System. Appl. Microbiol. 19, 312–321 (1996)

An *infB*-Homolog in *Sulfolobus acidocaldarius*

PATRICK J. KEELING[1], SANDRA L. BALDAUF[1], W. FORD DOOLITTLE[1], WOLFRAM ZILLIG[2], and HANS-PETER KLENK[2]

[1] Dalhousie University, Department of Biochemistry, Halifax, Nova Scotia, Canada B3H 4H7
[2] Max-Planck-Institut für Biochemie, 82152 Martinsried, Germany

Received February 13, 1996

Summary

We have identified an archaeal homologue of the bacterial translation initiation factor 2 (IF-2 or *inf*B) in a partial open reading frame situated upstream of the gene cluster coding for the large subunits of the DNA-dependent RNA polymerase (RNAP) in *Sulfolobus acidocaldarius*. Based on this similarity, a larger genomic clone of this region was isolated and sequenced. Although the putative *Sulfolobus* translation factor gene is highly similar to *inf*B, it shares an even higher degree of similarity with the recently described *FUN12* gene from *Saccharomyces cerevisiae*. Phylogenetic trees inferred from sequences of homologous translation initiation, elongation and termination factors confirm that both the new *Sulfolobus* gene and yeast *FUN12* are members of the IF-2 family and that the root of the IF-2 subtree determined within a 3-fold rooted universal treee of IF-2, EF-1α/Tu and EF-2/G strongly supports a close phylogenetic relationship between the archaea and the eukaryotes. The genomic context of the *Sulfolobus infB* also reveals links between two highly conserved bacterial gene clusters, the RNAP operon and the *nus*A-*inf*B operon. In bacteria these operons are not linked, but the location of the *Sulfolobus infB*- and *nus*A-homologues immediately upstream and downstream of the RNAP gene cluster, respectively, links the two conserved bacterial operons and may indicate an ancient genome reorganization.

Key words: Translation factors – Translation initiation – *Sulfolobus acidocaldarius* – Archaea – Phylogeny – Multiple gene duplication – Genome organization – *nus*A-*inf*B-operon

Introduction

The nucleotide sequence surrounding the genes coding for the largest subunits of the DNA-dependent RNA polymerase (RNAP) of *Sulfolobus acidocaldarius* (*Pühler* et al., 1989) contains eight ORFs, five of which were functionally identified during primary analysis. These are *rpo*B, *rpo*A1 and *rpo*A2, encoding the three largest RNAP subunits B, A′ and A″, respectively, and ORF118 (*rps*12) and ORF104 (*rpl*30), encoding ribosomal proteins. Two of the remaining ORFs were later characterized by routine database comparisons. ORF88 (now *rpo*H) encodes the small RNAP subunit H, which is a homologue of the eukaryotic RNAP subunit ABC27 and has no known counterpart in bacteria (*Klenk* et al., 1992). ORF130 is *nus*Ae, a homologue of the bacterial *nus*A-gene and a paralogue of *rps*3 (*Klenk* and *Zillig,* 1993; *Gibson* et al., 1993; *Klenk,* 1994). We show here that the last remaining partial ORF from this long DNA sequence is a homologue of bacterial *inf*B genes. Bacterial *inf*B genes code for one of three bacterial translation initiation factors, IF-2, whose function is to promote the binding of initiator tRNA to mRNA. However, sequence comparisons also show that the new *Sulfolobus* protein is even more closely related to the recently described (but still functionally uncharacterized) FUN12-protein of *Saccharomyces cerevisiae* (*Sutrave* et al., 1994).

Archaeal translation elongation factors have been studied a great deal (*Auer,* 1989; *Auer* et al., 1991; *Schröder* and *Klink,* 1991), but very little is known about initiation factors, and no factor has been identified based on its activity. One putative archaeal initiation factor, a homologue of the hypusine-containing eukaryotic IF-5A (*Bartig* et al. 1992), has been described, but the role of this factor in translation initiation is now very tenuous (*Kang* et al., 1994). In a systematic search for undetected archae-

al homologues of bacterial and eukaryotic translation initiation factors in the databases, *Keeling* and *Doolittle* (1995a) recently suggested that the genome of *Thermoplasma acidophilum* encodes a homologue of eIF-1A, a factor which increase the efficiency of interactions between the eukaryotic small ribosomal subunit, mRNA and the ternary complex (*Thomas* et al., 1980).

The *Sulfolobus infB* homologue is not only interesting for the elaboration of the so far poorly understood archaeal translation initiation machinery, but also provides a unique opportunity for inferring phylogenetic trees. The high degree of sequence similarity between the different translation factors made it possible to use EF-1α/Tu and EF-2/G as one of the first markers for rooting the universal phylogenetic tree (*Iwabe* et al., 1989). IF-2 has been used to root the EF-1α/Tu and EF-2/G subtrees since then (S.L.B., W.F.D., and J.D. Palmer, in press), and now by adding an archaeal IF-2 to this data set, a three-fold rooting of the universal tree can be inferred. This third tree further supports claims that archaea and eukaryotes are sister groups by demonstrating this relationship in all three protein families.

While these proteins may reveal a greater affinity between the archaea and eukaryotes, there are certainly many respects in which the archaea and bacteria are more akin. Genome organization and gene structure of archaea and bacteria show a great deal more similarity with each other than with that of eukaryotes. In particular the conserved gene order between some archaeal and bacterial operons or gene clusters, for instance *str, spc, rpo*BC and L11-L10, are remnants of a common ancestor that have vanished from eukaryotes (for review see *Zillig,* 1991; *Keeling* et al., 1994). The genomic organization context of *infB*, the *rpo* operon, and *nus*A in *Sulfolobus* allow us to draw some new inferences about ancient genome reorganizations.

Material and Methods

S. acidocaldarius was cultivated in Brock's salts and yeast-sucrose broth at 80 °C with aeration. 25 ml log-phase cultures were harvested and DNA purified by repeated phenol extractions and ethanol precipitations. 1 μg of genomic DNA was digested in a 20 μl volume with 5 units of *Hin*dIII overnight. This digested DNA was phenol extracted, ethanol precipitated once again and resuspended in a 50 μl ligation reaction to favor intramolecular ligations. This library of subgenomic circles was used as the template in PCR reactions using exact match primers designed to prime DNA synthesis in opposite directions at the *infB* locus. Amplification reactions consisted of 35 cycles of 92 °C denaturing for one minute, 50 °C annealing for one minute, and 72 °C extension for two minutes, all followed by a five minute extension at 72 °C in a 100 μl volume containing 10 mM TrisHCl, 50 mM KCl, 1.5 mM $MgCl_2$, 0.1% Triton-X100, and 0.2 mg/ml BSA at pH9. Inverse PCR products were separated by agarose gel electrophoresis, isolated, and cloned into the TA-tailed vector pCRII (In Vitrogen). Double stranded sequencing was carried out by cycle-sequencing using dye terminator biochemistry and electrophoresis on an ABI 373A.

Inferred amino acid sequences were aligned using the GCG program PILEUP (*Devereux* et al., 1984), with default gap penalties, followed by manual corrections. Parsimony trees were searched using PAUP version 3.1.1 (*Swofford,* 1993). Shortest tree searches consisted of 100 replicates of random additon and bootstrap analyses of 500 replicates using a single round of random addition each. Distance calculations based on the Dayhoff substitution matrix were made using the PROTDIST program from the PHYLIP package version 3.53c (*Felsenstein,* 1994). Once again, bootstrap analyses of distance trees consisted of 500 replicates.

Results

An Archaeal infB Homologue

In the initial characterization of a 10 kbp DNA sequence containing the genes for the large subunits of the DNA-dependent RNAP from *S. acidocaldarius* (*Pühler* et al., 1989), the authors noted a partial open reading frame, 'COOH-END', at the 5'-end of the sequence (Fig. 1), which they claimed to be transcribed in the same direction as the RNAP genes which follow downstream. A 2 kb mRNA was identified by Northern analysis with *Acc*I- and *Cla*I-fragments as probes, which was taken as proof for the expression of that region. However, no significant similarity to any entry of the protein sequence databases was found with the derived 101 amino acid sequence.

We have re-analyzed this region and find that it can also be translated in the opposite direction with a partial open reading frame of 187 codons (Fig. 1). Preceding this ORF at a suitable distance is a reasonably good ribosome binding site and an archaeal promoter BoxA-motif, TTTAAT. Moreover, the *Cla*I-fragment used as a hybridization probe by *Pühler* et al. shows a clear overlap with the open reading frame described here, but not with the reading frame which they initially described. It appears that the partial open reading frame proposed here represents a transcribed gene of up to 660 codons (estimated from the 2 kb mRNA detected by *Pühler* et al. in 1989), of which about one quarter is encoded in the previously sequenced DNA-fragment.

Most compelling is the comparison of the derived 187 amino acid sequence of the newly proposed partial ORF with the protein sequence databases. This resulted in significant BLASTP (*Altschul* et al., 1990) scores with the bacterial and mitochondrial translation initiation factor, IF-2, and numerous translation elongation and termination factors. A complementary search for sequence signatures with the PROSITE database (*Bairoch* and *Bucher,* 1994) also indicated a match with the ATP/GTP-binding site motif A (*Walker* et al., 1982) close to the N-terminus of the derived protein (Fig. 1). This motif is also found in the same region of translation factors, supporting the designation of the newly derived protein as the first archaeal member of the IF-2 translation initiation factor family.

In addition to 10 bacterial and mitochondrial IF-2 sequences, with BLASTP scores between 8.0 E-29 and 3.9 E-35, the newly proposed *Sulfolobus* protein was found to share an even higher degree of similarity (1.9 E-52) with the recently described FUN12-protein from *S. cerevisiae* (*Sutrave* et al., 1994). However, although *FUN12* has been shown in gene disruption experiments to be an essen-

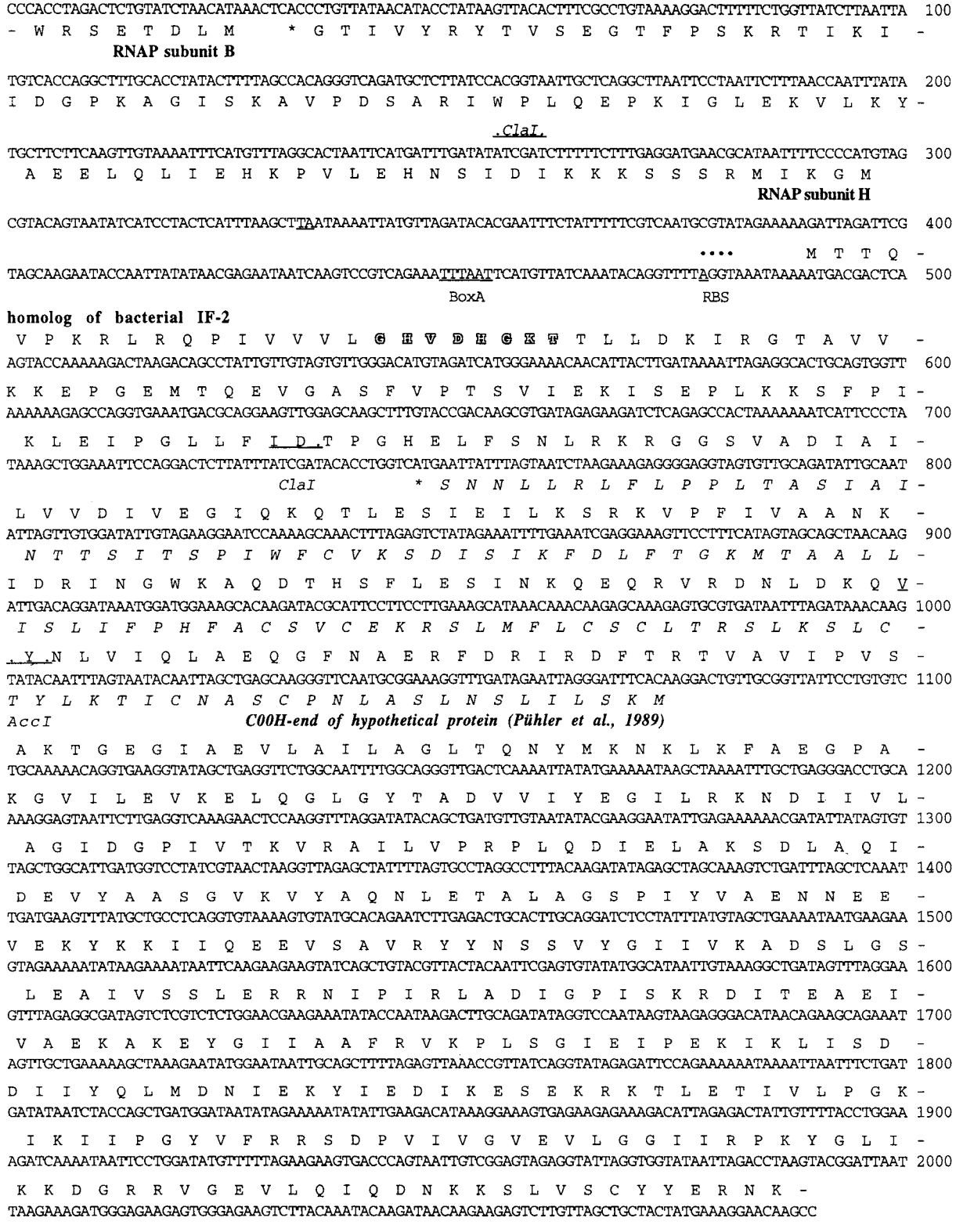

Fig. 1. Nucleotide sequence (5′ to 3′ direction) and derived amino acid sequences from the region upstream of the RNAP-gene cluster in *S. acidocaldarius*. The proposed *inf*B gene stretches from position 490 to the end of the known sequence. The translation predicted by *Pühler* et al. (1989) is shown below the DNA sequence in italics. Its RBS (ribosome binding site or 'Shine-Dalgarno' sequence), BoxA motif (archaeal promoter; *Hain* et al., 1992) and probable transcription initiation site are shown. *Acc*I and *Cla*I restriction sites are also shown. The end of the clone characterized by *Pühler* et al. (1989) corresponds to position 1050 of this figure, the remainder comes exclusively from the inverse PCR product reported here. The outlined amino acids GHVDHGKT near the N-terminus of the IF-2 homologue indicate the ATP/GTP-binding site motif (*Walker* et al., 1982). Sequence of the inverse PCR product has been deposited in GenBank under accession number U43413.

tial gene, nothing is known about its function, and no similarity to previously known proteins has been reported.

The high degree of similarity between this *Sulfolobus* open reading frame and members of this IF-2 family led us to seek a larger fragment from this region of the *Sulfolobus* chromosome by inverse PCR. Sequencing this inverse PCR clone proved our initial identification to be accurate, as the inferred amino acid sequence maintains a high similarity to FUN12 (BLAST score of 5 E-111) and bacterial *infB* genes (BLAST scores as low as 24 E-45) throughout its length, which covers about 90% of the *Saccharomyces* gene and its bacterial homologues (Fig. 2).

Relationships Between Translation Factors

The IF-2 family of sequences is part of a larger family of translation factors which are involved in initiation, elongation and termination. An alignment of representatives of these factors includes four blocks of highly conserved or invariable amino acids which are also conserved in the

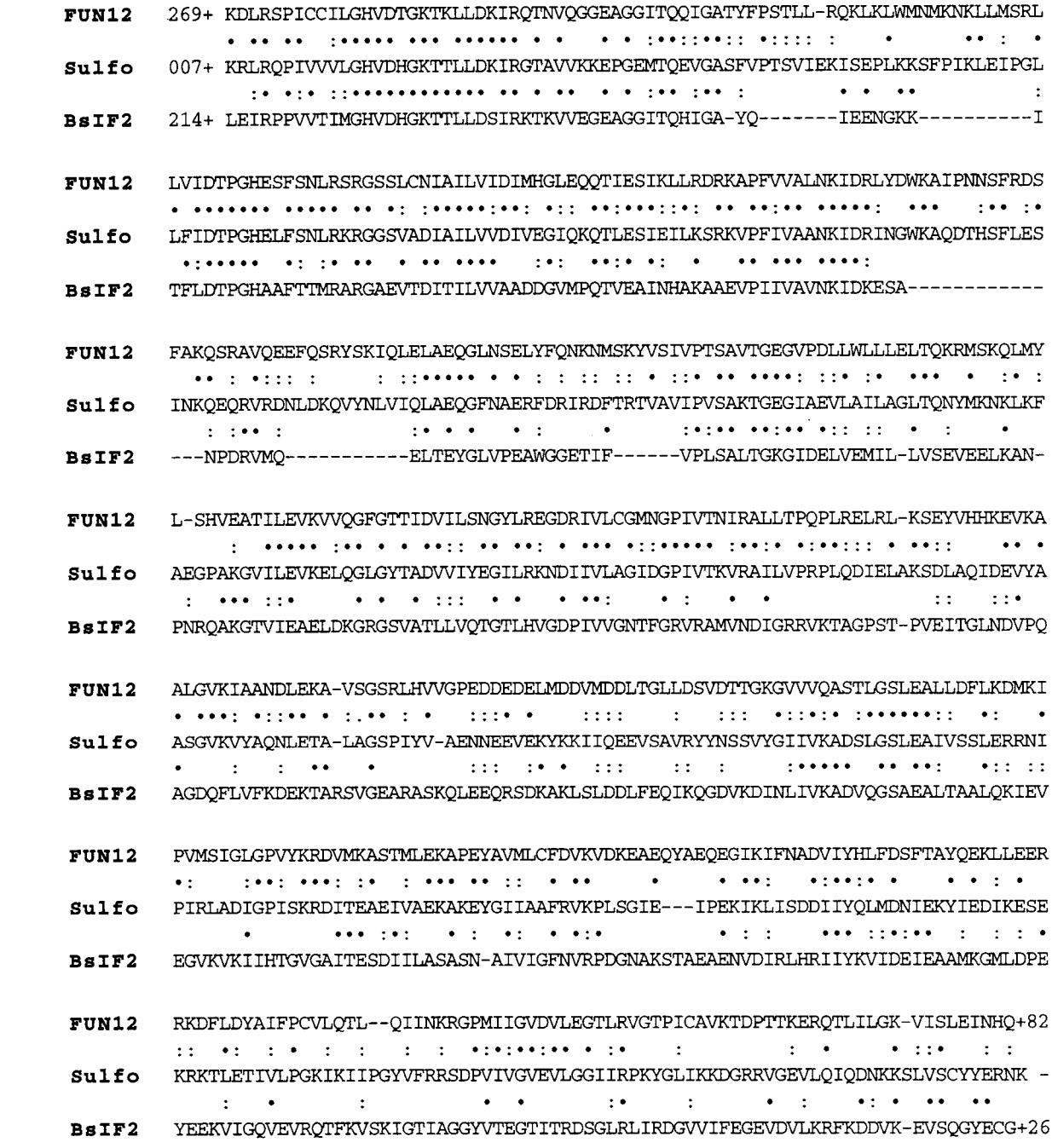

Fig. 2. Alignment of *Bacillus subtilis* IF-2, yeast FUN12 and the protein derived from *S. acidocaldarius infB*. Identical or conserved positions are indicated by a dot or colon between the sequences, respectively. Dashed lines indicate gaps in the alignment.

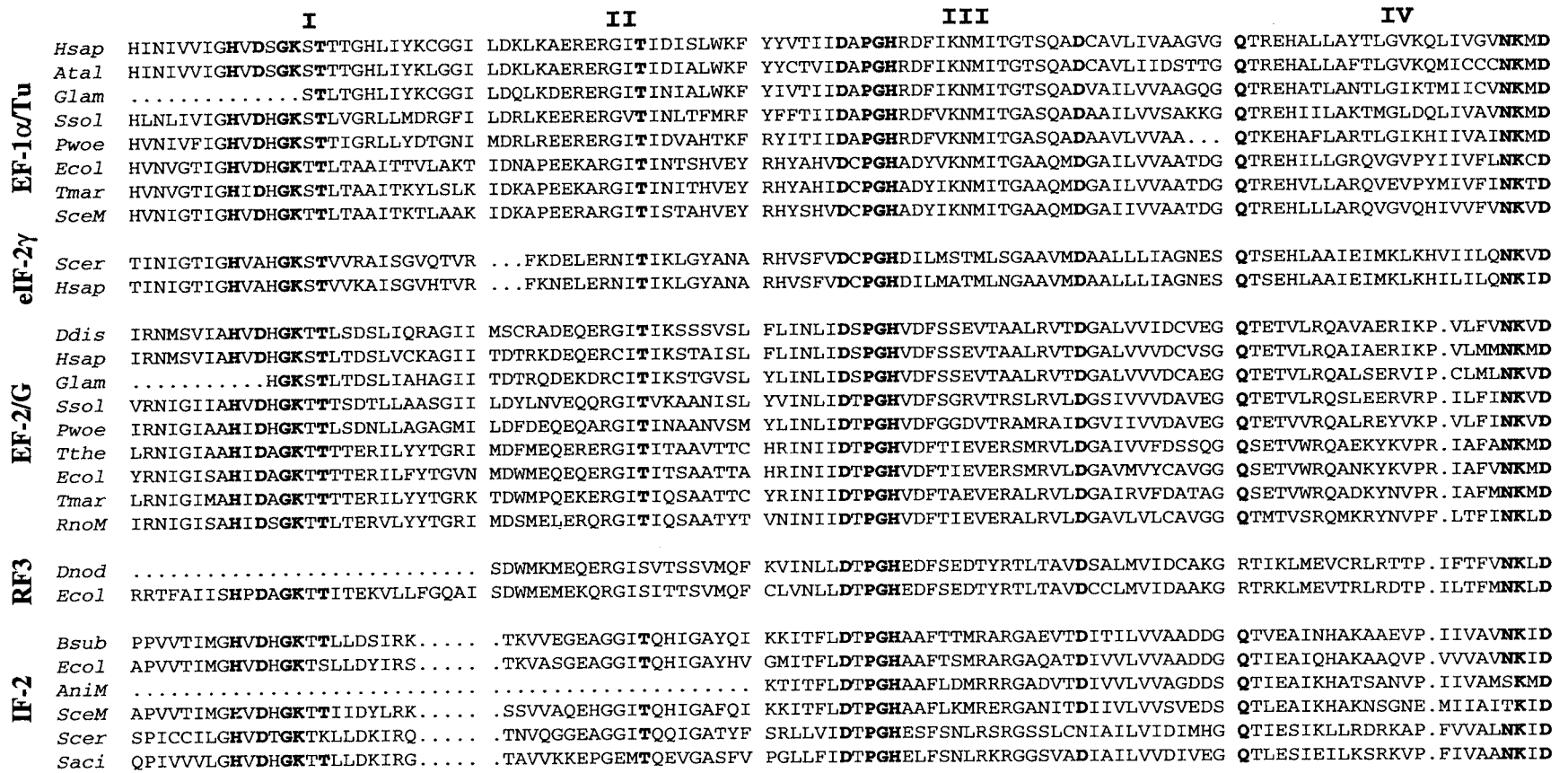

Fig. 3. Aligned amino acids sequences of a selection of homologous translation initiation (IF-2, eIF-2 gamma), elongation (EF-2/G, EF-1α/Tu and termination (RF3) factors. Blocks I–IV cover 111 positions with the highest degree of conservation between the different factors, including the new *Sulfolobus* sequence and yeast FUN12-protein. Abbreviations are: Saci, *S. acidocaldarius*; Ssol, *S. solfataricus*; Pwoe, *Pyrococcus woesei*; Ecol, *Escherichia coli*; Bsub, *B. subtilis*; Tmar, *Thermotoga maritima*; Tter, *Thermus thermophilus*; Dnod, *Dichelobacter nodosus*; SceM, *S. cerevisiae* mitochondrion; AniM, *Aspergillus niger* mitochondrion; RnoM, *Rattus norvegicus* mitochondrion; Glam, *Giardia lamblia*; Ddis, *Dictyostelium discoideum*; Scer, *S. cerevisiae*, Atal, *Arabidopsis thaliana*; Hsap, *Homo sapiens*. All sequences were fetched from the common protein sequence databases, SwissProt (*Bairoch* and *Boeckmann,* 1994) and PIR (*George* et al., 1994). Amino acids at the most conserved positions are highlighted in bold face.

amino acid sequence deduced from the *Sulfolobus infB* gene (Fig. 3). To see how the *Sulfolobus* protein is related to other members of the IF-2 family, phylogenetic trees were inferred using these regions of IF-2 and EF-2/G, the two families that contain the highest degree of sequence similarity to the newly identified *Sulfolobus* protein. The overall high similarity between the EF-2/G and IF-2 families allows the construction of a reciprocally rooted

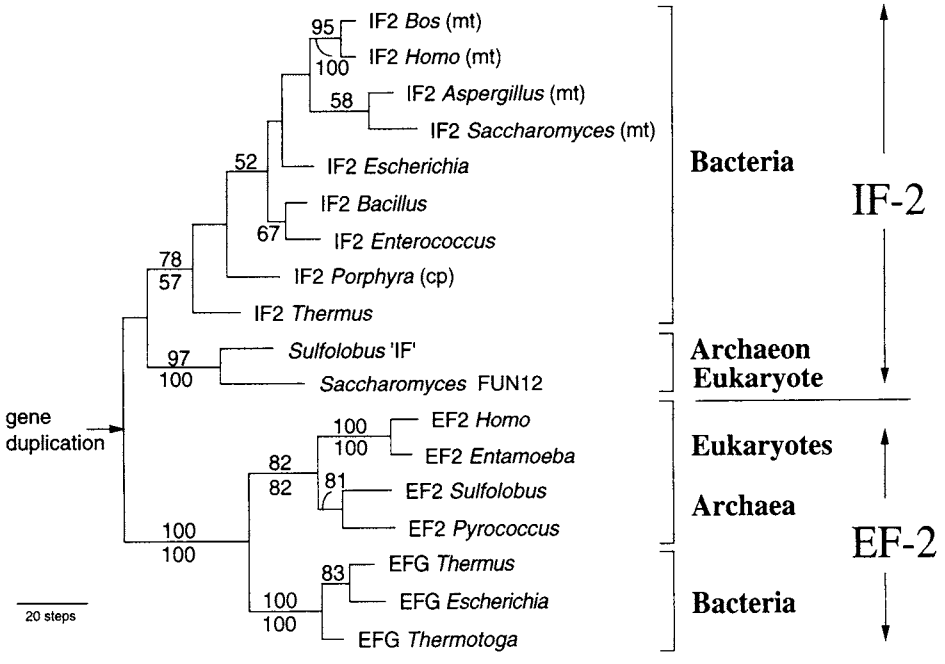

Fig. 4. Reciprocally rooted IF-2/EF-2 tree. The tree shown is one of two shortest trees found by parsimony analysis of the common amino acid positions indicated in Fig. 3. The tree is drawn to scale as indicated. Bootstrap values over 50% are shown above the nodes for parsimony analysis and below the nodes for distance analysis. The tree is 407 steps long, has a consistency index of 0.792 excluding uninformative characters and a retention index of 0.8. The other tree found at this length differs placing *E. coli* IF-2 as the next branch out from animal mitochondria.

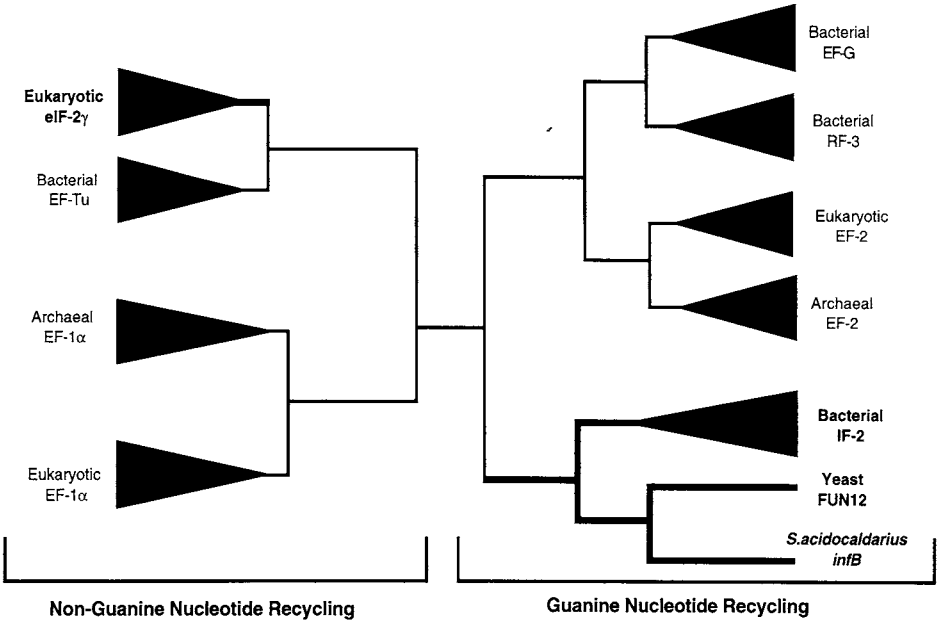

Fig. 5. Schematic phylogenetic tree of a broader variety of proteins involved in translation. Factors on the right half of the tree are able to recycle GTP, while those on the left require a guanine nucleotide exchange factor. The nine triangles indicate that several sequences of the same family have been used for the inference.

tree which reveals the position of the *Sulfolobus infB* sequence within the IF-2 subtree. This tree, shown in Fig. 4, shows that the *Sulfolobus* sequence is most closely related to the *Saccharomyces* FUN12-protein with high statistical confidence. The closer similarity of the new *Sulfolobus* sequence and yeast FUN12-protein to each other as compared to the bacterial IF-2's is further support for the archaea and eukaryotes as sister groups (*Iwabe* et al., 1989; *Gogarten* et al., 1989; *Klenk*, 1994; *Brown* and *Doolittle*, 1995).

The similarity of the IF-2 sequences to the EF-2/G and EF-1α/Tu families, the bacterial peptide release factor-3 (RF-3) and the gamma subunit of the eukaryotic translation initiation factor-2, allows a more comprehensive look at the evolution of translation factors, and also provides the opportunity for inferring a three-fold rooted phylogenetic tree. Fig. 5 shows a schematic tree inferred from the four conserved sequence blocks shown in Fig. 3. The tree confirms the location of the gene duplication separating the IF-2 and EF-2/G subtrees as shown in Fig. 4. Moreover, all three subtrees with representatives from all three domains (IF-2, EF-2/G and EF-1α/Tu) show the archaea and the eukaryotes as sister groups.

Genome Order in the Ancestor of Prokaryotes

A comparison of the archaeal and bacterial operons coding for the large RNAP subunits and *infB-nusA* suggests that both gene clusters might have been linked in the common ancestor of the prokaryotes. Fig. 6 shows representative examples of the gene order in the archaeal RNAP-gene-cluster (*S. acidocaldarius*), and the bacterial *rpoBC*- and *nusA-infB* operons (*E. coli*). The series *rpoH-rpoB* (*rpoB1* and *rpoB2* in methanogens and extreme halophiles) – *rpoA1* – *rpoA2* is invariable in all archaea analyzed so far. Also, the *nusA*-equivalent is found in all archaea downstream of this gene series. *rpl30* is missing only in *Thermoplasma* and *Halobacterium*, and *rps12* is situated immediately downstream of *nusAe* in all archaea but *Thermoplasma* (for a review see *Klenk*, 1994). The *infB*-homologue is known only in *Sulfolobus*, as there is little sequence data for the region upstream of *rpoH* for other species, so it remains open if the location of this gene is a general feature of the archaea. In bacteria, the genes coding for the two largest RNAP subunits in bacteria are always organized as an operon, *rpoBC*, but unlike archaea, this operon is always preceded by the L11-L12-gene cluster (*Liao* and *Dennis*, 1992), never by a homologue of *rpoH*, which is unknown in bacteria, and never by *infB*. Similarly, no bacterial *rpoBC* operon is followed by *nusA*. In fact, there is no discernible conservation of gene order downstream of *rpoC*: *Aquifex rpoC* is followed by *alaS* (*Klenk*, 1994) whereas *E. coli* has the *htrc*-gene (*Blattner* et al., 1993). Instead, *nusA* and *infB* are linked in *E. coli, H. influenzae, B. subtilis, M. genitalium* and *Thermus aquaticus thermophilus* (*Plumbridge* and *Springer*, 1983; *Fleischmann* et al., 1995; *Shazand* et al., 1990 and 1993; *Fraser* et al., 1995; *Vornlocher* and *Sprinzel*, 1995), suggesting that this is probably the ancestral state in bacteria. Both operons have been mapped in *E. coli* and are located rather distantly from each other: *rpoBC* at 90 minutes, *nusA-infB* at 68.5 minutes of the standard map (*Plumbridge* and *Springer*, 1983). Fig. 6 outlines a hypothetical ancestral state of archaea, and possibly all prokaryotes, where the *infB* and *nusA* genes were adjacent to those of the large RNAP subunits. An inversion involving the RNAP-genes together with *nusAe* would results in the order and orientations found in *Sulfolobus*.

Discussion

Based on sequence similarity and molecular phylogeny we have identified the last of the eight open reading frames

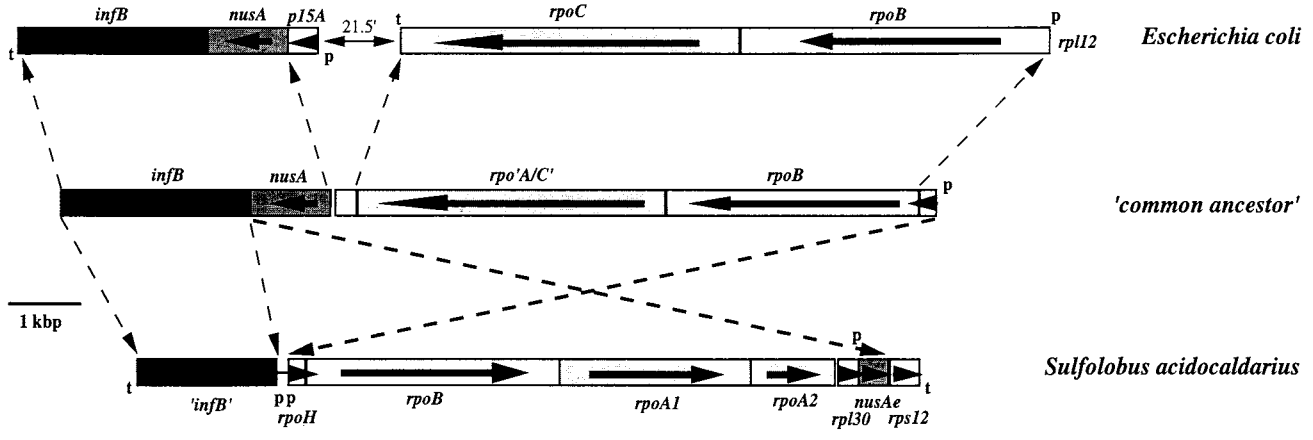

Fig. 6. Gene order around the RNAP-gene cluster in *S. acidocaldarius*, and the *rpoBC* and *nusA-infB* operons in *E. coli*. Abbreviations and gene names: *rpoA1* and *rpoA2* code for the archaeal RNAP subunits A′ and A″, respectively; *rpoC* codes for the homologous bacterial RNAP subunit β′. *rpoB* codes for the archaeal RNAP subunit B or the homologous bacterial RNAP subunit β, respectively. *rpoH* encodes the small archaeal RNAP subunit H. *infB* codes for the bacterial translation initiation factor IF–2, *nusA* for the NusA-protein, which is involved in transcription termination. *rps12*, *rpl12* and *rpl30* code for the ribosomal proteins S12, L12 and L30, respectively. p and t indicate promoters and terminators. The bar represents 1 kbp. Compared with *E. coli, B. subtilis* shows two additional small ORFs of unknown function between *nusA* and *infB* (not shown, *Shazand* et al., 1993).

in the RNAP gene cluster of *S. acidocaldarius* (*Pühler* et al., 1989). The proposed reading frame encodes the amino-terminus of an archaeal homologue of the bacterial translation initiation factor IF-2, a factor which plays a central role in sequestering the initiator-tRNA^{fMet} to the start site of the mRNA in bacteria. In addition we have identified the functionally uncharacterized FUN12-protein of yeast as another IF-2 homologue, and show that this protein is much more closely related to the archaeal ORF than either are to bacterial IF-2. These *inf*B homologues in archaea and eukaryotes raise some interesting questions about the evolution of translation initiation.

Most of the early information on archaeal translation machinery, for instance the size of the ribosomes and large rRNAs, the presence of Shine-Dalgarno-like ribosome binding sites on many mRNAs, and the lack of caps or long poly-A-tails on mRNAs, suggested that the archaeal translation system might be more akin to the bacterial system than to its eukaryotic counterpart (for review see *Brown* and *Daniels*, 1989). However, the higher degree of sequence similarity of the archaeal translation factors EF-1α and EF-2 to their eukaryotic homologues suggested instead that there may be a closer relationship between archaeal and eukaryotic translation elongation (*Iwabe* et al., 1989; S.L.B., W.F.D. and J.D. Palmer, in press). The issue is further complicated by the recent description of two putative archaeal translation initiation factors with homologues only in eukaryotes, which may be taken to indicate that translation initiation is also more similar between archaea and eukaryotes (*Keeling* and *Doolittle*, 1995a,b).

Given the current discrepancies, the significance of the *Sulfolobus infB* and FUN12 must be carefully considered. Since the yeast FUN12 protein has neither been demonstrated to be involved in translation nor been detected as part of the *Saccharomyces* translation initiation complex, it is unlikely that it acts in the same capacity as its bacterial orthologues. Indeed, in eukaryotes the role of IF-2 is assumed by eIF-2, a multi-subunit complex.

Interestingly however, the gamma subunit of eIF-2 is a paralogue of IF-2, the two factors having evolved independently from different subgroups of this family to perform similar, but not identical roles (*Keeling* and *Doolittle*, 1995). Furthermore, the relationship between the gamma subunit of eIF-2 and EF-Tu (supported by a handful of amino acids shared exclusively in Fig. 3 and the topology of the tree in Fig. 5) implies that eIF-2 arose from EF-Tu after the divergence of bacteria from archaea and eukaryotes. If this relationship is real, it casts some suspicion over the provenance of eIF-2 in the eukaryotic cell, possibly even suggesting that the gamma subunit was derived from a horizontally transferred bacterial EF-Tu. This uncertainty makes the role of FUN12 in eukaryotes all the more intriguing, and makes it unwise to speculate on the specific role of the *inf*B product in archaea until it is known whether archaea have an orthologue of the gamma subunit of eIF-2. This is especially true since there are also hints that archaea might contain factors homologous to other subunits of eIF-2: *Bazan* et al. (1994) have identified an archaeal homologue of a glycoprotein which interacts with the alpha subunit of eIF-2, protecting it from phosphorylation-inhibition (*Ray* et al., 1992).

Although the tree of translation factors remains silent in many ways on the evolution of translation, it does reveal a great deal about organismal relationships. Including the newly derived *Sulfolobus* sequence and the yeast FUN12 sequence in the IF-2 tree provides a rare opportunity for inferring a three-fold rooted universal tree with complete sets of orthologous proteins in all three domains. The location of the root within the IF-2 subtree confirms the location of the roots within the two previously described reciprocally rooted translation factor subtrees, EF-2/G and EF-1α/Tu (*Iwabe* et al., 1989; S.L.B., W.F.D., and J.D. Palmer, in press).

The sisterhood of archaea and eukaryotes implied by these trees allows some very precise inferences to be made about the last common ancestor of all cells by identifying homologous properties of archaea and bacteria. One instance where this has proved to be particularly illuminating is in the conservation of gene order within certain operons. In even very distantly related bacteria, the *nus*A-gene is part of the conserved *nus*A-*inf*B operon which is not tightly linked to the RNAP operon, being separated in *E. coli* by 21.5 minutes on the standard map (*Shazand* et al., 1993; *Vornlocher* and *Sprinzel*, 1995). However, the archaeal *nus*Ae, is part of the RNAP gene cluster in all archaea studied so far (*Klenk*, 1994), and in *Sulfolobus* at least, *inf*B is immediately upstream of the RNAP operon. It is tempting to speculate about an ancient order for these genes like the one shown in Fig. 6 which would provide the possibility for co-transcription of most of the RNA polymerase protein mass (NusA can in some fashion be thought of as a 'temporary' termination RNAP subunit). The separation to two clusters in bacteria and an inversion within the gene cluster in archaea would result in the present arrangements.

Acknowledgments. This work was supported by grants from the Medical Research Council of Canada (MT4467) and the Canadian Genome Analysis and Technology Program (GO12319). P.J.K. is a recipient of an MRC studentship and W.F.D. is a fellow of the Canadian Institute for Advances Research. We would like to thank *M. E. Schenk*, to whom we are indebted for assistance in sequencing.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J.: Basic local alignment search tool. J. Mol. Biol. *215,* 403–410 (1990)

Auer, J.: Studien zur molekularen Evolution des Translationsapparats. Dissertation, Ludwig-Maximilians-Universität München (1989)

Auer, J., Spicker, G., Mayerhofer, L., Pühler, G., Böck, A.: Organization and nucleotide sequence of a gene cluster comprising the translation elongation factor 1α from *Sulfolobus acidocaldarius*. System. Appl. Microbiol. *14,* 14–22 (1991)

Bairoch, A., Boeckmann, B.: The SWISS-PROT protein sequence data bank: current status. Nucleic Acids Res. *22,* 3578–3580 (1994)

Bairoch, A., Bucher, P.: PROSITE: recent developments. Nucleic Acids Res. *22*, 3583–3589 (1994)

Bartig, D., Lemkemeier, K., Frank, J., Lottspeich, F., Klink, F.: The archaebacterial hypusine-containing protein: Structural features suggest common anchestry with eukaryotic translation initiation factor 5A. Eur. J. Biochem. *204*, 751–758 (1992)

Bazan, J. F., Weaver, L. H., Roderick, S. L., Huber, R., Matthews, B. W.: Sequence and structure comparison suggest that methionine aminopeptidase, prolidase, aminopeptidase P, and creatinase share a common fold. Proc. Natl. Acad. Sci. USA *91*, 2473–2477 (1994)

Blattner, F. R., Burland, V., Plunkett, G., Sofia, H. J., Daniels, D. L.: Analysis of the Escherichia coli genome. IV. DNA sequences of the region 89.2–92.8 minutes. Nucleic Acids Res. *21*, 5408–5417 (1992)

Brown, J. W., Daniels, C. J.: Gene structure, organization, and expression in archaebacteria. CRC Crit. Rev. Microbiol. *16*, 287–335 (1989)

Brown, J. R., Doolittle, W. F.: Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. Proc. Natl. Acad. Sci. *92*, 2441–2445 (1995)

Devereux, J., Haeberli, P., Smithies, O.: A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. *12*, 387–395 (1984)

Felsenstein, J.: PHYLIP user manual version 3.53. University of Washington (1994)

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J.-F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C. A., Gocayne, J. D., Scott, J. D., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J. M., Weidman, J. F., Phillips, C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, M. C., Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L., McDonald, L. A., Small, K. V., Fraser, C. M., Smith, H. O., Venter, J. C.: Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science *269*, 496–512 (1995)

Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., Fritchman, J. L., Weidman, J. F., Small, K. V., Sandusky, M., Fuhrmann, J. L., Nguyen, D. T., Utterback, T. R., Saudek, D. M., Phillips, C. A., Merrick, J. M., Tomb, J.-F., Dougherty, B. A., Bott, K. F., Hu, P.-C., Lucier, T. S., Peterson, S. N., Smith, H. O., Hutchison, C. A. I., Venter, J. C.: The minimal gene complement of *Mycoplasma genitalium*. Science *270*, 397–403 (1995)

George, D. G., Barker, W. C., Mewes, H.-W., Pfeiffer, F., Tsugita, A.: The PIR-International sequence database. Nucleic Acids Res. *22*, 3569–3573 (1994)

Gibson, T. J., Thompson, D., Heringa, J.: The KH domain occurs in a diverse set of RNA-binding proteins that include the antiterminator NusA and is probably involved in binding to nucleic acid. FEBS Lett. *324*, 361–366 (1993)

Gogarten, J. P., Kibak, H., Dittrich, P., Taiz, P., Bowman, E. J., Manolson, M. F., Poole, R. J., Date, T., Oshima, T., Konishi, J., Denda, K., Yoshida, M.: Evolution of the vacuolar H+-ATPase: Implications for the origin of eukaryotes. Proc. Natl. Acad. Sci. *86*, 9355–9359 (1989)

Hain, J., Reiter, W.-D., Hüderpohl, U., Zillig, W.: Elements of an archaeal promoter defined by mutational analysis. Nucleic Acids Res. *20*, 5423–5428 (1992)

Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S., Miyata, T.: Evolutionary relationship of archaebacteria, eubacteria, and eucarkota inferred from trees of duplicated genes. Proc. Natl. Acad. Sci. *86*, 9355–9359 (1989)

Kang, H. A., Hershey, J. W. B.: Effect of initiation factor eIF-5A depletion on protein synthesis and proliferation of *Saccharomyces cerevisiae*. J. Biol. Chem. *269*, 3934–3940 (1994)

Keeling, P. J., Charlebois, R. L., Doolittle, W. F.: Archaebacterial genomes: eubacterial form and eukaryotic content. Current Opin. Genet. and Develop. *4*, 816–822 (1994)

Keeling, P. J., Doolittle, W. F.: An archaebacterial eIF-1A: New Grist for the mill. Molec. Microbiol. *17*, 399–400 (1995a)

Keeling, P. J., Doolittle, W. F.: Archaea: Narrowing the gap between Prokaryotes and Eukaryotes, Proc. Natl. Acad. Sci. *92*, 5761–5764 (1995b)

Klenk, H.-P., Palm, P., Lottspeich, F., Zillig, W.: Component H of the DNA-dependent RNA polymerases of Archaea is homologueous to a subunit shared by the three eucaryal nulear RNA polymerases. Proc. Natl. Acad. Sci. USA *89*, 407–410 (1992a)

Klenk, H.-P., Zilli,g W.: Archaea contain an open reading frame paralogous to the gene of the ribosomal protein S3. System. Appl. Microbiol. *16*, 22–24 (1993)

Klenk, H.-P.: Evolution der Organismen – DNA-abhängige RNA-Polymerasen als molekulare Chronometer für die Stammesgeschichte der drei Domänen der Lebewesen: Archaea, Eucarya und Bacteria. Deutsche Hochschulschriften 551, Egelsbach, Verlag Hänsel-Hohenhausen (1994)

Liao, D., Dennis, P. P.: The organiszation and expression of essential transcription translation component genes in the extremely thermophilic eubacterium *Thermotoga maritima*. J. Biol. Chem. *267*, 22787–22797 (1992)

Plumbridge, J. A., Springer, M.: Organization of the *Escherichia coli* chromosome around the genes for translation initiation factor IF2 (infB) and a transcription termination factor (nusA). J. Mol. Biol. *167*, 227–243 (1983)

Pühler, G., Lottspeich, F., Zillig, W.: Organization and nucleotide sequence of the genes encoding the large subunits A, B and C of the DNA-dependent RNA polymerase of the archaebacterium *Sulfolobus acidocaldarius*. Nucleic Acids Res. *17*, 4517–4534 (1989)

Ray, M. K., Datta, B., Chakraborty, A., Chattopadhyay, A., Meza-Keuthen, S., Gupta, N. K.: The eukaryotic intiation factor 2-associated 67-kDa polypeptide (p67) plays a critical role in regulation of protein synthesis initiation in animal cells. Proc. Natl. Acad. Sci. USA *89*, 539–543 (1992)

Schröder, H., Klink, F.: Gene for the ADP-ribosylatable elongation factor 2 from the extreme thermoacidophilic archaebacterium *Sulfolobus acidocaldarius*. Eur. J. Biochem. *195*, 321–327 (1991)

Shazand, K., Tucker, J., Chiang, R., Stansmore, K., Sperling-Peterson, H. U., Grunberg-Manago, M., Rabinowitz, J. C., Leighton, T.: Isolation and molecular genetic characterization of the *Bacillus subtilis* gene (infB) encoding protein synthesis initiation factor 2. J. Bacteriol. *172*, 2675–2687 (1990)

Shazand, K., Tucker, J., Grunberg-Manago, M., Rabinowitz, J. C., Leighton, T.: Similar organization of the nusA-infB operon in Bacillus subtilis and Escherichia coli. J. Bacteriol. *175*, 2880–2887 (1993)

Sutrave, P., Shafer, B. K., Strathern, J. N., Hughes, S. H.: Isolation, identification and characterization of the FUN12 gene of *Saccharomyces cereviseae*. Gene *146*, 209–213 (1994)

Swofford, D.: PAUP users manual version 3.1.1. Smithonian Institution (1993)

Thomas, A., Goumans, H., Voorma, H. O., Benne, R.: The mechanism of action of eukaryotic initiation factor 4C in protein synthesis. Eur. J. Biochem. *107*, 39–45 (1980)

Vornlocher, H., Sprinzel, M.: Molecular cloning of the Thermus thermophlus nusA/infB operon. unpublished NCBI sequence ID 642366 (1995)

Walker, J. E., Saraste, M., Runswick, M. J., Gay, G. J.: Distantly

related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. *1*, 945–951 (1982)

Zillig, W.: Comparative biochemistry of archaea and bacteria. Current Opinion in Genet. and Develop. *1*, 544–551 (1991)

Dr. *Hans-Peter Klenk*, Dalhousie University, Department of Biochemistry, Halifax, Nova Scotia, Canada B3H 3H7