

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)

Gerhard Mittenhuber*

Institut für Mikrobiologie und Molekularbiologie,
Ernst-Moritz-Arndt-Universität, F.-L. Jahnstr. 15,
D-17487 Greifswald, Germany

Abstract

Several GTP-binding proteins with poorly defined functions were previously identified in *Escherichia coli* (i.e. Era, ThdF (TrmE)), *Bacillus subtilis* (i.e. Obg) and *Neisseria gonorrhoeae* (i.e. EngA). In these species, every individual protein is encoded by an essential gene. BLAST searches were used to detect orthologs in genomes of various organisms. Alignments of orthologous sequences allowed the construction of phylogenetic trees and the definition of protein families. The BLAST searches also resulted in the identification of two additional families, the YchF and YihA families, named after the *ychF* and *yihA* genes of *E. coli*. Most families are not present in archaeal genomes, but representatives of each family were also detected in eukaryotic genomes. Only representatives of the YchF family are present in every genome sequenced to date, suggesting that YchF-like proteins might be involved in a fundamental life process. The GTP1/DRG family consisting of eukaryotic and archaeal proteins is related to the YchF family of GTP-binding proteins. The relationship of the six prokaryotic families of GTP-binding proteins and the GTP1/DRG family to eukaryotic GTPase families was also investigated: With the exception of the ARF family, a clear separation of the six prokaryotic families and the GTP1/DRG family with respect to eukaryotic (RAB, RAN, RAS and RHO) GTPases was observed.

Introduction

In eukaryotes, regulatory proteins called GTP-binding proteins or GTPases constitute a mechanism for controlling multiple biochemical pathways within the complex cellular environment. These proteins act as molecular switches in signal transduction pathways: In the active (GTP-bound) conformation, the protein interacts with cellular targets promoting a response, while GTP hydrolysis generates the

resting, inactive state. At least five different GTPase classes (ARF, RAB, RAN, RAS, and RHO) exhibiting different functions can be distinguished (for review, see Hall, 2000). The Era (*Escherichia coli* ras-like) protein was the first example of a prokaryotic gene product exhibiting sequence similarity to the RAS protein of *Saccharomyces cerevisiae* (Ahn *et al.*, 1986); a finding which caused great excitement because human RAS was the first human oncogene to be discovered (Parada *et al.*, 1982). However, later analysis of the homology between RAS and Era (Bardwell *et al.*, 1989; Chen *et al.*, 1990) revealed that most of the similarity between these proteins is confined to the GTP-binding domain (Bourne *et al.*, 1990, 1991; Kjeldgaard *et al.*, 1996). The *era* gene is essential in *E. coli* (Lerner and Inouye, 1991). The Era protein exhibits cycling between a cytoplasmic location and a membrane associated state (Lin *et al.*, 1994). Binding to 16 S rRNA (Meier *et al.*, 2000) and association with the 30S ribosomal subunit (Sayed *et al.*, 1999) was also demonstrated. However, the function of Era is still unknown. Cells carrying mutations in *era* exhibit pleiotropic phenotypes (Lerner and Inouye, 1991) including profoundly disturbed carbon metabolism (Pillutla *et al.*, 1996) and a lack of cell cycle progression (Gollop and March, 1991; Britton *et al.*, 1998). Cross-species complementation of *era* mutants has been performed (Pillutla *et al.*, 1995). Britton *et al.* (1998) also report on the identification of *era* homologs in the genomes of *Caenorhabditis elegans*, mouse and man.

The Obg (*spo0B* associated GTP-binding protein) protein of *Bacillus subtilis*, an essential GTP-binding protein of unknown function (Trach and Hoch, 1989), is involved in the signal transduction pathway resulting in activation of the stress sigma factor σ^B (Scott and Haldenwang, 1999) and associates with ribosomal protein L13 (Scott *et al.*, 2000). Obg is also required for sporulation in *B. subtilis* (Vidwans *et al.*, 1995), *Streptomyces griseus* (Okamoto *et al.*, 1997) and *Streptomyces coelicolor* (Okamoto and Ochi, 1998). The Obg ortholog of *Caulobacter crescentus*, CgtA (*Caulobacter* GTP-binding protein A) is also essential (Maddock *et al.*, 1997; Lin *et al.*, 1999).

In database entries of Era and Obg orthologs, these two protein families are classified as the "Era/ThdF family" and the "Obg/GTP1" family. ThdF (thiophen degradation) is an *E. coli* protein (Alam and Clark, 1991) involved in thiophen and furan oxidation. The sequence reported by these authors for the *thdF* gene (accession numbers A38160; AAB19981) does not exhibit similarity to any other protein (BLAST data not shown). This sequence was revised by Burland *et al.* (1993) (see accession number

Received August 25, 2000; accepted September 22, 2000.

*For correspondence. Email mittenhu@microbio4.biologie.uni-greifswald.de; Tel. +49-3834-864218; Fax. +49-3834-864202.

Table 1. Species Analyzed in this Study. An abbreviation is used in the tables and figures

Organism	Abbreviation	Organism	Abbreviation
<i>Aeropyrum pernix</i>	Ape	<i>Mycoplasma capricolum</i>	Mca
<i>Acholeplasma laidlawii</i>	Ala	<i>Mycoplasma genitalium</i>	Mge
<i>Antirrhinum majus</i>	Ama	<i>Mycoplasma pneumoniae</i>	Mpn
<i>Aquifex aeolicus</i>	Aae	<i>Neisseria gonorrhoeae</i>	Ngo
<i>Arabidopsis thaliana</i>	Ath	<i>Neisseria meningitidis</i>	Nme
<i>Archaeoglobus fulgidus</i>	Afu	<i>Oncorhynchus tshawytscha</i> (king salmon)	Ots
<i>Aspergillus nidulans</i>	Ani	<i>Oryza sativa</i> (rice)	Osa
<i>Bacillus halodurans</i>	Bha	<i>Pisum sativum</i> (pea)	Psa
<i>Bacillus subtilis</i>	Bsu	<i>Plasmodium falciparum</i>	Pfa
<i>Borrelia burgdorferi</i>	Bbu	<i>Pseudomonas aeruginosa</i>	Pae
<i>Bradyrhizobium japonicum</i>	Bja	<i>Pseudomonas putida</i>	Ppu
<i>Buchnera aphidicola</i>	Bap	<i>Pyrococcus abyssi</i>	Pab
<i>Caenorhabditis elegans</i>	Cel	<i>Pyrococcus horikoshii</i>	Pho
<i>Campylobacter jejuni</i>	Cje	<i>Rattus norvegicus</i> (rat)	Rno
<i>Candida albicans</i>	Cal	<i>Rhodobacter spaeroides</i>	Rsp
<i>Capsicum annum</i>	Can	<i>Rickettsia prowazekii</i>	Rpr
<i>Caulobacter crescentus</i>	Ccr	<i>Salmonella typhi</i>	Sty
<i>Chlamydia muridarum</i>	Cmu	<i>Saccharomyces cerevisiae</i>	Sce
<i>Chlamydia pneumoniae</i>	Cpn	<i>Schizosaccharomyces pombe</i>	Spo
<i>Chlamydia trachomatis</i>	Ctr	<i>Streptococcus agalactiae</i>	Sag
<i>Coxiella burnetii</i>	Cbu	<i>Streptococcus mutans</i>	Smu
<i>Cyanidium caldarium</i> (red alga)	Cca	<i>Streptococcus pneumoniae</i>	Spn
<i>Deinococcus radiodurans</i>	Dra	<i>Streptococcus pyogenes</i>	Spy
<i>Dictyostelium discoideum</i>	Ddi	<i>Streptococcus thermophilus</i>	Sth
<i>Discopyge ommata</i> (electric ray)	Dom	<i>Streptomyces coelicolor</i>	SCO
<i>Drosophila melanogaster</i>	Dme	<i>Streptomyces griseus</i>	Sgr
<i>Escherichia coli</i>	Eco	<i>Suberites domuncula</i> (a sponge)	Sdo
<i>Haemophilus influenzae</i>	Hin	<i>Sulfolobus solfataricus</i>	Sso
<i>Halobacterium cutirubrum</i>	Hcu	<i>Sullius bovinus</i> (a fungus)	Sbo
<i>Helicobacter pylori</i>	Hpy	<i>Synechococcus elongatus</i>	Sel
<i>Homo sapiens</i>	Hsa	<i>Synechococcus</i> sp. PCC 7942	Ssp7942
<i>Lactococcus lactis</i>	Lla	<i>Synechocystis</i> sp. PCC 6803	Ssp
<i>Leishmania major</i>	Lma	<i>Teladorsagia circumcincta</i> (a nematode)	Tci
<i>Leptospira interrogans</i>	Lin	<i>Thermoplasma acidophilum</i>	Tac
<i>Methanobacterium thermoautotrophicum</i>	Mth	<i>Thermotoga maritima</i>	Tma
<i>Methanococcus jannaschii</i>	Mja	<i>Thiobacillus ferrooxidans</i>	Tfe
<i>Mucor racemosus</i>	Mra	<i>Treponema pallidum</i>	Tpa
<i>Mus musculus</i> (mouse)	Mmu	<i>Ureaplasma urealyticum</i>	Uur
<i>Mycobacterium leprae</i>	Mle	<i>Vibrio cholerae</i>	Vch
<i>Mycobacterium tuberculosis</i>	Mtu	<i>Zymomonas mobilis</i>	Zmo

P25522 for the corrected *ThdF* entry). However, *thdF* is allelic with *trmE* (t-RNA modification E) (Cabedo *et al.*, 1999) which encodes an enzyme involved in the biosynthesis of 5-methylaminomethyl-2-thiouridine, a nucleoside found in the wobble position of some tRNAs (Elseviers *et al.*, 1984; Hagervall *et al.*, 1998). It was suggested by Leung *et al.* (1998) that *trmE* should be renamed *mnme* (methylaminomethyl E). Depending on the genetic background, *thdF* (*trmE*) is essential in *E. coli* (Cabedo *et al.*, 1999).

The GTP1 family refers to GTP-binding protein 1 of *Schizosaccharomyces pombe* (Hodson and Young, 1993). Members of this family are also called DRG proteins (developmentally regulated GTP-binding protein) and were identified in a variety of vertebrates: Kumar *et al.* (1993) analyzed the expression of *drg* during development of *Xenopus laevis* and the human DRG2 protein was identified as a protein whose synthesis is repressed in SV40 transformed fibroblasts (Schenker *et al.*, 1994). These authors showed tissue specificity of expression (highest levels in skeletal muscle, heart and kidney; low levels in colon, thymus, spleen, lung, small intestine and leukocytes), a cytoplasmic subcellular localization and concluded that DRG2 may play a role in cell proliferation,

differentiation and death. Two papers by Sazuka *et al.*, (1992a,b) investigate expression patterns of a murine *drg* ortholog. They find fairly high levels in liver, heart, kidney, testis and brain and very low levels lung, spleen and skeletal muscle. They also observe a cytoplasmic location of DRG and show that DRG is predominantly expressed in embryos and downregulated during development. Proteins interacting with DRG (SCL/TAL1, TAL2 and LYL1) were also identified in this study and by Mahajan *et al.* (1996). Binding of SCL to a human DRG has been also observed (Zhao and Aplan, 1998). DRG-like proteins were also detected in plants (Devitt *et al.*, 1999), exhibiting maximal transcription in growing parts of the plant; in *Drosophila melanogaster* (Sommer *et al.*, 1994) and in the archaeobacterium *Halobacterium cutirubrum* (Shimmin and Dennis, 1989).

Very recently, the first member of another, putative bacterial GTP-binding protein family was identified in *Neisseria gonorrhoeae* (Mehr *et al.*, 2000). This protein was named EngA (essential neisserial GTP-binding protein A). Interestingly, this protein contains two sets of GTP-binding motifs as defined by Bourne *et al.* (1991) which is in contrast to the other GTP-binding proteins which possess one set.

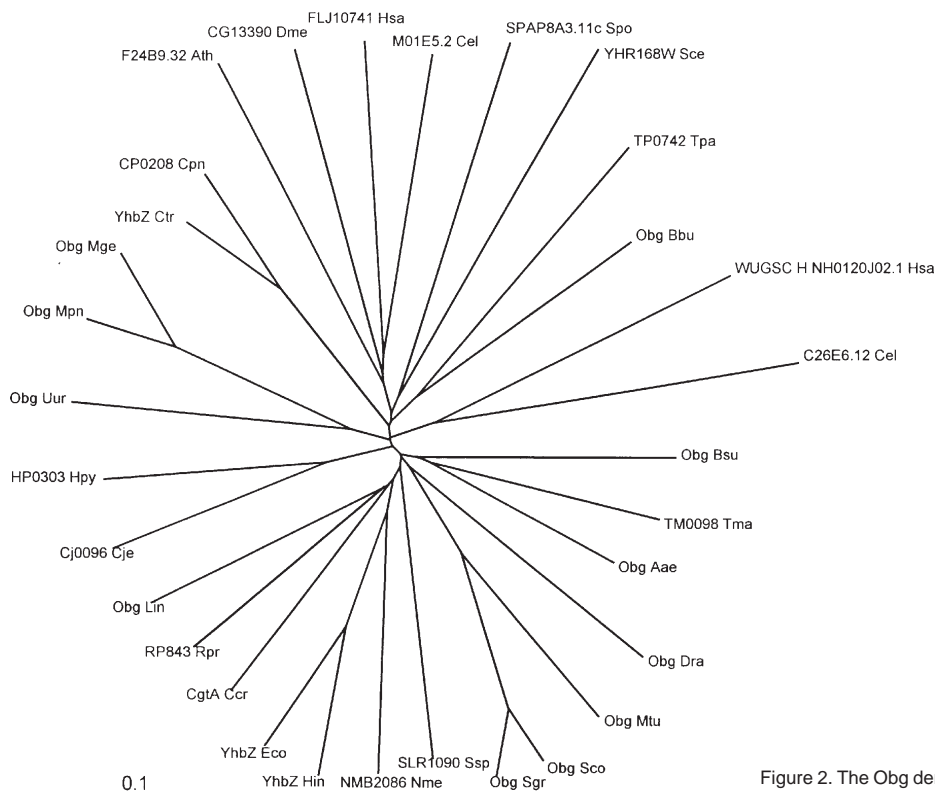


Figure 2. The Obg derived tree.

Table 3. Obg Orthologs. Individual proteins are sorted according to ascending E-values

Designation	Length	Accession number	E-value
Obg Bsu	428	P20964	1.8e-207
TM0098 Tma	435	Q9WXV3	9.7e-88
Obg Aae	343	O67849	2.0e-80
Obg Sco	478	P95722	4.5e-74
Obg Sgr	478	P95758	2.9e-72
Obg Mtu	479	P71909	2.0e-71
YhbZ Hin	390	P44915	1.6e-69
SLR 1090 Ssp	368	P72931	1.8e-69
YhbZ Eco	390	P42641	2.7e-69
NMB2086 Nme	384	AAF42403	3.7e-69
Obg Uur	435	AAF30873	7.5e-65
Obg Mpn	433	P75215	3.3e-64
CgtA Ccr	453	O30861	5.3e-64
Obg Mge	433	P47624	5.4e-64
RP843 Rpr	331	Q9ZCB6	1.5e-61
Obg Dra	438	Q9RY66	4.4e-60
YhbZ Ctr	335	O84423	2.9e-56
Cj0096 Cje	350	CAB72580	1.1e-54
CP0208 Cpn	343	AAF38079	7.9e-54
HP0303 Hpy	360	Q9ZMD3	2.1e-53
CG13390 Dme	381	Q9VLN0	3.7e-51
TP0742 Tpa	376	O83724	4.6e-49
F24B9.32 Ath	1029	AC007583	1.4e-48
FLJ10741 Hsa	401	BAA91783	1.6e-48
SPAP8A3.11c Spo	419	Q9UT06	5.2e-46
Obg Bbu	328	O51722	1.4e-42
M01E5.2 Cel	358	O45691	2.0e-41
YHR168W Sce	499	P38860	1.7e-35
WUGSC:H_NH0120J02.1 Hsa	206 (fragment)	Q9Y6T6	3.7e-24
C26E6.12 Cel	1802	Q18219	6.4e-20
Obg Lin	115 (fragment)	O87884	1.2e-13

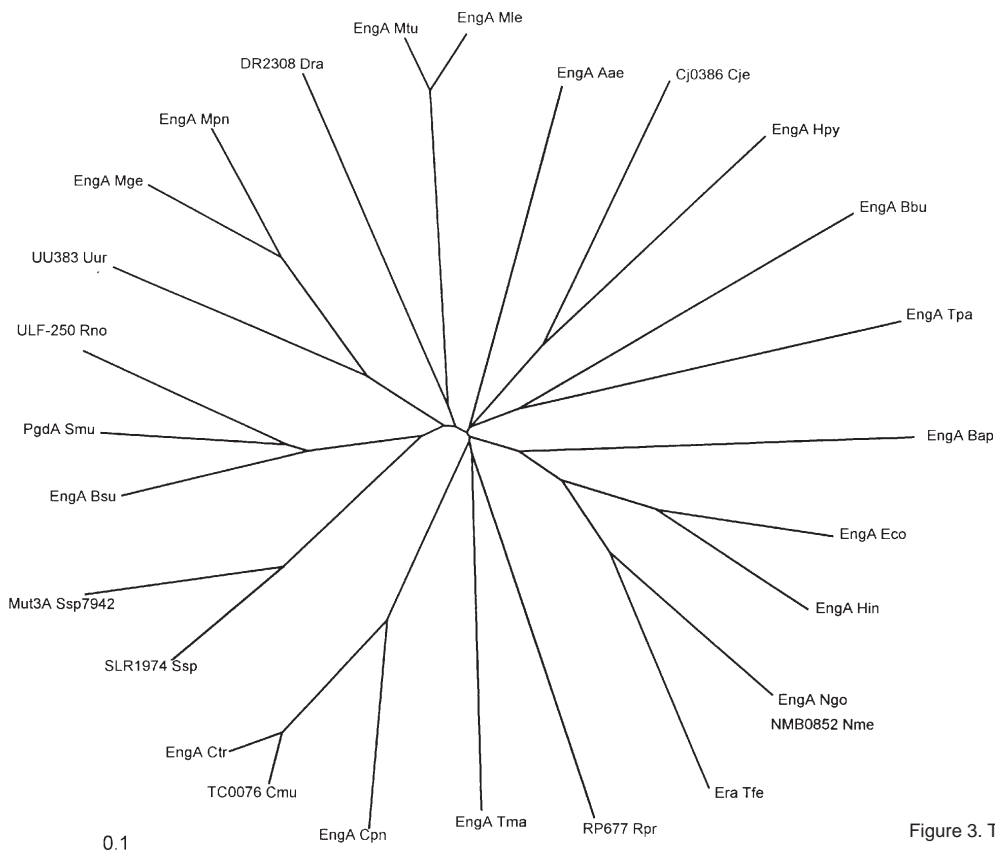


Figure 3. The EngA derived tree.

Table 4. EngA Orthologs. Individual proteins are sorted according to ascending E-values

Designation	Length	Accession number	E-value
EngA Eco (YfgK)	503	BAA16397	6.3e-251
EngA Hin	504	P44536	8.8e-174
EngA Ngo	485	O87407	3.2e-118
NMB0852	485	AE002438	3.2e-118
EngA Bap	453	O51881	5.1e-98
SLR1974 Ssp	452	P74120	6.6e-80
EngA Bsu	436	P50743	2.5e-78
Era Tfe	281 (fragment)	AAF64397	1.7e-76
PgdA Smu	436	Q9RHV5	1.1e-75
EngA Aae	433	O67749	3.7e-73
RP677 Rpr	447	Q9ZCP6	1.2e-68
UU383 Uur	442	AAF30793	2.9e-67
DR2308 Dra	438	Q9RS19	6.1e-67
EngA Mtu	463	O33212	4.8e-65
EngA Mge	448	P47571	2.9e-63
EngA Mle	461	Q49884	3.8e-63
TC0076 Cmu	490	AAF38958	2.0e-62
EngA Mpn	449	P75309	3.3e-62
EngA Ctr	490	O84709	5.4e-62
EngA Cpn	487	Q9Z762	1.4e-59
Cj0386 Cje	460	CAB74222	2.6e-55
EngA Bbu	433	O51461	1.3e-54
EngA Hpy	462	Q9ZL09	1.5e-52
EngA Tpa	460	P96128	7.7e-49
EngA Tma	439	Q9X1F8	6.2e-38
Mut3A Ssp7942	280	P72548	2.7e-27
ULF-250 Rno	188	Q9Z212	1.8e-25

With the availability of complete genomic DNA sequences of organisms, it became possible to start comparative genomics. The purposes of such analyses are manifold (Kanehisa, 2000). For example, ubiquitous genes present in every genome can be found, essential genes can be identified and a "minimal genome" can be described theoretically (Mushegian and Koonin, 1996) and experimentally (Hutchinson *et al.*, 1999). Functional genomics resulted in the identification of an essential *E. coli* gene encoding a previously uncharacterized GTP-binding protein, the *yihA* gene (Arigoni *et al.*, 1998; Dassain *et al.*, 1999).

In this paper, BLAST searches were performed in order to identify and to define prokaryotic GTP-binding protein families in a variety of organisms (see Table 1 for abbreviations of names used in Tables and Figures). Phylogenetic trees for these families are presented. The relationships between pro- and eukaryotic GTP-binding protein families was also investigated.

The Era Family

Using the deduced sequence of the *E. coli* Era protein as query, orthologous genes were identified in a variety of organisms (Table 2) and an unrooted phylogenetic tree (Figure 1) was obtained from an alignment. In agreement with the phylogenetic position of *E. coli*, the Era protein of this γ proteobacterium clusters with other members of this group whereas Era proteins of grampositive organisms are associated with the *Synechocystis* sp. ortholog. Archaea

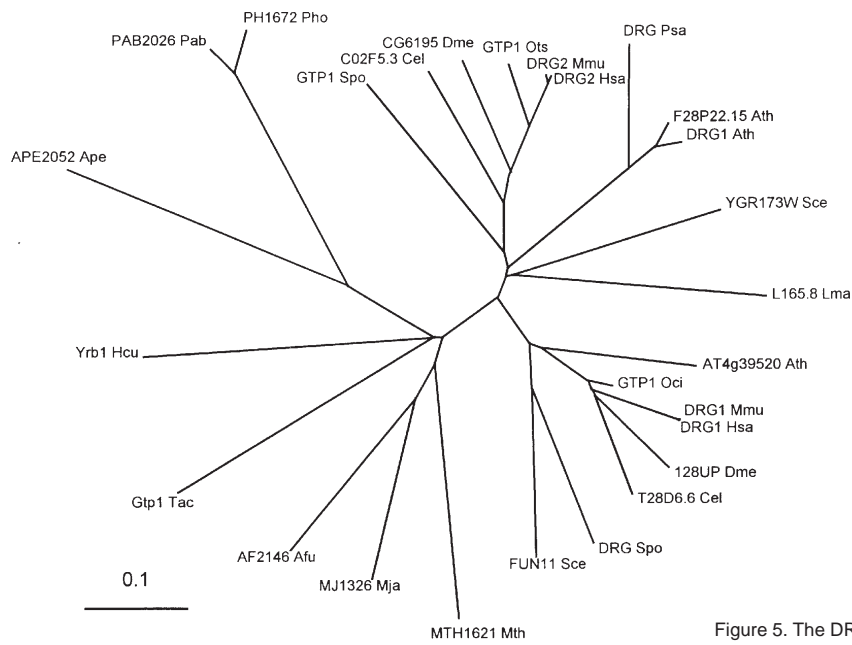


Figure 5. The DRG derived tree.

exception of an ORF expressed in rat uterus (ULF-250) which is related to EngA of grampositive bacteria, this family contains exclusively eubacterial orthologs. Curiously, the EngA ortholog of *S. mutans* was identified as phosphoglycerate dehydrogenase. The classification of an ORF found in *Thiobacillus ferrooxidans* as Era ortholog is probably an error since this ORF clusters with the EngA orthologs of the γ subdivision of proteobacteria.

The ThdF Family

A sequence named "ThdF" was readily available in the "SubtiList" database, but not in the corresponding "Colibri" database. Therefore, the deduced sequence of the *B. subtilis* ThdF (TrmE) protein was used as query and orthologous genes were identified in a variety of organisms (Table 5) and an unrooted phylogenetic tree (Figure 4) was obtained from an alignment. In accordance with Cabedo *et al.* (1999) archaeal representatives of this family were not identified. Orthologs identified in *Arabidopsis* and in the red alga *Cyanidium* are related to the cyanobacterial orthologs. The MSS1 gene of yeast encodes a nuclear-encoded GTP-binding protein involved in mitochondrial cytochrome c biosynthesis (Decoster *et al.*, 1993). This protein is associated with the small subunit of mitochondrial ribosomes and seems to play a role in mitochondrial translation.

The DRG Family

Using the deduced sequence of the *Homo sapiens* DRG1 protein as query, orthologous genes were identified in a variety of organisms (Table 6) and an unrooted phylogenetic tree was obtained from an alignment (Figure 5). This tree is deeply branched between Archaea and eukaryotes. No eubacterial representatives are included in this family. The eukaryotic lineage can be subdivided in two sections, each including orthologs of yeast, fission yeast and higher

eukaryotes. The DRG family was also described and identified as a novel GTP-binding protein family by Li and Trueb (2000).

The YchF Family

During the BLAST search using the Obg protein of *B. subtilis* as query, two archaeal proteins (*i.e.* MTH1515 of *Methanobacterium thermoautotrophicum* (E-value: 3.8e-16) and MJ1332 of *Methanococcus jannaschii* (E-value: 7.9e-14) which are unrelated to Obg/GTP1 family were

Table 6. DRG Orthologs. Individual proteins are sorted according to ascending E-values

Designation	Length	Accession number	E-value
DRG1 Hsa	367	Q9Y295	4.5e-177
DRG1 Mmu	367	P32233	3e-170
128UP Dme	368	P32234	3.4e-149
T28D6.6 Cel	366	Q9TVG4	2.7e-147
AT4g39520 Ath	369	Q9SVA6	3.6e-129
DRG Spo	366	P78786	7.0e-117
FUN11 Sce	369	P39729	2.1e-115
F28P22.15 Ath	399	Q9SGF9	1.9e-100
DRG1 Ath	399	O04174	1.3e-99
CG6195 Dme	363	Q9VDV2	4.6e-97
DRG2 Hsa	364	P55039	4.6e-97
DRG2 Mmu	364	Q9QXB9	4.5e-97
GTP1 Ots	364	O57346	6.7e-96
C02F5.3 Cel	573	P34280	2.3e-95
GTP1 Spo	364	P32235	4.5e-90
L165.8 Lma	374	AC009601	1.5e-89
YGR173W Sce	368	P53295	1.4e-86
DRG Psa	399	O22445	2.3e-86
MJ1326 Mja	391	Q58722	2.9e-79
MTH1621 Mth	380	O27658	7.0e-78
AF2146 Afu	355	O28136	8.2e-75
Gtp1 Tac	360	P96085	1.7e-58
Yrb1 Hcu	370	P17103	6.7e-57
PAB2026 Pab	387	Q9V1C8	7.1e-53
PH1672 Pho	387	O59329	3.9e-52
GTP1 Tci	140 (fragment)	O02458	3.5e-44
APE2052 Ape	368	Q9YA87	6.6e-34

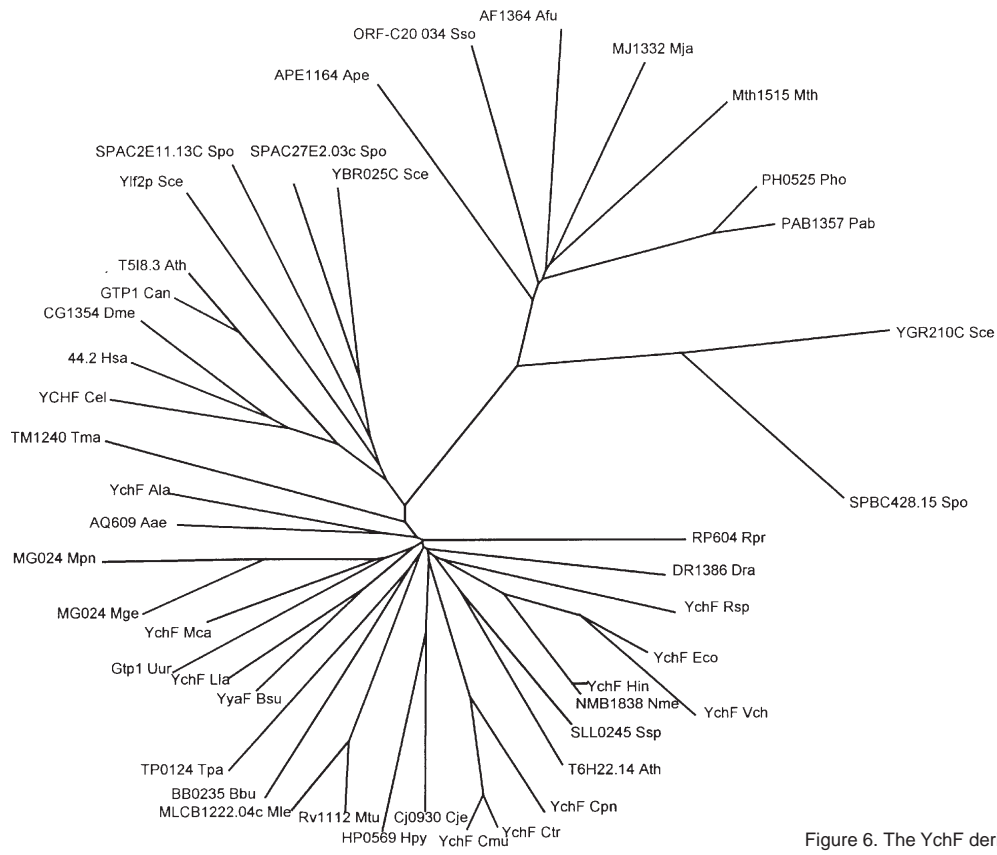


Figure 6. The YchF derived tree.

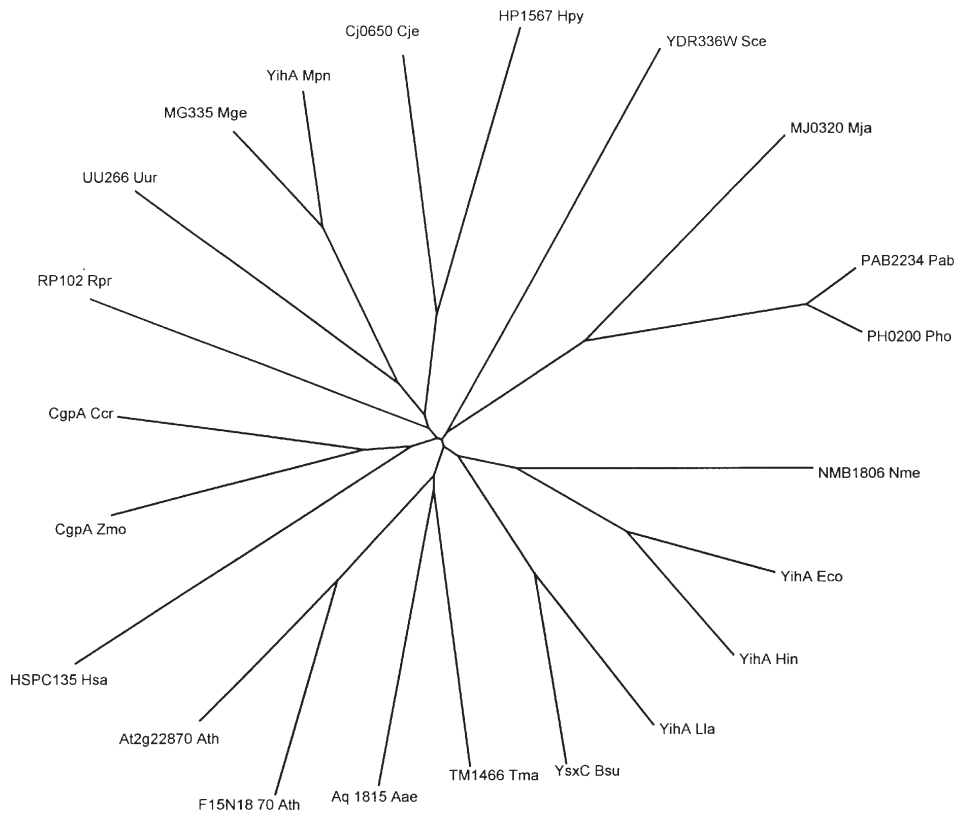


Figure 7. The YihA derived tree.

identified. Using the MTH1515 sequence as query, bacterial orthologs including the *yhfF* gene of *E. coli* were identified. These proteins are classified as YchF family of GTP-binding proteins in the databases. Using the deduced sequence of the *E. coli* YchF protein as query, orthologous genes were identified in a variety of organisms (Table 7) and an unrooted phylogenetic tree (Figure 6) was obtained from an alignment. This tree is deeply branched and also includes distantly related archaeal orthologs which are also related to two ORFs of yeast and fission yeast. Another branch which is more closely related to the eubacterial orthologs includes also orthologs found in genomes of higher eukaryotes. The YchF ortholog of *B. subtilis* is represented by an ORF of unknown function, *yfaF*.

The YihA Family

Using the deduced sequence of the essential (Arigoni *et al.*, 1998; Dassain *et al.*, 1999) YihA protein of *E. coli* as query, orthologous genes were identified in a variety of organisms (Table 8) and an unrooted phylogenetic tree

Table 7. YchF Orthologs. Individual proteins are sorted according to ascending E-values

Designation	Length	Accession number	E-value
YchF Eco	363	BAA36061	6.5e-185
YchF Hin	363	U32723	8.1e-155
NMB1838 Nme	363	AAF42173	9.8e-127
YyaF Bsu	363	P37518	1.1e-104
SLL0245 Ssp	363	P73886	3.9e-100
YchF Ctr	366	O84094	5.0e-100
YchF Cmu	364	AAF73549	8.1e-100
YchF Rsp	369	Q9X4F2	3.5e-99
Cj0930 Cje	367	CAB73187	2.2e-97
T6H22.14 Ath	419	Q9SGT3	1.8e-95
AQ609 Aae	370	O66865	2.9e-95
YchF Cpn	364	Q9Z8L7	2.9e-95
HP0569 Hpy	366	O25293	1.6e-94
DR1386 Dra	365	QRUK0	7.2e-92
Gtp1 Uur	368	AAF31009	1.3e-90
Rv1112 Mtu	357	O53459	2.9e-88
BB0235 Bbu	368	O51251	3.7e-88
RP604 Rpr	365	Q9ZCV5	7.9e-86
TP0124 Tpa	368	O83161	1.6e-85
MLCB1222.04c Mle	356	Q9X779	2.1e-85
MG024 Mge	367	P47270	3.1e-84
MG024 Mpn	362	P75088	2.2e-74
44.2 Hsa	396	Q9UNY9	8.5e-73
GTP1 Can	394	AAF65513	3.7e-72
T5l8.3 Ath	394	Q9SA73	1.6e-71
CG1354 Dme	399	Q9W317	3.3e-71
YCHF Cel	395	P91917	5.4e-71
SPAC27E2.03c Spo	392	O13998	5.3e-64
TM1240 Tma	357	Q9X0W7	1.4e-63
YBR025C Sce	394	P38219	2.6e-62
YchF Vch	145 (fragment)	O34225	1.2e-59
SPAC2E11.13C Spo	407	O14078	9.5e-58
YchF Lla	214 (fragment)	O54380	2.1e-53
YchF Ala	218 (fragment)	PC7024	1.0e-51
Ylf2p Sce	405	Z29089	3.7e-47
YchF Mca	178 (fragment)	Q48961	1.0e-44
AF1364 Afu	388	O28907	3.1e-21
PH0525 Pho	397	O58261	9.4e-20
PAB1357 Pab	397	Q9UYI3	2.9e-17
MJ1332 Mja	393	Q58728	4.8e-17
ORF-C20_034 Sso	408	Q9UXD2	1.0e-15
APE1164 Ape	410	Q9YCU8	5.6e-14
YGR210C Sce	411	P42942	4.1e-13
Mth1515 Mth	399	O27559	7.6e-13
SPBC428.15 Spo	409	O94362	2.6e-12

(Figure 7) was obtained from an alignment. This family includes mainly eubacterial orthologs and a few archaeal as well as eukaryotic representatives. In *B. subtilis*, the gene encoding the YihA ortholog YsxC - which is not essential in this organism - is located downstream of and cotranscribed with the *lon* gene (Riethdorf *et al.*, 1994; Schmidt *et al.*, 1994). Similarly as transcription of the *lon* gene encoding the ATP-dependent serine protease Lon, transcription of the *ysxC* gene (designated *orfX* in both papers) is inducible by various stresses (Riethdorf *et al.*, 1994). Homology between the *ysxC* and the *yihA* genes as well as the GTP-binding protein motif in both genes was noted previously (Schmidt *et al.*, 1994).

Relationships Between GTP-Binding Protein Families

Table 9 gives a summarizing overview on the occurrence of individual GTP-binding protein family members in the genomes of different species. The COG (Cluster of Orthologous Groups) number (Tatusov *et al.*, 1997; Tatusov *et al.*, 2000) is given for reference to this classification scheme.

In order to visualize the relationships between the bacterial GTP-binding protein families, all individual GTP-binding protein sequences were aligned (data not shown) and a summarizing tree was constructed (Figure 8). Several proteins are represented by numbers (Table 10) in this tree. A clear separation of the Era, Obg, EngA, ThdF and YihA families can be observed whereas the DRG and YchF families are related. Within the individual families, similar, if not identical branching patterns and levels of relatedness as observed in the trees shown in Figures 1 to 7 are apparent.

The relationships between the prokaryotic GTP-binding proteins and the eukaryotic GTP-binding proteins of the ARF, RAB, RAN, RAS and RHO families (Hall, 2000) were also investigated: Alignments comprising approximately 50 protein sequences of each individual

Table 8. YihA Orthologs. Individual proteins are sorted according to ascending E-values

Designation	Length	Accession number	E-value
YihA Eco	210	P24253	1.5e-107
YihA Hin	205	P46453	2.2e-74
NMB1806 Nme	211	AAF42143	1.4e-44
CgpA Zmo	212	Q9RNL6	4.7e-33
YihA Lla (ClpX)	195	AAF63739	3.8e-31
YsxC Bsu	195	P38424	1.3e-28
TM1466 Tma	195	Q9X1H7	4.4e-28
CgpA Ccr	199	Q9ZG89	8.3e-27
RP102 Rpr	214	Q9ZE46	1.4e-26
F15N18 70 Ath	318	CAB87708	1.7e-26
MG335 Mge	191	P47577	3.7e-24
YihA Mpn	193	P75303	3.7e-24
Cj0650 Cje	198	CAB75286	2.0e-23
At2g22870 Ath	219	O81004	3.3e-23
Aq 1815 Aae	183	O67679	8.8e-23
UU226 Uur	208	AAF30675	1.7e-21
HP1567 Hpy	208	O26087	2.2e-19
YDR335W Sce	314	Q05473	2.0e-12
MJ0320 Mja	219	Q57768	1.1e-08
PAB2234 Pab	190	Q9V288	3.3e-06
PH0200 Pho	190	O57939	9.2e-06
HSPC135 Hsa	307	AAF29099	1.9e-05

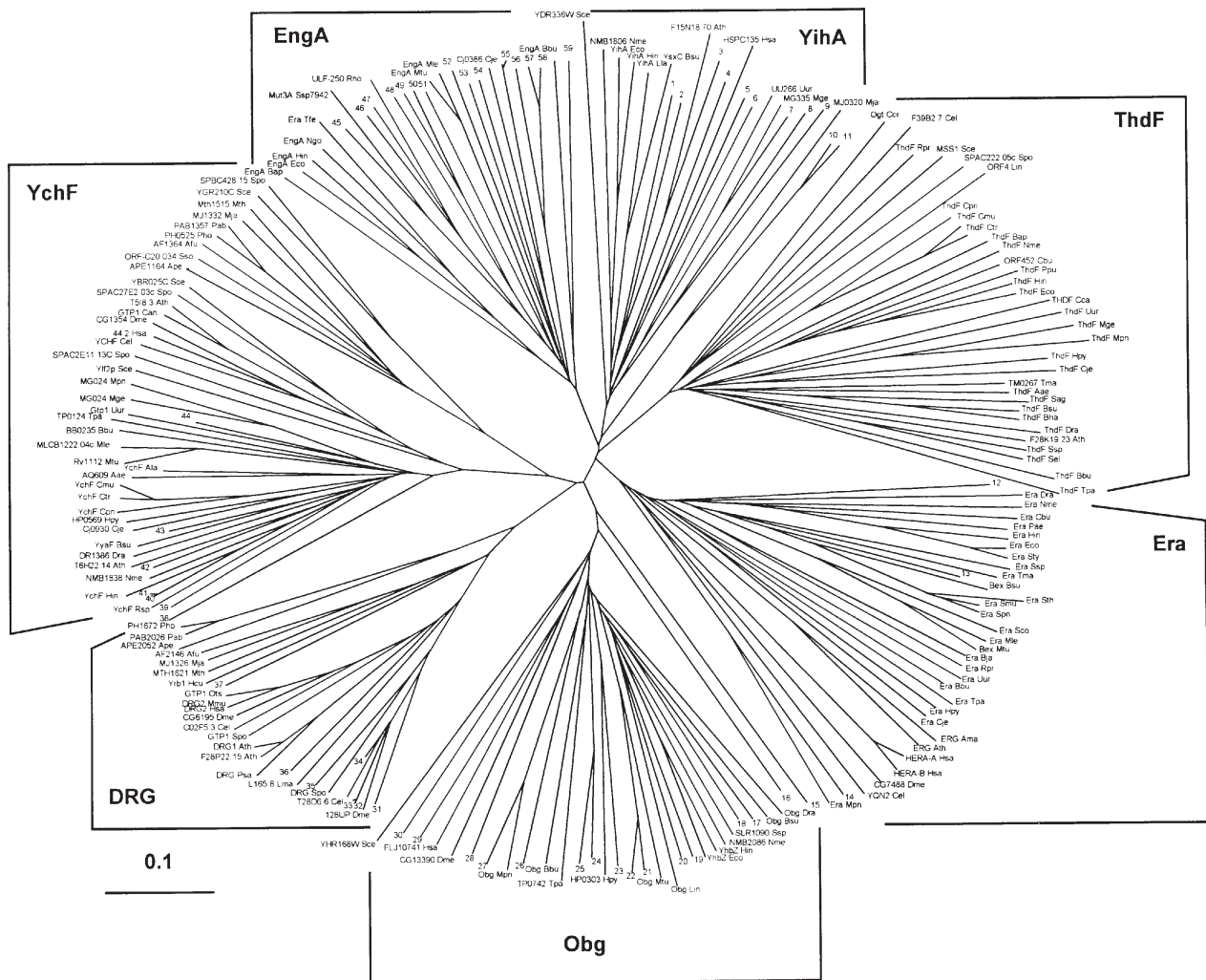


Figure 8. Summarizing tree showing the relatedness between the prokaryotic GTPase families investigated in this study. Brackets indicate the individual families as analyzed in Figures 1 to 7.

were performed (data not shown) and phylogenetic trees for each of these families were also constructed (data not shown) in order to identify distantly related proteins within each individual family. For each family, amino acid sequences representing several distantly related proteins (Table 11) were aligned and a phylogenetic tree was constructed (Figure 9). Several proteins are represented by numbers (Table 12) in this tree. Interestingly, clear clustering patterns of the individual families can be observed. The prokaryotic families exhibit similar patterns of relatedness as shown in Figure 8. With the exception of the ARF family, the related eukaryotic families form clusters and are more closely related to each other than to the prokaryotic families. Most importantly, the eukaryotic members of the prokaryotic families are associated with their eubacterial counterparts as shown in Figures 1 to 7 and do not show any relatedness to the eukaryotic families.

Discussion

By the construction of phylogenetic trees, prokaryotic GTP-binding protein families could be defined and the

relationships between these families and eukaryotic GTP-binding protein families were described. The trees shown in Figures 8 and 9 do not represent true phylogenetic relationships in evolutionary terms but rather serve as an aid and a tool to illustrate these relationships. It is probably not appropriate to define an "Era/ThdF family" and an "Obg/GTP1 family" as found in current database entries describing individual proteins, since clear distinctions separating these four families are evident.

Assuming that essential, conserved genes might possess conserved function in all organisms, prokaryotic GTP-binding proteins are likely to perform important, yet unknown functions in bacteria as well as in eukaryotes. At least the Era, Obg and ThdF proteins are associated with the translation machinery. The YhbZ and YihA proteins of *E. coli* were discovered in a search for essential genes as possible targets of essential bacterial genes for novel antibiotics (Arigoni *et al.*, 1998). The presence of genes encoding these targets in eukaryotic genomes might complicate this approach, since all antibiotics known so far interfere with unique aspects of bacterial physiology.

Theoretical (*i.e.* comparative genomics; Mushegian

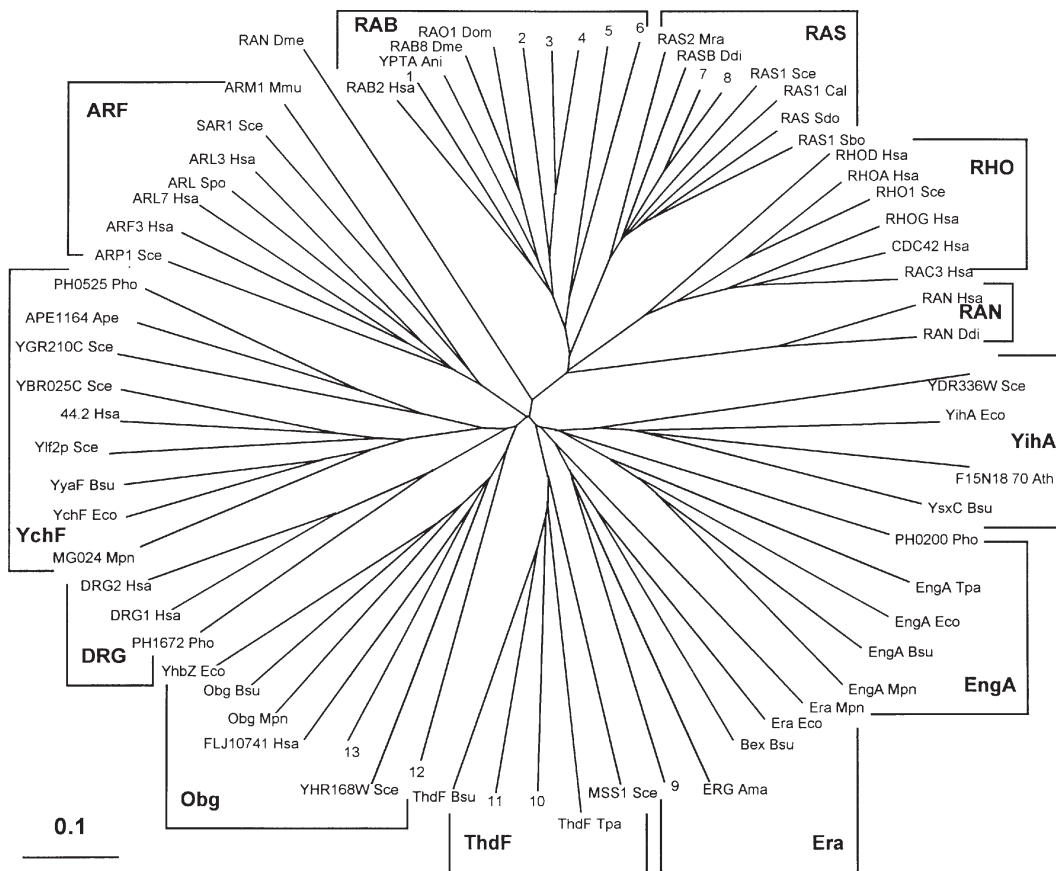


Figure 9. Summarizing tree showing the relatedness between the prokaryotic GTPase families investigated in this study and the eukaryotic ARF, RAB, RAN, RAS and RHO families. Brackets indicate the individual families.

Table 9. Distribution of Members of Bacterial GTPase Families in Completely Sequenced Genomes of Organisms. The COG number for each family is also given

Designation	Era	Obg	EngA	ThdF	DRG	YchF	YihA	Reference
COG number	1159	0536	1160	0486	1163	0012	0218	
Aae	+	+	+	+		+	+	Tatusov <i>et al.</i> , 2000
Afu					+	+		Deckert <i>et al.</i> , 1998
Ape					+	+		Klenk <i>et al.</i> , 1997
Bbu	+	+	+	+		+		Kawarabayasi <i>et al.</i> , 1999
Bsu	+	+	+	+		+	+	Fraser <i>et al.</i> , 1997
Cel	+	++		+	++	+		Kunst <i>et al.</i> , 1997
Cje	+	+	+	+		+	+	<i>C. elegans</i> sequencing consortium, 1998
Cpn		+	+	+		+		Parkhill <i>et al.</i> , 2000
Ctr		+	+	+		+		Kalman <i>et al.</i> , 1999
Dra	+	+	+	+		+		Stephens <i>et al.</i> , 1998
Dme	+	++			++	+		White <i>et al.</i> , 1999
Eco	+	+	+	+		+	+	Myers <i>et al.</i> , 2000
Hin	+	+	+	+		+	+	Blattner <i>et al.</i> , 1997
Hpy	+	+	+	+		+	+	Fleischmann <i>et al.</i> , 1997
Mge	+	+	+	+		+	+	Tomb <i>et al.</i> , 1997
Mja					+	+	+	Fraser <i>et al.</i> , 1995
Mpn	+	+	+	+		+	+	Bult <i>et al.</i> , 1996
Mth					+	+		Himmelreich <i>et al.</i> , 1996
Mtu	+	+	+			+		Smith <i>et al.</i> , 1997
Nme	+	+	+	+		+	+	Cole <i>et al.</i> , 1998
Pab					+	+	+	Tettelin <i>et al.</i> , 2000
Pho					+	+	+	Heilig, R., unpublished
Rpr	+	+	+	+		+	+	Kawarabayasi <i>et al.</i> , 1998
Sce		++		+	++	+++	+	Andersson <i>et al.</i> , 1998
Ssp	+	+	+	+		+		Mewes <i>et al.</i> , 1997
Tma	+	+	+	+		+	+	Kaneko <i>et al.</i> , 1996
Tpa	+	+	+	+		+		Nelson <i>et al.</i> , 1999
Uur	+	+	+	+		+	+	Fraser <i>et al.</i> , 1998
								Glass <i>et al.</i> , unpublished

Table 10. Identification of Proteins. Due to space limitations, the designation of individual GTP-binding proteins was replaced by numbers in Figure 8. This table serves to identify these proteins

Number in Figure 8	Designation	Number in Figure 8	Designation
1	Aq1815 Aae	31	AT4g39520 Ath
2	TM1466 Tma	32	DRG1 Hsa
3	At2g22870 Ath	33	DRG1 Mmu
4	RP102 Rpr	34	GTP1 Oci
5	CgpA Zmo	35	FUN11 Sce
6	CgpA Ccr	36	YGR173W Sce
7	YihA Mpn	37	Gtp1 Tac
8	Cj0650 Cje	38	TM1240 Tma
9	HP1567 Hpy	39	Rp604 Rpr
10	PAB2234 Pab	40	YchF Eco
11	PH0200 Pho	41	YchF Vch
12	Era1 Aae	42	SLL0245 Ssp
13	Era Spy	43	YchF Lla
14	Era Mge	44	YchF Mca
15	C26E6.12 Cel	45	SLR1074 Ssp
16	WUGSC H NH0120.J02.1 Hsa	46	EngA Bsu
17	TM0098 Tma	47	PdgA Smu
18	Obg Aae	48	UU383 Uur
19	CgtA Ccr	49	EngA Mge
20	RP843 Rpr	50	EngA Mpn
21	Obg Sco	51	DR2308 Dra
22	Obg Sgr	52	RP677 Rpr
23	Cj0096 Cje	53	EngA Aae
24	YhbZ Ctr	54	EngA Tma
25	CP0208 Cpn	55	EngA Hpy
26	Obg Uur	56	EngA Cpn
27	Obg Mge	57	TC0076 Cmu
28	F24B9.32 Ath	58	EngA Ctr
29	M01E2.2 Cel	59	EngA Tpa
30	SPAP8A3.11c Spo		

and Koonin, 1996) as well as experimental approaches (*i.e.* transposon mutagenesis; Hutchinson *et al.*, 1999) aim to characterize a “minimal gene set” or a “minimal genome”. The results from this work should help to solve the interesting question whether a living cell can be fully described in terms of molecular biology. The functions of many genes constituting the “minimal gene set” are unknown, indicating that some basic aspects of life were not yet addressed by experimental biology. With two exceptions, all members of the six families of prokaryotic GTP-binding proteins described here were included in the “minimal gene set” by both approaches: Surprisingly, the EngA family (COG 1160; represented by MG329 of *M. genitalium* and HI0136 of *H. influenzae*) is not included in the theoretical minimal genome described by Mushegian and Koonin (1996); representatives of all other families were included in the minimal gene set. Transposon insertions were isolated in the gene encoding the YchF orthologs MP128 of *M. pneumoniae* and MG024 of *M. genitalium*, whereas no insertions in the genes encoding the other mycoplasmal GTP-binding could be isolated (Hutchinson *et al.*, 1999).

It seems however, that archaeal genomes were not included in similar studies. Representatives of essential Era, Obg, EngA and ThdF families are absent in archaeal genomes. The absence of these proteins in most archaea might either indicate that these organisms (1) do not share the pathways regulated by bacterial GTP-binding proteins, (2) that sequence similarity of archaeal proteins is too low to be detected by the methods used in this paper or (3)

Table 11. Eukaryotic GTP-binding proteins of different families, whose sequences were included in the alignment to construct Figure 9. The assignment to a GTPase family is indicated in the last column

Designation	Length	Accession number	Family
ARP1 Sce	198	Q02804	ARF
ARF3 Hsa	181	M74493	ARF
ARL7 Hsa	192	P56559	ARF
ARL Spo	186	Q09767	ARF
ARL3 Hsa	182	P36405	ARF
SAR1 Sce	190	P20606	ARF
ARM1 Mmu	243	Q9QXJ4	ARF
RAN Dme	216	AAF30287	RAN
RAB2 Hsa	212	M28213	RAB
RAB4 Dme	213	Q9V3L5	RAB
YPTA Ani	201	AAF63333	RAB
RAB8 Dme	207	O18228	RAB
RAO1 Dom	200	P22127	RAN
GTP1 Psa	215	Q08145	RAB
YPT3 Spo	214	P17610	RAB
RIC2 Osa	217	P40393	RAB
RHA1 Ath	200	P31582	RAN
RAB6 Pfa	240	Q94663	RAN
RAS2 Mra	198	P22279	RAS
RASB Ddi	197	P32252	RAS
HRAS1 Hsa	189	P01112	RAS
RAS1 Dme	189	P08646	RAS
RAO1 Sce	390	K01970	RAS
RAS1 Cal	290	Q9UQX7	RAS
RAS Sdo	191	O97342	RAS
RAS Sbo	216	AAF65465	RAS
RHOD Hsa	210	O00212	RHO
RHOA Hsa	193	TVHU12	RHO
RHO1 Sce	209	P06780	RHO
RHOG Hsa	191	X61587	RHO
CDC42 Hsa	191	P21181	RHO
RAC3 Hsa	193	O14658	RHO
RAN Hsa	219	Q9JUEU9	RAN
RAN Ddi	212	P33519	RAN

that archaea regulate these pathways by other mechanisms (a phenomenon called “nonorthologous gene displacement” by Mushegian and Koonin, 1996). Interestingly, families of putative GTP-binding proteins mainly composed of archaeal representatives are listed in the COG database (*i.e.* COG1100; COG1084; COG1341). It will be interesting to determine whether these proteins are essential in archaea. Representatives of the YchF family however, are found in every completely sequenced genome. This conserved distribution pattern might indicate a pivotal role for the YchF family in living organisms.

Table 12. Due to space limitations, the designation of individual GTPases was replaced by numbers in Figure 9. This table serves to identify these proteins

Number in Figure 9	Designation	Number in Figure 9	Designation
1	RAB4 Dme	8	RAS1 Dme
2	GTP1 Psa	9	HERA-A Hsa
3	YPT3 Spo	10	ThdF Eco
4	RIC2 Osa	11	F28K19.23 Ath
5	RHA1 Ath	12	C26E6.12 Cel
6	RAB6 Pfa	13	WUGSC H NH0120J02.1 Hsa
7	HRAS1 Hsa		

Experimental Procedures

Protein sequences of *E. coli* were obtained at the "Colibri" database (<http://genolist.pasteur.fr/Colibri/>), *B. subtilis* sequences at the "Subtilist" website (<http://genolist.pasteur.fr/Subtilist>) (Moszer *et al.*, 1995, Moszer, 1998). Using the deduced protein sequences, BLAST searches (Altschul *et al.*, 1997) were performed at <http://dove.embl-heidelberg.de/Blast2/>. Alignments were generated using CLUSTALW (Thompson *et al.*, 1994) at <http://www.ebi.ac.uk/clustalw/>. Information about completely sequenced genomes was obtained at <http://www.ncbi.nlm.nih.gov/Genome>. The COG database can be assessed at <http://www.ncbi.nlm.nih.gov/COG>. Phylogenetic trees were constructed using the data from the alignments with the help of the program TreeView (<http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>; Page, 1996) and edited using the program Metafile Companion (<http://www.companionsoftware.com/>).

Acknowledgements

This work was performed in the lab of Michael Hecker, whose support is gratefully acknowledged. I thank Michael Hecker and Ulrich Zuber for comments on the manuscript. This paper is dedicated to Patrick J. Piggot on the occasion of his 60th birthday.

References

- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389-3402.
- Ahn, J., March, P.E., Takiff, H.E., and Inouye, M. 1986. A GTP-binding protein of *Escherichia coli* has homology to yeast RAS proteins. *Proc. Natl. Acad. Sci. U S A.* 83: 8849-8853.
- Alam, K.Y., and Clark, D.P. 1991. Molecular cloning and sequence of the *thdF* gene, which is involved in thiophene and furan oxidation by *Escherichia coli*. *J. Bacteriol.* 173: 6018-6024.
- Andersson, S.G., Zomorodipour, A., Andersson, J.O., Sicheritz-Ponten, T., Alsmark, U.C., Podowski, R.M., Naslund, A.K., Eriksson, A.S., Winkler, H.H., and Kurland, C.G. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 296: 133-140.
- Arigoni, F., Talabot, F., Peitsch, M., Edgerton, M.D., Meldrum, E., Allet, E., Fish, R., Jamotte, T., Curchod, M.L., and Loferer, H. 1998. A genome-based approach for the identification of essential bacterial genes. *Nature Biotechnol.* 16: 851-856.
- Bardwell, J.C.A., Regnier, P., Chen, S.-M., Nakamura, Y., Grunberg-Manago, M., and Court, D.L. 1989. Autoregulation of RNase III operon by mRNA processing. *EMBO J.* 8: 3401-3407.
- Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davies, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., and Shao, Y. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* 277: 1453-1474.
- Bourne, H.R., Sanders, D.A., and McCormick, F. 1990. The GTPase superfamily: a conserved switch for diverse cell functions. *Nature* 348: 125-132.
- Bourne, H.R., Sanders, D.A., and McCormick, F. 1991. The GTPase superfamily: conserved structure and molecular mechanism. *Nature* 349: 117-127.
- Britton, R.A., Powell, B.S., Dasgupta, S., Sun, Q., Margolin, W., Lupski, J.R., and Court, D.L. 1998. Cell cycle arrest in Era GTPase mutants: a potential growth rate-regulated checkpoint in *Escherichia coli*. *Mol. Microbiol.* 27: 739-750.
- Bult, C.J., White O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D., Kerlavage, A.R., Dougherty, B.A., Tomb, J., Adams, M.D., Reich, C.I., Overbeek, R., Kirkness, E.F., Weinstock, K.G., Merrick, J.M., Glodek, A., Scott, J.D., Geoghagen, N.S., Weidman, J.F., Fuhrmann, J.L., Nguyen, D.T., Utterback, T., Kelley, J.M., Peterson, J.D., Sadow, P.W., Hanna, M.C., Cotton, M.D., Hurst, M.A., Roberts, K.M., Kaine, B.B., Borodovsky, M., Klenk, H.P., Fraser, C.M., Smith, H.O., Woese, C.R., and Venter, J.C. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* 273: 1058-1073.
- Burland, V., Plunkett, G. III, Daniels, D.L. and Blattner, F.R. 1993. DNA sequence and analysis of 136 kb of the *Escherichia coli* genome: organizational symmetry around the origin of replication. *Genomics* 16: 551-561.
- Cabedo, H., Macian, F., Villarroya, M., Escudero, J.C., Martinez-Vicente, M., Knecht, E., and Armengod, M.E. 1999. The *Escherichia coli trmE* (*mmE*) gene, involved in tRNA modification, codes for an evolutionarily conserved GTPase with unusual biochemical properties. *EMBO J.* 18: 7063-7076.

- Caenorhabditis elegans* sequencing consortium. 1998. Genome sequence of the nematode *Caenorhabditis elegans*: A platform for investigating biology. *Science* 282: 2012-2018.
- Chen, S.-M., Takiff, H.E., Barber, A.M., Dubois, G.C., Bardwell, J.C.A., and Court, D.L. 1990. Expression and characterization of RNase III and Era proteins. *J. Biol. Chem.* 265: 2888-2895.
- Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skalton, S., Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S., and Barrell, B.C. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393: 537-544.
- Dassain, M., Leroy, A., Colosetti, L., Carole, S., and Bouche, J.P. 1999. A new essential gene of the 'minimal genome' affecting cell division. *Biochimie* 81: 889-895.
- Deckert, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L., Graham, D.E., Overbeek, R., Snead, M.A., Keller, M., Anjay, M., Huber, R., Feldman, R.A., Short, J.M., Olson G.J. and Swanson, R.V. 1998. The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* 392: 353-358.
- Decoster, E., Vassal, A., and Faye, G. 1993. MSS1, a nuclear-encoded mitochondrial GTPase involved in the expression of COX1 subunit of cytochrome c oxidase. *J. Mol. Biol.* 232: 79-88.
- Devitt, M.L., Maas, K.J. and Stafstrom, J.P. 1999. Characterization of DRG's, developmentally regulated GTP-binding proteins, from pea and *Arabidopsis*. *Plant Mol. Biol.* 39: 75-82.
- Elseviers, D., Petruccio, L.A., and Gallagher, P. 1984. Novel *E. coli* deficient in biosynthesis of 5-aminomethyl-2-thiouridine. *Nucleic Acids Res* 12: 3521-3534.
- Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G.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, H., Cotton, M.D., Utterback, T., Hanna, M.C., Nguyen, S.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen N.S., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O., and Venter, J.C. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. 1995. *Science* 269: 496-512.
- 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.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. III, and Venter, J.C. 1995. The minimal gene complement of *Mycoplasma genitalium*. *Science* 270: 397-403.
- Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R.A., Lathiagra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K., Gwinn, M., Dougherty, B., Tomb, J.-F., Fleischmann, R.D., Richardson, D., Peterson, J., Kerlavage, A.R., Quackenbush, J., Salzberg, S., Hanson, M., van-Vogt, R., Palmer, N., Adams, M.D., Gocayne, J.D., Weidman, J., Utterback, T., Wattley, L., McDonald, L., Artiach, P., Bowman, C., Garland, S., Fujii C., Cotton, M.D., Horst, K., Roberts, K., Hatch, B., Smith, H.O., and Venter, J.C. 1997. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* 390: 580-586.
- Fraser, C.M., Norris, S.J., Weinstock, G.M., White O., Sutton, G.G., Dodson, R., Gwinn, M., Hickey, E.K., Clayton, R., Ketchum, K.A., Sodergren, E., Hardham, J.M., McLeod, M.P., Salzberg, S., Peterson, J., Khalak, H., Richardson, D., Howell, J.K., Chidambaram, M., Utterback, T., McDonald, L., Artiach, P., Bowman, C., Cotton, M.D., Fujii, C., Garland, S., Hatch, B., Horst, K., Roberts, K., Wattley, L., Weidman, J., Smith, H.O., and Venter, J.C. 1998. Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science* 281: 375-388.
- Glass, J.I., Lewkowitz, E.J., Glass, J.S., Heiner, C.R., Chen, E.Y., and Cassell, G.H. unpublished. The complete sequence of *Ureaplasma urealyticum*: Alternate views of a minimal genome and sexually transmitted pathogen (<http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/framk?db=Genome&gi=151>).
- Gollop, N., and March, P.E. 1991. A GTP-binding protein (Era) has an essential role in growth rate and cell cycle control in *Escherichia coli*. *J. Bacteriol.* 173: 2265-2270.
- Hagervall, T.G., Pomerantz, S.C., and McCloskey, J.A. 1998. Reduced misreading of asparagine codons by *Escherichia coli* tRNA^{Asp} with hypomodified derivatives of 5-methylaminomethyl-2-thiouridine in the wobble position. *J. Mol. Biol.* 284: 33-42.
- Hall, A. (ed.) 2000. GTPases. Oxford University Press, Oxford, U.K.

- Heilig, R. unpublished. *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome structure and evolution. (<http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/framiki?db=Genome&gi=143>).
- Himmelreich, R., Hilbert, H., Plagens, H., Pirkel, E., Li, B.C., and Herrmann, R. 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res.* 24: 4420-4449.
- Hudson, J.D., and Young, P.G. 1993. Sequence of the *Schizosaccharomyces pombe gtp1* gene and identification of a novel family of putative GTP-binding proteins. *Gene* 125: 191-193.
- Hutchinson, C.A., Peterson, S.N., Gill, S.R., Chine, R.T., White, O., Fraser, C.M. Smith, H.O. and Venter, J.C. 1999. Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* 286: 2165-2169.
- Ingram, G.C., Simon, R., Carpenter, R., and Coen, E.S. 1998. The *Antirrhinum* ERG gene encodes a protein related to bacterial small GTPases and is required for embryonic viability. *Curr. Biol.* 8: 1079-1082.
- Kalman, S., Mitchell, W., Maranthe, R., Lammel, C., Fan, J., Hyman, R.W., Olinger, L., Grimwood, J., Davis, R.W., and Stephens, R.S. 1999. Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*. *Nature Genet.* 21: 385-389.
- Kanehisa, M. 2000. Post-genome informatics. Oxford University Press, Oxford, UK.
- Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiyama, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., and Tabata, S. 1996. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* 3: 103-136.
- Kawarabayashi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y., Sakai, M., Ogura, K., Otuka, R., Nakazawa, H., Takamiya, M., Ohfuku, Y., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida N., Oguchi, A., Aoki, K., Nakamura, Y., Robb, T.F., Horikoshi, K., Masuchi, Y., Shizuya, H., and Kikuchi, H. 1998. Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3. *DNA Res.* 5: 55-76.
- Kawarabayashi, Y., Hino, Y., Horikawa, H., Yamazaki, S., Haikawa, Y., Jinno, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A., Kosugi, H., Hosoyama, A., Fukui, S., Nagai, Y., Nishijama, K., Nakazawa, H., Takamiya, M., Masuda, S., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., Kubota, K., Nakamura, Y., Nomura, N., Sako, Y., and Kikuchi, H. 1999. Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1. *DNA Res.* 6: 83-101.
- Kjeldgaard, M., Nyborg, J., and Clark, B.F.C. 1996. The GTP-binding motif: variations on a theme. *FASEB J.* 10: 1347-1368.
- Klenk, H.P., Clayton, R.A., Tomb, J., White, O., Nelson, K.E., Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D., Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyrpides, N.C., Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S., Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B., Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A., Zhou, L., Overbeek, R., Gocayne J.D., Weidman, J.F., McDonald, L., Utterback, T., Cotton, M.D., Spriggs T., Artiach, P., Kaine, B.P., Sykes, S.M., Sadow, P.W., D'Andrea K.P., Bowman, C., Fujii, C., Garland, S.A., Mason, T.M., Olson, G.J., Fraser, C.M., Smith, H.O., Woese, C.R., and Venter J.C. 1997. The complete genome sequence of the hyperthermophilic sulfate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* 390: 364-370.
- Kumar, G.P., Laloraya, M., and Koide, S.S. 1996. Characterization of a uterine luminal fluid protein ULF-250 using N-terminal microsequencing and RT-PCR identifies a novel estrogen-regulated gene in the rat uterus. *FEBS Lett.* 399: 33-36.
- Kumar, S., Iwao, M., Yamagishi, T., Noda, M. and Asashima, M. 1993. Expression of GTP-binding protein gene *drg* during *Xenopus laevis* development. *Int. J. Dev. Biol.* 37: 539-546.
- Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C., Bron, S., Brouillet, S., Bruschi, W., Caldwell, B., Capuano, V., Carter, N.M., Chio, S.K., Codani, J.J., Connerton, I.F., Danchin, A., et al. 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* 390: 249-256.
- Lerner, C.G., and Inouye, M. 1991. Pleiotropic changes resulting from depletion of Era, an essential GTP-binding protein in *Escherichia coli*. *Mol. Microbiol.* 5: 951-957.
- Leung, H.C.E., Hagervall, T.G., Björk, G.R., and Winkler, M.E. 1998. Genetic locations and database accession numbers of RNA-modifying and -editing enzymes. In: *Modification and Editing of RNA*. H. Grosjean and R. Benne, eds. American Society for Microbiology, Washington, DC, p. 561-567.
- Li, B., and Trueb, B. 2000. DRG represents a family of two closely related GTP-binding proteins. *Biochim. Biophys. Acta* 149: 196-204.
- Lin, Y.P., Sharer, J.D., and March, P.E. 1994. GTPase-dependent signaling in bacteria: characterization of a membrane-binding site for Era in *Escherichia coli*. *J. Bacteriol.* 176: 44-49.
- Lin, B., Covalle, K.L., and Maddock, J.R. 1999. The *Caulobacter crescentus* CgtA protein displays unusual guanine nucleotide binding and exchange properties. *J. Bacteriol.* 181: 5825-5832.
- Maddock, J., Bhatt, A., Koch, M., and Skidmore, J. 1997. Identification of an essential *Caulobacter crescentus* gene encoding a member of the Obg family of GTP-binding proteins. *J. Bacteriol.* 179: 6426-6431.
- Mahajan, M.A., Park, S.T., and Sun, X.-H. 1996. Association of a novel GTP-binding protein, DRG, with TAL oncogenic proteins. *Oncogene* 12: 2343-2350.
- March, P.E. 1992. Membrane-associated GTPases in bacteria. *Mol. Microbiol.* 6: 1253-1257.
- Mehr, I.J., Long, C.D., Serkin, C.D., and Seifert, H.S. 2000. A homologue of the recombination-dependent growth gene, *rdgC*, is involved in gonococcal pili antigenic variation. *Genetics* 154: 523-532.
- Meier, T.I., Peery, R.B., McAllister, K.A., and Zhao, G. 2000. Era GTPase of *Escherichia coli*: binding to 16S rRNA and modulation of GTPase activity by RNA and carbohydrates. *Microbiology* 146: 1071-1083.
- Mewes, H.W., Albermann, K., Bahr, M., Frishman, D., Gleissner, A., Hani, J., Heumann, K., Kleine, K., Meierl, A., Oliver, S.G., Pfeiffer, F., and Zöllner, A. 1997. Overview of the yeast genome. *Nature* 387 (Suppl.): 7-65.
- Moszer, I., Glaser, P., and Danchin, A. 1995. Subtilist: a relational database for the *Bacillus subtilis* genome. *Microbiology* 141: 261-268.
- Moszer, I. 1998. The complete genome of *Bacillus subtilis*: From sequence annotation to data management and analysis. *FEBS Lett.* 430: 28-36.
- Mushegian, A.R., and Koonin, E.V. 1996. A minimal genome set for cellular life by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. USA* 93: 10268-10273.
- Myers, E.W., Sutton, G.G., Delcher, A.L., Dew, I.M., Fasulo, D.P., Flanigan, M.J., Kravitz, S.A., Mobarry, C.M., Reinert, K.H., Remington, K.A., Anson, E.L., Bolanos, R.A., Chou H.H., Jordan, C.M., Halpern, A.L., Lonardi, S., Beasley, E.M., Brandon, R.C., Chen, L., Dunn, P.J., Lai, Z., Liang, Y., Nusskern, D.R., Zhan, M., Zhang, Q., Zheng, X., Rubin, G.M., Adams, M.D., and Venter, J.C. 2000. A whole-genome assembly of *Drosophila*. *Science* 287: 2196-2204.
- Nelson, K.E., Clayton, R.A., Gills, R., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M., Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A., Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D., Eisen, J.A., White, O., Salzberg, S.L., Smith, H.O., Venter, J.C., and Fraser, C.M. 2000. Evidence for lateral gene transfer between archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399: 323-329.
- Okamoto, S., Itoh, M., and Ochi, K. 1997. Molecular cloning and characterization of the *obg* gene of *Streptomyces griseus* in relation to the onset of morphological differentiation. *J. Bacteriol.* 179: 170-179.
- Okamoto, S., and Ochi, K. 1998. An essential GTP-binding protein functions as a regulator for differentiation in *Streptomyces coelicolor*. *Mol. Microbiol.* 30: 107-119.
- Page, R.D.M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. *CABIOS*. 12: 357-358.
- Parada, L.F., Tabin, C.J., Shih, C., and Weinberg, R.A. 1982. Human EJ bladder carcinoma oncogene is homologue of Harvey sarcoma *ras* gene. *Nature* 297: 474-478.
- Parkhill, J., Wren, B.W., Mungall, K., Kettle, J.M., Churcher, C., Basham, D., Chillingworth, T., Davis, R.M., Feltwell, T., Holroyd, S., Jagels, K., Karlyshev, A., Moule, S., Pallen, M.J., Penn, C.W., Quail, M., Rajandream, M.A., Rutherford, K.M., VanVliet, A., Whitehead, S., and Barrell, B.G. 2000. The complete genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* 403: 665-668.
- Pillutla, R.C., Sharer, J.D., Gulati, P.S., Wu, E., Yamashita, Y., Lerner, C.G., Inouye, M., and March, P.E. 1995. Cross-species complementation of the indispensable *Escherichia coli era* gene highlights amino acid regions essential for activity. *J. Bacteriol.* 177: 2194-2196.
- Pillutla, R.C., Ahnn, J., and Inouye, M. 1996. Deletion of the putative effector region of Era, an essential GTP-binding protein in *Escherichia coli*, causes a dominant-negative phenotype. *FEMS Microbiol. Lett.* 143: 47-55.
- Riethdorf, S., Völker, U., Gerth, U., Winkler, A., Engelmann, S., and Hecker, M. 1994. Cloning, nucleotide sequence, and expression of the *Bacillus subtilis lon* gene. *J. Bacteriol.* 176: 6518-6527.
- Sayed, A., Matsuyama, S.-i., and Inouye, M. 1999. Era, an essential *Escherichia coli* small G-protein, binds to the 30S ribosomal subunit. *Biochem. Biophys. Res. Commun.* 264: 51-54.
- Sazuka, T., Tomooka, Y., Ikawa, Y., Noda, M. and Kumar, S. 1992a. DRG: a novel developmentally regulated GTP-binding protein. *Biochem.*

- Biophys. Res. Commun. 189: 363-370.
- Sazuka, T., Kinoshita, M., Tomooka, Y., Ikawa, Y., Noda, M. and Kumar, S. 1992b. Expression of DRG during murine embryonic development. *Biochem. Biophys. Res. Commun.* 189: 371-377.
- Schenker, T., Lach, C., Kessler, B., Calderara, S. and Trueb, B. 1994. A novel GTP-binding protein which is selectively repressed in SV40 transformed fibroblasts. *J. Biol. Chem.* 269: 25447-25453.
- Schmidt, R., Decatur, A.L., Rather, P.N., Moran, C.P., Jr., and Losick, R. 1994. *Bacillus subtilis* Lon protease prevents inappropriate transcription of genes under the control of the sporulation transcription factor σ^G . *J. Bacteriol.* 176: 6528-6537.
- Scott, J.M., and Haldenwang, W.G. 1999. Obg, an essential GTP binding protein of *Bacillus subtilis*, is necessary for stress activation of transcription factor σ^B . *J. Bacteriol.* 181: 4653-4660.
- Scott, J.M., Ju, J., Mitchell, T., and Haldenwang, W.G. 2000. The *Bacillus subtilis* GTP binding protein Obg and regulators of the σ^B stress response transcription factor cofractionate with ribosomes. *J. Bacteriol.* 182: 2771-2777.
- Shimmin, L.C., and Dennis P.P. 1989. Characterization of the L11, L1, L10 and L12 equivalent ribosomal protein gene cluster of the halophilic archaeobacterium *Halobacterium cutirubrum*. *EMBO J.* 8: 1225-1235.
- Smith, D.R., Doucette-Stamm, L.A., Deloughery, C., Lee, H.-M., Dubios, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lum W., Pothier, B., Qiu, D., Spadafora, R., Vicare, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiواني, N., Caruso, A., Bush, D., Safer, H., Patwell, D., Prahakar, S., McDougall, S., Shimer, G., Goyal, A., Pietrovski, S., Church, G.M., Daniels, C.J., Mao, J.-i., Rice, P., Nölling, J., and Reeve, J.N. 1997. Complete genome sequence of *Methanobacterium thermoautotrophicum* Δ H: functional analysis and comparative genomics. 1997. *J. Bacteriol.* 179: 7153-7155.
- Sommer, K.A., Petersen, G. and Bautz, E.K.F. 1994. The gene upstream of DmRP128 codes for a novel GTP-binding protein of *Drosophila melanogaster*. *Mol. Gen. Genet.* 242: 391-398.
- Stephens, R.S., Kalman, S., Lammel, C., Fan, J., Maranthe R., Aravind, L., Mitchell, W., Olinger, L., Tatusov, R.L., Zhao, Q., Koonin, E., and Davis, R.W. 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science* 282: 754-759.
- Tatusov, R.L., Koonin, E.V., and Lipman, D.J. 1997. A genomic perspective on protein families. *Science* 278: 631-637.
- Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28: 33-36.
- Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C., Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F., Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D., Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D., Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E., Citti, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H., Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Masignani, V., Pizza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R., Rappuoli, R., and Venter, J.C. Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. 2000. *Science* 187: 1809-1815.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673-4680.
- Tomb, J.-F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A., Nelson K., Quackenbush, J., Zhou, L., Kirkness, E.F., Peterson, S., Loftus, B., Richardson, D., Dodson R., Khalak, H.G., Glodek, A., McKenney, K., Fitzgerald, L.M., Lee, N., Adams, M.D., Hickey, E.K., Berg, D.E., Gocayne, J.D., Utterback, T.R., Peterson, J.D., Kelley, J.M., Karp, P.D., Smith, H.O., Fraser, C.M., and Venter, J.C. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 388: 539-547.
- Trach, K., and Hoch, J.A. 1989. The *Bacillus subtilis* *spo0B* sporulation operon encodes an essential GTP-binding protein. *J. Bacteriol.* 171: 1362-1371.
- Vidwans, S.J., Ireton, K., and Grossman, A.D. 1995. Possible role for the essential GTP-binding protein Obg in regulating the initiation of sporulation in *Bacillus subtilis*. *J. Bacteriol.* 177: 3308-3311.
- White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., Moffatt, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S., Smith, H.O., Venter, J.C., and Fraser, C.M. 1999. Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R 1. *Science* 286: 1571-1577.
- Zhao, X.F., and Aplan, P.D. 1998. SCL binds the human homologue of DRG in vivo *Biochim. Biophys. Acta* 1448: 109-114.

