

Research

# Gene discovery within the planctomycete division of the domain Bacteria using sequence tags from genomic DNA libraries

Cheryl Jenkins\*<sup>§</sup>, Vishram Kedar<sup>†§</sup> and John A Fuerst<sup>‡§</sup>

Addresses: \*Department of Microbiology, University of Washington, Seattle, WA 98195, USA. <sup>†</sup>Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, North Carolina, NC 27759-7260, USA. <sup>‡</sup>Department of Microbiology and Parasitology, School of Molecular and Microbial Sciences, University of Queensland, Brisbane, Queensland 4072, Australia. <sup>§</sup>All authors contributed equally to this work.

Correspondence: John A Fuerst. E-mail: j.fuerst@mailbox.uq.edu.au

Published: 30 May 2002

*Genome Biology* 2002, **3**(6):research0031.1-0031.11

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2002/3/6/research/0031>

© 2002 Jenkins et al., licensee BioMed Central Ltd  
(Print ISSN 1465-6906; Online ISSN 1465-6914)

Received: 1 February 2002

Revised: 15 April 2002

Accepted: 17 April 2002

## Abstract

**Background:** The planctomycetes comprise a distinct group of the domain Bacteria, forming a separate division by phylogenetic analysis. The organization of their cells into membrane-defined compartments including membrane-bounded nucleoids, their budding reproduction and complete absence of peptidoglycan distinguish them from most other Bacteria. A random sequencing approach was applied to the genomes of two planctomycete species, *Gemmata obscuriglobus* and *Pirellula marina*, to discover genes relevant to their cell biology and physiology.

**Results:** Genes with a wide variety of functions were identified in *G. obscuriglobus* and *Pi. marina*, including those of metabolism and biosynthesis, transport, regulation, translation and DNA replication, consistent with established phenotypic characters for these species. The genes sequenced were predominantly homologous to those in members of other divisions of the Bacteria, but there were also matches with nuclear genomic genes of the domain Eukarya, genes that may have appeared in the planctomycetes via horizontal gene transfer events. Significant among these matches are those with two genes atypical for Bacteria and with significant cell-biology implications - integrin alpha-V and inter-alpha-trypsin inhibitor protein - with homologs in *G. obscuriglobus* and *Pi. marina* respectively.

**Conclusions:** The random-sequence-tag approach applied here to *G. obscuriglobus* and *Pi. marina* is the first report of gene recovery and analysis from members of the planctomycetes using genome-based methods. Gene homologs identified were predominantly similar to genes of Bacteria, but some significant best matches to genes from Eukarya suggest that lateral gene transfer events between domains may have involved this division at some time during its evolution.

## Background

The planctomycetes (order Planctomycetales) comprise a distinct group of the domain Bacteria that forms a separate kingdom-level division on the basis of 16S rRNA analyses [1-6]. Planctomycetes are currently represented by only a few cultured and characterized heterotrophic members

isolated from aquatic habitats; however, the presence of planctomycete sequences in 16S rRNA clone libraries constructed from environmental samples reveals that these organisms occupy diverse ecological niches including wastewater and waste-treatment bioreactors [7-9], marine sediments and organic aggregates [10-12], and oxic and anoxic

terrestrial habitats [13-15]. In many of these environments planctomycetes make up a significant proportion of the microbial population [7,16], indicating that they may have a significant role in the cycling of organic or inorganic compounds. The recent discovery that the “missing lithotroph” responsible for the anaerobic oxidation of ammonium (anammox process) is an autotrophic planctomycete [8], and the existence of other planctomycetes with similar activities [9], highlights the potential importance of these bacteria for the flux of nutrients in the environment and indicates the potentially wide physiological diversity of the division.

Consistent with their phylogenetic distinctiveness, the planctomycetes possess a series of unusual phenotypic characteristics common to members of the division [5]. These include budding reproduction, peptidoglycan-less (proteinaceous) cell walls and a complex internal ultrastructure [17]. Most notably, cells of at least three species exhibit eukaryote-like membrane-bounded nuclear regions: the genomic DNA of *Gemmata obscuriglobus* is enclosed by two membranes [17,18], whereas that of *Pirellula marina* and *Pi. staleyi* is enveloped by a single membrane [19]. Membranes surrounding the nucleoid are unique to the planctomycetes among members of the domain Bacteria. This feature is only one aspect of a unique type of cell organization shared by all planctomycetes examined so far, involving compartmentalization via intracytoplasmic membranes [17].

Despite the interesting ecological and cell biological aspects of the planctomycetes, molecular studies on this group have been relatively few. To date, DNA sequencing studies on the planctomycetes have involved genes for small subunit (SSU) and large subunit (LSU) rRNA [1,20,21], and a small number of protein-coding genes, including those for the  $\beta$ -subunit of ATPase from *Pi. marina* [22], the DnaK heat-shock protein (HSP70) from *Pirellula* and *Planctomyces* species [23] and the gene *rpoN* for sigma factor 54 from *Planctomyces limnophilus* [24]. Phylogenetic analyses using these gene sequences have failed to elucidate the evolutionary relationship of the planctomycetes relative to the other divisions of the Bacteria. Some sequence analyses show that the planctomycetes form a sister-group of the chlamydiae [1,25], while others place them as a deeply branching division [1,22]. In more recent analyses using the gene for the conserved protein, elongation factor-Tu, inconsistency in the branch position of the planctomycetes was attributed to long-branch attraction effects [26].

In this study we applied a random-sequencing approach to the genomes of two planctomycete species, *G. obscuriglobus* and *Pi. marina*, in order assist discovery of genes that may be relevant to their ecology, phylogeny and cell biology. These species are members of phylogenetically distinct genera within the planctomycete division as judged by 16S rRNA sequence analysis [20]. While it is noteworthy that members of both the genus *Pirellula* (*Pirellula* sp. strain 1)

and the genus *Gemmata* (*Gemmata* strain Wa1-1) are currently the subject of whole-genome studies by the REGX project and Integrated Genomics, respectively [27,28], to date, no reports have been published concerning analyses of the resulting genome sequence data. Furthermore, sequence data from these whole-genome projects is not currently accessible to the public. Thus the random-sequencing approach used in this study provides the first insight into the genomes of planctomycetes. Such sequence tag approaches have proved an effective method of gene discovery in other prokaryotes [29-32].

## Results and discussion

Sequence tags from *G. obscuriglobus* and *Pi. marina* that represent putative protein-coding genes were identified by comparison of individual clone nucleotide sequence translated in all reading frames against protein-sequence databases using the BLASTX algorithm (Tables 1,2). Only sequence matches with expected (*e*) values below  $E^{-4}$  (as determined by BLASTX) were considered to be significant [33] and are presented here. Of the clones sequenced from *G. obscuriglobus*, 27% (43/160) showed significant matches with known proteins or hypothetical proteins, whereas 32% (29/91) of *Pi. marina* sequences had significant database matches. In addition to these matches, nucleotide sequences for several clones were shown by BLASTN analyses against nucleotide-sequence databases to represent SSU and LSU rRNA gene sequences (data not shown). Best matches for protein-coding genes were predominantly with gene homologs from Bacteria, but there were also some best matches with genes from Archaea and most notably from Eukarya (see below). Within the Bacteria, best matches did not cluster within any one division, although matches with members of Gram-positive and cyanobacterial divisions were common in the case of *G. obscuriglobus*.

Random sequence tags from the two planctomycete species displayed homology with proteins of diverse function, including metabolic enzymes (30 and 38% of total for *G. obscuriglobus* and *Pi. marina* respectively), and proteins involved with transport (7 and 10%), regulation (9 and 10%) and the central processes of translation (9% and 7%) and DNA replication (16 and 10%). Selected proteins from these functional groups are described in detail below.

## Metabolism and biosynthesis

A large proportion of the metabolism and biosynthesis genes identified in *G. obscuriglobus* and *Pi. marina* were homologous to those coding for enzymes involved in amino-acid and vitamin biosynthesis. For *G. obscuriglobus*, close database matches were found to the enzymes isopropyl malate dehydratase, diamino-pimelate epimerase, asparagine synthetase and 2-amino-3-ketobutyrate coenzyme A ligase, which are involved in the biosynthesis of leucine, lysine, asparagine and glycine, respectively. In addition, a homolog

Table 1

Putative protein-coding genes identified in *Gemmata obscuriglobus*

Clone number	GenBank accession	Best BLASTX match	E value of best match	Species of best match (GenBank accession number)	Domain of best match*
<b>Metabolism and biosynthesis</b>					
G63	BH001096	Putative isopropylmalate dehydratase	9.00E-09	<i>Methanococcus jannaschii</i> (Q58673)	B
G80	BH001097	Heptaprenyl diphosphate synthase component	6.00E-18	<i>Bacillus stearothermophilus</i> (P55785)	B
G83	BH001098	Peptide methionine sulfoxide reductase	4.00E-11	<i>Deinococcus radiodurans</i> (Q9RTB6)	B
G84	BH001099	N4-(beta-N-acetylglucosaminy)-L-asparagine precursor	6.00E-27	<i>Flavobacterium meningosepticum</i> (Q47898)	B
G87	BH001100	3-Oxoadipate enol-lactonase	2.00E-08	<i>Acinetobacter calcoaceticus</i> (P00632)	B
G103	BH001101	Diaminopimelate epimerase	2.00E-43	<i>Synechocystis</i> PCC6803 (P74667)	B
G108	BH001102	FDHD protein homolog	2.00E-45	<i>Streptomyces coelicolor</i> (Q9ZBWO)	B
G110	BH001103	Acyl-CoA fatty-acid delta (9) desaturase	3.00E-05	<i>Rattus norvegicus</i> (P07308)	E
G136	BH001104	1,4-Alpha-glucan branching enzyme	4.00E-11	<i>Bacillus caldolyticus</i> (P30537)	B
G179	BH001105	Putative asparagine synthetase (glutamine-hydrolyzing)	2.00E-07	<i>Mycobacterium tuberculosis</i> (O10374)	B
G184	BH001106	2-Amino-3-ketobutyrate coenzyme A ligase	5.00E-25	<i>Bacillus subtilis</i> (O31777)	B
G222	BH001107	Threonine dehydratase biosynthetic (threonine deaminase)	9.00E-53	<i>Burkholderia cepacia</i> (P53607)	B
G240	BH001108	Ribonucleotide reductase large chain	1.00E-05	<i>Schizosaccharomyces pombe</i> (P36602)	E
<b>Transport</b>					
G74	BH001109	Large conductance mechanosensitive channel	8.00E-10	<i>Haemophilus influenzae</i> (P44789)	B
G139	BH001110	Cadmium efflux system accessory protein	4.00E-05	<i>Listeria monocytogenes</i> (Q56405)	B
G185	BH001111	Sulfate transport ATP-binding protein CysA	6.00E-35	<i>Synechocystis</i> PCC6803 (P74548)	B
<b>Regulation and signal transduction</b>					
G99	BH001112	OmpR homolog	6.00E-11	<i>Guillardia theta</i> chloroplast (AF041468)	E(c)
G114	BH001113	Nitrogen regulation protein NtrB	2.00E-16	<i>Bradyrhizobium</i> sp. (P10578)	B
G217	BH001114	Transcriptional regulatory protein AfsQ1	3.00E-09	<i>Streptomyces coelicolor</i> (Q04942)	B
G223	BH001115	Transcriptional repressor SmtB	2.00E-05	<i>Synechococcus</i> PCC7942 (P30340)	B
<b>DNA replication/insertion</b>					
VG33	BH001116	Replicative DNA helicase	2.00E-11	<i>Synechocystis</i> PCC6803 (Q55418)	B
G86	BH001117	H repeat-associated protein in RHSC-PHRB intergenic region	1.00E-17	<i>Escherichia coli</i> (P75741)	B
G142	BH001118	Tral protein (DNA helicase I)	2.00E-12	<i>Escherichia coli</i> (P14565)	B
G165	BH001119	Insertion element IS402 hypothetical 24 kd	6.00E-05	<i>Burkholderia cepacia</i> (P24536)	B
G182	BH001120	Putative helicase HelY	5.00E-05	<i>Mycobacterium tuberculosis</i> (Q10701)	B
G196	BH001121	Tral protein (DNA helicase I)	3.00E-12	<i>Escherichia coli</i> (P14565)	B
G235	BH001122	Probable transposase for insertion sequence element IS702	3.00E-18	<i>Fremyella diplosiphon</i> (Q00462)	B
<b>Translation</b>					
G51	BH001123	50S ribosomal protein L23	1.00E-13	<i>Aquifex aeolicus</i> (O66433)	B
G154	BH001124	50S ribosomal protein L4	3.00E-10	<i>Bacillus stearothermophilus</i> (P28601)	B
G156	BH001125	Arginyl-tRNA synthetase	4.00E-21	<i>Escherichia coli</i> (P11875)	B
G230	BH001126	Probable tRNA methyl transferase	5.00E-05	<i>Synechocystis</i> PCC6803 (P73755)	B
<b>Miscellaneous</b>					
VG17	BH001127	Serine-threonine protein kinase	1.00E-22	<i>Myxococcus xanthus</i> (P54737)	B
G92	BH001128	Cell division protein FtsH	4.00E-04	<i>Lactococcus lactis</i> subsp. <i>lactis</i> (P46469)	B
G149	BH001129	Integrin alpha-V (vitronectin receptor alpha subunit precursor)	9.00E-05	<i>Homo sapiens</i> (P06756)	E
G159	BH001130	Serine protease DO-like precursor	7.00E-10	<i>Chlamydia trachomatis</i> (P18584)	B
G189	BH001131	H repeat-associated protein in RHSE-NARV intergenic region	2.00E-13	<i>Escherichia coli</i> (P28917)	B
G191	BH001132	Ribonuclease inhibitor	6.00E-05	<i>Sus scrofa</i> (P10775)	E
<b>Hypothetical proteins</b>					
VG21	BH001133	Hypothetical 47.5 kDa protein Cy9c4.09	1.00E-04	<i>Mycobacterium tuberculosis</i> (Q50739)	B
G76	BH001134	Hypothetical protein in DHLA 3' region	3.00E-17	<i>Xanthobacter autotrophicus</i> (P22644)	B
G96	BH001135	Hypothetical 42.1 kDa protein	5.00E-04	<i>Bacillus subtilis</i> (O31712)	B
G137	BH001136	Hypothetical protein in DHLA 3' region	4.00E-31	<i>Xanthobacter autotrophicus</i> (P22644)	B
G180	BH001137	Hypothetical 30.2 kDa protein SLR1717	1.00E-29	<i>Synechocystis</i> strain PCC6803 (P73846)	B
G231	BH001138	Hypothetical protein MJ0682	2.00E-35	<i>Methanococcus jannaschii</i> (Q58095)	A

\*B, Bacteria; A, Archaea; E, Eukarya; E(c), eukaryal chloroplast.

**Table 2****Putative protein-coding genes identified in *Pirellula marina***

Clone Number	GenBank accession	Best BLASTX match	E value of best match	Species of best match	Domain of best match*
<b>Metabolism and biosynthesis</b>					
B8	BH001139	Transaldolase	4.00E-14	<i>Mycobacterium tuberculosis</i> (O06812)	B
D10	BH001140	Flagellar biosynthesis protein FlhA	4.00E-05	<i>Aquifex aeolicus</i> (O67265)	B
E9	BH001141	Orotate phosphoribosyltransferase	4.00E-07	<i>Methanococcus jannaschii</i> (Q58509)	A
pI2-A7	BH001142	Glutamate synthase	5.00E-25	<i>Bacillus subtilis</i> (P39812)	B
pI2-B11	BH001143	Uroporphyrin-III C-methyltransferase	1.00E-22	<i>Bacillus subtilis</i> (P42437)	B
pI2-C9	BH001144	Glutamate-1-semialdehyde 2,1-aminomutase (GSA)	7.00E-14	<i>Xanthomonas campestris</i> (Q06741)	B
pI2-D11	BH001145	Phosphoribosylamine-glycine ligase (GARS)	4.00E-21	<i>Homo sapiens</i> (P22102)	E
pI2-E2	BH001146	Cytochrome c oxidase polypeptide I	4.00E-08	<i>Squalus acanthias</i> mitochondrion (Q9ZZ52)	E(m)
pI2-F2	BH001147	1-Deoxy-D-xylulose 5-phosphate reductoisomerase	9.00E-44	<i>Zymomonas mobilis</i> (Q9X5F2)	B
pI2-F9	BH001148	Putative sulfatase	1.00E-09	<i>Escherichia coli</i> (P31447)	B
pI2-G7	BH001149	Glucosamine fructose-6-phosphate aminotransferase	6.00E-05	<i>Pyrococcus horikoshii</i> (O57981)	A
<b>Transport</b>					
A2	BH001150	Vitamin B12 receptor precursor	2.00E-20	<i>Escherichia coli</i> (P06129)	B
C5	BH001151	ABCA protein	3.00E-14	<i>Aeromonas salmonicida</i> (Q07698)	B
pI2-A3	BH001152	O antigen export system ATP-binding protein RfbB	1.00E-06	<i>Klebsiella pneumoniae</i> (Q48476)	B
<b>Regulation and signal transduction</b>					
C9	BH001154	Protein-glutamate methyltransferase CheB	2.00E-08	<i>Bacillus subtilis</i> (Q05522)	B
pI2-A11	BH001155	Sensor protein BarA	5.00E-10	<i>Escherichia coli</i> (P26607)	B
pI2-G4	BH001156	Chemotaxis protein CheC	2.00E-05	<i>Bacillus subtilis</i> (P40403)	B
<b>DNA replication/insertion</b>					
C6	BH001157	DNA polymerase I (Pol I)	3.00E-37	<i>Haemophilus influenzae</i> (P43741)	B
C8	BH001158	Transposase for insertion sequence element	1.00E-05	<i>Lactococcus lactis</i> subsp. <i>lactis</i> (P35878)	B
G1	BH001159	ATP-dependent DNA helicase HUS2	4.00E-10	<i>Schizosaccharomyces pombe</i> (Q09811)	E
<b>Translation</b>					
B2	BH001160	Leucyl-tRNA synthetase (leucine tRNA ligase)	1.00E-25	<i>Thermotoga maritima</i> (Q9WY15)	B
pI2-A10	BH001161	Alanyl tRNA synthetase	6.00E-14	<i>Bacillus subtilis</i> (Q34526)	B
<b>Miscellaneous</b>					
F3	BH001162	Ribonuclease E (RNase E)	1.00E-38	<i>Haemophilus influenzae</i> (P44443)	B
H6	BH001163	Inter-alpha-trypsin inhibitor heavy chain H2 precursor	2.00E-06	<i>Mus musculus</i> (Q61703)	E
pI2-D4	BH001164	Putative methionine aminopeptidase B (MAP) peptidase M	3.00E-08	<i>Rickettsia prowazekii</i> (Q9ZCD3)	B
<b>Hypothetical proteins</b>					
A4	BH001153	Hypothetical transcriptional regulator HT1052	6.00E-05	<i>Haemophilus influenzae</i> (P45008)	B
pI2-E10	BH001165	Hypothetical protein MJ1612	4.00E-15	<i>Methanococcus jannaschii</i> (Q59007)	A
pI2-F11	BH001166	Hypothetical 25.5 kDa protein in HupB-COF intergenic region	2.00E-09	<i>Escherichia coli</i> (P77756)	B
pI2-G10	BH001167	Hypothetical 17.3 kDa protein CYIA11.16C	1.00E-13	<i>Mycobacterium tuberculosis</i> (Q50606)	B

\*B, Bacteria; A, Archaea; E, Eukarya; E(m), eukaryal mitochondrion.

of threonine deaminase was identified. This enzyme catalyzes the formation of  $\alpha$ -ketobutyrate from threonine, an intermediary step in isoleucine biosynthesis [34]. In *Pi. marina*, a gene putatively coding for the NADH/NADPH-dependent enzyme glutamate synthase was found. This enzyme, a glutamine oxoglutarate aminotransferase (GOGAT), is important in the incorporation of inorganic nitrogen into cell material by conversion of 2-oxoglutarate and L-glutamine to L-glutamate [35].

Homologs of genes involved in the biosynthesis of vitamins and cofactors were also identified in *Pi. marina*. These include glutamate-1-semialdehyde-2,1-aminomutase, which catalyzes the final step in the conversion of glutamate to 4-aminolevulinic acid (the precursor of tetrapyrrole synthesis), uroporphyrin-III C-methyltransferase and 1-deoxy-xylulose 5-phosphate reductoisomerase (DXP). Uroporphyrin-III C-methyltransferase is involved in the synthesis of both cobalamin (vitamin B12) and siroheme (a cofactor for sulfite



and nitrite reductases) [36], whereas DXP catalyzes the initial reactions in biosynthesis of isoprenoids, which are precursors of some vitamins such as vitamin A [37].

The large proportion of amino-acid and vitamin biosynthesis genes identified in the two planctomycete species suggests that they may be able to synthesize many of these growth factors *de novo*. This finding is consistent with the ability of planctomycetes to grow in oligotrophic conditions and habitats [38,39]. Many planctomycetes, including *Pi. marina* can be cultured on minimal media with only a modest or no requirement for the addition of vitamins [38-41].

Other sequences homologous to metabolic genes of interest that were identified include heptaprenyl diphosphate synthase from *G. obscuriglobus* and glucosamine fructose-6-phosphate aminotransferase (GlmS) from *Pi. marina*. Heptaprenyl diphosphate synthase has been shown in *Bacillus* species to catalyze the synthesis of the prenyl side chain of menaquinone-7 [42]. The presence of a menaquinone biosynthesis gene in *G. obscuriglobus* is consistent with a previous report demonstrating that the planctomycetes possess menaquinones [43], despite the fact that ubiquinones rather than menaquinones are typically associated with aerobic bacteria.

GlmS reversibly catalyses the formation of D-glucosamine-6-phosphate and L-glutamate from D-fructose-6-phosphate, using L-glutamine as the ammonia source [44]. This is the initial step in a pathway that produces N-acetyl-D-glucosamine as the final product, a compound that is a major component of the bacterial cell-wall constituent peptidoglycan. Interestingly, peptidoglycan is not a component of the predominantly protein cell walls of planctomycetes, including those of *Pirellula* and *Gemmata* species [20,45,46]. Therefore, it is possible that the net reaction catalyzed by GlmS in planctomycetes is the reverse reaction - the conversion of glucosamine-6-phosphate to fructose-6-phosphate (an intermediate compound in the glycolysis pathway). Reactions flowing in this direction would be consistent with the ability of planctomycetes to utilize N-acetylglucosamine as a sole source of carbon and nitrogen [39]. Alternatively, it is possible that N-acetylglucosamine is synthesized but utilized in the formation of glycoprotein or polysaccharide rather than in that of peptidoglycan.

Homologs of the genes for the enzymes phosphoribosyl-amine-glycine ligase (GARS) and orotate phosphoribosyl transferase (OPRT) were also found in *Pi. marina*. These enzymes are involved in the pathways of pyrimidine and purine biosynthesis, respectively. Furthermore, a sequence homolog of ribonucleotide reductase, an enzyme central to deoxyribonucleotide synthesis, was discovered in *G. obscuriglobus*. Significantly, two out of three of these gene sequences had their closest BLASTX matches with organisms from domains other than the Bacteria. The closest

match to the putative *Pi. marina* OPRT gene was with *Methanococcus jannaschii*, a member of the Archaea, whereas the putative gene for ribonucleotide reductase had its closest match with a eukaryotic organism. The GARS sequence homolog had a top match with a cyanobacterium; however, some of the next best matches were with eukaryotes and other bacteria (data not shown). Certain enzymes involved in the eukaryotic purine and pyrimidine biosynthesis pathways are believed to have been derived from prokaryotes by horizontal gene transfer [47,48]. If the genes involved in these pathways are subject to frequent horizontal transfers, this may also explain the presence of the eukaryote- and archaea-like genes in the planctomycetes.

### Transporters

A number of genes homologous to those involved in transport across membranes were sequenced from both *G. obscuriglobus* and *Pi. marina*.

A homolog of a gene encoding a large-conductance mechanosensitive channel (MscL) was identified in *G. obscuriglobus*. In *Escherichia coli*, MscL forms a channel in the inner membrane that has been shown to gate in response to tension in the lipid bilayer. The MscL is believed to protect cells from lysis upon osmotic shock [49]. MscL channels may have a particularly important role in planctomycetes, which frequently inhabit osmotically stressful oligotrophic environments but lack the rigidifying peptidoglycan component in their cell walls.

Several transport protein gene homologs were also identified in *Pi. marina*. Of particular interest were two genes putatively involved in the synthesis and transport of capsular and O polysaccharide, RfbB [50] and ABCA protein [51], suggesting that this organism may be able to produce and secrete extracellular polysaccharides or lipopolysaccharides, in a similar manner to Gram-negative organisms. This is consistent with previous studies on planctomycete lipids, which suggested the presence of lipopolysaccharide lipid A in *Planctomyces*, *Pirellula* [52] and *Isosphaera* [53] species on the basis of 3-hydroxy fatty acid detection.

Also in *Pi. marina*, a gene homologous to that encoding the FlhA protein was identified. In various bacteria this protein is believed to be essential for the export of the flagellar hook protein and hook-capping protein to the periplasm [54]. This protein may play a similar role in *Pi. marina*, as it possesses a single sheathed flagellum [40].

### Regulation and signal transduction

Several genes putatively involved in signal transduction and regulation of transcription were identified. Homologs of the osmotic shock response regulator OmpR and the nitrogen regulation sensor kinase NtrB were identified in *G. obscuriglobus*. Thus, in this organism, there is evidence for signal transduction proteins that form 'two-component

systems', simple regulatory systems operating through protein phosphorylation cascades that allow adaptation to environmental changes.

In *Pi. marina*, there is evidence for the presence of a chemotaxis signal transduction system. Homologs of two proteins involved in bacterial chemotaxis in *Bacillus subtilis*, CheB and CheC, were identified. CheB has been shown to display methylesterase activity and is involved in modification of the cytoplasmic domains of the chemoreceptor methyl-accepting chemotaxis proteins [55], while CheC is involved in determining the direction of flagellar rotation [56].

### DNA replication

A number of sequence tags with matches to enzymes involved in DNA replication were identified, including several genes putatively encoding helicases. A gene for a replicative DNA helicase (DnaB), which acts at the replication fork by disrupting the hydrogen bonds between the complementary base pairs [57], was identified in *G. obscuriglobus*. *Pi. marina* also contains a putative DNA helicase that had its closest match with the Hus2 helicase of the eukaryote *Schizosaccharomyces*. The Hus2 helicase is highly homologous to the RecQ family of helicases of bacteria and other eukaryotes [58]. Bacterial RecQ helicases were also among the top BLASTX hits for this sequence tag (data not shown); thus, the eukaryote match for this sequence may not be highly significant.

Two additional helicases were putatively identified from *G. obscuriglobus*. These included an RNA helicase (HelY) and the TraI protein. TraI has been shown to possess both site-specific nicking and DNA unwinding activity, and is specifically involved in the conjugal DNA transfer of plasmid DNA [59]. The presence of plasmids in *G. obscuriglobus* has not yet been investigated; however, another member of the planctomycetes, *Pl. limnophilus*, has been shown by pulse-field gel electrophoresis to contain an extrachromosomal element [60]. The transfer of plasmid DNA may thus be a useful function in planctomycetes as in other bacteria, consistent with the demonstrated ability of a planctomycete, *Pl. maris*, to acquire conjugative plasmids [61].

A second class of enzyme involved in DNA replication and repair was identified in *Pi. marina*. DNA polymerase I is responsible for processing Okazaki fragments on the lagging strand during replication [62]. In bacteria, this enzyme also possesses 5'-3' exonuclease activity and in some organisms, including *E. coli*, 3'-5' exonuclease activity [63]. The 3'-5' exonuclease activity is encoded by the central portion of the *polA* gene and organisms deficient in this activity lack three essential protein motifs in this region [64]. The DNA polymerase I sequence tag from *Pi. marina* encompasses two out of three of these motifs (NLKYD and YAAE), both of which are conserved relative to the *E. coli polA* gene. This suggests that the *Pi. marina* DNA polymerase I may also

possess the 3'-5' exonuclease activity, an hypothesis testable by enzyme assay.

### Translation

Several sequence tags putatively coding for proteins central to translation were identified. These included two classes of proteins: ribosomal proteins and aminoacyl-tRNA synthetases. In *G. obscuriglobus*, two homologs of ribosomal protein genes were found, one for ribosomal protein L23 and one for L4. In the assembled ribosome, L23 is located within the A site of the 50S subunit [65] and is one of the few ribosomal proteins that directly binds LSU rRNA. In bacteria, the L4 protein is implicated in both ribosomal peptidyltransferase activity and in some cases, autoregulation of the S10 ribosomal protein operon [66].

Three sequence tags putatively coding for aminoacyl-tRNA synthetases, enzymes catalyzing the esterification of amino acids to their respective tRNA molecules, were identified in *G. obscuriglobus* and *Pi. marina*. Two of these belong to the class I tRNA synthetases (arginyl- and leucyl- tRNA synthetase), which are typically monomeric, and one belongs to class II tRNA synthetases (alanyl-tRNA synthetase), which are di- or tetrameric.

Both the ribosomal proteins and the aminoacyl-tRNA synthetases are central to the translational process; thus, they have been employed as phylogenetic markers in a number of studies. In particular, ribosomal proteins (when used in concatenated analyses in conjunction with rRNAs) and leucyl-tRNA synthetases appear to uphold the ribosomal RNA-based phylogeny of the three domains as well as a large proportion of the inter-bacterial relationships [67,68]. From this perspective, the ribosomal protein sequences from *G. obscuriglobus* and the leucyl-tRNA synthetase sequence from *Pi. marina* may prove useful in re-examining the phylogenetic position of the planctomycetes within the Bacteria. In contrast, both the alanyl- and arginyl- tRNA synthetases do not uphold the rRNA tree, displaying multiple horizontal gene transfers between lineages [68]. Nonetheless, the alanyl-tRNA synthetase of certain bacterial taxa, including the *Chlamydia* group, is believed to contain an important amino-terminal signature sequence [69]. The alanyl-tRNA synthetase sequence tag generated from *Pi. marina* does not encompass this region, and the retrieval of the remaining portion of this gene made possible by using the sequence tag as basis for a probe may aid in determining whether the planctomycetes share this signature with the chlamydiae and thus whether the purported relationship between these two groups [1,25] is supported.

### Eukaryote-like genes

Several sequence tags from *G. obscuriglobus* and *Pi. marina* displayed closest BLASTX matches with proteins from members the domain Eukarya, including an acyl-CoA fatty-acid delta(9) desaturase, a ribonucleotide reductase large

chain, and a ribonuclease inhibitor in the case of *G. obscuriglobus*, and a phosphoribosylamine-glycine ligase and the ATP-dependent DNA helicase HUS2 in the case of *Pi. marina*. However, only two of these matches were with proteins that are considered atypical in Bacteria. These include a homolog of integrin alpha-V in *G. obscuriglobus* and a homolog of inter-alpha-trypsin inhibitor protein in *Pi. marina*.

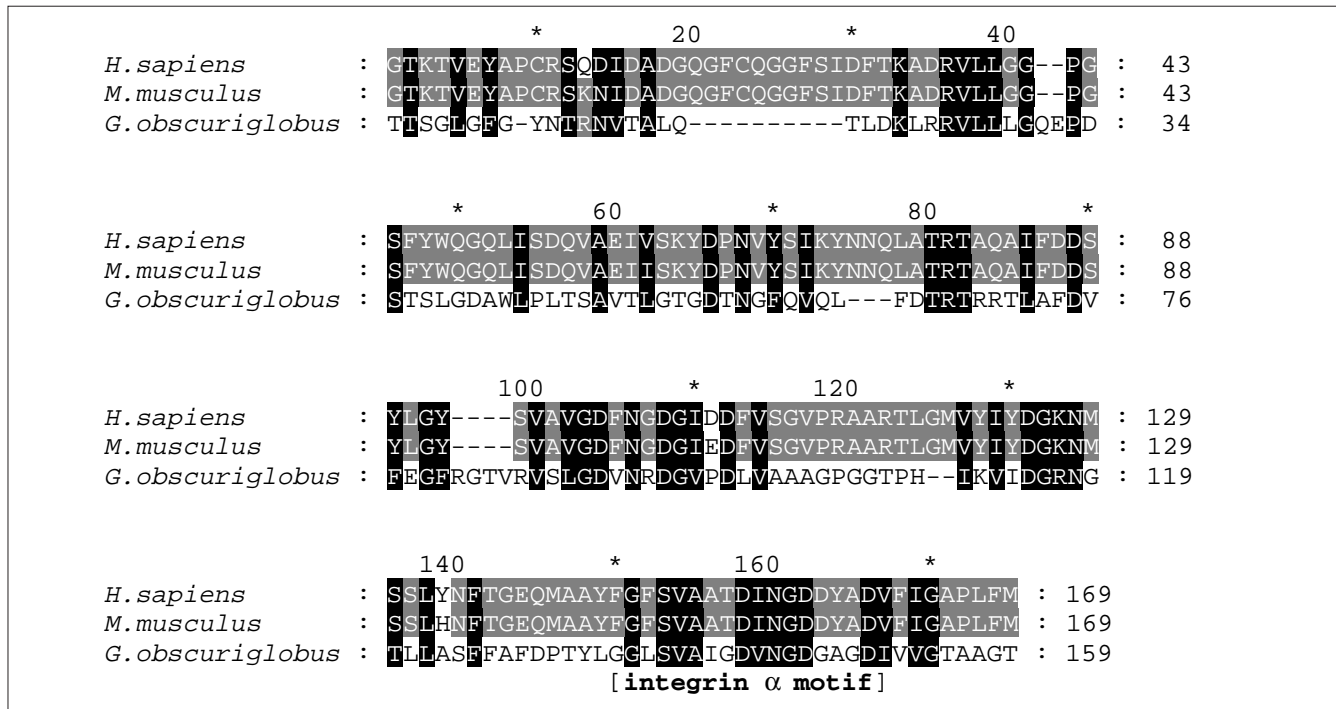
In higher eukaryotes, integrins are important in transmembrane signal transduction from the extracellular matrix and in organization of the cytoskeleton [70], and members of the inter-alpha-trypsin inhibitor protein family play a major role in extracellular matrix stability [71].

Additional evidence for homology between the planctomycete sequence tags and integrin alpha-V and inter-alpha-trypsin inhibitor was found by conducting iterative database searches with the position-specific iterative program PSI-BLAST. In both cases, PSI-BLAST converged on the integrin and inter-alpha-trypsin inhibitor proteins respectively. Furthermore, multiple sequence alignment of the translated sequence tags with these proteins revealed a number of regions of conservation (Figures 1 and 2). A motif characteristic of integrin alpha-chain proteins, G..[ILV]...D..[DN]...[FILMV][FILMV][ILMV] (single-letter

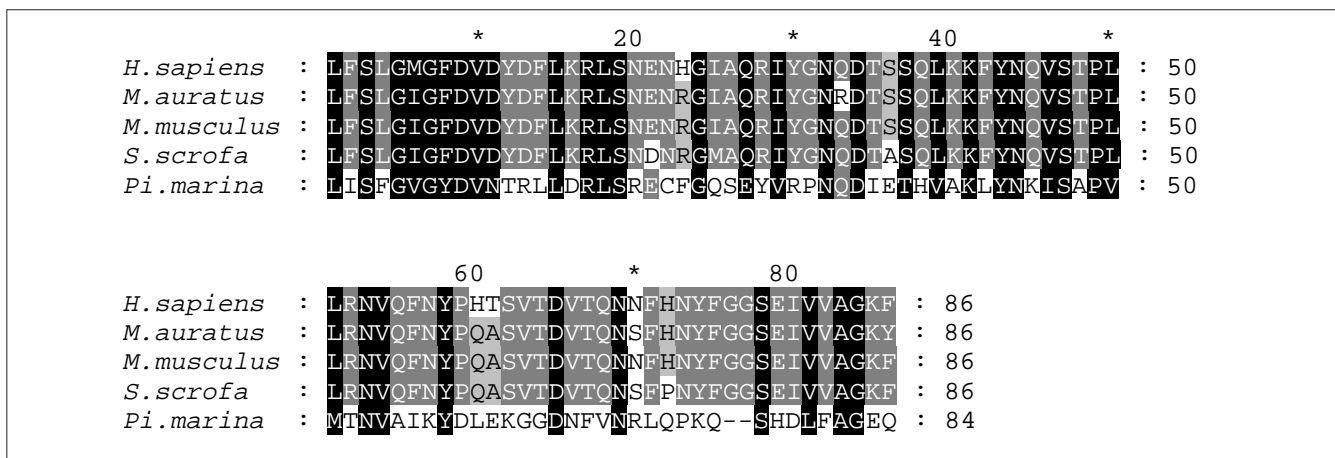
amino-acid code), was found in the *G. obscuriglobus* homolog as GLSVAIGDVNGDGAGDIVV (Figure 1). This integrin alpha-chain motif is part of one of the putative Ca<sup>2+</sup>-binding regions of the alpha subunit of the vitronectin receptor and other integrins [72].

Genes homologous with integrins (beta4 and alpha6) have also been detected in the genome of the cyanobacterium *Synechocystis* sp. PCC6803 [73] and are assumed to be the result of horizontal transfer from eukaryotes. Homologs of the inter-alpha-trypsin inhibitor heavy chain were also detected by us via the ERGO genome sequence database [74] in the genomes of the cyanobacteria *Nostoc punctiforme* (sequence ID RNPU05803) and *Anabaena* sp. (sequence ID RAN03682). Integrin and inter-alpha-trypsin inhibitor homologs thus appear to be present in two separate divisions of the Bacteria, the planctomycetes and the cyanobacteria. This may suggest a lateral gene transfer from eukaryotes to bacteria that occurred before the separation of these two divisions, or retention of such genes only in Eukarya and these two bacterial divisions accompanied by gene loss in others.

Studies of bacterial genome sequences routinely reveal genes homologous to those from other domains, for example, the relatively large proportion of archaeal and eukaryal homologs in the genomes of *Thermotoga maritima* [75] and



**Figure 1**  
Alignments for clone M3G149 from *Gemmata obscuriglobus* with integrin alpha-V sequences from the eukaryotes *Homo sapiens* and *Mus musculus*. Reference sequences are from SWISS-PROT database (P06756 for *H. sapiens*, P43406 for *M. musculus*). Shading indicates level of conservation of amino acids. Black, 100% similarity; dark gray, ≥ 80% similarity; light gray, ≥ 60% similarity; white, < 60% similarity. Asterisks indicate every 10th amino-acid position where this is not already indicated by a number.

**Figure 2**

Alignment for clone H6 of *Pirellula marina* with inter-alpha-trypsin inhibitor H2 precursor from the eukaryotes *H. sapiens*, *Mesocricetus auratus* (Syrian hamster), *M. musculus* (mouse) and *Sus scrofa* (pig). Reference sequences are from the SWISS-PROT database (P19823 for *H. sapiens*, P97279 for *M. auratus*, Q61703 for *M. musculus*, and O02668 for *S. scrofa*). Shading indicates level of conservation of amino acids. Black, 100% similarity; dark gray,  $\geq 80\%$  similarity; light gray,  $\geq 60\%$  similarity; white,  $< 60\%$  similarity. Asterisks indicate every 10th amino-acid position where this is not already indicated by a number.

*Chlamydia* species [76], respectively. Horizontal gene transfer has been a favored explanation for occurrence of such inter-domain homologs [77], and may also be invoked to explain the occurrence of eukaryal homologs in planctomycetes. This hypothesis could be tested by consideration of gene organization, base composition and codon usage, which may be deduced only with a more extensive dataset derived from a planctomycete genome sequence.

## Conclusions

In summary, the random-sequence-tag approach is a valuable and economic means of generating genomic sequence information for members of the distinctive planctomycete division of the Bacteria, and such information is largely consistent with known phenotypic properties of this group. Gene homologs identified were predominantly similar to genes of Bacteria, but some significant best matches to genes from Eukarya suggest lateral gene transfer events between domains may have involved this division at some time during its evolution. The construction of genomic DNA libraries from *G. obscuriglobus* and *Pi. marina* and the sequence data presented here provide a basis for further studies on the planctomycetes, a distinctive and widely distributed bacterial division that is potentially of great value in understanding the diversity and evolution of the domain Bacteria by comparative microbial genomics.

## Materials and methods

### Bacterial strains and culture

*G. obscuriglobus* (Australian Collection of Microorganisms (ACM) 2246) was cultured on M1 agar plates [39] at 28°C.

*Pi. marina* (ACM 3344) was cultured on half-strength marine agar [19] at 28°C.

### DNA extraction

Genomic DNA was extracted from *G. obscuriglobus* using a modification of the DNAzol technique (Gibco BRL). *G. obscuriglobus* cells were harvested from plate cultures and suspended in 10 ml TE buffer. The cells were pelleted by centrifugation, resuspended in 10 ml DNAzol and incubated at 65°C for 30 min. Cell debris was pelleted by centrifugation at 14,000g for 15 min. An equal volume of absolute ethanol was added to the supernatant and DNA precipitated overnight at -20°C. Precipitated DNA was recovered by centrifugation at 14,000g for 10 min at 4°C. The DNA pellet was washed with 70% ethanol and air-dried. DNA was resuspended in TE buffer containing 20 µg/ml RNase A.

Genomic DNA was extracted from *Pi. marina* as follows. Growth from half-strength marine agar plates was harvested and suspended in 10 ml STE buffer (0.75 M sucrose, 50 mM Tris-HCl pH 8.3, 40 mM EDTA). The cells were treated with 2% SDS for 1 h at 42°C followed by 50 µg/ml proteinase K at 55°C for 30 min. An equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) was added to the lysate and the sample centrifuged at 7,000g for 45 min. DNA contained in the aqueous phase was precipitated with 0.1 volumes of 3 M sodium acetate and 2 vols absolute ethanol. DNA was spooled onto a glass rod, washed with 70% ethanol, dried and resuspended in TE buffer for storage at -20°C.

### Clone library preparation

Plasmid (*G. obscuriglobus*) and lambda phage (*Pi. marina*) clone libraries were prepared from restricted



genomic DNA. *G. obscuriglobus* DNA was digested with *Sau3AI* to give fragments of 400 base-pairs (bp) to 3 kilobases (kb), and ligated into the *Bam*HI site of the pBlue-script II SK(-) phagemid. The ligated DNA was electroporated into *E. coli* XL1-Blue. Recombinant *E. coli* were selected on LB agar supplemented with ampicillin. Plasmid DNA was isolated using a High Pure plasmid isolation kit (Boehringer Mannheim).

Genomic DNA from *Pi. marina* was restricted with the enzyme *Bcl*I to give DNA fragments in the range of 2-12 kb. The restricted DNA fragments were ligated into the *Bam*HI arms of the lambda Zap Express vector (Stratagene). The vector DNA, containing inserts, was packaged into lambda phage heads using Gigapack III Gold packaging extract (Stratagene). *E. coli* cells (strain XL1-Blue MRF<sup>+</sup>) were infected with the packaged lambda phage according to the manufacturer's instructions. Recombinant phage were blue-white screened and titered on LB agar plates overlaid with LB top agarose containing isopropylthiogalactose and X-gal. The phage library was amplified according to the manufacturer's instructions. Plaques of recombinant phage were picked randomly and stored in 96-well microtiter plates. The vector inserts were amplified by PCR according to the protocol of the Filarial Genome Network [78].

### Cycle sequencing

Plasmids and PCR products were sequenced from the respective *G. obscuriglobus* and *Pi. marina* clone libraries using the BigDye dideoxy terminator sequencing mix (Applied Biosystems) and T3 or T7 primers. A total of 160 *G. obscuriglobus* clones and 91 *Pi. marina* clones were sequenced. Sequencing products were electrophoresed and visualized using an ABI 377 automated sequencer operated by the Australian Genome Research Facility.

### Sequence analysis

The resulting single-pass sequences were edited and analyzed to identify the represented genes using BLASTX or BLASTN algorithms for sequence comparison with databases available within the National Center for Biotechnological Information (NCBI) website [79]. Selected sequence tags were analyzed with the position-specific-iterative search algorithm PSI-BLAST (NCBI) and aligned with homologous sequences using PILEUP within Bionavigator (BioNavigator.com provided by Entigen Corporation) within the Australian National Genomic Information Service (ANGIS). GeneDoc [80] was used to edit final alignments and produce residue shading. Nucleotide accession numbers for nucleotide sequences from the clones are given in Tables 1 and 2.

### Acknowledgements

We thank the Australian Research Council for funding research on planctomycetes in J.A.F.'s laboratory, and the Australian Postgraduate Award scheme for support for C.J.

### References

- Liesack W, Söller R, Stewart T, Haas H, Giovannoni S, Stackebrandt E: **The influence of tachytelically (rapidly) evolving sequences on the topology of phylogenetic trees - intrafamily relationships and the phylogenetic position of Planctomycetaceae as revealed by comparative analysis of 16S ribosomal RNA sequences.** *Syst Appl Microbiol* 1992, **15**:357-362.
- Schlesner H, Stackebrandt E: **Assignment of the genera Planctomyces and Pirella to a new family Planctomycetaceae fam. nov. and description of the order Planctomycetales ord. nov.** *Syst Appl Microbiol* 1986, **8**:174-176.
- Woese CR, Stackebrandt E, Macke TJ, Fox GE: **A phylogenetic definition of the major eubacterial taxa.** *Syst Appl Microbiol* 1985, **6**:143-151.
- Embley TM, Hirt RP, Williams DM: **Biodiversity at the molecular level: the domains, kingdoms and phyla of life.** *Phil Trans R Soc Lond B* 1994, **345**:21-33.
- Fuerst JA: **The planctomycetes: emerging models for microbial ecology, evolution and cell biology.** *Microbiology* 1995, **141**:1493-1506.
- Hugenholtz P, Pitulle C, Herschberger KL, Pace NR: **Novel division level bacterial diversity in a Yellowstone hot spring.** *J Bacteriol* 1998, **180**:366-376.
- Neef A, Amann R, Schlesner H, Schleifer K-H: **Monitoring a widespread bacterial group: in situ detection of planctomycetes with 16S rRNA-targeted probes.** *Microbiology* 1998, **144**:3257-3266.
- Strous M, Fuerst JA, Kramer EHM, Logemann S, Muyzer G, van de pas-Schoonen KT, Webb R, Kuenen JG, Jetten MSM: **Missing lithotroph identified as new planctomycete.** *Nature* 1999, **400**:446-449.
- Schmid M, Twachtman U, Klein M, Strous M, Juretschko S, Jetten M, Metzger JW, Schleifer K-H, Wagner M: **Molecular evidence for genus level diversity of bacteria capable of catalyzing anaerobic ammonium oxidation.** *Syst Appl Microbiol* 2000, **23**:93-106.
- Gray JP, Herwig RP: **Phylogenetic analysis of the bacterial communities in marine sediments.** *Appl Environ Microbiol* 1996, **62**:4049-4059.
- DeLong EF, Franks DG, Alldredge AL: **Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages.** *Limnol Oceanogr* 1993, **38**:924-934.
- Vergin KL, Urbach E, Stein JL, DeLong EF, Lanoil BD, Giovannoni SJ: **Screening of a fosmid library of marine environmental genomic DNA fragments reveals four clones related to members of the order Planctomycetales.** *Appl Environ Microbiol* 1998, **64**:3075-3078.
- Liesack W, Stackebrandt E: **Occurrence of novel groups of the domain Bacteria as revealed by analysis of genetic material isolated from an Australian terrestrial environment.** *J Bacteriol* 1992, **174**:5072-5078.
- Borneman J, Triplett EW: **Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation.** *Appl Environ Microbiol* 1997, **63**:2647-2653.
- Derakshani M, Lukow T, Liesack W: **Novel bacterial lineages at the (sub)division level as detected by signature nucleotide-targeted recovery of 16S rRNA genes from bulk soil and rice roots of flooded rice microcosms.** *Appl Environ Microbiol* 2001, **67**:623-631.
- Zarda B, Hahn D, Chatzinotas A, Schönhuber W, Neef A, Amann RI, Zeyer J: **Analysis of bacterial community structure in bulk soil by in situ hybridization.** *Arch Microbiol* 1997, **168**:185-192.
- Lindsay MR, Webb RI, Strous M, Jetten MS, Butler MK, Forde RJ, Fuerst JA: **Cell compartmentalisation in planctomycetes: novel types of structural organisation for the bacterial cell.** *Arch Microbiol* 2001, **175**:413-429.
- Fuerst JA, Webb RI: **Membrane-bounded nucleoid in the eubacterium Gemmata obscuriglobus.** *Proc Natl Acad Sci USA* 1991, **88**:8184-8188.
- Lindsay MR, Webb RI, Fuerst JA: **Pirellosomes: a new type of membrane-bounded cell compartment in planctomycete bacteria of the genus Pirellula.** *Microbiology* 1997, **143**:739-748.
- Stackebrandt E, Wehmeyer U, Liesack W: **16S ribosomal RNA- and cell wall analysis of Gemmata obscuriglobus, a new member of the order Planctomycetales.** *FEMS Microbiol Lett* 1986, **37**:289-292.

21. Ward NL, Rainey FA, Hedlund BP, Staley JT, Ludwig W, Stackebrandt E: **Comparative phylogenetic analyses of members of the order Planctomycetales and the division Verrucomicrobia: 23S rRNA gene sequence analysis supports the 16S rRNA gene sequence-derived phylogeny.** *Int J Syst Evol Microbiol* 2000, **50**:1965-1972.
22. Rönner S, Liesack W, Wolters J, Stackebrandt E: **Cloning and sequencing of a large fragment of the atpD-gene of *Pirellula marina* - a contribution to the phylogeny of Planctomycetales.** *Endocytobios Cell Res* 1991, **7**:219-229.
23. Ward-Rainey N, Rainey FA, Stackebrandt E: **The presence of a dnaK (HSP70) multigene family in members of the orders Planctomycetales and Verrucomicrobiales.** *J Bacteriol* 1997, **179**:6360-6366.
24. Leary BA, Ward-Rainey N, Hoover TM: **Cloning and characterization of *Planctomyces limnophilus* rpoN: complementation of a *Salmonella typhimurium* rpoN mutant strain.** *Gene* 1998, **221**:151-157.
25. Weisburg WG, Hatch TP, Woese CR: **Eubacterial origin of chlamydiae.** *J Bacteriol* 1986, **167**:570-574.
26. Jenkins C, Fuerst JA: **Phylogenetic analysis of evolutionary relationships of the planctomycete division of the domain Bacteria based on amino acid sequences of elongation factor-Tu.** *J Mol Evol* 2001, **52**:405-418.
27. **The Red Environment Genomics (REGX) project** [<http://www.regx.de>]
28. **GOLD: genomes online database homepage** [<http://wit.IntegratedGenomics.com/GOLD/gold.html>]
29. Choi IG, Kim SS, Ryu J, Han YS, Bang W, Kim S, Yu YG: **Random sequence analysis of genomic DNA of a hyperthermophilic *Aquifex pyrophilus*.** *Extremophiles* 1997, **1**:125-134.
30. Fitz-Gibbon S, Choi AJ, Miller JH, Stetter KO, Simon MI, Swanson R, Kim UJ: **A fosmid-based genomic map and identification of 474 genes of the hyperthermophilic archaeon *Pyrobaculum aerophilum*.** *Extremophiles* 1997, **1**:36-51.
31. Kim CW, Markiewicz P, Lee JJ, Schierle CF, Miller JH: **Studies of the hyperthermophile *Thermotoga maritima* by random sequencing of cDNA and genomic libraries. Identification and sequencing of the trpEG(D) operon.** *J Mol Biol* 1993, **231**:960-981.
32. Peterson SN, Hu P, Bott KF, Hutchison III CA: **A survey of the *Mycoplasma genitalium* genome by using random sequencing.** *J Bacteriol* 1993, **175**:7918-7930.
33. Koonin EV, Mushegian AR, Rudd KE: **Sequencing and analysis of bacterial genomes.** *Curr Biol* 1996, **6**:404-416.
34. Guillouet S, Rodal AA, An G, Lessard PA, Sinskey AJ: **Expression of the *Escherichia coli* catabolic threonine dehydratase in *Corynebacterium glutamicum* and its effect on isoleucine production.** *Appl Env Microbiol* 1999, **65**:3100-3107.
35. Vanoni MA, Curti B: **Glutamate synthase: a complex iron-sulfur flavoprotein.** *Cell Mol Life Sci* 1999, **55**:617-638.
36. Fazio TG, Roth JR: **Evidence that the CysG protein catalyzes the first reaction specific to B12 synthesis in *Salmonella typhimurium*, insertion of cobalt.** *J Bacteriol* 1996, **178**:6952-6959.
37. Grolle S, Bringer-Meyer S, Sahn H: **Isolation of the dxr gene of *Zymomonas mobilis* and characterization of the 1-deoxy-D-xylulose 5-phosphate reductoisomerase.** *FEMS Microbiol Lett* 2000, **191**:131-137.
38. Bauld J, Staley JT: ***Planctomyces maris* sp. nov.: a marine isolate of the Planctomyces-Blastocaulis group of budding bacteria.** *J Gen Microbiol.* 1976, **97**:45-55.
39. Schlesner H: **The development of media suitable for microorganisms morphologically resembling *Planctomyces* spp., *Pirellula* spp., and other Planctomycetales from various aquatic habitats using dilute media.** *Syst Appl Microbiol* 1994, **17**:135-145.
40. Schlesner H: ***Pirella marina* sp. nov., a budding, peptidoglycan-less bacterium from brackish water.** *Syst Appl Microbiol* 1986, **8**:177-180.
41. Staley JT, Fuerst JA, Giovannoni S, Schlesner H: **The order Planctomycetales and the genera *Planctomyces*, *Pirellula*, *Gemmata* and *Isosphaera*.** In *The Prokaryotes: a Handbook on the Biology of Bacteria: Ecophysiology, Isolation, Identification, Applications*. Edited by Balows A, Trüper H, Dworkin M, Harder W, Schleifer K-H. Vol IV, 2nd edn. New York: Springer-Verlag, 1992: 3710-3731.
42. Koike-Takeshita A, Koyama T, Ogura K: **Identification of a novel gene cluster participating in menaquinone (vitamin K2) biosynthesis. Cloning and sequence determination of the 2-heptaprenyl-1,4-naphthoquinone methyltransferase gene of *Bacillus stearothersophilus*.** *J Biol Chem* 1997, **272**:12380-12382.
43. Sittig M, Schlesner H: **Chemotaxonomic investigation of various prosthecae and/or budding bacteria.** *Syst Appl Microbiol* 1993, **16**:92-103.
44. Teplyakov A, Obmolova G, Badet-Denisot MA, Badet B: **The mechanism of sugar phosphate isomerization by glucosamine-6-phosphate synthase.** *Prot Sci* 1999, **8**:596-602.
45. Liesack W, König H, Schlesner H, Hirsch P: **Chemical composition of the peptidoglycan-free cell envelopes of budding bacteria of the *Pirella/Planctomyces* group.** *Arch Microbiol* 1986, **145**:361-366.
46. König H, Schlesner H, Hirsch P: **Cell wall studies on budding bacteria of the *Planctomyces/Pasteuria* group and on a *Prosthecomicrobium* sp.** *Arch Microbiol* 1984, **138**:200-205.
47. Senecoff JF, Meagher RB: **Isolating the *Arabidopsis thaliana* genes for de novo purine synthesis by suppression of *Escherichia coli* mutants I. 5'-phosphoribosyl-5-aminoimidazole synthetase.** *Plant Physiol* 1993, **102**:387-399.
48. Nara T, Hashimoto T, Aoki T: **Evolutionary implications of the mosaic pyrimidine-biosynthetic pathway in eukaryotes.** *Gene* 2000, **257**:209-222.
49. Spencer RH, Chang G, Dees DC: **"Feeling the pressure": structural insights into a gated mechanosensitive channel.** *Curr Opin Struct Biol* 1999, **9**:448-454.
50. Guo D, Bowden MG, Pershad R, Kaplan HB: **The *Myxococcus xanthus* rfbABC operon encodes an ATP-binding cassette transporter for O-antigen biosynthesis and multicellular development.** *J Bacteriol* 1996, **178**:1631-1639.
51. Chu S, Noonan B, Cavaignac S, Trust TJ: **Endogenous mutagenesis by an insertion sequence element identifies *Aeromonas salmonicida* AbcA as an ATP-binding cassette transport protein required for biogenesis of smooth lipopolysaccharide.** *Proc Natl Acad Sci USA* 1995, **92**:5754-5758.
52. Kerger BD, Manusco CA, Nichols PD, Whitte DC, Langworthy T, Sittig M, Schlesner H, Hirsch P: **The budding bacteria, *Pirellula* and *Planctomyces*, with atypical 16S rRNA and absence of peptidoglycan, show eubacterial phospholipids and uniquely high proportions of long chain beta-hydroxy fatty acids in the lipopolysaccharide lipid A.** *Arch Microbiol* 1988, **149**:255-260.
53. Giovannoni SJ, Godchaux III W, Schabtach E, Castenholtz RW: **Cell wall and lipid composition of *Isosphaera pallida*, a budding eubacterium from hot springs.** *J Bacteriol* 1987, **169**:2702-2707.
54. Minamino T, Macnab RM: **Components of the *Salmonella* flagellar export apparatus and classification of export substrates.** *J Bacteriol* 1999, **181**:1388-1394.
55. Kirby JR, Niewold TB, Maloy S, Ordal GW: **CheB is required for behavioural responses to negative stimuli during chemotaxis in *Bacillus subtilis*.** *Mol Microbiol* 2000, **35**:44-57.
56. Rosario MM, Kirby JR, Bochar DA, Ordal GW: **Chemotactic methylation and behavior in *Bacillus subtilis*: role of two unique proteins, CheC and CheD.** *Biochemistry (Moscow)* 1995, **34**:3823-3831.
57. Chang P, Marians KJ: **Identification of a region of *Escherichia coli* DnaB required for functional interaction with DnaG at the replication fork.** *J Biol Chem* 2000, **275**:26187-26195.
58. Stewart E, Chapman CR, Al-Khodairy F, Carr AM, Enoch T: ***rqh1+*, a fission yeast gene related to the Bloom's and Werner's syndrome genes, is required for reversible S phase arrest.** *EMBO J* 1997, **16**:2682-2692.
59. Traxler BA, Minkley EG: **Evidence that DNA helicase I and oriT site-specific nicking are both functions of the F Tra I protein.** *J Mol Biol* 1988, **204**:205-209.
60. Ward-Rainey N, Rainey FA, Wellington EMH, Stackebrandt E: **Physical map of the genome of *Planctomyces limnophilus*, a representative of the phylogenetically distinct planctomycete lineage.** *J Bacteriol* 1996, **178**:1908-1913.
61. Dahlberg C, Bergström M, Andreassen M, Christensen BB, Molin S, Hermansson M: **Interspecies bacterial conjugation by plasmids from marine environments visualized by gfp expression.** *Mol Biol Evol* 1998, **15**:385-390.
62. Moolenaar GF, Moorman C, Goosen N: **Role of the *Escherichia coli* nucleotide excision repair proteins in DNA replication.** *J Bacteriol.* 2000, **182**:5706-5714.

63. Derbyshire V, Grindley ND, Joyce CM: **The 3'-5' exonuclease of DNA polymerase I of *Escherichia coli*: contribution of each amino acid of the active site to the reaction.** *EMBO J* 1991, **10**:17-24.
64. Tvermyr M, Kristiansen BE, Kristensen T: **Cloning, sequence analysis and expression in *E. coli* of the DNA polymerase I gene from *Chloroflexus aurantiacus*, a green nonsulfur eubacterium.** *Genet Anal* 1998, **14**:75-83.
65. Chenail A, Soliganc M, Bernard M: **Evolution of the large-subunit ribosomal RNA binding site for protein L23/25.** *Mol Biol Evol* 1997, **14**:578-588.
66. Allen T, Shen P, Samsel L, Liu R, Lindahl L, Zengel JM: **Phylogenetic analysis of L4-mediated autogenous control of the S10 ribosomal protein operon.** *J Bacteriol* 1999, **181**:6124-6132.
67. Bocchetta M, Gribaldo S, Sanangelantoni A, Cammarano P: **Phylogenetic depth of the bacterial genera *Aquifex* and *Thermotoga* inferred from analysis of ribosomal protein, elongation factor and RNA polymerase subunit sequences.** *J Mol Evol* 2000, **50**:366-380.
68. Woese CR, Olsen GJ, Ibba M, Söll D: **Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process.** *Microbiol Mol Biol Rev* 2000, **64**:202-236.
69. Gupta R: **The phylogeny of proteobacteria: relationships to other eubacterial phyla and eukaryotes.** *FEMS Micro Rev* 2000, **24**:367-402.
70. Miyamoto S, Teramoto H, Coso OA, Gutkind JS, Burbelo PD, Akiyama SK, Yamada KM: **Integrin function: molecular hierarchies of cytoskeletal and signalling molecules.** *J Cell Biol* 1995, **131**:791-805.
71. Bost F, Diarra-Mehrpour M, Matin J-P: **Inter- $\alpha$ -trypsin inhibitor proteoglycan family. A group of proteins binding and stabilizing the extracellular matrix.** *Eur J Biochem* 1998, **252**:339-346.
72. Fitzgerald LA, Ponez M, Steiner B, Rall SC, Bennett JS, Phillips DR: **Comparison of cDNA-derived protein sequences of the human fibronectin and vitronectin receptor  $\alpha$ -subunits and platelet glycoprotein IIb.** *Biochemistry* 1987, **26**:8158-8165.
73. May AP, Ponting CP: **Integrin  $\alpha$ - and  $\beta$ 4 subunit-domain homologues in cyanobacterial proteins.** *Trends Biochem Sci* 1999, **24**:12-13.
74. **Integrated Genomics** [<http://www.integratedgenomics.com>]
75. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA, et al.: **Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*.** *Nature* 1999, **399**:323-329.
76. Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, Aravind L, Mitchell W, Olinger L, Tatusov RL, Zhao Q, et al.: **Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*.** *Science* 1998, **282**:754-759.
77. Doolittle WF: **Lateral genomics.** *Trends Biochem Sci* 1999, **24**:M5-M8.
78. **Filarial Genome Network** [<http://circuit.neb.com/fgn/methods/phage.html>]
79. **NCBI BLAST** [<http://www.ncbi.nlm.nih.gov/BLAST/>]
80. Nicholas KB, Nicholas HB Jr, Deerfield DW II: **GeneDoc: analysis and visualization of genetic variation.** *EMB News* 1997, **4**:14.