylase and IMP cyclohydrolase activities, respectively. All the enzymes of *de novo* pyrimidine biosynthesis are found in *P. abyssi.* Like in many species, the genes encoding the catalytic and regulatory subunits of aspartate carbamoyltransferase are adjacent. The thymidylate synthase of *P. abyssi* (PAB0861) belongs to a newly described family of these proteins (ThyX) widely distributed in Archaea and Bacteria (Myllykallio *et al*., 2002). ThyX proteins have no structural similarity and exhibit a new enzymatic mechanism compared to classical thymidylate synthases (ThyA). The complete set of elements for the reduction of nucleoside triphosphates to deoxynucleoside triphosphates were found, glutaredoxin (PAB2245), several thioredoxin-like proteins, and an S-adenosylmethionine-dependent anaerobic ribonucleoside triphosphate reductase (PAB2337).

A more detailed analysis of the nucleotide biosynthesis pathways is available (see *Supplementary material).*

Vitamin biosynthesis. Pyrococcus abyssi genome does not contain genes encoding for the biosynthesis pathways of biotin, riboflavin and haem. In contrast, it encodes a complete set of enzymes for pyridine nucleotide biosynthesis and most (but not all) of the homologues of bacterial genes for thiamine biosynthesis. A more detailed analysis is available and can be found as supplementary data.

Comparative genomics

Previous studies focusing on the comparison of the genomes of *P. abyssi*, *P. horikoshii* and *P. furiosus* (Ettema *et al*., 2001; Lecompte *et al*., 2001) revealed that a high amount of differential gains and losses of genes occurred since the divergence of the three species. The fraction of genes absent in one or two *Pyrococcus* species includes well-characterized operons involved in amino acid biosynthesis, maltose transport and phosphate uptake. In fact, proteins conserved in the three species represent only two-thirds of each proteome, which highlights the great genomic and metabolic plasticity of these three free-living Archaea.

To gain insight into the evolutionary history of *Pyrococci*, we have compared the *P. abyssi* proteome to those of other completely sequenced genomes. The taxonomic distribution (Fig. 2) of the closest homologues of *P. abyssi* proteins, excluding the orthologues found in *P. horikoshii* and *P. furiosus*, shows significant differences among single-gene phylogenies as previously observed in other genome scale studies (Doolittle, 1999; Sicheritz-Ponten and Andersson, 2001). *Pyrococcus*-specific genes represent 16% of the *P. abyssi* proteome and 50% of the ORFs are most similar to proteins of Euryarchaeota. At the functional level, most of the *Pyrococcus*-specific proteins are uncharacterized whereas the latter set includes most of the housekeeping genes, i.e. genes coding for proteins

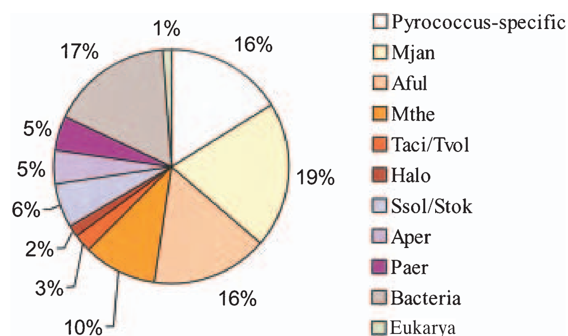

Fig. 2. Comparison of the *P. abyssi* ORFs with those of other completely sequenced genomes, excluding the *P. horikoshii* and *P. furiosus* genomes. The percentage of *P. abyssi* ORFs with greatest per cent identity (>20%) to each complete archaeal proteome is indicated. Abbreviations used for Euryarchaeota: Mjan: *M. jannaschii*, Aful: *A. fulgidus*, Mthe: *M. thermoautotrophicum*, Taci: *T. acidophilum*, Tvol: *T. volcanium*, Halo: *Halobacterium* sp. and for Crenarchaeota: Ssol: *S. solfataricus*, Stok: *S. tokodai*, Aper: *A. pernix*, Paer: *P. aerophilum*. The *P. abyssi* ORFs most similar to bacterial proteins were cumulated in a unique category for illustration purposes, as well as ORFs most similar to eukaryotic proteins.

involved in DNA metabolism, transcription and translation. Among Euryarchaeota, the proteomes of two methanogenic species, *Methanococcus jannaschii* and *Methanobacterium thermoautotrophicum*, as well as the sulphate reducer *Archaeoglobus fulgidus,* appear significantly closer to *Pyrococcus* than that of *Thermoplasma* and *Halobacterium* species, which is in agreement with the archaeal tree based on SSU rRNA analysis (Woese, 1996) or concatenated ribosomal proteins (Matte-Tailliez *et al*., 2002).

Although *P. abyssi* belongs to the phylum Euryarchaeota, 16% of the proteins have a crenarchaeal counterpart as a closer homologue. The position of *Pyrococcus* as the deepest branched Euryarchaeota sequenced to date in the archaeal tree singularly complicates the interpretation of the close relationship between *Pyrococcus* and Crenarchaeota. Some ancestral genes present in the common ancestor of Euryarchaeota and Crenarchaeota could have been lost or could have strongly diverged in the euryarchaeal lineage after the emergence of *Pyrococcus*, but the possibility of a lateral gene transfer cannot be excluded.

In addition, the *P. abyssi* proteome contains a substantial set (17%) of bacterial-related proteins. One-third of these proteins (103) are most similar to *T. maritima* proteins and 38 are closely related to *A. aeolicus*. The presence of a high number of archaeal-like genes in *A. aeolicus* and *T. maritima*, two deeply branched hyperthermophilic bacterial species, has been previously reported (Aravind *et al*., 1998; Nelson *et al*., 1999). Although the corresponding evolutionary scenario is subject to intense debates (Aravind *et al*., 1998, 1999; Kyrpides *et al*., 1999), lateral gene transfer between hyperthermophiles of the two domains has been well documented in several cases (Doolittle, 1999; Forterre *et al*., 2000). Another set (102 proteins) of bacterial-like genes is most similar to bacteria belonging to Firmicutes with a high number of proteins closely related to *Bacillus* and *Clostridium* proteins (48 and 25 respectively). The last significant group consists of 47 proteins closely related to proteins encoded by genomes of Proteobacteria.

We observed that crenarchaeal- and bacterial-like genes in *P. abyssi* are not uniformly distributed among the functional categories. A large number appear to function in transport, resulting in a largely heterogeneous repertoire of 104 transport genes in *P. abyssi* with only 30% of genes most similar to euryarchaeal counterparts. Remarkably, the proteins belonging to the four ABC-type sugar transport systems as well as to the five ABC-type peptide transport systems identified in *P. abyssi* are all most similar to *P. aerophilum*, *A. pernix* or *T. maritima* counterparts with the same operonic organization. Besides organic substrate transporters, many genes in *P. abyssi* that are involved in organic substrate degradation are bacterial or crenarchaeal-like. For instance, most of the glycosyl hydrolase genes, which are frequently clustered with ABC transporters, have no euryarchaeal counterparts or only very divergent ones. The orthologues of the endo-beta-glucanase (PAB0632) are exclusively bacterial and the closest orthologues of the alpha-glucan phosphorylase (PAB2414) and alpha-amylase (PAB0118) are found in *T. maritima* and in the thermophilic bacterium *Dictyoglomus thermophilum* respectively. Additionally, we detected genes, present in *P. furiosus* but absent in *P. horikoshii* that are involved in phospholipid degradation and glycerol metabolism. They consist of a phospholipase (PAB1050) closely related to *T. maritima* and a stretch of bacterial-like genes including a glycerolphosphoryl diester phosophodiesterase (PAB0180), a glycerol kinase (PAB2406) and a glycerol-3-phosphate dehydrogenase (PAB0183). Other proteins involved in degradation pathways also exhibit atypical phylogenetic relationships as seven peptidases are bacterial- or crenarchaeal-like, including two D-aminopeptidases (PAB1969 and PAB0045) and a D-aminoacylase (PAB0090). Two amino acid racemases are bacterial-like as well. Thus, the metabolism of D-amino acids, whose function and distribution in Archaea are as yet unclear, appears to be of bacterial origin in *Pyrococcus.* Some of the peptidases (Fig. 3) are restricted to *Pyrococcus* and to various bacteria dispersed in a wide phylogenetic spectrum but excluding the hyperthermophilic bacteria, *A. aeolicus* and *T. maritima*. Such a phylogenetic distribution may reflect a lateral gene transfer from Bacteria as the alternative hypothesis supposes multiple and massive gene losses in Crenarchaeota, Euryarchaeota and hyperthermophilic Bacteria. In conclusion,

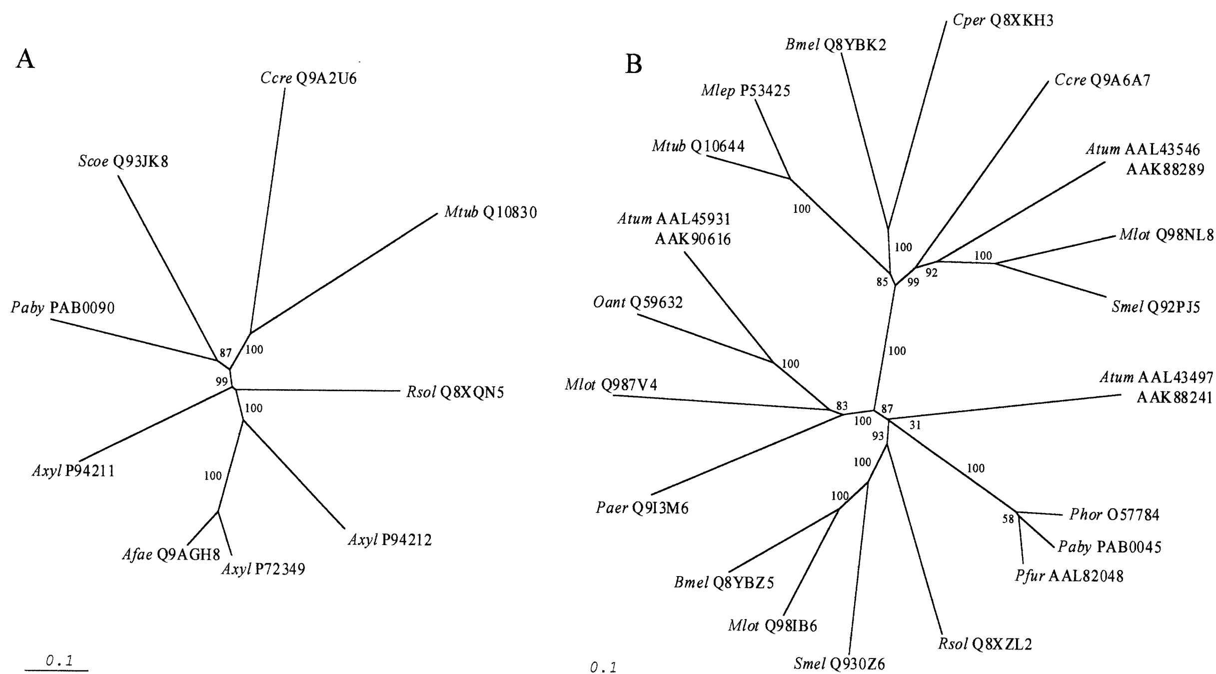

Fig. 3. Phylogenetic trees for proteins of a D-aminoacylase (A) and a D-aminopeptidase (B) family including PAB0090 and PAB0045 respectively. Phylogenetic trees are based on DbClustal (Galtier *et al.*, 1996) multiple alignments and constructed with the neighbor-joining method using the PHYLO_WIN package (Thompson *et al.*, 2000). Bootstrap values are indicated for 100 replicates. For each sequence, the organism name is indicated using an abbreviated form followed by the accession number. Abbreviations alphabetically listed: *Afae*: *Alcaligenes faecalis*, *Atum*: *Agrobacterium tumefaciens*, *Axyl*: *Alcaligenes xylosoxydans*, *Bmel*: *Brucella melitensis*, *Ccre*: *Caulobacter crescentus*, *Cper*: *Clostridium perfringens*, *Mlep*: *Mycobacterium leprae*, *Mlot*: *Mesorhizobium loti*, *Mtub*: *Mycobacterium tuberculosis*, *Oant*: *Ochrobactrum anthropi*, *Paby*: *P. abyssi*, *Paer*: *Pseudomonas aeruginosa*, *Pfur*: *P. furiosus*, *Phor*: *P. horikoshii*, *Rsol*: *Ralstonia solanacearum*, *Scoe*: *Streptomyces coelicolor*, *Smel*: *Sinorhizobium meliloti*.

our analysis reveals that many genes directly linked to the heterotrophic metabolism of *P. abyssi* have a complex and singular evolutionary history in the Euryarchaeota kingdom. This may confer to *Pyrococcus* the ability to inhabit variable environments by transporting and metabolising a wide range of organic substrates.

Thermophily adaptation and gene transfer

Whereas the transfer of hyperthermophilic proteins from Bacteria to Archaea and *vice versa* is now well established, it is not known whether or not proteins from mesophilic or moderately thermophilic organisms can be successfully transferred to hyperthermophiles. Such possibility cannot be dismissed, as many bacterial proteins identified in the *Pyrococcus* genome are presently only known in mesophilic bacteria. It has been shown that structural adaptation of proteins to higher temperatures is accompanied by sequence composition modification at the genome scale level (Cambillau and Claverie, 2000). Hyperthermophilic organisms such as *pyrococci*, exhibit large proportion differences of charged versus polar (noncharged) amino acids, as compared to moderate thermophiles and mesophiles (Danson and Hough, 1998; Yip *et al.*, 1998). We thus use this signature to scrutinise the thermophilic fitting of genes supposedly acquired by genetic transfer from mesophilic organisms. Figure 4 displays in parallel the distribution of bacterial related genes and their charged-polar (CP) amino acids content along the *P. abyssi* chromosome. The CP distribution shows a featureless alternation of high (good thermophilic adaptation) and low (poor adaptation) values. The highest CP region (PAB2137-PAB2119, point A in Fig. 4) coincides exactly with the main ribosomal cluster, a set of most highly expressed genes. Low value points of the CP distribution (points B, C, D, G in Fig. 4) on the contrary, should represent recently acquired 'non-thermophilic' genes, and indeed in each case these regions correspond to bacterial-like clusters. This could indicate a direct lateral gene transfer from mesophilic bacteria to *Pyrococcus*. Most of them belong to transport and amino acids metabolism or energy metabolism (B: PAB00286-0243 aro/shikimate pathway; C: PAB1902-0498 ion transport/energy metabolism; D: PAB1850-0538 proline or nucleotide transport; G: PAB1411-1389 hydrogenase-4). Except for the latter, these clusters of gene might represent recently acquired functions conferring new (non-essential) prototrophic phenotypes. The hydrogenase four

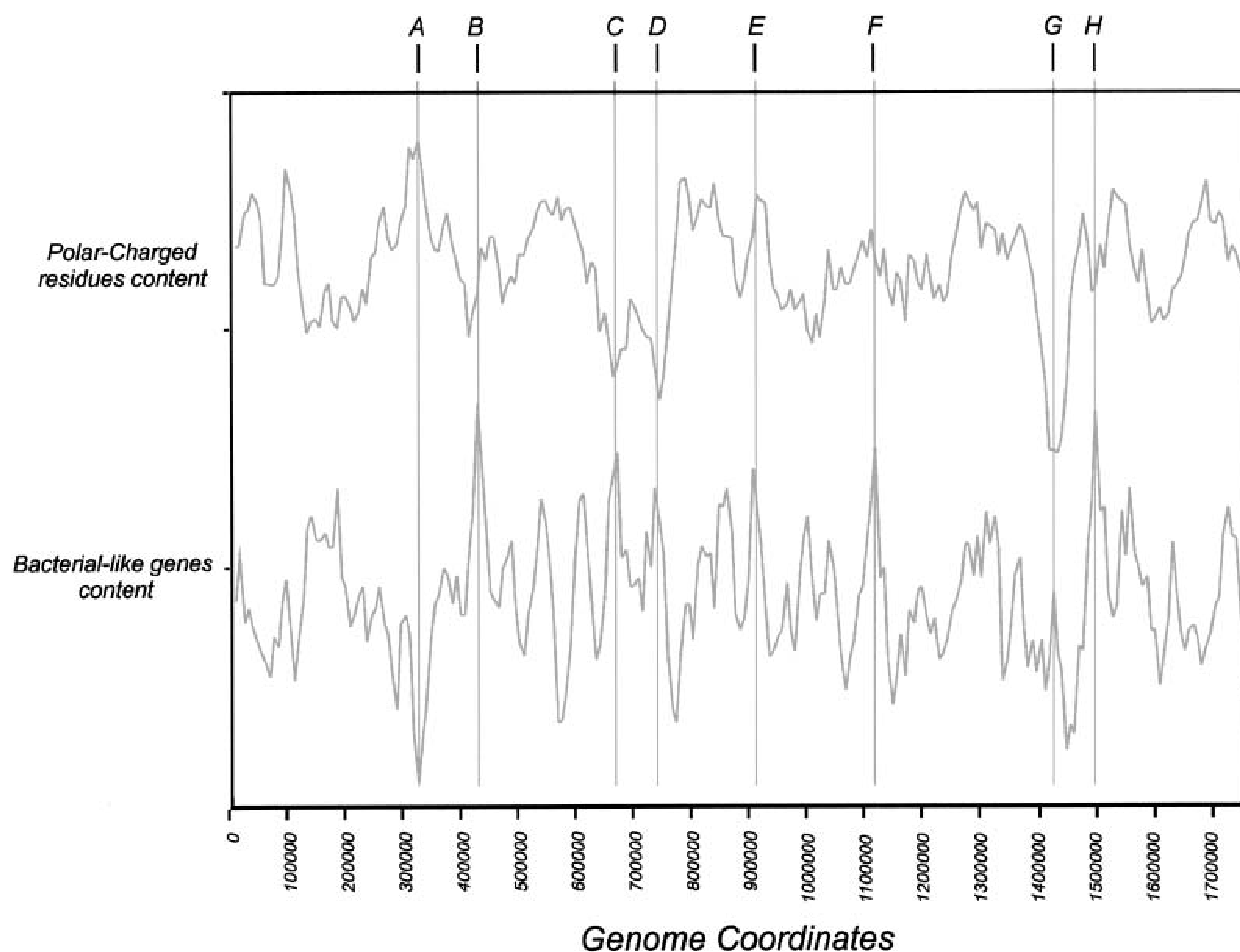

Fig. 4. Distribution of polar-charged residues as a function of genome co-ordinates
Upper: The difference of percentage of charged (K,R,D,E) and polar amino (N,Q,S,T) acids is calculated for each gene, and the averaged value over a fixed 1/50 th genome length window and plotted as function of 1/240th genome length increments (grey curve). Lower: distribution of the number of genes most closely related to bacteria (see *Experimental procedures*). Values are averaged over a fixed 1/50 th genome length window and plotted as function of 1/240th genome length increments as above (red curve). Points A to F are discussed in the text.

cluster (PAB 1396–1389; point G in Fig. 4) is unexpected, as it belongs to the hydrogen metabolism pathway, an important component of sulphur metabolizing organisms, and should therefore be well adapted to thermophily. However, this operon seems partly duplicated in *P. abyssi* with an hydrogenase related operon (PAB1894-1888), that could represent an ancestral counterpart in this pathway. Nevertheless, there are cases where bacteria-like clusters exhibit good thermophilic adaptation, as exemplified by points F (PAB0764-0790 cell envelope/UDPGP glycosyl transferase) and H (PAB1348-1027 ABC transporter/chemotaxis) in Fig. 4. This could correspond to a pool of ancestral genes, common to bacteria and archaea, that have adapted to their host during evolution, or alternatively, the good CP value of these genes might indicate that gene transfer has taken place in the opposite direction, i.e. from thermophiles to mesophiles. It should be interesting to check *in vitro* if CP value indeed correlates with thermophilicity of the corresponding proteins that could be either in the process of thermoadaptation. In addition, one cannot dismiss the possibility that some mesophilic proteins could be useful for *Pyrococcus* in specific environmental conditions as it is known that *P. abyssi* can survive for long period at low tem-perature, becoming relatively oxygen tolerant (Erauso *et al.*, 1993).

Conclusion

The sequencing of three *Pyrococcus* genomes in the late years of the last century explains why they are presently the most studied anaerobic hyperthermophiles (668 entries in Medline October 20, 2002). The *P. abyssi* genome has been completely re-annotated, taking into

account the most recent biochemical and structural data available.

As a result, several new functions for both informational and operational proteins have been proposed. In addition, careful sequence analysis in combination with recent biochemical data has provided an almost complete map of the key metabolic pathways: (i) the variant Embden-Meyerhof pathway with several unique carbohydrate-converting enzymes; (ii) the archaeal-type network of amino acid fermentation; (iii) the anticipated role of ferredoxin as main electron carrier to maintain the intracellular redox balance; (iv) the anticipated but yet to be verified protein complexes that may generate an electrochemical potential across the cytoplasmic membrane (identification suggesting anaerobic respiration as energy-transducing system in addition to the well established substrate-level phosphorylation); and (v) most of the anticipated amino acid and nucleotide biosynthesis pathways (Fig. 1, Table 2). This study should serve as a basis for designing post-genomic programs with the final aim to get a full understanding of the proteome encoded by the relatively small genome of *P. abyssi*.

Comparative genomic analysis has shown that about 80% of the *Pyrococcus* proteins (both informational and operational) have closer relatives in Archaea. This indicates the existence of an archaeal core of orthologous proteins that becomes only evident after the sequencing of a sufficiently high number of archaeal genomes. This fits well with the finding that the 'core' of conserved genes useful to construct whole genome trees also includes operational proteins (Daubin *et al*., 2002). However, *Pyrococcus* genome also contains genes that have been exchanged with other Archaea and Bacteria living in the same environment. Furthermore, identification of 'mesophilic signatures' in amino-acid composition in certain regions of the *P. abyssi* genome suggests recent transfer from mesophilic bacteria to this hyperthermophilic Archaea, expending the possibilities for gene sharing between the two prokaryotic domains.

Experimental procedures

Sequencing of the 1765 118 bp-long genome of *P. abyssi* was performed using a pairwise global approach. A plasmid library, of 5–6 kb insert size, was constructed in pBAM3 vector (a Bluescript derivative) after partial digestion of genomic DNA by CviJI (PuG/CPy), followed by agarose gel purification. Sequencing at both ends of inserts on a Licor 4200 type of sequence analyser produced 19 300 sequences (mean size 900 bp), corresponding to 10 × coverage. In order to extend the scaffolding over the most part of the genome, 1635 end sequences of cosmid clones (38–46 kb) were added. The cosmid library was constructed by *Sau*3AI partial digestion of genomic DNA prepared in agarose plugs, followed by gel purification and ligation to the cosmid arms of the cosRH3 vector (Heilig, unpublished), before encapsidation. Assembly, using PHRED and PHRAP software (Ewing and Green, 1998; Ewing *et al*., 1998), resulted in 69 contigs which ranged in size between 1.5 and 215 kb (mean 25.6 kb), covering more than 95% of the *P. abyssi* genome and scaffolded into 12 supercontigs. Gap filling between contigs was performed by primer walking using the linking subclones, and between supercontigs by sequencing PCR products or by direct genomic sequencing. Polishing reactions (3200) were performed to ensure an error rate $<10^{-4}$ and to conform to the Bermuda rules. Validation of the overall assembly was accomplished first by comparing the predicted versus the experimental *Not*I and *Asc*I maps of the whole *P. abyssi* genome, second by the fingerprinting (six restriction enzymes) of 4O cosmid clones covering about 65% of the genome, completed by sizing PCR products for the remaining part.

Compared to initial genome sequence, most gene's descriptions have been re-examined and corrected in order to conform to new experimental data. Similarity searches were performed using BLASTP tool (Altschul *et al*., 1997) in public databases, benefiting from many new complete genomic sequences made available – mostly from the bacteria domain – as well as numerous additions of data not derived from genome sequencing projects.

For each *P. abyssi* gene, the CP index (charged–polar residues) is calculated as the ratio of percentage of charged (K,R,D,E) to polar amino (N,Q,S,T) acids, normalized by taking the difference of this value with the average *P. abyssi* CP value. Closest phylogenetical relatives for each gene were determined as follows: all gene sequences were blasted against the most recent NR database (NCBI) using as threshold an expect value of P^2 1^{e-05}. Matched hits, ordered by increasing *P*-values, were tagged according to the taxonomic phylum of the organism from which they originated, yielding a pattern of phylogenetical best relatives (archaea, bacteria or eucarya), for each *P. abyssi* gene. Thus, *P. abyssi* 'Bacterial-like' genes are those for which the phylogenetical domain pattern is: archaea > bacteria hits only, or archaea > bacteria > eucarya hits.

Supplementary material

The following material is available from http://www.blackwellpublishing.com/products/journals/suppmat/mole/mole3381/mmi3381sm.htm and also http://www.archbac.u-psud.fr/genomes/newpab/supplementary.html

Further analysis – DNA replication, chromosome segregation and cell division; Transcription; Motility; Amino acid biosynthesis; Nucleotide synthesis; Vitamin biosynthesis; Isoprenoid biosynthesis and utilization.

Fig. S1. *Pyrococcus abyssi* trinucleotides bias.

References

Adams, M.W.W. (1999) The biochemical diversity of life near and above 100 degrees C in marine environments. *J Appl Microbiol* **85** (Suppl. S): 108S–117S.

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.

Aravind, L., Tatusov, R.L., Wolf, Y.I., Walker, D.R., and Koonin, E.V. (1998) Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. *Trends Genet* **14:** 442–444.

Aravind, L., Tatusov, R.L., Wolf, Y.I., Walker, D.R., and Koonin, E.V. (1999) Reply. *Trends Genet* **15:** 299–300.

Bell, S.D., and Jackson, S.P. (2001) Mechanism and regulation of transcription in archaea. *Curr Opin Microbiol* **4:** 208–213.

Berthet-Colominas, C., Seignovert, L., Hartlein, M., Grotli, M., Cusack, S., and Leberman, R. (1998) The crystal structure of asparaginyl-tRNA synthetase from *Thermus thermophilus* and its complexes with ATP and asparaginyl-adenylate: the mechanism of discrimination between asparagine and aspartic acid. *EMBO J* **17:** 2947–2960.

Bocquier, A.A., Liu, L., Cann, I.K., Komori, K., Kohda, D., and Ishino, Y. (2001) Archaeal primase. bridging the gap between RNA and DNA polymerases. *Curr Biol* **11:** 452–456.

Brinkman, A.B., Dahlke, I., Tuininga, J.E., Lammers, T., Dumay, V., de Heus, E., *et al.* (2000) An Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus* is negatively autoregulated. *J Biol Chem* **275:** 38160–38169.

Brinkman, A.B., Bell, S.D., Lebbink, R.J., De Vos, W.M., and Van der Oost, J. (2002) The *Sulfolobus solfataricus* Lrp-like protein LysM regulates lysine biosynthesis in response to lysine avalability. *J Biol Chem* **277:** 29537–29549.

Cambillau, C., and Claverie, J.M. (2000) Structural and genomic correlates of hyperthermostability. *J Biol Chem* **275:** 32383–32386.

Cann, I.K.O., and Ishino, Y. (1999) Archaeal DNA replication: Identifying the pieces to solve a puzzle. *Genetics* **152:** 1249–1267.

Cann, I.K.O., Komori, K., Toh, H., Kanai, S., and Ishino, Y. (1998) A heterodimeric DNA polymerase: Evidence that members of Euryarchaeota possess a distinct DNA polymerase. *Proc Natl Acad Sci USA* **95:** 14250–14255.

Cann, I.K., Ishino, S., Hayashi, I., Komori, K., Toh, H., Morikawa, K., and Ishino, Y. (1999) Functional interactions of a homolog of proliferating cell nuclear antigen with DNA polymerases in Archaea. *J Bacteriol* **181:** 6591–6599.

Cann, I.K., Ishino, S., Yuasa, M., Daiyasu, H., Toh, H., and Ishino, Y. (2001) Biochemical Analysis of Replication Factor C from the Hyperthermophilic Archaeon *Pyrococcus furiosus*. *J Bacteriol* **183:** 2614–2623.

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.

Constantinesco, F., Forterre, P., and Elie, C. (2002) NurA, a novel 5′-3′ nuclease gene linked to rad50 and mre11 homologs of thermophilic Archaea. *EMBO Rep* **3:** 537–542.

Danson, M.J., and Hough, D.W. (1998) Structure, function and stability of enzymes from the Archaea. *Trends Microbiol* **6:** 307–314.

Daubin, V., Gouy, M., and Perrière, G. (2002) A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. *Genome Res* **12:** 1080–1090.

De Vos, W.M., Voorhorst, W.G., Dijkgraaf, M., Kluskens, L.D., Van der Oost, J., and Siezen, R.J. (2001) Purification, characterization, and molecular modeling of pyrolysin and other extracellular thermostable serine proteases from hyperthermophilic microorganisms. *Methods Enzymol* **330:** 383–393.

Delaney, J.R., McDuff, R.E., and Lupton, J.E. (1984) Hydrothermal fluid temperatures of 400 degrees C on the Endeavour Segment, northern Juan de Fuca. *Eos, Transactions, Am Geophys Union* **65:** 973.

Dennis, P.P., Omer, A., and Lowe, T. (2001) A guided tour: small RNA function in Archaea. *Mol Microbiol* **40:** 509–519.

Dib, R., Chobert, J.M., Dalgalarrondo, M., Barbier, G., and Haertle, T. (1998) Purification, molecular properties and specificity of a thermoactive and thermostable proteinase from *Pyrococcus abyssi*, strain st 549, hyperthermophilic archaea from deep-sea hydrothermal ecosystem. *Febs Lett* **431:** 279–284.

DiRuggiero, J., Santangelo, N., Nackerdien, Z., Ravel, J., and Robb, F.T. (1997) Repair of extensive ionizing-radiation DNA damage at 95 degrees C in the hyperthermophilic archaeon *Pyrococcus furiosus*. *J Bacteriol* **179:** 4643–4645.

Doolittle, W.F. (1999) Phylogenetic classification and the universal tree. *Science* **284:** 2124–2129.

Enzlin, J.H., and Scharer, O.D. (2002) The active site of the DNA repair endonuclease XPF-ERCC1 forms a highly conserved nuclease motif. *EMBO J* **21:** 2045–2053.

Erauso, G., Charbonnier, F., Barbeyron, T., Forterre, P., and Prieur, D. (1992) Caractérisation préliminaire d'une archaebactérie hyperthermophile possédant un plasmide, isolée d'une source hydrothermale du bassin Nord Fidjien. *C R Acad Sci Paris* **314:** 387–393.

Erauso, G., Reysenbach, A.L., Godfroy, A., Meunier, J.-R., Crump, B., Partensky, F., *et al.* (1993) *Pyrococcus abyssi* sp. nov., a new hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. *Arch Microbiol* **160:** 338–349.

Erauso, G., Godfroy, A., Raguénès, G., and Prieur, D. (1995) Plate cultivation technique for strictly anaerobic, thermophilic, sulfur-metabolizing archae. In *Protocol for Archaebacterial Research*. Fleischmann, E.M., Place, A.R., Robb, F.T and Scheir, H.J., (eds). Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, pp. 25–29.

Ettema, T.J., Van der Oost, J., and Huynen, M.A. (2001) Modularity in the gain and loss of genes in the *Pyrococci* genus. *Trends Genet* **17:** 485–487.

Ewing, B., and Green, P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8:** 186–194.

Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8:** 175–185.

Forterre, P., Bouthier, D.L.T.C., Philippe, H., and Duguet, M. (2000) Reverse gyrase from hyperthermophiles: probable transfer of a thermoadaptation trait from archaea to bacteria. *Trends Genet* **16:** 152–154.

Galperin, M.Y., Aravind, L., and Koonin, E.V. (2000) Aldola-

ses of the DhnA family: a possible solution to the problem of pentose and hexose biosynthesis in archaea. *FEMS Microbiol Lett* **183:** 259–264.

Galtier, N., Gouy, M., and Gautier, C. (1996) SEAVIEW and PHYLO_WIN. two graphics tools for sequence alignment and molecular phylogeny. *Comput Appl Biosci* **12:** 543–548.

Gaspin, C., Cavaille, J., Erauso, G., and Bachellerie, J.P. (2000) Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the *Pyrococcus* genomes. *J Mol Biol* **297:** 895–906.

Gerard, E., Labedan, B., and Forterre, P. (1998) Isolation of a minD-like gene in the hyperthermophilic archaeon *Pyrococcus* AL585, and phylogenetic characterization of related proteins in the three domains of life. *Gene* **222:** 99–106.

Gerard, E., Jolivet, E., Prieur, D., and Forterre, P. (2001) DNA protection mechanisms are not involved in the radioresistance of the hyperthermophilic archaea *Pyrococcus abyssi* and *P. furiosus*. *Mol Genet Genomics* **266:** 72–78.

Godfroy, A., Raven, N.D.H., and Sharp, R.J. (2000) Physiology and continuous culture of the hyperthermophilic deep-sea vent archaeon *Pyrococcus abyssi* ST549. *FEMS Microbiol Lett* **186:** 127–132.

Gonzalez, J.M., Masuchi, Y., Robb, F.T., Ammerman, J.W., Maeder, D.L., Yanagibayashi, M., *et al.* (1998) *Pyrococcus horikoshii* sp. nov., a hyperthermophilic archaeon isolated from a hydrothermal vent at the Okinawa Trough. *Extremophiles* **2:** 123–130.

Gueguen, Y., Rolland, J.J.L., Lecompte, O., Azam, P., Le, R.G., Flament, D., *et al.* (2001) Characterization of two DNA polymerases from the hyperthermophilic euryarchaeon *Pyrococcus abyssi*. *Eur J Biochem* **268:** 5961–5969.

Hanzelka, B.L., Darcy, T.J., and Reeve, J.N. (2001) TFE, an archaeal transcription factor in *Methanobacterium thermoautotrophicum* related to eucaryal transcription factor TFIIEalpha. *J Bacteriol* **183:** 1813–1818.

Hase, C.C., Fedorova, N.D., Galperin, M.Y., and Dibrov, P.A. (2001) Sodium ion cycle in bacterial pathogens: evidence from cross-genome comparisons. *Microbiol Mol Biol Rev* **65:** 353–370.

Hausner, W., Lange, U., and Musfeldt, M. (2000) Transcription factor S, a cleavage induction factor of the archaeal RNA polymerase. *J Biol Chem* **275:** 12393–12399.

Hayashi, I., Morikawa, K., and Ishino, Y. (1999) Specific interaction between DNA polymerase II (PolD) and RadB, a Rad51/Dmc1 homolog. *Pyrococcus furiosus*. *Nucleic Acids Res* **27:** 4695–4702.

Henneke, G., Gueguen, Y., Flament, D., Azam, P., Querellou, J., Dietrich, J., *et al.* (2002) Replication factor C from the hyperthermophilic Archaeon *Pyrococcus abyssi* does not need ATP hydrolysis for clamp-loading and contains a functionally conserved RFC PCNA-binding domain. *J Mol Biol* **323:** 795–810.

Hiramatsu, T., Kodama, K., Kuroda, T., Mizushima, T., and Tsuchiya, T. (1998) A putative multisubunit Na^+/H^+ antiporter from *Staphylococcus aureus*. *J Bacteriol* **180:** 6642–6648.

Hopfner, K.P., Karcher, A., Craig, L., Woo, T.T., Carney, J.P., and Tainer, J.A. (2001) Structural biochemistry and interaction architecture of the DNA double-strand break repair Mre11 nuclease and Rad50-ATPase. *Cell* **105:** 473–485.

Huynen, M.A., Dandekar, T., and Bork, P. (1999) Variation and evolution of the citric-acid cycle: a genomic perspective. *Trends Microbiol* **7:** 281–291.

Hwang, K.Y., Chung, J.H., Kim, S.H., Han, Y.S., and Cho, Y. (1999) Structure-based identification of a novel NTPase from *Methanococcus jannaschii*. *Nat Struct Biol* **6:** 691–696.

Jordan, I.K., Kondrashov, F.A., Rogozin, I.B., Tatusov, R.L., Wolf, Y.I., and Koonin, E.V. (2001) Constant relative rate of protein evolution and detection of functional diversification among bacterial, archaeal and eukaryotic proteins. *Genome Biol* **2:** 1–9.

Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., *et al.* (1998) Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3 (Suppl.). *DNA Res* **5:** 147–155.

Komori, K., and Ishino, Y. (2001) Replication protein A in *Pyrococcus furiosus* is involved in homologous DNA recombination. *J Biol Chem* **276:** 25654–25660.

Komori, K., Miyata, T., DiRuggiero, J., Holley-Shanks, R., Hayashi, I., Cann, I.K., *et al.* (2000a) Both RadA and RadB are involved in homologous recombination in *Pyrococcus furiosus*. *J Biol Chem* **275:** 33782–33790.

Komori, K., Sakae, S., Fujikane, R., Morikawa, K., Shinagawa, H., and Ishino, Y. (2000b) Biochemical characterization of the hjc holliday junction resolvase of *Pyrococcus furiosus*. *Nucleic Acids Res* **28:** 4544–4551.

Koning, S.M., Elferink, M.G., Konings, W.N., and Driessen, A.J. (2001) Cellobiose uptake in the hyperthermophilic archaeon *Pyrococcus furiosus* is mediated by an inducible, high-affinity ABC transporter. *J Bacteriol* **183:** 4979–4984.

Koonin, E.V., Wolf, Y.I., and Aravind, L. (2001) Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach. *Genome Res* **11:** 240–252.

Kyrpides, N.C., and Woese, C.R. (1998) Archaeal translation initiation revisited: the initiation factor 2 and eukaryotic initiation factor 2B alpha-beta-delta subunit families. *Proc Natl Acad Sci USA* **95:** 3726–3730.

Kyrpides, N.C., Olsen, G.J., and Woese, C.R. (1999) Archaeal and bacterial hyperthermophiles: horizontal gene exchange or common ancestry? *Trends Genet* **15:** 298–299.

Ladrat, C., Alayse-Danet, A.-M., Dietrich, J., and Barbier, G. (1996) A new thermostable glucose activated b-glucosidase from the hyperthermophilic marine archaebacterium *Pyrococcus abyssi:* purification and characterization. *J Mar Biotechnol* **4:** 192–199.

Langer, D., Hain, J., Thuriaux, P., and Zillig, W. (1995) Transcription in archaea: similarity to that in eucarya. *Proc Natl Acad Sci USA* **92:** 5768–5772.

Lecompte, O., Ripp, R., Puzos-Barbe, V., Duprat, S., Heilig, R., Dietrich, J., *et al.* (2001) Genome evolution at the genus level: comparison of three complete genomes of hyperthermophilic archaea. *Genome Res* **11:** 981–993.

Liu, X.Q. (2000) Protein spling intein: genetic mobility, origin and evolution. *Annu Rev Genet* **34:** 61–76.

Liu, L., Komori, K., Ishino, S., Bocquier, A.A., Cann, I.K., Kohda, D., and Ishino, Y. (2001) The archaeal DNA primase: biochemical characterization of the p41-p46 complex from *Pyrococcus furiosus*. *J Biol Chem* **276:** 45484–45490.
Lucas, S., Toffin, L., Zivanovic, Y., Charlier, D., Moussard, H., Forterre, P., *et al.* (2002) Construction of a shuttle vector and spheroplasts transformation of the hyperthermophilic archaeon *Pyrococcus abyssi*. *Appl Environ Microbiol* **68:** 2258–5536.
MacNeill, S.A. (2001) Understanding the enzymology of archaeal DNA replication: progress in form and function. *Mol Microbiol* **40:** 520–529.
Maeder, D.L., Weiss, R.B., Dunn, D.M., Cherry, J.L., Gonzalez, J.M., DiRuggiero, J., and Robb, F.T. (1999) Divergence of the hyperthermophilic archaea *Pyrococcus furiosus* and *P. horikoshii* inferred from complete genomic sequences. *Genetics* **152:** 1299–1305.
Mai, X., and Adams, M.W. (1996) Purification and characterization of two reversible and ADP-dependent acetyl coenzyme A synthetases from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J Bacteriol* **178:** 5897–5903.
Makino, S., Suzuki, M., Zivanovic, Y., Myllykallio, H., and Forterre, P. (2001) Bacterial genomic reorganization upon DNA replication. *Science* **292:** 803a.
Marck, C., and Grosjean, H. (2002) tRNomics. Analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA* **8:** 1189–1232.
Marguet, E., and Forterre, P. (1994) DNA stability at temperatures typical for hyperthermophiles. *Nucleic Acids Res* **22:** 1681–1686.
Marteinsson, V.T., Watrin, L., Prieur, D., Caprais, J.C., Raguenes, G., and Erauso, G. (1995) Phenotypic characterization, DNA similarities, and protein profiles of twenty sulfur-metabolizing hyperthermophilic anaerobic archaea isolated from hydrothermal vents in the southwestern Pacific Ocean. *Int J Syst Bacteriol* **45:** 623–632.
Marteinsson, V.T., Moulin, P., Birrien, J.L., Gambacorta, A., Vernet, M., and Prieur, D. (1997) Physiological responses to stress conditions and barophilic behavior of the hyperthermophilic vent archaeon *Pyrococcus abyssi*. *Appl Environ Microbiol* **63:** 1230–1236.
Matsuda, T., Morikawa, M., Haruki, M., Higashibata, H., Imanaka, T., and Kanaya, S. (1999) Isolation of TBP-interacting protein (TIP) from a hyperthermophilic archaeon that inhibits the binding of TBP to TATA-DNA. *FEBS Lett* **457:** 38–42.
Matsunaga, F., Forterre, P., Ishino, Y., and Myllykallio, H. (2001) *In vivo* interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. *Proc Natl Acad Sci USA* **98:** 11152–11157.
Matsunaga, F., Norais, C., Forterre, P., and Myllykallio, H. (2003) Identification of short eukaryotic Okazaki fragments synthesized from a prokaryotic replication origin. *EMBO reports* (in press).
Matte-Tailliez, O., Brochier, C., Forterre, P., and Philippe, H. (2002) Archaeal phylogeny based on ribosomal proteins. *Mol Biol Evol* **19:** 631–639.
Mayer, C., Suck, D., and Poch, O. (2001) The archaeal homolog of the Imp4 protein, a eukaryotic U3 snoRNP component. *Trends Biochem Sci* **26:** 143–144.
Mojica, F.J., Diez-Villasenor, C., Soria, E., and Juez, G. (2000) Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol* **36:** 244–246.
Morales, A.J., Swairjo, M.A., and Schimmel, P. (1999) Structure-specific tRNA-binding protein from the extreme thermophile *Aquifex aeolicus*. *EMBO J* **18:** 3475–3483.
Myllykallio, H., Lopez, P., Lopez-Garcia, P., Heilig, R., Saurin, W., Zivanovic, Y., *et al.* (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.
Natale, D.A., Galperin, M.Y., Tatusov, R.L., and Koonin, E.V. (2000) Using the COG database to improve gene recognition in complete genomes. *Genetica* **108:** 9–17.
Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., *et al.* (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399:** 323–329.
d'Orval, B.C., Bortolin, M.L., Gaspin, C., and Bachellerie, J.P. (2001) Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNATrp intron guides the formation of two ribose-methylated nucleosides in the mature tRNATrp. *Nucleic Acids Res* **29:** 4518–4529.
Poterszman, A., Delarue, M., Thierry, J.C., and Moras, D. (1994) Synthesis and recognition of aspartyl-adenylate by *Thermus thermophilus* aspartyl-tRNA synthetase. *J Mol Biol* **244:** 158–167.
Prangishvili, D., Klenk, H.P., Jakobs, G., Schmiechen, A., Hanselmann, C., Holz, I., and Zillig, W. (1998) Biochemical and phylogenetic characterization of the dUTPase from the archaeal virus SIRV. *J Biol Chem* **273:** 6024–6029.
Purcarea, C., Erauso, G., Prieur, D., and Herve, G. (1994) The catalytic and regulatory properties of aspartate transcarbamoylase from *Pyrococcus abyssi*, a new deep-sea hyperthermophilic archaeobacterium. *Microbiology* **140:** 1967–1975.
Robb, F.T., Maeder, D.L., Brown, J.R., DiRuggiero, J., Stump, M.D., Yeh, R.K., *et al.* (2001) Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: implications for physiology and enzymology. *Methods Enzymol* **330:** 134–157.
Sanchez, L.B., Galperin, M.Y., and Muller, M. (2000) Acetyl-CoA synthetase from the amitochondriate eukaryote *Giardia lamblia* belongs to the newly recognized superfamily of acyl-CoA synthetases (Nucleoside diphosphate-forming). *J Biol Chem* **275:** 5794–5803.
Sandigursky, M., and Franklin, W.A. (2000) Uracil-DNA glycosylase in the extreme thermophile *Archaeoglobus fulgidus*. *J Biol Chem* **275:** 19146–19149.
Schut, G.J., Zhou, J., and Adams, M.W. (2001) DNA microarray analysis of the hyperthermophilic archaeon *Pyrococcus furiosus*: evidence for anNew type of sulfur-reducing enzyme complex. *J Bacteriol* **183:** 7027–7036.
Sicheritz-Ponten, T., and Andersson, S.G. (2001) A phylogenomic approach to microbial evolution. *Nucleic Acids Res* **29:** 545–552.

Silva, P.J., d. B.E.C., Wassink, H., de Haaker, H.C.B., Robb, F.T., and Hagen, W.R. (2000) Enzymes of hydrogen metabolism in *Pyrococcus furiosus*. *Eur J Biochem* **267:** 6541–6551.

Tatusov, R.L., Koonin, E.V., and Lipman, D.J. (1997) A genomic perspective on protein families. *Science* **278:** 631–637.

Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., *et al.* (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* **29:** 22–28.

Thomm, M. (1996) Archaeal transcription factors and their role in transcription initiation. *FEMS Microbiol Rev* **18:** 159–171.

Thompson, J.D., Plewniak, F., Thierry, J., and Poch, O. (2000) DbClustal: rapid and reliable global multiple alignments of protein sequences detected by database searches. *Nucleic Acids Res* **28:** 2919–2926.

Tumbula, D.L., Becker, H.D., Chang, W.Z., and Soll, D. (2000) Domain-specific recruitment of amide amino acids for protein synthesis. *Nature* **407:** 106–110.

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.

Verhees, C.H. (2002) *Molecular characterization of glycolysis in* Pyrococcus furiosus. Thesis Wageningen University.

Vijayvargia, R., and Biswas, I. (2002) MutS2 family protein from *Pyrococcus furiosus*. *Curr Microbiol* **44:** 224–228.

Ward, D.E., Kengen, S.W., Van der Oost, J., and De Vos, W.M. (2000) Purification and characterization of the alanine aminotransferase from the hyperthermophilic Archaeon *Pyrococcus furiosus* and its role in alanine production. *J Bacteriol* **182:** 2559–2566.

Watrin, L., Lucas, S., Purcarea, C., Legrain, C., and Prieur, D. (1999) Isolation and characterization of pyrimidine auxotrophs, and molecular cloning of the pyrE gene from the hyperthermophilic archaeon *Pyrococcus abyssi*. *Mol Gen Genet* **262:** 378–381.

Watrin, L., and Prieur, D. (1996) UV and ethyl methanesulfonate effects in hyperthermophilic archaea and isolation of auxotrophic mutants of *Pyrococcus* strains. *Curr Microbiol* **33:** 377–382.

Watrin, L., and Prieur, D. (1998) Puromycin-resistant mutants of the hyperthermophilic archaeon *Pyrococcus abyssi*. *J Mar Biotechnol* **6:** 124–128.

Watrin, L., Martin-Jézéquel, V., and Prieur, D. (1995) Amino acid requirements of *Pyrococcus abyssi*, strain GE5. *Appl Environ Microbiol* **61:** 1138–1140.

Woese, C.R. (1996) Whither microbiology? Phylogenetic trees. *Curr Biol* **6:** 1060–1063.

Yip, K.S., Britton, K.L., Stillman, T.J., Lebbink, J., De Vos, W.M., Robb, F.T., *et al.* (1998) Insights into the molecular basis of thermal stability from the analysis of ion-pair networks in the glutamate dehydrogenase family. *Eur J Biochem* **255:** 336–346.

Zappa, S., Rolland, J.L., Flament, D., Gueguen, Y., Boudrant, J., and Dietrich, J. (2001) Characterization of a highly thermostable alkaline phosphatase from the euryarchaeon *Pyrococcus abyssi*. *Appl Environ Microbiol* **67:** 4504–4511.

Zivanovic, Y., Lopez, P., Philippe, H., and Forterre, P. (2002) *Pyrococcus* genome comparison evidences chromosome shuffling-driven evolution. *Nucleic Acids Res* **30:** 1902–1910.