PROTEINS: Structure, Function, and Genetics 52:585–597 (2003)

Survey for G-Proteins in the Prokaryotic Genomes: Prediction of Functional Roles Based on Classification

Shashi B. Pandit and N. Srinivasan*
Molecular Biophysics Unit, Indian Institute of Science, Bangalore, India

ABSTRACT **The members of the family of G-proteins are characterized by their ability to bind and hydrolyze guanosine triphosphate (GTP) to guanosine diphosphate (GDP). Despite a common biochemical function of GTP hydrolysis shared among the members of the family of G-proteins, they are associated with diverse biological roles. The current work describes the identification and detailed analysis of the putative G-proteins encoded in the completely sequenced prokaryotic genomes. Inferences on the biological roles of these G-proteins have been obtained by their classification into known functional subfamilies. We have identified 497 G-proteins in 42 genomes. Seven small GTP-binding protein homologues have been identified in prokaryotes with at least two of the diagnostic sequence motifs of G-proteins conserved. The translation factors have the largest representation (234 sequences) and are found to be ubiquitous, which is consistent with their critical role in protein synthesis. The GTP_OBG subfamily comprises of 79 sequences in our dataset. A total of 177 sequences belong to the subfamily of GTPase of unknown function and 154 of these could be associated with domains of known functions such as cell cycle regulation and t-RNA modification. The large GTP-binding proteins and the α-subunit of heterotrimeric G-proteins are not detected in the genomes of the prokaryotes surveyed. Proteins 2003;52:585–597.**

Key words: proteins; genome analysis; GTPases; prokaryotes; protein Families; protein structures

INTRODUCTION

The family of G-proteins is comprised of regulatory Guanosine Triphosphate (GTP) hydrolases. They are known to bind and hydrolyze GTP to result in Guanosine Diphosphate (GDP). G-proteins act as molecular switches in cellular signalling pathways. The GTP- and GDP-bound states of these proteins determine their affinity for other proteins in downstream signalling events.[1–3] These proteins are involved in essential functions of the cell such as differentiation and proliferation, transmembrane signalling mediated by hormones and light, protein trafficking, secretion, cytoskeletal organization, cell motility, endocytosis, and protein synthesis.[1–3]

The biological activities of certain members of the G-protein family are regulated by various accessory proteins. These include (1) the families of Guanine nucleotide Exchange Factor (GEF), which catalyzes the exchange of GTP for GDP, (2) GTPase Activating Protein (GAP), which enhances the rate of hydrolysis of GTP, and (3) Guanine nucleotide Dissociation Inhibitor (GDI), which is involved in the inhibition of the release of GDP from certain classes of G-proteins such as the Rab class.[1–3]

The G-protein family can be classified broadly into four subfamilies on the basis of their biological functions and molecular weights. These are (1) small GTP-binding proteins, (2) translational GTPases, (3) α-subunit of heterotrimeric G-proteins, and (4) large GTP-binding proteins.

Small GTP-binding proteins are conserved across all the eukaryotes from yeast to human.[4] This subfamily can be divided further, based on similarities in their effector binding region and downstream targets. The five classes of small GTPases are Ras, Rho/Rac/Cdc42, Rab, Arf/Sar1, and Ran.[3–6] The members of these subfamilies are involved in diverse functions such as signalling, vesicular trafficking, and cytoskeletal organization.[3–8] These proteins appear to elicit their functions through their mutual cross-talks and multiple downstream effectors in a variety of cellular events. Small GTP-binding proteins, with the exception of Ran and Sar1, undergo post-translational modification that localizes them to the membrane. The sites of post-translational modification vary within the subfamily. The C-terminal of the Ras, Rho/Rac/Cdc42, and Rab class of proteins is covalently attached to various lipid moieties such as the farnesyl, geranylgeranyl, palmitoyl, and methyl groups, while the members of the Arf class undergo modification at the N-terminal with myristic acid.[4,9]

The α-subunit of heterotrimeric G-protein in its GDP-bound state forms a ternary complex with the β and γ subunits in unstimulated cells. On perception of signals, from the receptor, GTP replaces the GDP with the help of GEFs, and this step is followed by the dissociation of

Grant sponsor: Council of Scientific and Industrial Research, India; Grant sponsor: Wellcome Trust, London; Grant sponsor: Department of Biotechnology, New Delhi.

*Correspondence to: N. Srinivasan, Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India. E-mail: ns@mbu.iisc.ernet.in

Received 30 July 2002; Accepted 22 January 2003

α-subunit from the βγ-subunits. The heterotrimeric G-proteins are currently known to occur only in eukaryotes.[10]

The translational GTPase subfamily is comprised of those members of G-proteins that are involved in the various steps during protein.[11] This subfamily includes initiation factors (Initiation factor 2 (IF2), eIF2γ, elongation factor-Tu (EF-Tu), selenocysteine-specific elongation factor B (SELB), elongation factor-G (EF-G), and release factor 3 (RF3). Most of these are known to occur both in prokaryotes and eukaryotes. This might be due to their involvement in the most conserved mechanism of protein synthesis.[11]

The large GTP-binding protein subfamily is known to occur only in eukaryotes and it is comprised of proteins of high molecular weight. These proteins are associated with high intrinsic rates of GTPase activity and are involved in a variety of functions including endocytosis[12–14] and viral resistance.[15,16] Despite low-sequence similarity shared among the members of this subfamily, they share similar biochemical and structural properties. This subfamily includes proteins such as dynamin, interferon γ induced Guanylate Binding Protein (GBP), Mx proteins, yeast Mgm1, Vps1, and Dnm1.[14]

In addition to these well-characterized subfamilies, the family of G-proteins also includes a few new subfamilies. Members of these new subfamilies share low sequence similarity with the members of the well-characterized subfamilies. The subfamily of Developmentally regulated GTP binding proteins (DRG)[17–19] is known to be involved in the regulation of cell differentiation. *Escherichia coli* Ras-like GTP-binding protein (Era)[20] mediates the coordination of the cell growth rate with cytokinesis[21,22] and hence plays a crucial role in cell cycle regulation. Obg (spoOB associated GTP binding protein) plays a critical role in regulating replication and differentiation and has also been shown to be essential for cell viability in *B. subtilits*.[23]

The classification of the G-proteins based on sequence and functional similarity as described in databases such as PRODOM,[24] SMART,[25] and Pfam[26] is largely consistent with the classification described above.

The three-dimensional (3-D) structures of various G-proteins solved to date reveal a common structural core belonging to the P-loop nucleotide hydrolase fold. The most conserved guanine nucleotide-binding site is formed by residues located far in the primary structure, which come close to one another in the 3-D structure.[27] Five polypeptide loops that form the guanine nucleotide-binding pocket mark the most highly conserved elements that characterize the G-protein family. The five loops are designated G-1 through G-5. G-1, also referred to as P-loop, is characterized by the consensus sequence G-X-X-X-X-G-K-S/T (X is any amino acid) and is known to interact with the α and β phosphates of guanine nucleotide. The G-2 motif is X-T-X, where threonine is involved in the coordination of Mg^{2+}. The G-3 motif has the characteristic D-X-X-G motif wherein the Asp co-ordinates with Mg^{2+} and Gly is hydrogen bonded to the γ phosphate of GTP through the backbone amide.[28,29] The N/T-K-X-D motif in the G-4 loop is responsible for the recognition of the guanine ring over other bases.[30] The G-5 loop with consensus sequence T/G-C/S-A also helps in the recognition of the guanine base.[27,31] The G-1, G-3, and G-4 motifs are highly conserved among G-proteins and hence they form the characteristic features of G-proteins.

TABLE I. Classification Scheme Followed for Functional Assignment of the Putative G-Proteins in Prokaryotes

G-protein subfamily	Class within subfamily
Small GTP-binding protein	1. Ras
	2. Rho/Rac/Cdc42
	3. Ran
	4. Arf
	5. Rab
Translational GTPases	1. Initiation factor includes Initiaiton factor 2 (IF2) and Eukaryotic initiation factor 2 γ(eIF2 γ)
	2. Elongation factor Tu (EF-Tu)
	3. Elongation factor G (EF-G)
	4. Selenocysteine sepcific elongation factor B (SelB)
	5. Release factor 3 (RF3)
	6. LepA
	7. TypA/BipA
GTPase of unknown function	1. Era
	2. EngA
	3. TrmE/ThdF
	4. Hypothetical GTP binding protein
GTP_OBG	
α-subunit of heterotrimeric G-protein	
Large GTP-binding protein	1. Dynamin
	2. Mx

In the present analysis, we have surveyed the G-proteins encoded in the 42 completely sequenced prokaryotic genomes using sensitive sequence search methods. The G-proteins thus identified have been grouped into various functional classes following the classification scheme of Leipe et al.[32] who have studied the evolution of the GTPases and related ATPases (that belong to the broad class of P-loop nucleotide hydrolases) identified from the nonredundant sequence databases. The classification scheme followed in the analysis for the functional assignment of putative G-proteins is tabulated in Table I. The present analysis particularly focuses on the comprehensive collection of prokaryotic members identified using sequence search methods. The current work also provides an in-depth perspective on the varieties of G-proteins that occur in specific prokaryotes with complete genome data. Further analysis presented in this study is on the occurrence of other functional domains in the gene products containing the GTPase domains. Such analysis on the combinations of domains provides hints on the specific biological roles of those G-proteins.

MATERIALS AND METHODS

Databases

The completely sequenced genomes of 8 archaebacteria and 34 eubacteria have been surveyed for the occurrence of

G-proteins. The translated amino acid sequence data, of all the proteins encoded in these genomes, has been obtained from the National Centre for Biotechnology Information (NCBI) genome resource. In addition, sequences and alignments of proteins in various homologous families have been obtained from the Pfam database[26] available at the Sanger Centre (http://www.sanger.ac.uk/Software/Pfam). The nonredundant sequence database (NRDB) has been obtained from NCBI (ftp://ncbi.nlm.nih.gov/blast/db).

Sequence Analysis

Amino acid sequences of all the putative proteins encoded in these genomes have been analyzed in order to identify the G-proteins in these genomes. The forward and reverse family profile matching programs, PSI-BLAST[33] and IMPALA,[34] obtained from the NCBI resource have been used for this purpose. The Hidden Markov Model matching program HMMER2[35] has been obtained from the University of Washington (http://hmmer.wustl.edu/) and this program has also been used in the current analysis. All the computations have been performed on several single processor linux-driven PCs and on a six-node multi-processor linux cluster.

IMPALA searchable profiles[36] or position-specific scoring matrices (PSSMs) were generated for a collection of diverse members of every family represented in Pfam using the multiple sequence alignments given in Pfam. All the sequences from the genome data have been queried against the PSSM database with an E-value cutoff of 10^{-8}. This E-value cutoff has been extrapolated from the one reported by Schaffer et al.[34] as well as based on benchmarking (Mhatre and Srinivasan, unpublished results) using a database of structure-based sequence alignments of similarly folded proteins.[37]

The amino acid sequences from the genome data were queried, also against family specific Hidden Markov Models obtained from Pfam, using the HMMER2 algorithm (E-value cutoff: 10^{-2}) for domain assignment. For most of the gene products, domain assignments have been made by consulting the results from both IMPALA and HMMER searches. The domain boundaries have been assigned, mostly based on IMPALA results although HMMER alignments have also been considered. The hits were carefully analyzed manually for the presence of characteristic motifs of G-protein (G-1, G-3, and G-4) specifically to remove false positives from the analysis.

Further PSSMs specific to individual subfamilies has been generated using the multiple sequence alignments of members that are known to constitute the subfamily. These have been used in assigning the gene products to specific subfamilies of G-proteins. The PSSMs for the subfamilies of G-proteins were searched using the program, IMPALA, in order to associate the query sequence with a subfamily.

The regions of at least 50 residues length in the GTPase containing gene products, which could not be assigned a functional domain, have been searched against the NRDB non-redundant database. For this purpose, PSI-BLAST has been used with the E-value cutoff of 10^{-4} until the convergence is reached or for 20 cycles whichever is earlier. This has been performed to explore the possibility of occurrence of domains that may not be identified in IMPALA and HMMER searches.

The phylogenetic trees of G-protein family members have been generated using PHYLIP package[38] using the dissimilarity measures given by the multiple sequence alignment program MALIGN.[39] Table II lists the number of different G-proteins of various subfamilies in the genomes analyzed.

RESULTS AND DISCUSSION

The occurrence of members of the G-protein family has been investigated in 42 completely sequenced prokaryotic genomes. We have used mainly sensitive tools such as IMPALA, HMMER2, and PSI-BLAST in our survey for G-proteins (see Materials and Methods). The hits have been evaluated in terms of their statistical significance (e-value) and the occurrence of G-protein-specific motifs (G-1, G-3, and G-4). The subjective decision has been taken in some ambiguous cases with reference to classification into subfamilies. The list of various genomes of the complete data surveyed and the number of hits of various subfamilies of G-proteins are given in Table II. A comprehensive list of all the GTPase domains containing gene products has been provided along with their gene codes and domain assignments in the Supplementary material available with the on-line version of this article.

Small GTP-Binding Proteins

Small GTP-binding proteins are known to be involved in specialized functions of eukaryotes like cellular differentiation, cytoskeletal organization, and cell motility and hence their occurrence has been known to be largely restricted to eukaryotes.[4,8] Recently, ras-like GTPase has been identified in *Myxococcus xanthus*,[40] a Gram-negative bacterium known to have specialized developmental stages in its life cycle. However, the occurrences and roles of small G-proteins in other bacteria have been less well studied. We have systematically surveyed for the presence of homologues of small GTP-binding protein in prokaryotes. In our analysis, we have identified 7 homologues (3 in archaebacteria, and 4 in eubacteria) as listed in Table II. The identified gene products have further been classified into functional classes tabulated in Table III. Among the various classes of small G-proteins described earlier,[6] the Arf, Ras, and Rab class members have been identified in the prokaryotic genomes analyzed in the current study. Although the best pairwise sequence identity among these putative small G-proteins with the eukaryotic counterparts is 26%, the 3-D fold of P-loop nucleotide hydrolase, shared among all the eukaryotic small G-proteins, could be reliably associated to these homologues by GenTHEADER.[41] Subsequently, we have analyzed the conservation of GTP-binding motifs in these hits and the results are summarized in Table III. Only one of the homologues, from *M. loti* (gi13472827), has all the 3 characteristic GTP-binding motifs present while the G-1 and G-4 motifs are well conserved among all the prokaryotic homologues.

TABLE II. Number of Occurrences of G-Proteins and Their Subfamilies in Prokaryotic Genomes*

Organism	No. of ORFs	Small GTP-binding protein	IF2	EF-Tu	EF-G	LepA	TypA /BipA	SelB	RF3	GTPase of Uknown function				GTP_ OBG
										Era	EngA	TrmE	Hypo	
Archaebacteria														
Aeropyrum pernix	2,694	—	2	1	1	—	—	—	—	1	—	—	1	2
Archaeglobus fuldigus	2,407	—	2	2	1	—	—	—	—	2	—	1	—	2
Methanobacterium thermoautotrophicum	1,869	2	2	2	1	—	—	—	—	1	—	—	—	2
Methanococcus jannschii	1,715	—	2	1	1	—	—	1	—	4	—	—	1	2
Pyrococcus abyssii	1,765	—	2	1	1	—	—	—	—	2	—	—	1	2
Pyrococcus horikoshii	2,064	—	2	1	1	—	—	—	—	2	—	—	1	2
Halobacterium sp.	2,058	—	2	2	1	—	—	—	—	3	—	1	—	2
Thermoplasma acidophilum	1,478	1	2	1	1	—	—	—	—	1	—	—	—	2
Eubacteria														
Aquifex aeolicus	1,522	2	1	2	1	1	—	1	—	4	1	1	—	2
Bacillus halodurans	4,066	—	1	1	1	1	1	—	—	3	1	1	2	2
Bacillus subtilis	4,099	—	1	1	1	1	1	—	—	2	1	1	2	2
Borrelia burgdorrferi	850	—	1	1	2	1	—	—	—	1	1	1	1	2
Buchnera sp.	564	—	1	1	1	1	1	—	1	2	1	1	—	2
Campylobacter jejuni	1,634	—	1	1	1	1	1	1	—	2	1	1	—	2
Chlamydia muridarum	818	—	1	1	1	1	—	—	—	1	1	0	—	0
Chlamydia trachomatis	894	—	1	1	1	1	—	—	—	0	1	1	—	2
Chlamydia pneuomiae (CWL029)	1,052	—	1	1	1	1	—	—	—	1	1	1	—	2
Chlamydia pneuomiae (AR39)	997	—	1	1	1	1	—	—	—	1	1	0	—	1
Chlamydia pneuomiae (J138)	1,070	—	1	1	1	1	—	—	—	1	1	1	—	2
Deinococcus radiodurans	2,580	1	1	2	2	1	1	—	1	3	1	1	—	2
Haemophilus influenzae	1,709	—	1	2	1	1	1	1	1	2	1	1	—	2
Escherichia coli	4,289	—	1	2	1	1	1	1	1	2	1	1	—	2
Helicobacter pylori	1,553	—	1	1	1	1	1	—	—	1	1	1	1	2
Helicobacter pylori (stJ99)	1,491	—	1	1	1	1	1	—	—	1	1	1	—	2
Mycoplasma genitalium	480	—	1	1	1	1	—	—	—	3	1	1	—	2
Lactococcus lactis	2,266	—	1	1	1	1	1	—	1	2	1	1	3	2
Mycoplasma pneumoniae	677	—	1	1	1	1	—	—	—	2	1	1	1	2
Mycobacterium tuberculosis (H37Rv)	3,918	—	1	1	2	1	1	—	—	1	1	1	—	1
Mycobacterium leprae	1,605	—	1	1	1	1	1	—	—	2	1	1	—	1
Mesorhizobium loti	6,752	1	1	2	2	1	1	—	1	3	1	1	1	2
Neisseria meningitidis	2,025	—	1	2	1	1	1	—	1	2	1	1	1	2
Pasterulla muticoda	2,014	—	1	2	1	1	1	1	1	3	1	1	—	2
Pseudomonas aeruginosa PA01	5,565	—	1	2	2	1	1	1	1	3	1	1	—	2
Rickettsia prowazekii	834	—	1	1	1	1	1	—	—	2	1	1	—	2
Synechocystis sp.	3,169	—	1	1	3	1	1	—	1	3	1	2	1	2
Staphylococcus aureus	2,595	—	1	1	1	1	1	—	1	2	1	1	2	2
Streptococcus sp	1,696	—	1	1	1	1	1	—	1	2	1	1	2	2
Treponema pallidum	1,031	—	1	1	2	1	—	—	—	1	1	1	—	2
Thermotoga maritima	1,846	—	1	1	2	1	—	—	—	4	1	1	1	2
Ureaplasma urealyticum	611	—	1	1	1	1	—	—	—	2	1	1	1	2
Vibrio cholerae	2,736	—	1	2	2	1	1	—	1	3	1	1	—	2
Xylella fastidiosa	2,766	—	1	2	1	1	1	—	1	2	1	1	—	2

*Dashes indicate that no homologues have been identified. In order to emphasize that no homologues could be detected, in some interesting cases, zero occurence has been explicitly indicated.

TABLE III. Analysis of Bacterial Small GTP-Binding Subfamily Hits for the Presence of G-1, G-3, and G-4 Motifs Corresponding to the Various GTP-Binding Sequence Motifs in the Classical G-Proteins†

Gene product	Organism	G-1	G-3	G-4	Class	Fold prediction
gi2984130	*A. aeolicus*	+	−	+	Ras	P-loop nucleotide hydrolase
gi2983918	*A. aeolicus*	+	−	+	Arf	P-loop nucleotide hydrolase
gi6458569	*D. radiodurans*	+	−	+	Arf	P-loop nucleotide hydrolase
gi13472827	*M. loti*	+	+	+	Rab	P-loop nucleotide hydrolase
gi10640503	*T. acidophilum*	+	−	+	Ras	P-loop nucleotide hydrolase
gi2621855	*M. thermoautotrophicum*	+	−	+	Rab	P-loop nucleotide hydrolase
gi2621673	*M. thermoautotrophicum*	+	−	+	Arf	P-loop nucleotide hydrolase

†+ and − indicate the presence and absence, respectively, of a given motif in a given gene product.

The careful examination of putative bacterial homologues lacking the G-3 motif revealed two distinct variants of the G-3 motif with TXXG or GXXG replacing the conventional DXXG motif. The known 3-D structures of G-proteins show that the side chain of aspartate in DXXG motif interacts with the Mg^{2+} through a water molecule in the GTP-bound form and directly with the Mg^{2+} in the GDP-bound form.[28,29] The side chain of threonine can potentially play the same role as aspartate in DXXG motif and can be accommodated in place of aspartate. Although the glycyl residue is less likely to be involved in the co-ordination of Mg^{2+}, the involvement of its main chain in a similar role cannot be precluded.

The absence of cysteine-rich motifs, namely CXC and CXXC, the sites of post-translational modifications, at the C-terminus in the bacterial small GTPases suggests that their subcellular localisation may be mainly cytosolic. However, the adapter-mediated translocation of these bacterial small GTPases is still an open question.

Three of the hits, namely gi2649400 (*A. fuldigus*), gi1591981 (*M. jannaschii*), and gi2623036 (*M.thermoautotrophicum*) have been identified previously as members of a small GTP-binding protein subfamily in a separate study.[42] While these relationships have been made with a broad set of small GTP-binding protein subfamilies, the present analysis goes a step further and shows that these hits are more closely related to the Era-like protein (which belongs to GTPase of an unknown function subfamily).

The length of all the putative bacterial small G-protein homologues is confined to less than 200 residues with the only exception (gi2621673) a putative GTPase from *M. thermoautotrophicum*. This gene product contains the GTP-binding domain of 153 residues at the C-terminus. The N-terminus of this gene product shows significant similarity to the KaiC protein involved in circadian rhythm in cyanobacteria.[43] The experimental evidence suggests that KaiC is involved in the regulation of cell division, nitrogen fixation, and photosynthesis.[43] The GTPase domain along with KaiC probably allows fine-tuning of the function of KaiC, acting as a switch between the active and inactive states of the protein.

We have obtained a phylogenetic tree for these hits along with some of the classical small GTP-binding proteins (Fig. 1). All bacterial putative GTPases are separately clustered in the dendrogram with the exception of the putative G-protein from *M. loti*. This gene product also contains the maximal number of conserved GTP-binding sequence motifs that is characteristic of the eukaryotic counterpart and is clustered with eukaryotic small GTPases. Since these putative GTP-binding proteins are distantly related to eukaryotic Ras proteins, there could be profound structural differences and these distinct regions,

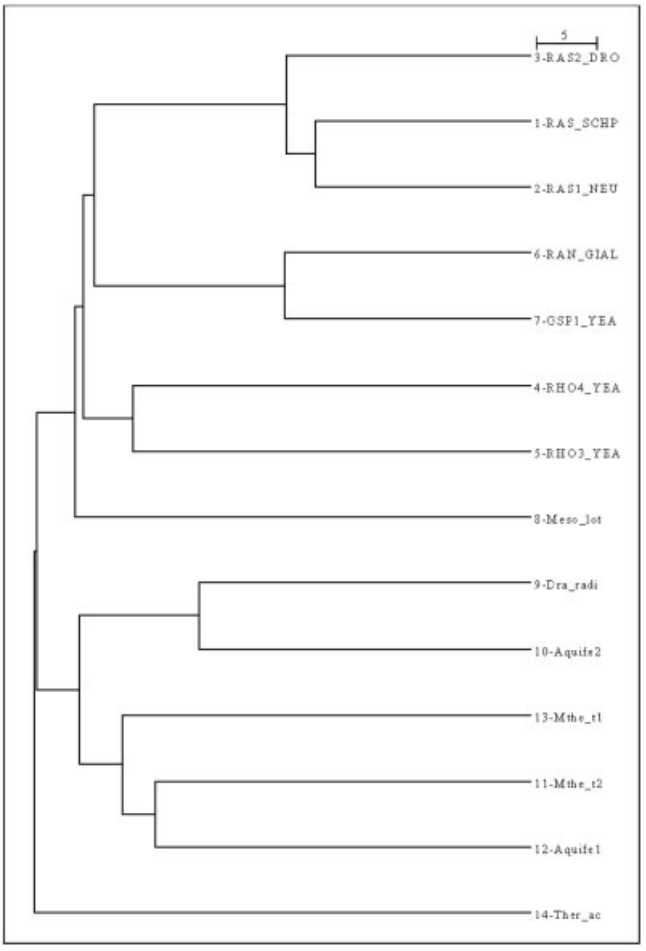

Fig. 1. Dendrogram showing the relationship between eukaryotic and prokaryotic small GTP-binding proteins. RAS2_DRO, RAS_SCHO, RAS1_NEU, RAN_GIAL, GSP1_YEA, RHO4_YEA, and RHO3_YEA are bona fide eukaryotic small GTP-binding proteins; Meso_loti, Dra_radi, Aquife2, Aquife1, Mthe_t1, Mthe_t2, and Ther_ac are from *M. loti, D. radiodurans, A. aeolicus, M. theroautotricum*, and *T. acidophilum*.

apart from the characteristic GTP-binding motifs, might influence their activity.

The prokaryotic genomes have also been surveyed for the occurrences of eukaryotic-like GAP and GEF proteins that are known to regulate GTPase activity. Our search for these accessory proteins did not identify any proteins to be similar to the eukaryotic counterparts. However, the occurrence of prokaryotic GAP and GEF, which are known to be distinct in sequence and structure to the eukaryotic GAP and GEFs, has been reported previously by experimental studies in some of the pathogenic bacteria.[44–47]

Translational GTPases

The members of this subfamily are involved in protein synthesis, which is an essential process for cell survival. These have been well conserved during the evolution and most of them are ubiquitous.[11,48] We have classified the bacterial homologues belonging to this subfamily further into different classes based on their function as Initiation factor, Elongation factor Tu, Elongation factor G, SelB, and Release factor 3. The homologues of LepA and TypA/BipA have also been included as two distinct classes in the above subfamily (Table I). The biological roles of the prokaryotic putative translational GTPases have been suggested based on their similarity in domain combination (Fig. 2) and similarities in inter-domain regions of previously well-characterized translational GTPases of known function.

The initiation factor class includes eIF2γ and IF2/eIF5B, which are involved in the initiation of early events in protein biosynthesis. These proteins promote the formation of the initiation complex by bringing fMet-tRNA$_i$ to the ribosome 30S subunit at the start codon of mRNA.[49] The eubacterial IF2s are functionally equivalent to the eIF2s occurring in archaea as well as eukaryotes. The GTPase domain of IF2 shows sequence similarity with the eIF2γ-subunit of the eukaryotic and archaeal eIF2 complex.[50] Recently, the homologue of IF2, also referred to as eIF5B, has been identified in archae[51] and in eukaryotes such as yeast[52] and human.[53] Although the function of eIF5B/IF2 in the eukaryotes and archaea is not clear, experimental evidence suggests that it is required during the translational initiation and association of ribosomal subunits.[54] The deletion mutant of eIF5B/IF2 results in a slow growth phenotype in yeast and is not lethal, suggesting a less critical role compared to the well-studied eIF2s in the yeast.[51,52] Every archaeal genome analyzed encodes two homologues of this class. One of these homologues is closely related to IF2 and the other homologue shares significant sequence similarity with eIF2γ. However in all eubacterial genomes surveyed, only one IF2 has been identified and no eIF2γ homologues were detected (Table II). The dendrogram of the identified IF2 and eIF2γ homologues shows distinct clusters though they are involved in similar functions (Fig. 3).

Further, we have analyzed the extent of the conservation of GTP-binding motifs (G-1, G-3, G-4) in the prokaryotic IF2s and eIF2γs identified. While G-3 and G-4 motifs are well conserved in all the initiation factors, a variation has been observed in the G-1 motif of a few of the initiation factors. The following putative initiation factors, gi5458562 (*P. abysii*), gi1590990 (*M. jannaschii*), and gi3257511 (*P. horikosii*) have their G-1 motif modified as GXXXXGKC in place of the more typical GXXXXGK(T/S) motif. The G-1 motif in these cases is also located at an unusual distance (in the primary structure) with respect to G-2 with a huge insertion of about 400 amino acids between the G-1 and G-2 motifs. The substitution of cysteine for threonine in the G-1 motif has been observed in these exceptional cases. Threonine in the canonical motif mediates the interactions with α and β oxygen of GTP through the main chain amide. This substitution is, hence, unlikely to be drastic and can be well accommodated without the disruption of polar interactions. In addition, the weakly polar thiol group of cysteine has potential to form a weak hydrogen bond with β phosphate oxygen and can coordinate metal ions similar to the interaction mediated by the side chains of Ser/Thr. The G-4 motif occurs as NKXE, instead of the canonical NKXD, in gi3258137 (*P. horikosii*), which has been identified as putative eIF2γ. Here the replacement of Asp by Glu is unlikely to affect this interaction since both are acidic.

The elongation factor Tu (EF-Tu) class includes EF-Tu/EF1A, which participates in the elongation step wherein they are involved in carrying the aminoacylated tRNA to the mRNA programmed ribosome P-site.[55] We have identified members of this class to be present in all the archaebacteria and eubacteria (Table II) as these are involved in the critical function of protein synthesis. In the following genomes, *A. fuldigus, M. thermoautotrophicum, M. jannaschii, Halobacterium sp, A. aeolicus, D. radiodurans, E. coli, H. influenzae, M. loti, N. meningitidis, P. multicoda, P. aeruginosa, V. cholerae*, and *X. fastidiosa*, more than one copy of EF-Tu is present. The putative EF-Tu identified in these genomes showed the conserved G-1, G-3, and G-4 motifs except in some of the sequences where asparagine of the G-4 motif is substituted by threonine or serine. This substitution can be accommodated since Asn is mostly involved in hydrogen bond interaction and Ser/Thr also has the potential to form such hydrogen bonds.

The elongation factor G (EF-G) class includes EF-G/EF2 and catalyzes the translocation step, wherein after peptide bond formation, a peptidyl-tRNA moves from the A site to the P site with the release of deacylated t-RNA from the P-site.[56] Members of this class are identified in all the genomes surveyed (Table II). This can be understood from the fact that they are essential for the survival of the cell. In *B. burgdorrferi, D. radiodurans, L. lactis, M. tuberculosis, M. loti, P. aeruginosa, Synechocystis sp., T. pallidum, T. maritima*, and *V. cholerae*, we could identify EF-G paralogues, which would perform a similar function as EF-G. These putative EF-Gs have all the GTP-binding motifs conserved except in some variants where asparagine of the G-4 motif is substituted by threonine or serine. One of the putative EF-G, gi10582035 (*Halobacterium sp.*), does not have a G-1 motif. The mode of binding of this gene product to guanine nucleotides is, hence, unclear.

The SelB class includes selenocysteine-specific elongation factor SelB, which is required for incorporation of the

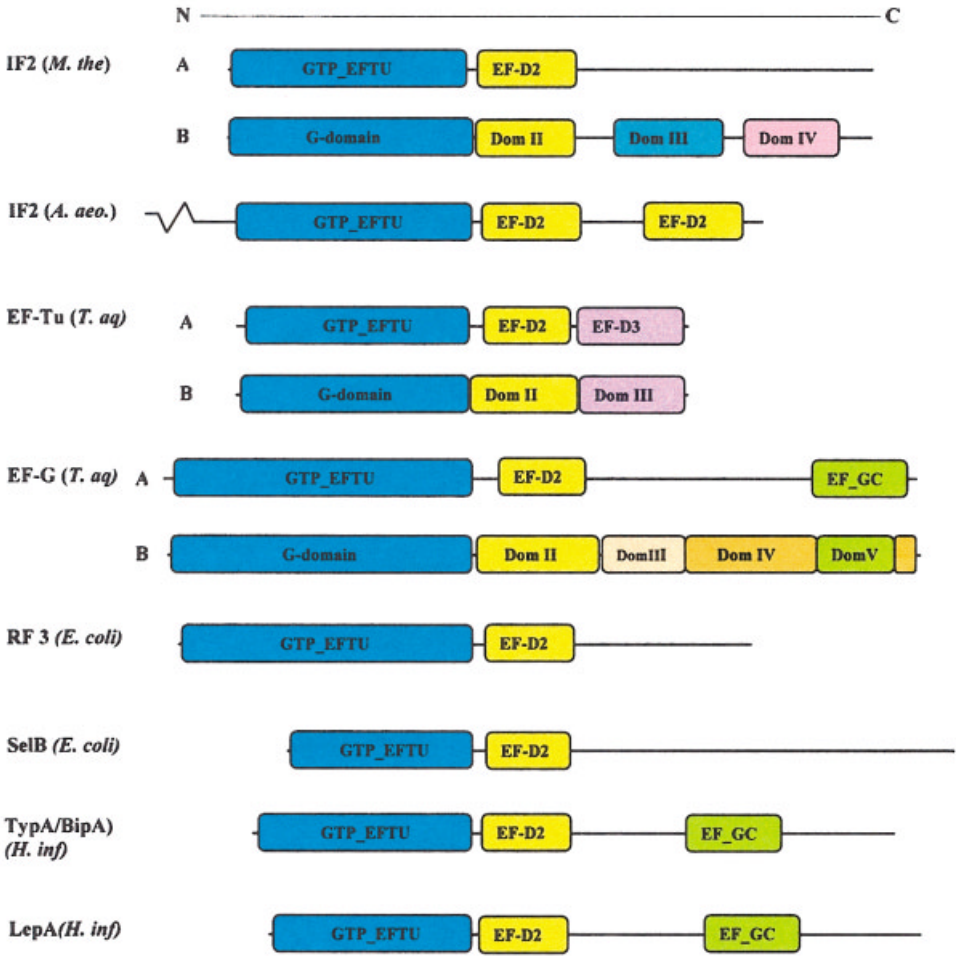

Fig. 2. Schematic representation showing the domain arrangement of various translational GTPases. The domain arrangement shown here corresponds to the functional domain given in Pfam database. For the proteins with known three-dimensional structures, the structural domain boundaries[87] are also shown below the functional domain organization. Some of the functional domains may correspond to the structural domain. For example, G-domain (structural domain) corresponds to GTP_EFTU (Pfam domain), domain II (structural domain) corresponds to GTP_EF-TU_ D2 (Pfam domain), GTP_EF-TU_D3 (Pfam domain) corresponds to domain III (structural domain) in EF-TU. EFG_C (Pfam domain) corresponds to domain V (structural domain) in EF-G. But, in many cases the structural domain may not correspond to the functional domain. For example, domain III and domain IV of EF-G does not have an equivalent Pfam domain. The same color shaded box represents the same or an equivalent domain. The IF2 of *A. aeo.* has a long N-terminal domain (shown as a wavy line) followed by the G-domain. Due to the insertion of the G′ domain in EFG and RF3, the G-domain in these proteins is longer than that of other translational GTPases. *A. aeo., Aquifex aeolicus*; *M. the*, *Methanobacterium thermoautotropicum*; *T. aq*, *Thermus aquaticus*; *E. coli*, *Escherichia coli*; *H. inf*, *Haemophilus influenzae*. IF2, Initiation factor 2; EF-Tu, Elongation factor Tu; EF-G, Elongation factor G; SelB, Selneocysteine factor specific elongation factor; RE 3, release factor3; TypA/BipA, Tyrosine phosphorylated protein A. EFD3 is EF-TU C-terminal domain and EF_GC is EFG C-terminal domain.

unusual amino acid selenocysteine at the in-frame (UGA) codon in mRNA during translation of selenoproteins.[57] The SelB shows sequence similarity with EF-Tu at the N-terminus and is longer than EF-Tu. The C-terminus of SelB is known to be involved in the recognition of mRNA secondary structures specific for selenocyteine insertion.[58] We have identified SelB only in 7 of the bacterial genomes (1 archae and 6 eubacteria) as tabulated in Table II. The synthesis of selenoproteins is restricted to either anaerobic growth condition or chemical environment protected from oxygen, since selenocyteine gets readily oxidized.[57] The absence of SelB from most of the genomes suggests that the SelB gene is selectively transferred or retained during the course of evolution in the organisms capable of growth in an anaerobic or chemical environment protected from oxygen where selenoproteins become indispensable for survival. Interestingly, we have identified the DEP domain towards the C-terminus in two of the putative SelB proteins from *H. influenzae* (gi1573710) and *P. multicoda* (gi12722183). While the functional role of DEP is unknown, in Drosophila it rescues polarity defects and induces JNK signaling.[59] The DEP domain most often occurs in signalling proteins that contain PH, rasGEF, rhoGEF, rhoGAP, RGS, and PDZ domains. Hence the

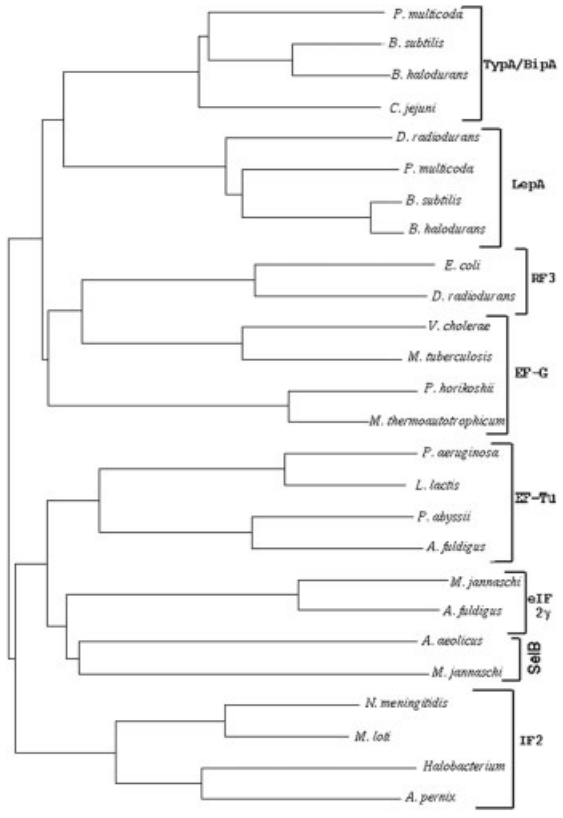

Fig. 3. The dendrogram of representative translational GTPases identified in prokaryotes shows different class members group into separate clusters. This tree is constructed considering solely the GTP-binding region.

presence of the DEP domain in prokaryotes is unexpected since it is known to occur only in the eukaryotic signalling proteins mentioned above. All the putative SelB proteins identified have well-conserved characteristic GTP-binding motifs except for a few cases where the G-4 motif has Thr/Ser in place of asparagine in the typical G-4 motif.

The release factor class includes release factor 3 (RF3), which is involved in the termination step of the protein synthesis in prokaryotes.[60] RF3-GTP promotes the release of RF1 and RF2 from the ribosome following peptidyl t-RNA hydrolysis at termination.[61] The translation termination step is highly regulated in eukaryotes with eRF3 performing the same function as RF3. The eRF3 has been shown to be essential for the survival of the eukaryotes.[60,62] In the current analysis, RF3 have been identified in 14 eubacterial genomes (Table II). None of the archaebacterial genomes surveyed have any representatives from this class of translational GTPases. The RF3 shares significant sequence similarity with EF-G, an elongation factor, and probably employs a similar mechanism as EF-G[11,63] for the release of the polypeptide at the termination step. The absence of RF3s in archae and a significant number (20) of the eubacterial genomes surveyed, therefore, suggest the EF-Gs could play a role of RF3 in the release of polypeptides.[63] All putative RF3 proteins identified in the present analysis have GTP-binding motifs conserved except that the G-1 motif has been conserved as SXXXXGK[T/S] instead of GXXXXGK[T/S]. This change is unlikely to affect the ability of these proteins to bind to GTP.

The LepA class of proteins includes LepA proteins, which shows sequence similarity to EF-G, but the functions of these are not well characterized experimentally.[64] It is probably involved in the regulation of translation although it has been shown not to be critical for the survival of the cell.[65] Among eukaryotes, yeast has been shown to have LepA homologues.[66] We have identified LepA members only in eubacterial genomes (Table II). Most of these putative LepA homologues have the GTP-binding motifs present except in a few cases where the G-4 motif occurs as [T/S]KXD instead of the NKXD motif. One of the homologues, gi2687964 (*B. burgdorrferi*), lacks the G-1 motif.

The TypA (Tyrosine phosphorylated protein A)/BipA class is comprised of GTP-binding proteins that get phosphorylated in the cell but the details of their functions are not completely understood.[67–69] This class of proteins also shares overall sequence similarity to EF-G. Certain experimental evidence suggests that it interacts with ribosome in a GTP-dependent manner and shows a novel mechanism of regulation of the expression of the target protein.[69] We have identified this protein in most of the eubacterial genomes. The representatives of this class of proteins have not been identified in the genomes of *A. aeolicus, B. burgdorrferi, C. muridarum, C. trachomatis, C. pneuomiae* (*CWL 1089*), *C. pneuomiae* (*AR39*), *Chlamydia pneuomiae* (*J138), M. genitalium, M. pneumoniae, T. pallidum, T. maritima* and *U. urealyticum* (Table II) and the archaebacterial genomes. Most of the putative TypA/BipA protein showed the presence of GTP-binding motifs. The phylogenetic tree (Fig. 3) construction using representative sequences of the identified translation GTPases shows that various class members are closely related and the classes form distinct clusters.

The switching ability of G-proteins with GTP hydrolysis, and the concomitant conformational change, have been exploited in protein synthesis and all the critical steps have proteins with conserved GTP-binding domains. These translational GTPases, apart from the GTP-binding domain, also contain one other conserved structural domain (Domain II),[70] which is equivalent to the EF-Tu 2 domain described in Pfam. These two domains are also known to interact with ribosome.[56] The tandem occurrence of these two domains in all the translational GTPases suggests them to be conserved evolutionarily as a structural unit[70] for interaction with ribosomes. The other domains associated with this structural unit in the various translational GTPases may confer the specificity towards various target proteins involved in translation.

New Subfamilies of G-Proteins

The other GTP-binding proteins, which show no significant sequence similarity with the currently known subfamilies, form new subfamilies of G-proteins. We have classified the prokaryotic representatives of these subfamilies

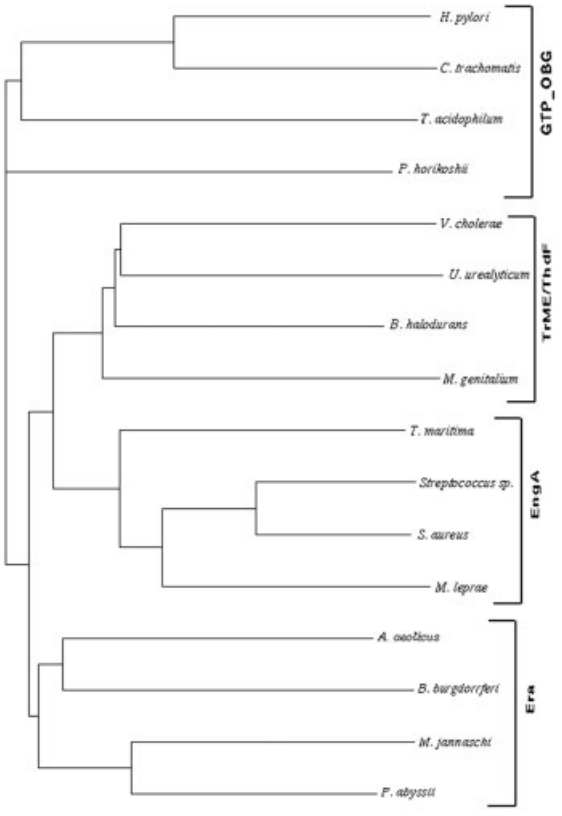

Fig. 4. The phylogenetic tree of representative prokaryotic members, identified in the present study, that are assigned to two new subfamilies. The two subfamilies, GTP_OBG and GTPase of unknown function, form distinct clusters. Note that Era, TrmE/ThdF, and EngA correspond to the sub-clusters in the subfamily of GTPase of unknown function. Hypothetical GTP-binding protein has been omitted from tree construction because of a circularly permutated motif that effects the quality of the multiple sequence alignment. This tree is constructed considering solely the GTP-binding region.

into GTPase of unknown function and GTP_OBG as described in the Pfam[26] classification.

The subfamily of GTPase of unknown function has been classified into four classes namely Era (*Escherichia coli* Ras-like), EngA (Essential neisserial GTP-binding protein), TrmE (t-RNA modification E)/ThdF, and hypothetical GTP-binding proteins. Using profiles of these classes, homologues from prokaryotic genomes have been identified and classified appropriately. In general, greater representation of these homologues has been observed in the eubacterial genomes compared to the archaeal genomes. The dendrogram of the identified homologues shows that different classes of this subfamily form distinct groups within the broad cluster of the subfamily of GTPase of unknown function and the GTP_OBG subfamily forms a separate cluster (Fig. 4).

The function of the Era class of proteins remains elusive but experimental evidence suggests that Era regulates the cell cycle by coupling cell growth rate with cytokinesis.[21,22,64] There is other experimental evidence that indicates it is also involved in regulating carbon/nitrogen metabolism[20,71] and are also proposed to be essential for cell survival.[21,22,71] The Era C-terminal has a pseudo KH domain (which is present in many RNA-binding proteins) and has been shown to bind to 16S RNA specifically.[72–74] The Era-like proteins are also present in eukaryotes such as *C. elegans*, yeast, human, and mouse.[75] We could identify the Era-like protein in all the genomes surveyed (Table II) as expected, since loss of the Era gene in *E.coli* is shown to be lethal.[21,22,71] We have also identified paralogues of the Era-like protein in many genomes, which is suggested to have resulted from gene duplication events. Era-like proteins, previously known to occur only in two archaeal genomes,[32] have now been identified in all the 8 archaeal genomes surveyed. Their genome-wide occurrence in archaea as well as eubacteria reiterates their critical roles in important functions, although to the best of our knowledge there have been no experimental studies on Era-like functions in archaea. The identified archaebacterial homologues cluster with the Era class of proteins as shown in Figure 4. We also surveyed for the occurrence of the KH domain in all the putative Era-protein-like gene products. Using HMMER, we could identify one gene product from each of *B. burgdorrferi*, *B. holandurans*, *B. subtilis*, *C. jejuni*, *D. radiodurans*, *H. pylori*, *H. pylori* (*st 99*), *L. lactis*, *M. tuberculosis*, *N. meningitis*, *S. aureus*, *Streptococcus*, and *T. martiama* that has the KH domain. The absence of the KH domain in other homologues of Era proteins suggests that either the occurrence may not be essential for the activity of Era proteins or the sequence has diverged beyond recognition by our search procedures.

The EngA class of proteins has been shown to be essential for growth in *N. gonorrohoeae*.[48,76] It has two tandem GTP-binding domains, in contrast to all other G-proteins that have one GTP-binding domain, which gives rise to the possibility of a tandem switch and fine regulation of cellular events by EngA.[64] We have identified EngA-like proteins in all the eubacterial genomes. The absence of EngA homologues from archae and eukaryotes[64] suggests them to be unique to the eubacterial lineage with their probable participation in specific functions restricted to eubacteria.

Members of the subfamily TrmE/ThdF are probably involved in tRNA modification.[48,77] It encodes an enzyme involved in the biosynthesis of 5-methylaminomethyl-2-thiouridine, a nucleoside found in the wobble position of some t-RNAs.[48,78,79] The GTP-binding domain in TrmE/ThdF is located at the C-terminus and it was shown that the amino terminus could be removed without affecting GTPase activity.[77] The last four amino acids of the TrmE/ThdF gene products are conserved either as a CIGK or CLGK motif that resembles the C-Al-Al-X (where Al is any aliphatic amino acid and X is any amino acid) motif, raising the possibility of lipid modification.[64] We have identified TrmE/ThdF homologues in all eubacterial genomes except in *C. pneumoniae* (*AR39*). In addition, homologues have been identified in two archaebacterial genomes: gi10581976 (*Halobacterium sp.*) and gi2650203 (*A. fuldigus*). In archaebacteria, to our knowledge, so far there is no report of the presence of homologues of TrmE/ThdF. The presence of this gene in archae suggests the emergence of this class of proteins before the diversifica-

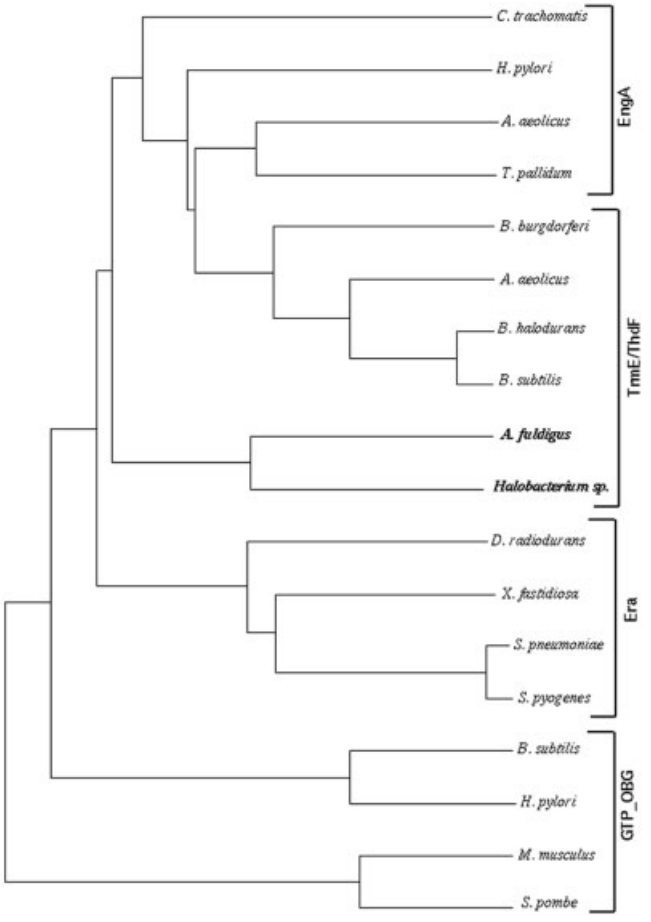

Fig. 5. Dendrogram showing the two TrmE/ThdF homologues identified in archae (shown in bold) that cluster with other members of the subfamily of GTPase of unknown function. The dendrogram has been constructed by considering the experimentally studied prokaryotic and eukaryotic homologues of GTPase of unknown function and GTP_OBG. Sequences of the GTP-binding regions alone are used in generating this dendrogram.

tion of archaebacterial and eubacterial lineages. The conserved C-terminal C-Al-Al-X motif found in all other eubacterial homologues is absent in the archaeal sequences. The relatedness of two identified archaebacterial members has been explored further using phylogenetic analysis (Fig. 5), which shows that these two homologues cluster with the subfamily of GTPase of unknown function.

In addition to the homologues of members belonging to the G-protein subfamilies described above, 23 gene products have been identified that cannot be assigned to any of the known subfamilies. They belong to the class of hypothetical GTP-binding proteins whose function is not yet known. The organisms encoding these gene products and their distribution are listed in Table II.

The GTP-binding motifs of the various members of the different classes of GTPase of unknown function are well conserved with a few exceptions. The variations are predominantly seen in the G-4 motif wherein Ser/Thr replaces the Asn in the NKXD motifs. The absence of the NKXD motif has also been observed in some cases.

An interesting variation in the order of occurrence of the various GTP-binding motifs along the primary structure of the putative hypothetical GTP-binding proteins has been observed in the analysis. The order of the conserved GTP binding motif is circularly permutated with order being G-4–G-1–G-3 instead of G-1–G-3–G-4. The homologues with such circularly permutated motifs include: gi1592105 (*M. jannaschi*), gi5458812 (*P. abyssii*), gi3257053 (*P. horikoshii*), gi10173939 (*B. halodurans*), gi10175096 (*B. halodurans*), gi12724271 (*L. lactis*), gi12723078 (*L. lactis*), gi7225967 (*N. meningitides*), gi1653032 (*Synechocystis sp.*), gi13701395 (*S. aureus*), gi13701043 (*S. aureus*), gi13621572 (*Streptococcus sp.*), gi13622297 (*Streptococcus sp.*), gi4981381 (*T. martiama*), gi4981296 (*T. martiama*), and gi6899603 (*U. urealyticum*). The occurrences of such a circularly permutated motif containing protein have been reported recently by Leipe et al.[32] where they group them into the YawG/YlqF subfamily of proteins. The function of these proteins is presently unknown in prokaryotes. The conservation of the 3-D disposition of the residues forming the GTP-binding pocket in these proteins may be possible despite variations in the order of occurrence at the sequence level.

The GTP_OBG subfamily includes Obg and DRG proteins. The Obg proteins are involved in the initiation of replication[80] and the initiation of sporulation or differentiation.[81] In *Streptomyces coeliocolor,* Obg has been shown to be a regulator for the onset of cellular differentiation.[82] In addition, Obg is also proposed to be the sensor of levels of GTP in the cell[83] and is implicated in regulation of the stress-induced genes. This is exemplified by the Obg of *B. subtilis* that activates σ^B, a transcription factor that controls the general stress response regulon[84] and also binds ribosomal protein L13 specifically.[85] DRG proteins are eukaryotic homologues of Obg and they play a critical role in cell proliferation, differentiation, and death.[17–19] We have identified members of this subfamily in all the prokaryotic genomes (Table II). The presence of Obg-like proteins in all the prokaryotic genomes suggests that Obg act as a general indicator of the cell developmental status in all the prokaryotes. Interestingly, the archaebacterial homologues contain an additional domain referred to as TGS domain, at their C-terminus. A similar domain organization is seen in Obg homologue in eukaryotes. The function of the TGS domain is unclear, but its presence in regulatory proteins such GTPases and guanosine polyphosphate phosphohydrolase/sythetases suggests a ligand (most likely nucleotide)-binding regulatory role.[86] These putative Obg-like proteins have GTP-binding motifs with a small variation of the G-4 motif where the Asp of the NKXD is replaced by the Glu residue. Variations have also been observed in the G-3 motif with a Glu residue replacing the Asp residue in the DXXG motif. Since the nature of substitution involves amino acids of similar chemical properties, the recognition and accommodation of the guanine nucleotide are not likely to be affected. The GTP_OBG subfamily forms separate clusters in the dendrogram generated using the identified representatives of the

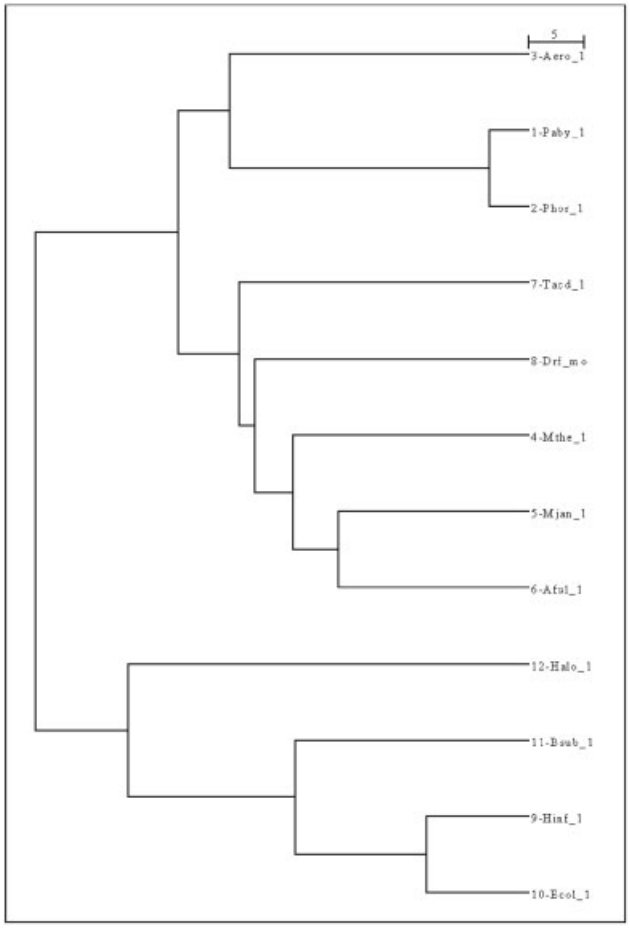

Fig. 6. The dendrogram of Obg-like proteins showing their separate clustering of eubacterial proteins away from archaebacterial and eukaryotic orthologues. Aero_1, Paby_1, Phor_1, Tacd_1, Mthe_1, Mjan_1, Aful_1, and Halo_1 are Obg-like proteins from archeabacteria *A. pernix, P. abysii, P. horikosii, T. acidophilum, M. thermoautotricum, M. jannaschii, A. fuldigus, Halobacterium sp.* respectively. Bsub_1, Hinf_1, Ecol_1 are Obg-like proteins from eubacteria *B. subtilis, H. influenzae,* and *E. coli*, respectively. Drf_mo is an eukaryotic Obg orthologue.

two subfamilies, namely, GTPase of unknown function and GTP_OBG (Fig. 4).

The phylogenetic analysis of this subfamily of proteins was carried out and the dendrogram is shown in Figure 6. The dendrogram shows that archae and eukaryote Obg-like proteins cluster together while the eubacterial homologues form a distinct group. The only exception is the variant from *Halobacterium sp.* that clusters with eubacterial Obg-like proteins. The domain arrangement of Obg from archae is the same as in the DRG of eukaryotes and it suggests a close evolutionary relationship between the two.

We could not identify homologues of the α-subunit of heterotrimeric G-protein and the large GTP protein subfamily of G-protein in prokaryotes. The probable absence of these proteins in prokaryotes could be explained from the fact that they are involved in very specific functions in the eukaryotes, which might not be required in prokaryotes.

CONCLUSIONS

The present study suggests a lack of correlation between the extent of the occurrence of G-protein members in a particular prokaryotic genome and their proteome size. For example *Buchnera sp.* comprises 12 G-proteins in a total of 564 ORFs, while *C. jejuni* has 12 G-proteins in 1,634 ORFs. This indicates that evolution and expansion of G-proteins have been restricted to certain organisms based on their functional requirements.

The G-proteins are involved in various functions such as signal transduction, cellular differentiation, and protein synthesis. Our survey for their homologues in prokaryotes has enabled us to understand the functional repertoire of G-proteins in these lower organisms. The translational GTPase have the largest representation in prokaryotes. Among the translational GTPases IF2, EF-Tu, and EF-G have representation in all the prokaryotes analyzed while SelB, RF3, and TypA/BipA are represented in some genomes and LepA members are restricted to eubacteria. The small GTP-binding proteins are poorly represented with only seven homologues identified. In the small GTP-binding proteins, the Ras, Arf, and Rab classes have been represented in prokaryotes while Ran and Rho/Rac/Cdc42 homologues are not detected. This observation is consistent with their specialized function in eukaryotes. One of the small GTP-binding protein homologues from *M. thermoautotrophicum* has N-terminal sharing sequence similarity to the KaiC protein involved in circadian rhythm. This might allow fine-tuning of circadian rhythm with GTP hydrolysis provided by the GTP-binding domain towards the C-terminus. Our survey also led to the identification of Era homologues in all archaebacterial genomes, which were previously known to occur in two archaea. This suggests that Era homologues have been retained in archaea and are involved in important functions in archaea as well. The TrmE/ThdF members have been identified in two archaebacterial genomes and these occurrences were not known to us previously. Archaea might have acquired these genes from eubacteria or been selectively retained in these genomes during the divergence of eubacteria from archaebacteria.

The phylogenetic analysis of identified G-proteins in the present study such as small GTPase and Obg proteins suggests that archaeal members are more closely related to those from eukaryotes than to eubacterial members. This is also evident from the domain combinations of Obg homologues that are similar in eukaryotes and archaebacterial homologues. Dendrograms of identified homologues of translational factors and GTPase of unknown function shows that various classes forms distinct clusters.

Following the classification scheme and evolution analysis of Leipe et al.[32] on the GTPases and related ATPases and the classification of G-proteins in Pfam, the present study reports many new variations in the G-protein family. As our analysis is confined solely to those prokaryotic genomes that have complete genomic data, it provides a comprehensive collection of G-proteins in these organisms. The deductions of their functions provide an opportunity for the experimentalists focusing on specific prokaryotes or

G-proteins to explore previously unidentified G-proteins in prokaryotes. These studies will enable a better understanding of the functional diversity of G-proteins in prokaryotes.

SUPPLEMENTARY INFORMATION

The complete list of all the G-proteins analyzed along with their gene identification (gi) codes, regions of domains, as well as subfamily and class assignments are provided in Supplementary Information (http://www.interscience.wiley.com/jpages/0887-3585/suppmat/).

ACKNOWLEDGMENTS

We thank Ms. S. Sujatha for her help in the initial stages of this work and Ms. A. Krupa for a critical reading of the manuscript. S.B.P. is supported by a fellowship from the Council of Scientific and Industrial Research, India. This research is supported by an International Senior Fellowship to N.S. by the Wellcome Trust, London, and by the computational genomics project sponsored by the Department of Biotechnology, New Delhi.

REFERENCES

1. Bourne HR, Sanders DA, McCormick F. The GTPase superfamily: a conserved switch for diverse cell functions. Nature 1990;348:125–132.
2. Bourne HR, Sanders DA, McCormick F. The GTPase superfamily: conserved structure and molecular mechanism. Nature 1991;349:117–127.
3. Boguski MS, McCormick F. Proteins regulating Ras and its relatives. Nature 1993;366:643–654.
4. Matozaki T, Nakanishi H, Takai Y. Small G-protein networks: their crosstalk and signal cascades. Cell Signal 2000;12:515–524.
5. Downward J. The ras superfamily of small GTP-binding proteins. Trends Biochem Sci 1990;15:469–472.
6. Valencia A, Chardin P, Wittinghofer A, Sander C. The ras protein family: evolutionary tree and role of conserved amino acids. Biochemistry 1991;30:4637–4648.
7. Gorlich D. Transport into and out of the cell nucleus. EMBO J 1998;17:2721–2727.
8. Takai Y, Sasaki T. Matozaki T. Small GTP-binding proteins. Physiol Rev 2001;81:153–208.
9. Magee AI, Newman CM, Giannakouros T, Hancock JF, Fawell E, Armstrong J. Lipid modifications and function of the ras superfamily of proteins. Biochem Soc Trans 1992;20:497–499.
10. Gilman AG. G proteins: transducers of receptor-generated signals. Annu Rev Biochem 1987;56:615–649.
11. Laalami S, Grentzmann G, Bremaud L, Cenatiempo Y. Messenger RNA translation in prokaryotes: GTPase centers associated with translational factors. Biochimie 1996;78:577–589.
12. Hinshaw JE. Dynamin and its role in membrane fission. Annual Rev Cell Dev Biol 2000;16:483–519.
13. Damke H, Baba T, Warnock DE, Schmid SL. Induction of mutant dynamin specifically blocks endocytic coated vesicle formation. J Cell Biol 1994;127:915–934.
14. Van der Bliek AM. Functional diversity in the dynamin family. Trends in Cell Biol 1999;9:96–102.
15. Cheng YS, Colonno RJ, Yin FH. Interferon induction of fibroblast proteins with guanylate binding activity. J Biol Chem 1983;258:7746–7750.
16. Staeheli P, Pitossi F, Pavlovic J. Mx proteins: GTPases with antiviral activity. Trends Cell Biol 1993;3:268–272.
17. Schenker T, Lach C, Kessler B, Calderara S, Trueb BA. Novel GTP-binding protein which is selectively repressed in SV40 transformed fibroblasts. J Biol Chem 1994;269:447–453.
18. Sazuka T, Tomooka Y, Ikawa Y, Noda M, Kumar S. DRG: a novel developmentally regulated GTP-binding protein. Biochem Biophys Res Commun 1992;189:363–370.
19. Sazuka, T, Kinoshita M, Tomooka Y, Ikawa Y, Noda M, Kumar S. Expression of DRG during murine embryonic development. Biochem Biophys Res Commun 1992;189:371–377.
20. Lerner CG, Inouye M. Pleiotropic changes resulting from depletion of Era, an essential GTP-binding protein in *Escherichia coli*. Mol Microbiol 1991;5:951–957.
21. Britton RA, Powell BS, Dasgupta S, Sun Q, Margolin W, Lupski JR, Court DL. Cell cycle arrest in Era GTPase mutants: a potential growth rate-regulated checkpoint in *Escherichia coli*. Mol Microbiol 1998;27:739–750.
22. Gollop N, March PE. A GTP-binding protein (Era) has an essential role in growth rate and cell cycle control in *Escherichia coli*. J Bacteriol 1991;173:2265–2270.
23. Lin B, Covalle KL, Maddock JR. The Caulobacter crescentus CgtA protein displays unusual guanine nucleotide binding and exchange properties. J Bacteriol 1999;181:5825–5832.
24. Corpet F, Servant F, Gouzy J. Kahn D. ProDom and ProDom-CG: tools for protein domain analysis and whole genome comparisons. Nucleic Acids Res 2000;28:267–269.
25. Schultz J, Copley RR, Doerks T, Ponting CP, Bork P. SMART: a web-based tool for the study of genetically mobile domains. Nucleic Acids Res 2000;28:231–234.
26. Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL. The Pfam protein families database. Nucleic Acids Res 2000;28:263–266.
27. Sprang SR. G protein mechanisms: insights from structural analysis. Annual Rev Biochem 1997;66:639–678.
28. Pai EF, Krengel U, Petsko GA, Goody RS, Kabsch W, Wittinghofer A. Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 A resolution: implications for the mechanism of GTP hydrolysis. EMBO J 1990;9:2351–2359.
29. Schlichting I, Almo SC, Rapp G, Wilson K, Wittinghofer A, Goody RS. et al. Time-resolved X ray crystallographic study of the conformational change in Ha-Ras p21 protein on GTP hydrolysis. Nature 1990;345:309–315.
30. Zhong JM, Chen-Hwang MC, Hwang YW. Switching nucleotide specificity of Ha-Ras p21 by a single amino acid substitution at aspartate 119. J Biol Chem 1995;270:10002–10007.
31. Kjeldgaard M, Nyborg J, Clark BF. The GTP binding motif: variations on a theme. FASEB J 1996;10:1347–1368.
32. Leipe DD, Wolf YI, Koonin EV, Aravind L. Classification and evolution of P-loop GTPases and related ATPases. J Mol Biol 2002;317:41–72.
33. Altschul SF, Madden ML, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman, DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
34. Schaffer AA, Wolf YI, Ponting CP, Koonin EV, Aravind L, Altschul SF. IMPALA: matching a protein sequence against a collection of PSI-BLAST-constructed position-specific score matrices. Bioinformatics 1999;12:1000–1011.
35. Eddy SR. Profile hidden Markov models. Bioinformatics 1998;14:755–763.
36. Pandit SB, Gosar D, Abhiman S, Sujatha S, Dixit SS, Mhatre NS, Sowdhamini R, Srinivasan N. SUPFAM: a database of potential protein superfamily relationships derived by comparing sequence-based and structure-based families: implications for structural genomics and function annotation in genomes. Nucleic Acids Res 2002;30:289–293.
37. Balaji S, Sujatha S, Kumar SSC, Srinivasan N. PALI-a database of Phylogeny and ALIgnment of homologous protein structures. Nucleic Acids Res 2001;29:61–65.
38. Felsenstein J. PHYLIP (Phylogeny Inference Package), version 3.5c. Distributed by the author. 1993. Department of Genetics, University of Washington, Seattle.
39. Johnson MS, Overington JP, Blundell TL. Alignment and searching for common protein folds using a data bank of structural templates. J Mol Biol 1993;231:735–752.
40. Hartzell PL. Complementation of sporulation and motility defects in a prokaryote by a eukaryotic GTPase. Proc Natl Acad Sci USA 1997;94:9881–9886.
41. Jones DT. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. J Mol Biol 1999;287:797–815.
42. Ponting CP, Aravind L, Schultz J, Bork P, Koonin EV. Eukaryotic signalling domain homologues in archaea and bacteria: ancient ancestry and horizontal gene transfer. J Mol Biol 1999;289:729–745.
43. Lorne J, Scheffer J, Lee A, Painter M, Miao VP. Genes controlling

circadian rhythm are widely distributed in cyanobacteria. FEMS Microbiol Lett 2000;189:129–133.
44. Nagai H, Kagan JC, Zhu X, Kahn RA, Roy CR. A bacterial guanine nucleotide exchange factor activates ARF on Legionella phagosomes. Science 2002;295:679–682.
45. Friebel A, Ilchmann H, Aepfelbacher M, Ehrbar K, Machleidt W, Hardt WD. SopE and SopE2 from Salmonella typhimurium activate different sets of Rho GTPases of the host cell. J Biol Chem 2001;276:34035–34040.
46. Stebbins CE, Galan JE. Structural mimicry in bacterial virulence. Nature 2001;412:701–705.
47. Evdokimov AG, Tropea JE, Routzahn KM, Waugh DS. Crystal structure of the Yersinia pestis GTPase activator YopE. Protein Sci 2002;11:401–408.
48. Mittenhuber G. Comparative genomics of prokaryotic GTP-binding proteins (the Era, Obg, EngA, ThdF (TrmE), YchF and YihA families) and their relationship to eukaryotic GTP-binding proteins (the DRG, ARF, RAB, RAN, RAS and RHO families). J Mol Microbiol Biotechnol 2001;3:21–35.
49. Laalami S, Sacerdot C, Vachon G, Mortensen K, Sperling-Petersen HU, Cenatiempo Y, Grunberg-Manago M. Structural and functional domains of *E. coli* initiation factor IF2. Biochimie 1991;73:1557–1566.
50. Pain VM. Initiation of protein synthesis in eukaryotic cells. Eur J Biochem 1996;236:747–771.
51. Kyrpides NC, Woese CR. Archaeal translation initiation revisited: the initiation factor 2 and eukaryotic initiation factor 2B alpha-beta-delta subunit families. Proc Natl Acad Sci USA 1998;95:3726–3730.
52. Choi SK, Lee JH, Zoll WL, Merrick WC, Dever TE. Promotion of met-tRNAiMet binding to ribosomes by yIF2, a bacterial IF2 homolog in yeast. Science 1998;280:1757–1760.
53. Lee JH, Choi SK, Roll-Mecak A, Burley SK, Dever TE. Universal conservation in translation initiation revealed by human and archaeal homologs of bacterial translation initiation factor IF2. Proc Natl Acad Sci USA 1999;96:4342–4347.
54. Roll-Mecak A, Shin BS, Dever TE, Burley SK. Engaging the ribosome: universal IFs of translation. Trends Biochem Sci 2001; 26:705–709.
55. Krab IM, Parmeggiani A. EF-Tu, a GTPase odyssey. Biochim Biophys Acta 1998;1443:1–22.
56. Jurnak F. The ABC of EF-G. Structure 1994;2:785–788.
57. Low SC, Berry MJ. Knowing when not to stop: selenocysteine incorporation in eukaryotes. Trends Biochem Sci 1996;21:203–208.
58. Kromayer M, Wilting R, Tormay P, Bock A. Domain structure of the prokaryotic selenocysteine-specific elongation factor SelB. J Mol Biol 1996;262:413–420.
59. Boutros M, Paricio N, Strutt DI, Mlodzik M. Dishevelled activates JNK and discriminates between JNK pathways in planar polarity and wingless signaling. Cell 1998;94:109–118.
60. Poole E, Tate W. Release factors and their role as decoding proteins: specificity and fidelity for termination of protein synthesis. Biochim Biophys Acta 2000;1493:1–11.
61. Cameron DM, Thompson J, March PE, Dahlberg AE. Initiation factor IF2, thiostrepton and micrococcin prevent the binding of elongation factor G to the *Escherichia coli* ribosome. J Mol Biol. 2002;319:27–35.
62. Grentzmann G, Brechemier-Baey D, Heurgue-Hamard V. Buckingham RH. Function of polypeptide chain release factor RF-3 in *Escherichia coli*. RF-3 action in termination is predominantly at UGA-containing stop signals. J Biol Chem 1995;270:10595–10600.
63. Nissen P, Kjeldgaard M, Thirup S, Polekhina G, Reshetnikova L, Clark BF, Nyborg J. Crystal structure of the ternary complex of Phe-tRNAPhe, EF-Tu, and a GTP analog. Science 1995;270:1464–1472.
64. Caldon CE, Yoong P, March PE. Evolution of a molecular switch: universal bacterial GTPases regulate ribosome function. Mol Microbiol 2001;41:289–297.
65. Dibb NJ, Wolfe PB. lep operon proximal gene is not required for growth or secretion by *Escherichia coli*. J Bacteriol 1986;166:83–87.
66. Kiser GL. Weinert TA. GUF1, a gene encoding a novel evolutionarily conserved GTPase in budding yeast Yeast 1995;11:1311–1316.
67. Freestone P, Grant S, Trinei M, Onoda T, Norris V. Protein phosphorylation in *Escherichia coli* L. form NC-7. Microbiology 1998;144:3289–3295.
68. Freestone P, Trinei M, Clarke SC, Nystrom T, Norris V. Tyrosine phosphorylation in *Escherichia coli*. J Mol Biol 1998;279:1045–1051.
69. Farris M, Grant A, Richardson TB, O'Connor CD. BipA: a tyrosine-phosphorylated GTPase that mediates interactions between enteropathogenic *Escherichia coli* (EPEC) and epithelial cells. Mol Microbiol 1998;28:265–279.
70. Ævarsson A. Structure-based sequence alignment of elongation factors Tu and G with related GTPases involved in translation. J Mol Evol. 1995;41:1096–1104.
71. Powell BS, Court DL, Inada T, Nakamura Y, Michotey V, Cui X, Reizer A, Saier MH Jr, Reizer J. Novel proteins of the phosphotransferase system encoded within the rpoN operon of *Escherichia coli*. Enzyme IIANtr affects growth on organic nitrogen and the conditional lethality of an erats mutant. J Biol Chem 1995;270: 4822–4839.
72. Meier TI, Peery RB, Jaskunas SR. Zhao G. 16S rRNA is bound to era of Streptococcus pneumoniae. J Bacteriol 1999;181:5242–5249.
73. Sayed A, Matsuyama S. Inouye M. Era, an essential *Escherichia coli* small G-protein, binds to the 30S ribosomal subunit. Biochem Biophys Res Commun 1999;264:51–54.
74. Johnstone BH, Handler AA, Chao DK, Nguyen V, Smith M, Ryu SY, Simons E.L, Anderson PE, Simons RW. The widely conserved Era G-protein contains an RNA-binding domain required for Era function in vivo. Mol Microbiol 1999;33:1118–1131.
75. Britton RA, Chen SM, Wallis D, Koeuth T, Powell BS, Shaffer LG, Largaespada D, Jenkins NA, Copel NG, Court DL, Lupski JR. Isolation and preliminary characterization of the human and mouse homologues of the bacterial cell cycle gene era. Genomics 2000;67:78–82.
76. Mehr IJ, Long CD, Serkin CD, Seifert HS. A homologue of the recombination-dependent growth gene, rdgC, is involved in gonococcal pilin antigenic variation. Genetics 2000;154:523–532.
77. Cabedo H, Macian F, Villarroya M, Escudero JC, Martinez-Vicente M, Knecht E, Armengod ME. The *Escherichia coli* trmE (mnmE) gene, involved in tRNA modification, codes for an evolutionarily conserved GTPase with unusual biochemical properties. EMBO J 1999;18:7063–7076.
78. Alam KY, Clark DP. Molecular cloning and sequence of the thdF gene, which is involved in thiophene and furan oxidation by *Escherichia coli*. J Bacteriol 1991;173:6018–6024.
79. Zabel MD, Bunch PK, Clark DP. Regulation of the thdF gene, which is involved in thiophene oxidation by *Escherichia coli* K-12. Microbios 2000;101:89–103.
80. Kok J, Trach KA, Hoch JA. Effects on Bacillus subtilis of a conditional lethal mutation in the essential GTP-binding protein Obg. J Bacteriol 1994;176:7155–7160.
81. Vidwans SJ, Ireton K, Grossman AD. Possible role for the essential GTP-binding protein Obg in regulating the initiation of sporulation in Bacillus subtilis. J Bacteriol 1995;177:3308–3311.
82. Okamoto S, Ochi K. An essential GTP-binding protein functions as a regulator for differentiation in *Streptomyces coelicolor*. Mol Microbiol 1998;30:107–119.
83. Welsh KM, Trach KA, Folger C, Hoch JA. Biochemical characterization of the essential GTP-binding protein Obg of Bacillus subtilis. J Bacteriol 1994;176:7161–7168.
84. Scott JM, Haldenwang WG. Obg, an essential GTP binding protein of Bacillus subtilis, is necessary for stress activation of transcription factor sigmaB). J Bacteriol 1999;181:4653–4660.
85. Scott JM, Ju J, Mitchell T, Haldenwang WG. The Bacillus subtilis GTP binding protein obg and regulators of the sigma (B) stress response transcription factor cofractionate with ribosomes. J Bacteriol 2000;182:2771–2777.
86. Wolf YI, Aravind L, Grishin NV, Koonin EV. Evolution of aminoacyl-tRNA synthetases—analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. Genome Res 1999;9:689–710.
87. Ævarsson A, Brazhnikov E, Garber M, Zheltonosova J, Chirgadze Y, al-Karadaghi S, Svensson LA, Liljas A. Three-dimensional structure of the ribosomal translocase: elongation factor G from Thermus thermophilus. EMBO J. 1994;13:3669–3677.