J Mol Evol (2004) 58:19–39
DOI: 10.1007/s00239-003-2523-5

JOURNAL OF MOLECULAR EVOLUTION

Molecular Evolution of FtsZ Protein Sequences Encoded Within the Genomes of Archaea, Bacteria, and Eukaryota

Sue Vaughan,[1] Bill Wickstead,[1] Keith Gull,[1] Stephen G. Addinall[2]

[1] Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK
[2] University of Manchester, School of Biological Sciences, Oxford Road, Manchester M13 9PT, UK

Received: 17 February 2003 / Accepted: 24 July 2003

Abstract. The FtsZ protein is a polymer-forming GTPase which drives bacterial cell division and is structurally and functionally related to eukaryotic tubulins. We have searched for FtsZ-related sequences in all freely accessible databases, then used strict criteria based on the tertiary structure of FtsZ and its well-characterized *in vitro* and *in vivo* properties to determine which sequences represent genuine homologues of FtsZ. We have identified 225 full-length FtsZ homologues, which we have used to document, phylum by phylum, the primary sequence characteristics of FtsZ homologues from the Bacteria, Archaea, and Eukaryota. We provide evidence for at least five independent *ftsZ* gene-duplication events in the bacterial kingdom and suggest the existence of three ancestoral euryarchaeal FtsZ paralogues. In addition, we identify "FtsZ-like" sequences from Bacteria and Archaea that, while showing significant sequence similarity to FtsZs, are unlikely to bind and hydrolyze GTP.

Key words: FtsZ — Tubulin — Division — Septation — Cytokinesis — GTPase — FtsA

Introduction

In 1980, the *ftsZ* gene was first identified as being involved in division of *Escherichia coli* cells (Lutkenhaus et al. 1980). Since then it has been demonstrated that FtsZ is critical for the process of cell division in *E. coli* and *Bacillus subtilis* (Lutkenhaus and Addinall 1997). Identification of FtsZ from a species of Archaea and demonstration that this protein was involved in cell division, indicated that FtsZ might have a universal role in prokaryotic cell division (Wang and Lutkenhaus 1996). Data from genome sequencing projects have predominantly supported this view. Genes encoding proteins with significant sequence identity to FtsZ have been found in the genomes of virtually all species of bacteria and all euryarchaeota for which sequence data are available. Amongst bacteria, only members of the Chlamydiaceae family and *Ureaplasma urealyticum* lack FtsZ homologues (Brown and Rockey 2000; Glass et al. 2000; Read et al. 2000). While the former are obligate intracellular bacteria (Schachter and Wyrick 1994) and may therefore use host mechanisms for division, *U. urealyticum* remains the only bacterial species known to lack FtsZ which can grow and divide in culture.

FtsZ homologues can also be found in a number of eukaryotes. There is good evidence that these eukaryotic FtsZ proteins are required for division of either chloroplasts (Mori et al. 2001; Osteryoung and McAndrew 2001; Vitha et al. 2001) or mitochondria (Beech and Gilson 2000; Beech et al. 2000; Gilson and Beech 2001). Indeed, phylogenetic analyses show that chloroplast and mitochondrial FtsZs are most closely related to FtsZ from Cyanobacteria and α-proteobacteria respectively (Beech et al. 2000; Osteryoung and McAndrew 2001), giving elegant support to the

Correspondence to: Stephen G. Addinall; *email:* stephen.addinall@man.ac.uk

endosymbiont model for the origins of these organelles. However, the absence of FtsZ homologues from a number of completed eukaryotic genomes (Goffeau et al. 1996; Consortium 1998; Adams et al. 2000; Wood et al. 2002) illustrates that not all species use FtsZ for mitochondrial division. Instead, it appears that dynamin-related proteins have taken over this role in many eukaryotes (Erickson 2000; Margolin 2000a; Smirnova et al. 2001; Arimura and Tsutsumi 2002; Shaw and Nunnari 2002).

The function of FtsZ is most extensively characterized in the γ-proteobacterium *E. coli*. Prior to the onset of cell division, FtsZ polymerises into a ring structure at the centre of the cell, on the inner surface of the cytoplasmic membrane. This Z-ring then reduces in diameter, driving invagination of the division septum (Lutkenhaus and Addinall 1997; Rothfield et al. 1999; Margolin 2000b). During this process at least nine proteins, which are essential for cell division, are recruited to the Z-ring in a heirarchical fashion (see Chen and Beckwith [2001] and Hale and de Boer [2002] and references therein). These include early-acting proteins such as ZipA and FtsA, which interact directly with FtsZ through its C-terminal region (Ma et al. 1997; Wang et al. 1997; Liu et al. 1999; Ma and Margolin 1999; RayChaudhuri 1999; Hale et al. 2000; Yan et al. 2000; Haney et al. 2001) and are themselves required for the recruitment of later-acting proteins (Chen and Beckwith 2001; Hale and de Boer 2002). The phylogenetic distribution of genes for accessory division proteins has been described previously (Margolin 2000b). Interestingly, ZipA is restricted to a small number of γ-proteobacteria, whereas, amongst prokaryotes, FtsA is lacking only from *Actinobacteria*, *Cyanobacteria*, Mollicutes, *Chloroflexus*, and Archaea (Margolin 2000b; our unpublished data).

FtsZ is of wider interest due to its evolutionary relationship with eukaryotic tubulins, which form microtubules in eukaryotic cells. FtsZ and the monomers of α/β-tubulin show remarkable similarity in their tertiary structures, despite limited primary sequence identity (Lowe and Amos 1998; Nogales et al. 1998b). FtsZ also shows functional similarity to tubulin, being a co-operative GTPase, which polymerizes in a nucleotide-dependent manner into linear, unbranched polymers *in vitro*. The Z-ring is therefore thought to be a prokaryotic antecedent of microtubules, however, the *in vivo* structure of polymers within the Z-ring is presently unknown (Addinall and Holland 2002).

We have carried out a comprehensive search for FtsZ-related sequences in publicly available sequence databases and used robust criteria to determine which open reading frames (ORFs) encode genuine FtsZ proteins. We present a summary of the primary sequence features of the predicted translation products of 225 full-length *ftsZ* sequences, along with phylogenetic analysis within and between phylogenetic groups. We demonstrate the presence of FtsZ paralogues which have arisen on at least five independent occasions in Bacteria and that these events differ from the appearance of FtsZ paralogues in Archaea and Eukaryota. In addition, we present further characterisation of the pattern of occurrence and phylogenetic relationships of FtsZ homologues in Archaea. We examine sequence conservation at the FtsZ C-terminus and redefine a consensus for this region, which is involved in protein–protein interactions. Finally, we document 10 "FtsZ-like" sequences from Bacteria and Archaea.

Materials and Methods

Database Searches and FtsZ Homologue Assignment

Potential FtsZ ORFs were initially identified by sequence homology searches of the translated nr database at the National Centre for Biotechnology Information (NCBI) using *Escherichia coli* FtsZ amino acid sequence NP_414637. All completed genome projects submitted to the National Centre for Biotechnology Information before 10 June 2003 (NCBI: www.ncbi.nlm.nih.gov) were checked for the presence of an FtsZ homologue by similar homology searches. Data from incomplete genome projects that have been submitted to NCBI and also genome projects in progress at The Institute for Genomic Research (www.tigr.org) and The Sanger Centre (www.sanger.ac.uk) were also searched. Supplementary data containing links to Genbank entries or genome projects, as well as amino acid sequences for all FtsZ homologues and FtsZ-like ORFs are available at the authors' website, http://www.biomed2.man.ac.uk/addinall/Vaughan_supp.html.

Following the initial homology search, we then defined an ORF as an FtsZ homologue if it met the following criteria: (1) it made a match to the FtsZ PRINTS fingerprint (www.bioinf.man.ac.uk/dbbrowser/PRINTS) (Attwood et al. 2002), (2) it contained the tubulin signature motif (PROSITE motif PS00227 [S/A/G]GGTG[S/A/T]G [Bucher and Bairoch 1994] with a maximum of two mismatches, and (3) it contained conserved residues that are involved in GTP binding and GTP hydrolysis (Lowe and Amos 1998; Nogales et al. 1998a). The last requirement was tested using multiple alignments which were prepared using the clustalW algorithm in Megalign (DNAstar) and manually adjusted as applicable. The NCBI taxonomy database (Benson et al. 2000; Wheeler et al. 2000) was a helpful guide to current taxonomic classification.

Phylogenetic Analysis

Bayesian maximum likelihood (ML) trees were inferred using the Metropolis-coupled Markov chain Monte Carlo (MCMCMC) method as implemented by the program MrBayes 3.0b4 (Huelsenbeck and Ronquist 2001). Trees were constructed from protein alignments of the FtsZ core region (Figs. 5 and 6) or the FtsZ core region plus the C-terminal motif (Fig. 4). A gamma-distributed, substitution rate variation was assumed (Yang 1993), with four gamma categories and shape parameter estimated from the data and the WAG substitution matrix (Whelan and Goldman 2001) was used. MCMCMC analyses were initiated on a random starting tree and run for 50,000 (Figs. 4 and 5) or 100,000 (Fig. 6) gener-

ations, with four chains run in parallel and a ''temperature'' of 0.4. Chains were sampled every 100 generations. Only generations after the ''burn-in'' for which the MCMCMC calculation had reached stationary phase (>5000 generations) were retained and used to calculate posterior probabilities and consensus trees.

Neighbor-joining (NJ) and maximum parsimony (MP) trees were inferred from the same protein alignments used for ML trees with the program PAUP 4.0b10 (Swofford 1998). Full heuristic searches were initiated on random starting trees with the implementation of a tree bisection and reconstruction (TBR) algorithm. In the case of MP, gaps in sequences were treated as a 21st amino acid state. PAUP defaults were used for other parameters. Datasets were resampled using 1000 bootstrap replications and bootstrap values were used in support of the inferred ML tree, as indicated in the figures.

Chi-square tests showed that four FtsZ sequences included in the analyses (*Magnetospirillum magnetotacticum* Fts1c, *Thermococcus kodakaraensis* FtsZ3, *Colwellia psychrerythraea 34H* FtsZ2, and one of the *Halobacterium nrc1* FtsZ3s) had significantly biased amino acid compositions. Excluding these sequences from the relevant phylogenetic analyses had no effect on remaining topology of the inferred tree for any of the methods used (ML, NJ, or MP). Using full-length protein sequences for NJ and MP tree construction also had no significant effect on inferred topology.

C-Terminal Motif Analysis

Visual inspection of all complete FtsZs revealed 154 sequences with putative 9-amino acid C-terminal motifs, all from Bacteria or Eukaryota. These motifs were used as a training set to construct a position-specific score matrix using the motif discovery tool Multiple EM for Motif Elicitation (MEME [Bailey and Elkan 1994]). The resulting score matrix could then be used to assign scores, and hence position-specific *p*-values (Bailey and Gribskov 1997), for any specified 9-amino acid sequence. All but 1 of the 154 identified motifs had *p*-values < 0.001. The weak motif identified in *Mycoplasma pulmonis* ($p = 0.0025$) also lacked the basic/amide tail of many other species. Moreover, other *Mycoplasma* FtsZs lack C-terminal motifs. For these reasons, this motif was discarded. Contrastingly, the 10 sequences immediately below the $p = 0.001$ cutoff possessed basic/amide tails and were also found in closely related FtsZ sequences, where available.

From the 71 complete FtsZ sequences classified by visual inspection as lacking C-terminal motifs, all possible motifs (characterized as any 9-amino acid word-containing proline at position 6) were extracted and scored against the motif position-specific matrix. All but 4 sequences had *p*-values > 0.001 and were discarded. These four sequences (two bacterial, *Porphyromonas gingivalis* and *Haemophilis ducreyi*, and two archaeal, *Methanococcus kandleri* FtsZ1 and *Pyrococcus abyssi* FtsZ1) were characterized as false positives on the basis of lack of basic/amide tail and lack of conservation in closely related species. Thus, we believe that the 153 sequences classified here as possessing the C-terminal motif constitute a consistent and complete set.

Sequence Logos

Sequence logos are a graphical way to display patterns of an alignment of DNA, RNA or protein sequences (Schneider and Stephens 1990). The height of each amino acid is proportional to the frequency this amino acid residue at that position. An alignment was produced of 153 C-terminus regions of FtsZ homologues and then put into a FASTA format and pasted into the sequence logo website: www.bio.cam.ac.uk/cgi-bin/seqlogo/logo.cgi. Predictions of the presence of mitochondrial or chloroplast import sequences were carried out using TargetP (Emanuelsson et al. 2000) (www.cbs.dtu.dk/services/TargetP). Amino acid sequences were analyzed for signature motifs using the PRINTS database: http://www.bioinf.man.ac.uk/fingerPRINTScan (Attwood and Beck 1994; Attwood et al. 1994, 1998, 2002). Figure 1B was produced using Cn3D molecular structure viewer on NCBI (Hogue 1997).

Results and Discussion

Distribution of FtsZ Homologues

We carried out sequence homology searches for FtsZ-related sequences in all freely available sequence databases. Using stringent criteria (see Materials and Methods) we ascertained which ORFs encode genuine FtsZ homologues. For the sake of completeness, we also examined sequences previously annotated as either ''FtsZ'' or ''cell division GTPase,'' even if these did not immediately meet our criteria. In this way we identified 235 ORFs encoding FtsZ homologues, 225 of which were full-length. The number and distribution of FtsZ homologues are outlined in Table 1 and discussed in the following section. All sequences described in this paper (together with partial FtsZ sequences) and links to Genbank entries are available as supplementary data from the authors' website, http://www.biomed2.man.ac.uk/addinall/Vaughan_supp.html.

Any phylogeny is biased by the availability of sequence data. Hence, the majority of bacterial FtsZ sequences found in our analysis come from the Firmicutes and Proteobacteria (the phyla from which most sequence data has been obtained thus far). Only just over half of the phyla in the Bacteria superkingdom are represented by organisms which have had their complete genome sequenced. Of these, FtsZ is completely lacking from only the Chlamydiales phylum. The only other bacterial species which lacks an FtsZ homologue is *Ureaplasma urealyticum* (Glass et al. 2000). This organism is from the *Mycoplasmataceae* family, which is composed of the two genera *Mycoplasma* and *Ureaplasma* and complete FtsZ orthologues have been identified from five *Mycoplasma* organisms. It is unclear how *U. urealyticum*, which can be cultured both in liquid and on solid media (Teng et al. 1994), is able to divide without FtsZ.

The Archaea superkingdom comprises three phyla—Euryarchaeota, Crenarchaeota, and Korarchaeota. Our searches identify FtsZ homologues from all seven classes of Euryarchaea, but no homologues have been identified from Crenarchaea (three completed genome projects [Kawarabayasi et al. 2001; She et al. 2001; Fitz-Gibbon et al. 2002]). No organisms from the Korarchaeota have been cultured to date.

In eukaryotes, FtsZ homologues have been identified in organisms from green plants and red, green, brown, and golden algae. However, completed genome projects from *Saccharomyces cerevisiae*, *Drosophila*

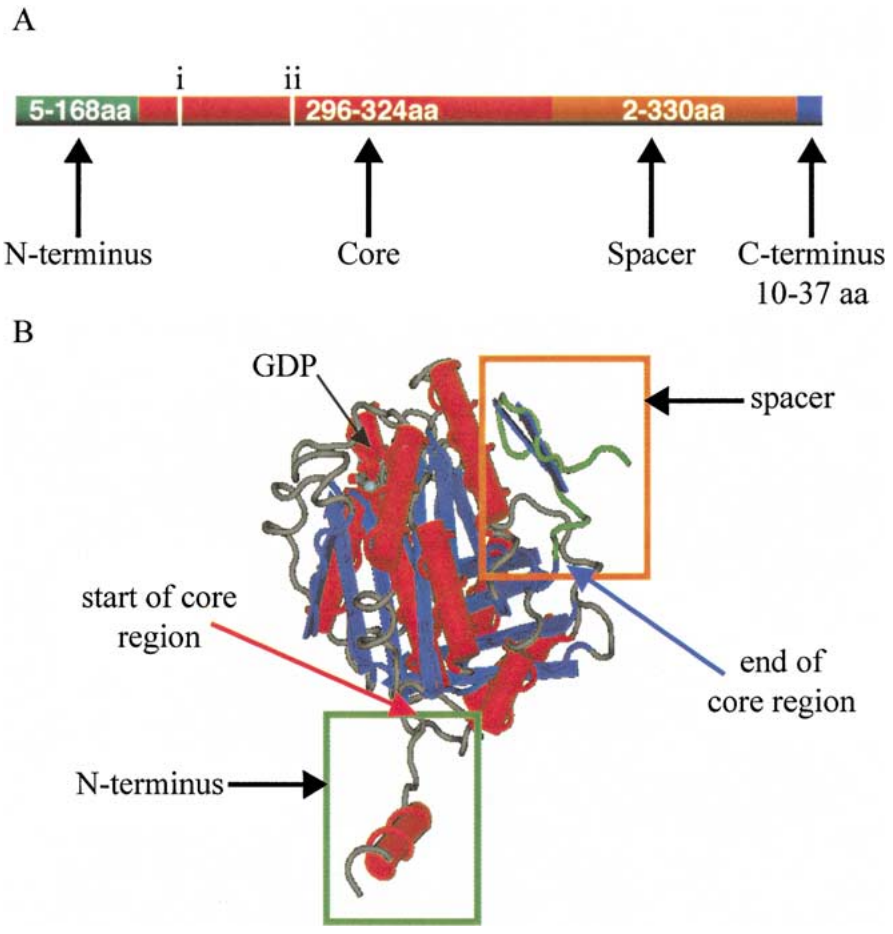


Fig. 1. **A** We define four distinct regions of FtsZ sequences: N-terminus, core, spacer, and C-terminus. Minimum and maximum lengths of each region (aa—amino acids) represent all FtsZ homologues, except *Mycoplasma pneumoniae*, *M. genitalium*, and the FtsZ3 sequences of Archaea, which have poorly conserved core regions (see text and Fig. 3). The position of insertions in the core regions of (i) *Wolbachia pipientis* and some Euryarchaeota FtsZ2s and (ii) Pasteurellales FtsZs, are indicated by white lines. **B** Tertiary structure of FtsZ1 from the Archaea *Methanococcus jannaschii* (Lowe and Amos 1998) illustrating the position of regions. The N-terminus (boxed in green), the start and end of the core region (red and blue arrows, respectively), and the spacer (boxed in orange) are highlighted. FtsZ homologues from Archaea organisms do not have the C-terminus region, which is found in most bacterial and eukaryotic FtsZ homologues (see text).

melanogaster, *Caenorhabditis elegans*, and *Schizosaccharomyces pombe* (Goffeau et al. 1996; Consortium 1998; Adams et al. 2000; Wood et al. 2002) reveal that these organisms lack FtsZ homologues. The mycetozoan, *Dictyostellium discoideum*, is the only non-chloroplast-containing eukaryotic organism found so far to possess FtsZ homologues (see later).

General Features of FtsZ Sequences

Here we present a general outline of FtsZ primary sequence structure, followed in the next section by a more detailed examination of sequence features found within taxonomic divisions.

With the exception of the FtsZ3 sequences from Archaea (see later) and FtsZ orthologues from two *Mycoplasma* species (*M. pneumoniae* and *M. genitalium*), we can describe the amino acid sequence of all FtsZ homologues using four distinct regions. These are a variable N-terminus, the highly conserved core region, a variable spacer region, and the C-terminus region (Fig. 1A). These regions are defined individually below and are illustrated on the FtsZ tertiary structure (Fig. 1B) and in a multiple alignment of representative FtsZ sequences (Fig. 2).

N-Terminus. Secondary structure alignments between α/β-tubulin and *Methanococcus jannaschii* FtsZ1 (Nogales et al. 1998a) reveal an N-terminal extension proximal to the core domain in FtsZ, whereas the Rossman fold of tubulins starts at the very N-terminus of the protein. We find that 218/225 full-length FtsZs have at least five amino acids proximal to the core domain, the only exceptions being seven FtsZ3 sequences from Archaea (see later). This represents a major difference between tubulins and FtsZs and is one of the major differences between FtsZ3 and the rest of the FtsZs identified to date. The N-terminus region of *M. jannaschii* FtsZ1 forms an α-helix and loop that projects out from the core of the molecule (Lowe and Amos 1998) (Fig. 1B). We find that the FtsZ N-terminus is not conserved in length or amino acid sequence between taxonomic divisions, but there is some similarity within taxonomic divisions, unlike the spacer region (see below). There is a lack of data relating to a possible function for this region, since characterized N-terminus deletions of *E. coli* FtsZ also removed part of the core region (Huang et al. 1996; Ma et al. 1996). The only function ascribed to the FtsZ N-terminus thus far is targeting of the protein to organelles in eukaryotic

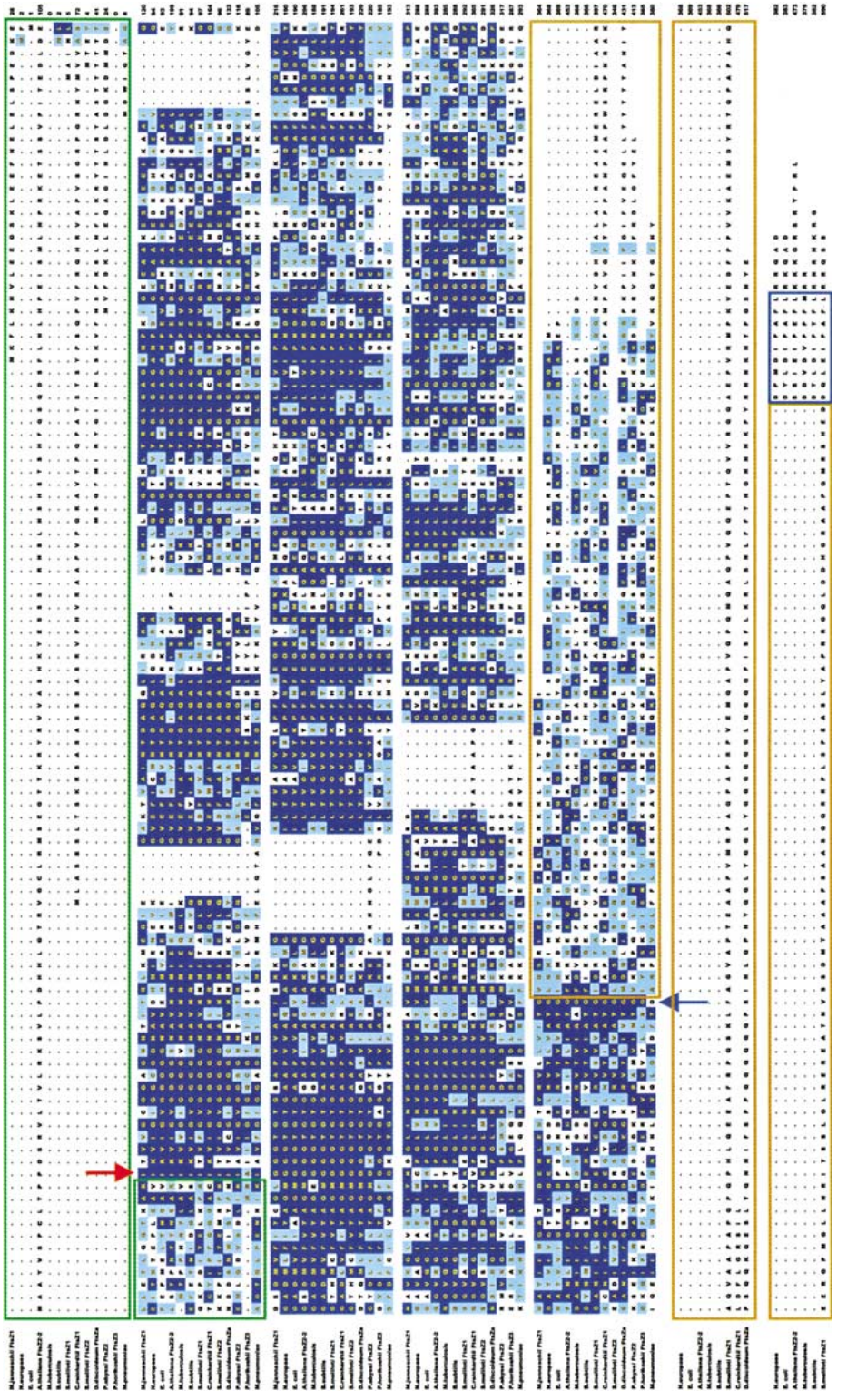


Fig. 2. Multiple sequence alignment of a selection of FtsZ homologues. Thirteen FtsZ sequences have been chosen to illustrate points covered in the text. *Nitrosomonas europeae*, *Escherichia coli*, *Mycobacterium tuberculosis*, *Bacillus subtilis* all possess a single FtsZ homologue. *Sinorhizobium meliloti* is an α-proteobacteria with two FtsZ paralogues—FtsZ2 is included here. FtsZ homologues from the Viridiplantae form two clades and one orthologue from each clade (*Arabidopsis thaliana* FtsZ2-2, *Chlamydomonas reinhardtii* FtsZ1) is shown. FtsZ homologues from the Archaea form three separate clades and one orthologue from each clade is shown (*Methanococcus jannaschii* FtsZ1, *Pyrococcus abyssi* FtsZ2, and *P. horikoshii* FtsZ3). *Mycoplasma pneumoniae* is included as a divergent FtsZ homologue, and, finally, one of the two FtsZs of the mycetozoan *Dictyostelium discoideum* (FtsZa) is shown. Dark blue shading highlights identical residues and light blue shading highlights similar residues. The four regions of FtsZ are highlighted as follows: N-terminus (boxed in green), spacer (boxed in orange), and the C-terminus motif (blue, when present). The start and end of the core region are indicated by arrows (red and blue, respectively).

organisms (Beech et al. 2000; Fujiwara and Yoshida 2001). As the targeting sequences comprise only a portion of this N-terminal extension (data not shown), it is likely that these properties are in addition to any other functions provided by this region.

Core Region. The core region (Erickson 1995) is a stretch of approximately 300 amino acids which contains all of the residues required for GTP-binding and GTP-hydrolysis (Fig. 1). We define the core region as starting at a conserved isoleucine (only 14 changes at this position in 225 sequences, with 5 of

these being conservative substitutions; Fig. 2, red arrow) which coincides with the start of the Rossman fold. We define the core region as ending at the equivalent of amino acids 335–339 (LVITG) from the *Methanococcus jannaschii* FtsZ1 sequence (Fig. 2; blue arrow). While there are too many substitutions at these positions to define a motif, only 12 FtsZ sequences lack a recognisable version of this sequence. Three of these (*Dictyostelium discoideum* FtsZb, *Fusobacterium nucleatum* FtsZ, and *Magnetospirillium magnetotacticum* FtsZ1c) can be aligned easily using neighboring amino acids. The other nine sequences (seven Archaea FtsZ3s, *M. pneumoniae*, and *M. genitalium* FtsZs) remain the only ones which diverge significantly in the core region (Fig. 3; see below).

The tertiary structure of the FtsZ core region is almost superimposable with two regions that are characterized for the α/β-tubulin tertiary structure—the GTP binding domain (containing the Rossman fold) and the "second" domain. This second domain is of interest because, although it is well conserved in α- and β-tubulin, it is less well conserved in other tubulin families. For example, δ-, ε-, and ζ-tubulin are predicted to have insertions and deletions in this region (McKean et al. 2001). In contrast, we find that insertions or deletions in the core of FtsZ homologues are very rare (individual instances will be discussed later), indicating that FtsZ proteins, like α- and β-tubulins, are relatively more structurally constrained.

The core region of all FtsZs contains a sequence related to the highly conserved tubulin signature motif (PROSITE motif PS00227: [S/A/G]GGTG[S/A/T]G [Bucher and Bairoch 1994]). We have determined a consensus sequence of this motif from the 225 full-length FtsZ homologues presented here—G[G/K/N][G/A/S][T/A/S]G[T/S/A/N][G/V]. The vast majority of FtsZ sequences, however, contain the sequence GGGTGTG (thus fitting the original PROSITE consensus). Changes at position 2 of this motif occur only in *Chlorobium tepidum*, FtsZ2 sequences from the Thermococcales order, and three *Mycoplasma* FtsZs. Substitution at position 3 of the motif occurs only in *Magnetospirillum* FtsZ1c and *Halobacterium* species FtsZ3b. Changes at position 4 are only found in *Streptococcus* species and *Corynebacterium flavum*. The conservative changes at position 6 are commonplace in both FtsZ and tubulin, however, the majority of tubulin homologues contain a serine at this position and the majority of FtsZ homologues contain a threonine at this position. Interestingly, all Archaea FtsZ3 orthologues (see later) have serine or alanine at this position. Only *Halobacterium* species FtsZ3b has a V at position 7 and is consequently the only FtsZ sequence presented here with two differences from the original PROSITE consensus.

Spacer. The spacer region (Din et al. 1998) is highly variable in length (2–330 amino acids; Figs. 1 and 2), and we define it as running from the C-terminal end of the core region to the start of the C-terminus region (see below). The spacer varies enormously in its amino acid composition between FtsZ homologues and shows little amino acid sequence conservation even within taxonomic divisions. X-Ray crystallography determined that the 25 amino acid spacer of *M. jannaschii* FtsZ1 consists of 16 amino acids forming two β-strands separated by a loop, followed by nine disordered residues (Lowe and Amos 1998) (Fig. 1B). Deletion analysis in *Caulobacter crescentus* has demonstrated that the spacer region is not required for polymerisation of FtsZ into protofilaments (Din et al. 1998).

C-Terminus. A conserved C-terminus motif has been described previously (Ma and Margolin 1999; Erickson 2001; Osteryoung and McAndrew 2001) and has been shown to be important for interactions with FtsA and ZipA (Ma et al. 1997; Wang et al. 1997; Liu et al. 1999; Ma and Margolin 1999; RayChaudhuri 1999; Hale et al. 2000; Yan et al. 2000; Haney et al. 2001). Our analysis of 225 full-length FtsZ homologues reveals that although this motif is variable, it is clearly absent from many FtsZ homologues (described in detail in a later section). When the C-terminus motif is present, we observe that it is followed by a short stretch of up to 28 amino acids rich in amide and basic residues. Consequently we have treated these two sequences together as a single C-terminus region (Figs. 1A and 2).

Bacterial FtsZ Sequences

Previous phylogenetic analyses of FtsZ sequences (Faguy and Doolittle 1998; Gilson and Beech 2001) have shown that bacterial FtsZs generally group in a way that reflects the known taxonomic relationships between species. Our analysis of all available FtsZ sequences demonstrates only a small number of exceptions to this. These paralogous sequences will be discussed individually as we describe detailed analyses of FtsZ primary sequence features both within and between taxonomic divisions. A summary of the features of full-length FtsZ sequences from each phyla is presented in Fig. 3.

Phylum: Proteobacteria. The proteobacteria phylum comprises five classes—α-, β-, γ-, δ-, and ε-proteobacteria. In α-proteobacteria, we find that 25 of the 31 FtsZs are well conserved and possess all four regions described above (Fig. 3). Amongst these, the FtsZ sequence from *Wolbachia pipientis* is unique in

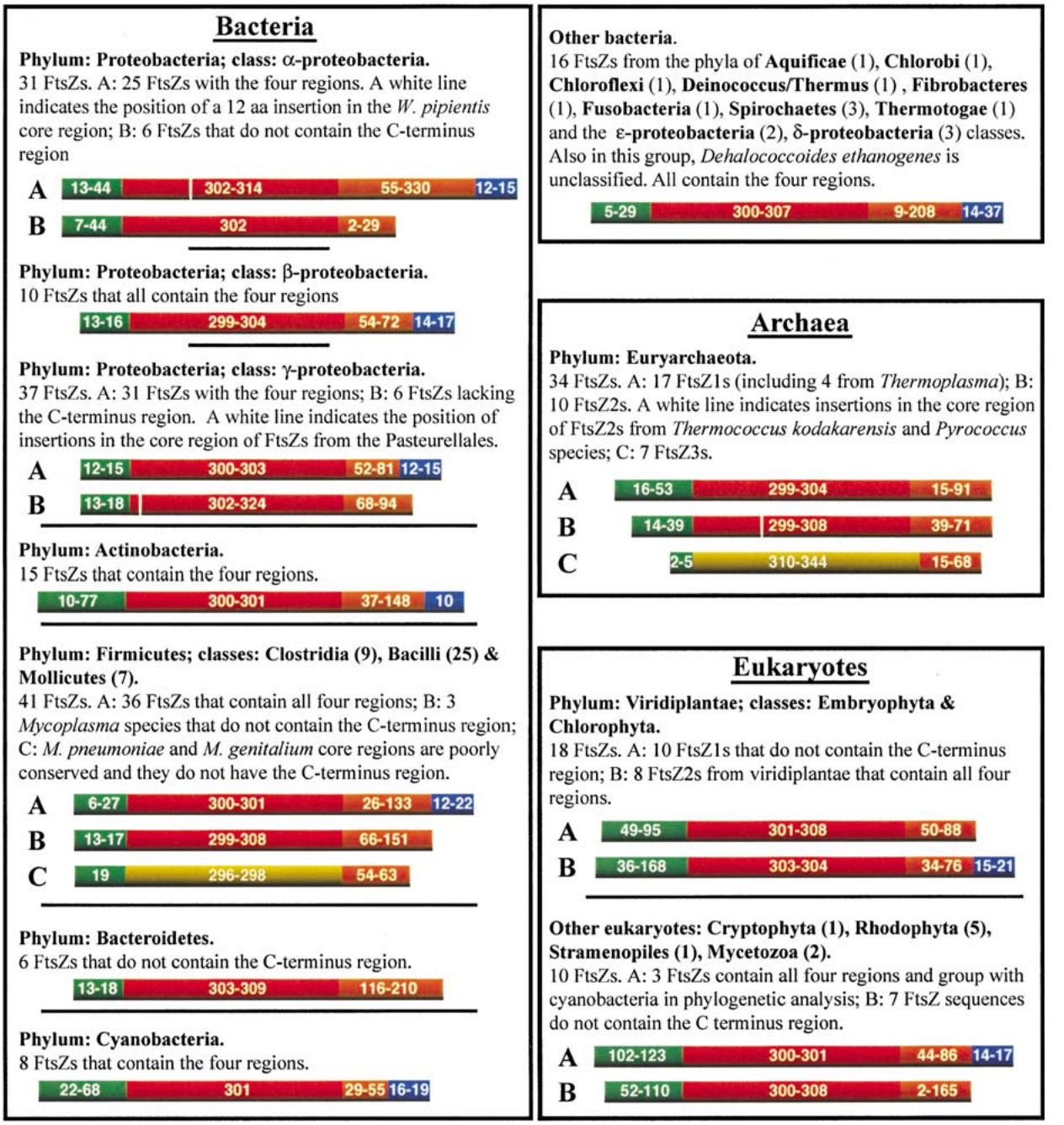


Fig. 3. Diagrammatic representation of multiple alignments of FtsZ sequences from Bacteria, Archaea, and Eukaryota. Only full-length sequences (n = 225) were used in this analysis. The four regions outlined in the text are indicated as follows: N-terminus (green), core (red), spacer (orange), and C-terminus (blue). Poorly conserved and therefore ill-defined core regions of divergent FtsZ homologues are colored yellow. Notable insertions in the core region are indicated by white lines.

having a 12-amino acid insertion in the core region, which lies in the loop between the equivalent of H3 and S5 of the *Methanococcus jannaschii* FtsZ tertiary structure. Interestingly, the six remaining FtsZ orthologues from α-proteobacteria lack the C-terminus motif and are all from organisms which possess two or more FtsZ homologues in their genomes (see later). FtsZ homologues of β-proteobacteria (10 in total) are well conserved and contain all four regions (Fig. 3).

Of the 37 FtsZ homologues of γ-proteobacteria, 31 are well conserved and possess all four regions (Fig. 3). The six remaining FtsZ sequences all lack a conserved C-terminus region. These include the *Dichelobacter nodosus* FtsZ, one of two FtsZ paralogues we have identified from the unfinished genome of *Colwellia psychrerythraea* 34H (see later), and FtsZs from species of the Pasteurellales order (*Actinobacillus actinomycetemcomitans*, *Haemophilus influenzae*, *H. ducroyi*, and *Pasteurella multocida*). The latter four

sequences are also unusual in that they all have an insertion (of 4 amino acids in *H. ducroyi* and 19 amino acids in the other three) at the same position in the core region (Fig. 3). Secondary structure alignments predict this insertion to occur between the H1 and S2 loop in the *M. jannaschii* FtsZ tertiary structure. Although there is no specific functional information about this region of FtsZ, the corresponding region of α- and β-tubulin is situated in the lumen of microtubules and is a common site for insertions and deletions in α/β-tubulin (Nogales et al, 1998a). For this reason we would predict that these insertions do not disrupt the tertiary structures of the Pasturellales FtsZ core domains, which are otherwise well conserved.

FtsZ homologues from three δ-proteobacteria and two ε-proteobacteria are well conserved and contain all four regions described above. Because of the small number of sequences from these divisions, we have treated them along with "other bacteria" for the purposes of presentation in Fig. 3.

Phylum: Actinobacteria. Actinobacteria FtsZ orthologues are well conserved and contain all four regions defined in this paper (Fig. 3).

Phylum: Firmicutes. The Mollicutes class of Firmicutes contains the two most divergent FtsZ homologues identified so far in the Bacteria superkingdom (Erickson 1995). *Mycoplasma pneumoniae* and *M. genitalium* FtsZs are 53% identical to each other but are only between 17 and 21% identical to other FtsZ homologues in this group, including three from the same genus, *M. pulmonis*, *M. penetrans*, and *M. hominis* (data not shown). Our comparisons with all other FtsZ homologues reveal very little sequence identity following the tubulin signature motif in the core region of these two sequences (see *M. pneumoniae* in Fig. 2). The motif, which demarks the C-terminal end of the core region (Fig. 2; blue arrow) is poorly conserved, making it difficult to split these sequences into the regions defined in this paper. We have found that FtsZ orthologues from bacterial species within the same genus generally show high levels of sequence identity at the amino acid level (data not shown). However, this is not the case for the seven Mollicutes for which we have data. Perhaps as a result of reductive evolution and the possession of unique internal cytoskeletal structures (Trachtenberg 1998), some Mollicutes have begun to develop modes of cell division which rely less or not at all on FtsZ.

Phylum: Bacteroidetes. The six FtsZ homologues of this phylum are well conserved, but all lack the C-terminus motif (Fig. 3).

Phylum: Cyanobacteria. The eight FtsZ homologues of this phylum are well conserved and contain all four regions (Fig. 3).

Other Bacteria. Eleven FtsZ homologues identified from the Aquificae, Chlorobi, Chloroflexi, Deinococcus/Thermus, Fibrobacteres, Fusobacteria, Spirochaetes, and Thermotogae phyla plus the unclassified *Dehalococcoides ethanogenes* contain all four regions outlined here.

In summary, 163 full-length FtsZ homologues from the Bacteria superkingdom were analyzed. Apart from the highly divergent FtsZs from *M. pneumoniae* and *M. genitalium*, the core region is highly conserved. The only other exceptions to this are large insertions interrupting the core regions of *W. pipientis*, *A. actinomycetemcomitans*, *H. influenzae*, and *P. multocida* FtsZs.

Bacteria with Multiple FtsZ Sequences

Previously in the published literature, only one bacterial species, the α-proteobacterium *Sinorhizobium meliloti*, was known to contain more than one FtsZ homologue in its genome. These were named FtsZ1 and FtsZ2 according to sequence similarity with the FtsZs from *E. coli* and *B. subtilis*. FtsZ1 is the more closely related to *E. coli* and *B. subtilis* FtsZs and was found to be essential for viability of *S. meliloti*. When overexpressed in *E. coli*, FtsZ1 from *S. meliloti* resulted in the filamentous phenotype indicative of a heterologous protein able to interfere with the endogenous FtsZ function. In contrast, overexpression of *S. meliloti* FtsZ2 resulted in filamentous cells that coiled, with some filaments bulging at one pole. FtsZ2 was found to be nonessential for viability in *S. meliloti* (Margolin et al. 1991; Margolin and Long 1994). The authors speculated that this could be because FtsZ2 may be expressed in the bacteroid, rather than the free-living form of *S. meliloti* (Margolin and Long 1994). During the searches outlined in this paper we have identified seven additional bacterial species that possess two or more FtsZs. These data are presented below and in Figs. 4 and 6.

Rhizobiales Order (*Phylum, Proteobacteria; Class, α-Proteobacteria*). We have found that five organisms within the Rhizobiales order (including *S. meliloti*; above) encode two *ftsZ*s in their completed genome. Three are from the Rhizobiaceae family—*S. meliloti*, *Agrobacterium tumefaciens*, and *Rhizobium leguminosarum* (Kaneko et al. 2000a, b; Galibert et al. 2001; Goodner et al. 2001); one is from the Bradyrhizobiaceae family—*Bradyrhizobium japonicum* (Kaneko et al. 2002); and one is from the Phyllo-

bacteriaceae family—*Mesorhizobium loti*. We present a phylogenetic tree (Fig. 4) of all full-length FtsZs from α-proteobacteria to highlight two independent gene duplication events in these three families of the Rhizobiales order. FtsZ2s from the Rhizobiaceae and Phyllobacteriaceae families form a single clade with strong support in all of the phylogenetic approaches used. The maximum likelihood analysis presented in Fig. 4 suggests that the FtsZ1s from these families are polyphyletic. However, this region of the tree is poorly supported by neighbor-joining and maximum parsimony approaches. Moreover, the monophyletic nature of the FtsZ2s suggests a gene duplication event in these taxa prior to divergence into families.

Neither of the two FtsZs of *B. japonicum* group with FtsZ2s of the Rhizobiaceae and Phyllobacteriaceae families, but both are contained within the paraphyletic group of FtsZ1s of the Rhizobiales order. These *B. japonicum* FtsZs (which we designate FtsZ1a and 1b) appear to be the products of a separate gene-duplication event to the more ancestral FtsZ1/2 split. The remainding α-proteobacterial FtsZs are referred to as FtsZ1s, since none of them group with the FtsZ2 clade.

The FtsZs of *S. meliloti*, *M. loti*, *A. tumefaciens*, *R. leguminosarum*, and *B. japonicum* all contain similar length N-termini, however, the FtsZ1s and *B. japonicum* FtsZ1a possess a very long spacer (226–269 amino acids, compared to 3–29 for FtsZ2 and 90 for *B. japonicum* FtsZ1b). In the case of *S. meliloti*, *M. loti*, *A. tumefaciens*, and *R. leguminosarum*, only FtsZ1 contains the conserved C-terminus motif. It is not these differences per se which cause the distribution into two separate clades, since phylogenetic trees constructed with only the core region of each sequence give trees with the same topology (data not shown). Both FtsZs from *B. japonicum* contain the C-terminus region, consistent with our annotation of them as FtsZ1a and 1b.

Other α-proteobacteria for which genome projects have been completed—*Rickettsia conorii*, *R. prowazekii*, and also *Brucella melitensis* (from the Rhizobiales)—have only one FtsZ homologue. Interestingly, the genera of *Rhizobium*, *Sinorhizobium*, *Bradyrhizobium*, and *Mesorhizobium* can either be free-living or invade the root nodule cells of leguminous host plants, where they differentiate into a bacteroid stage (for review see Gage and Margolin 2000). Since these genera are found in four different taxonomic families (Garrity and Holt 2001), which contain other genera that do not differentiate into a bacteroid stage, it appears that evolution of this differentiated stage is a polyphyletic trait that has occurred independently on at least two occasions. Margolin and Long (1994) suggested that FtsZ2 may be functional in the bacteroid stage in *S. meliloti* and the case for this is strengthened by the fact that the other α-proteobacteria with FtsZ2s (*M. loti*, *R. leguminosarum*, and *A. tumefaciens*) also form bacteroids. Significantly, we find that *B. japonicum*, which appears to have independently developed bacteroid differentiation, has also independently duplicated its FtsZ repertoire (Fig. 4).

Magnetospirillum magnetotacticum
(Phylum, Proteobacteria; Class, α-Proteobacteria; Order, Rhodospirillales; Family, Rhodospirillaceae)

We found four predicted ORFs with significant sequence similarity to FtsZ in the incomplete genome of the α-proteobacterium *Magnetospirillum magnetotacticum* (NCBI accession No. NZ_AAAP00000000). One of these ORFs (encoded on contig 3816) encodes a protein that has a nonconservative (D-to-N) substitution of the equivalent of D_{212} from *E. coli* FtsZ. This amino acid residue is in the synergy loop and is critical for GTP-hydrolysis, therefore, using our criteria, this ORF cannot be classed as a genuine FtsZ and we group it with other "FtsZ-like" sequences (see later). A second sequence (contig 3817) has two frameshifts, which make it impossible to reconstruct the entire ORF unambiguously. We therefore excluded this sequence from phylogenetic analysis, however, it is important to point out that this sequence contains well conserved N-terminus, core, and C-terminus motifs. For these reasons we believe the frameshifts may be sequencing errors. The remaining two sequences (contigs 3801, 3838) are genuine FtsZs that lack the C-terminus region. We therefore consider that *M. magnetotacticum* has FtsZ1a (3838), FtsZ1b (3817), and FtsZ1c (3801), named based on decreasing similarity to other α-proteobacteria FtsZ1s, and an "FtsZ-like" sequence (3816; see later). As both full-length sequences group together, this represents an *ftsZ* gene duplication independent of those in the Rhizobiales order (Fig. 4).

Colwellia psychrerythraea 34H (*Phylum, Proteobacteria; Class, γ-Proteobacteria; Family, Alteromonadaceae*). *Colwellia* species are psychrophilic members of the Alteromonadaceae family. A genome sequencing project is under way for species 34H (NC_003910), from which we have identified two predicted full-length ORFs on contig 1720 matching our criteria for FtsZ homologues. The ORF starting at 1,044,393 bp of contig 1720, which we designate FtsZ1, is closely related to other γ-proteobacterial FtsZ homologues. The second ORF (FtsZ2), starting at 3,039,251 bp, is only 30–33% identical to other FtsZs, including *C. psychrerythraea* FtsZ1; nevertheless, both FtsZ1 and FtsZ2 contain all four regions outlined above

(Fig. 3). *Shewanella oneidensis MR-1* is the closest relative of *C. psychrerythraea* for which complete genome sequence is available and has only a single FtsZ homologue (NC_004347). *C. psychrerythraea* 34H therefore is the only γ-proteobacteria currently known to possess multiple FtsZ homologues and, as such, represents an independent occurrence of multiple FtsZs in bacteria (Fig. 6). The grouping of *C. psychrerythraea* FtsZ2 in Fig. 6 makes this a candidate product of horizontal gene transfer, although a potential donor has yet to be identified.

Chloroflexus aurantiacus (*Phylum, Chloroflexi; Order, Chloroflexales; Family, Chloroflexaceae*). *Chloroflexus aurantiacus* is an anoxygenic filamentous gliding phototroph and is reported to be the earliest branching of the phototrophs (Oyaizu et al. 1987). This makes it an important organism in studies of the origin and evolution of photosynthesis. A genome project is under way (NCBI accession number NZ_AAAH00000000) and we have identified one full-length ORF (contig No. 1079; Fig. 3) and one partial ORF (contig 356 starts partway through this ORF) with sequence similarity to FtsZ. Although it is missing its N-terminal third (including the tubulin signature motif), the partial ORF has excellent conservation of the region around the synergy loop and a conserved C-terminal motif. Subject to completion of the genome, *C. auranticus* therefore represents a fifth independent duplication of FtsZ sequences in the bacterial kingdom.

In summary, we have demonstrated that multiple FtsZ homologues have arisen in eight bacterial species from two different phyla (Proteobacteria and Chloroflexi). Apart from the four related species from the Rhizobiaceae and Phyllobacteriaceae (see above), the accumulation of multiple FtsZ sequences appears to have occurred independently in each case since none of the multiple FtsZs cluster together between phyla, or even between classes (Figs. 4 and 6).

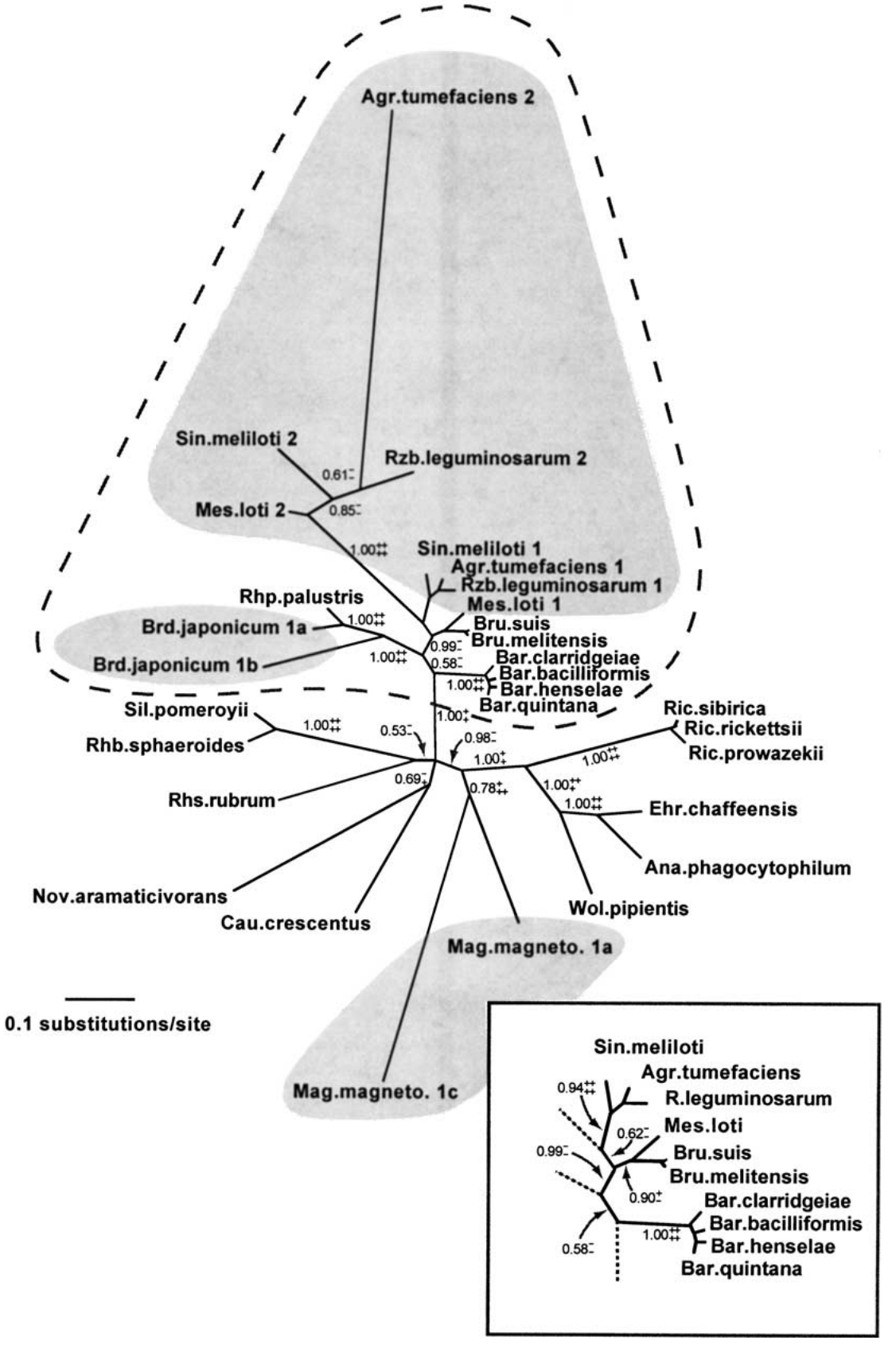
Multiple FtsZs in Archaea

Our searches identified a total of 34 Archaea FtsZ homologues, all from the Euryarchaeota phylum (Table 1). It has previously been noted that the completed genome project of the Crenarchaeon *Aerophyrum pernix* lacks an FtsZ homologue (Margolin 2000b) and we have found the same to be true for a further three Crenarchaeota—*Sulfolobus solfataricus*, *S. tokadaii*, and *Pyrobaculum aerophilum* (data not shown).

Some euryarchaeal organisms are known to possess multiple FtsZ sequences (Faguy and Doolittle 1998; Gilson and Beech 2001). Originally paralogues 1 and 2 (FtsZ1 and FtsZ2) were thought to share a common ancestor and form two separate clades (Faguy and Doolittle 1998), however, with the discovery of a third group (paralogue 3, FtsZ3) the relationships between these groups became unclear (Gilson and Beech 2001). Our searches have identified 34 FtsZ sequences from the Euryarchaeota and we find that all but 4 sequences can be clearly classified as FtsZ1s, FtsZ2s (Faguy and Doolittle 1998), or FtsZ3s (Fig. 5). The exceptions are FtsZs from two species in the Thermoplasmata class (*Thermoplasma acidophilum* and *T. volcanium*). Each *Thermoplasma* species has two FtsZs which do not group with the other archaeal FtsZ1s, FtsZ2s or FtsZ3s. Instead, they may form two separate clades between species, indicating divergence before speciation (Fig. 5). The four sequences lack the C-terminus conserved region, as we find for all Archaea FtsZs, but have an N-terminal domain, central core, and spacer region. Therefore we have designated these proteins FtsZ1a and FtsZ1b (Fig. 5) and have grouped them with FtsZ1s for the purposes of presentation in Fig. 3.

Faguy and Doolittle (1998) originally proposed a gene duplication event in Archaea to create FtsZ1 and FtsZ2 paralogues. The phylogeny of Gilson and Beech (2001), however, suggested FtsZ2 to be polyphyletic. Our phylogenetic analyses of FtsZs

Fig. 4. Phylogenetic analysis of all full-length α-proteobacterial FtsZs demonstrating at least three independent gene duplication events (highlighted in gray). The maximum likelihood (ML) tree shown was inferred from protein alignments of the FtsZ core region and C-terminal motif using the Metropolis-coupled Markov chain Monte Carlo method implemented by the program MrBayes 3.0b4 (Huelsenbeck and Ronquist 2001). Posterior probabilities (the proportion of time any topology is visited during tree inference) are shown next to the respective branches. Support for the ML topology from neighbor-joining (NJ; upper symbols) and maximum parsimony (MP; lower symbols) methods is also indicated next to the posterior probabilities. Bootstrap support for the ML topology: ≥90% (++),≥50% (+), <50% (−), as indicated. **Inset:** Topology support for short branches near center of tree. Scale bar indicates estimated number of changes per site used in analysis. Two independent gene duplications are predicted to have occurred in the Rhizobiales order (dotted outline). One of these predates the divergence of the Rhizobiaceae and Phyllobacteriaceae families, creating FtsZ1 and FtsZ2 paralogues in the species *Mesorhizobium loti* (Phyllobacteriaceae) and *Sinorhizobium meliloti*, *Rhizobium leguminosarum*, and *Agrobacterium tumefaciens* (Rhizobiaceae). The second is predicted to have occurred in the Bradyrhizobiaceae family (*Bradyrhizobium japonicum*) to form FtsZ1a and 1b. Separate from the gene duplications in Rhizobiales, gene duplications have occurred in the ancestor to *Magnetospirillum magnetotacticum*, (FtsZ1a, b, and c). FtsZ1b is not included, as its GenBank entry likely contains frameshift mistakes (see text). Genus names: *Anaplasma* (Ana), *Agrobacterium* (Agr), *Bartonella* (Bar), *Bradyrhizobium* (Brd), *Brucella* (Bru), *Caulobacter* (Cau), *Ehrlichia* (Ehr), *Magnetospirillum* (Mag), *Mesorhizobium* (Mes), *Novosphingobium* (Nov), *Rickettsia* (Ric), *Rhizobium* (Rzb), *Rhodobacter* (Rhb), *Rhodopseudomonas* (Rhp), *Rhodospirillum* (Rhs), *Silicibacter* (Sil), *Sinorhizobium* (Sin), and *Wolbachia* (Wol).


Agr.tumefaciens 2
Sin.meliloti 2
Rzb.leguminosarum 2
0.61
Mes.loti 2
0.85
1.00
Sin.meliloti 1
Agr.tumefaciens 1
Rzb.leguminosarum 1
Mes.loti 1
Rhp.palustris
Brd.japonicum 1a
1.00
Bru.suis
Bru.melitensis
0.99
Brd.japonicum 1b
1.00
0.58
Bar.clarridgeiae
Bar.bacilliformis
1.00
Bar.henselae
Bar.quintana
1.00
Sil.pomeroyii
1.00
Rhb.sphaeroides
0.53
0.98
1.00
Ric.sibirica
Ric.rickettsii
Ric.prowazekii
1.00
0.69
0.78
1.00
1.00
Rhs.rubrum
Ehr.chaffeensis
Ana.phagocytophilum
Nov.aramaticivorans
Wol.pipientis
Cau.crescentus
Mag.magneto. 1a
0.1 substitutions/site
Mag.magneto. 1c
Sin.meliloti
Agr.tumefaciens
0.94
R.leguminosarum
Mes.loti
0.62
0.99
Bru.suis
Bru.melitensis
0.90
Bar.clarridgeiae
Bar.bacilliformis
1.00
0.58
Bar.henselae
Bar.quintana

strongly supports a monophyletic origin for the euryarcheael FtsZ2s and FtsZ3s (Fig. 5). The maximum likelihood tree presented in Fig. 5 does not show a monophyletic origin for the FtsZ1 sequences. However, a monophyletic root to the archaeal FtsZ1s (excluding those of the Thermoplasmata) is predicted when using a neighbor-joining approach. Combined with the distribution of FtsZ2 and FtsZ3 paralogues (e.g., *Archeoglobus fulgidus* and *Pyrococcus abysii*), this makes single ancestral gene duplication events the most likely explanation for the occurrence of FtsZ2s and FtsZ3s in Archaea.

We have found that all Euryarchaeota for which complete genome sequence is available have at least one FtsZ1 orthologue. In fact (Thermoplasmata aside) they have either FtsZ1 only (*Methanobacterium thermoautotrophicum* and *Methanopyrus kandleri*), FtsZ1 and FtsZ2 (Class:Methanococci), or FtsZ1, 2, and 3 (*Archaeoglobus fulgidis*, *Halobacterium* species *NRC-1*, *Pyrococcus abysii*, *P. furiosus*, *P. horikoshii*). We observe no alternative combinations (e.g., FtsZ1 and FtsZ3) in completed genome projects (data not shown). Of the seven classes of the Euryarchaeota phylum, only Methanococci, Thermococci, and Thermoplasmata have more than one completed genome project. In each case we find that the genomes within these orders contain the same number and type of FtsZ homologues, and so far, none of the incomplete genome projects contradict this pattern (data not shown). Thus the distribution of FtsZs appears to be conserved within euryarchaeal phyla.

We find that euryarchaeal FtsZ1 and FtsZ2 orthologues are well conserved and contain the conserved core region and variable N-terminus (Figs. 2 and 3). However, we find that none of the Archaea FtsZ homologues contain the conserved C-terminus motif found in Bacteria (see later). By comparison to the secondary structure of the core region of *M. jannaschii* FtsZ1, we find that there is an eight-amino acid insertion in the H3-S5 loop of FtsZ2s from *Thermococcus kodakarensis* and *Pyrococcus* species. Intriguingly, the α-proteobacteria *Wolbachia pipientis* FtsZ has a 12-amino acid insertion in exactly the same place (see above).

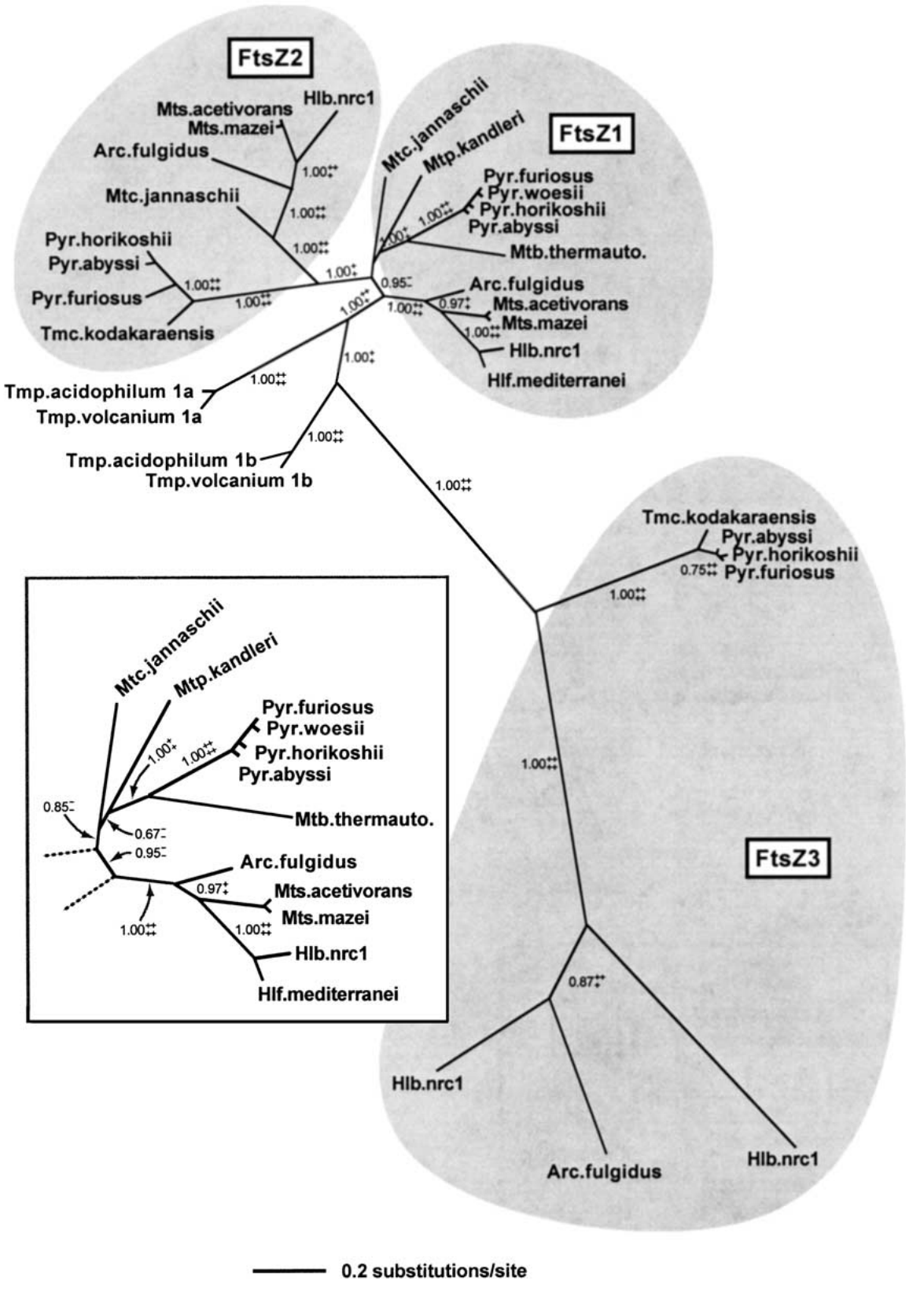
FtsZ3 orthologues are divergent compared to Archaea FtsZ1s and FtsZ2s (Figs. 3 and 5). The N-terminus region is particularly short, being between two and five amino acids before the start of the core region. Despite their highly divergent sequences we have used multiple alignments to delineate the FtsZ3 core region. The otherwise highly conserved isoleucine at the start of the core region is missing, but the next four amino acids are well conserved. The highly conserved glycine at the end of the core region is present in six out of seven FtsZ3s. Residues involved in GTP binding are well conserved, however, the otherwise totally conserved aspartic acid required for GTP hydrolysis in *E. coli* FtsZ (D_{212}) is changed to glutamic acid in two of the FtsZ3 orthologues. Interestingly, at the corresponding position, α-tubulin has a glutamic acid which contributes to GTP hydrolysis (for review see Nogales et al. 1998a).

It is difficult to speculate on the function of multiple FtsZ sequences in Euryarchaeota, as little is known about cell division in Archaea and a limited number of studies have been carried out on *ftsZ* genes from Archaea. Expression of the FtsZ1a of *T. acidophilum* in *E. coli* resulted in filamentation similar to that seen when overexpressing heterologous bacterial FtsZ proteins (Yaoi et al. 2000). In contrast, expression of *Halobacterium salinarium* FtsZ2 in *E. coli* did not induce an observable phenotype. However, when FtsZ2 was overexpressed in *H. salinarium*, cells lost their rod shape and exhibited spherical or triangular shapes (Margolin et al. 1996). This represents the only experimental work carried out on an FtsZ2 from Archaea. The implication of these experiments is that *T. acidophilum* FtsZ1a is able to form copolymers with *E. coli* FtsZ and, in doing so, disrupts the division process. In contrast, *H. salinarium* FtsZ2 cannot interact with *E. coli* FtsZ and hence shows no phenotype. This indicates that the *Thermoplasma* FtsZ1a is functionally more similar to *E. coli* FtsZ than are the Archaea FtsZ2s. An analogous situation has been demonstrated with the FtsZ1 and FtsZ2 of *S. meliloti* (see above).

Multiple FtsZs in Eukaryotes

At present, the only completed eukaryotic genome which contains FtsZ homologues is that of *Arabidopsis thaliana*—a member of the Viridiplantae phylum.

Fig. 5. Phylogenetic analysis of archaeal FtsZs. ML tree inferred as in Fig. 4. Bootstrap support for the topology from NJ (upper symbols) and MP (lower symbols) is shown next to posterior probabilities: ≥90% (++), ≥50% (+), < 50% (−). **Inset:** Topology support for short branches near center of tree. Scale bar indicates estimated number of changes per site used in analysis. Faguy and Doolittle (1998) have previously reported two FtsZ paralogues (FtsZ1 and FtsZ2; highlighted in gray), to which a third has been added (FtsZ3; highlighted in gray). Phylogenetic analysis strongly endorses duplications of the ubiquitous FtsZ1 near the euryarchaeal root to be responsible for FtsZ2 and FtsZ3 paralogues. The ML method used suggests that the archaeal FtsZ1s are not a monophyletic clade. However, this is not supported by either NJ or MP approaches, and given the distribution of FtsZ2 and 3 paralogues (e.g., *Archaeoglobus fulgidus*), it is unlikely that this part of the ML tree is correct. Genus names: *Archaeoglobus* (Arc), *Halobacterium* (Hlb), *Haloferax* (Hlf), *Methanobacterium* (Mtb), *Methanococcus* (Mtc), *Methanopyrus* (Mtp), *Methanosarcina* (Mts), *Pyrococcus* (Pyr), *Thermococcus* (Tmc), and *Thermoplasma* (Tmp).



A. thaliana has three FtsZ homologues, which are all proposed to be involved in chloroplast division and are closely related to FtsZ homologues from Cyanobacteria (Mori et al. 2001; Osteryoung and McAndrew 2001; Vitha et al. 2001). Multiple FtsZ homologues have been identified in a number of other Viridiplantae and they form two separate clades, called FtsZ1 and FtsZ2 (Osteryoung et al. 1998)

(Fig. 6). However, the number of FtsZ1 and FtsZ2 orthologues contained within an organism, differs between organisms. For example, *A. thaliana* has one FtsZ1 orthologue and two FtsZ2 orthologues, whereas *Nicotania tabacum* has three FtsZ1s and one FtsZ2. A major difference between plant FtsZ1s and FtsZ2s at the amino acid level is the lack of the C-terminus motif in FtsZ1s (Osteryoung and McAndrew 2001) (Fig. 3). We find that Viridiplantae FtsZ1 and FtsZ2 group into two clades whether complete sequences, the conserved core regions alone (data not shown), or core regions plus C-terminus regions (Fig. 6) are compared. This demonstrates a fundamental genetic difference between the two paralogue types, rather than modular addition and removal of FtsZ domains.

Multiple FtsZ homologues have also been identified in non-plant, chloroplast-containing organisms (see Table 1). In phylogenetic analysis (Fig. 6) some group with Viridiplantae FtsZ sequences and others group with FtsZ homologues from the α-proteobacteria (thought to be the progenitors of mitochondria). Indeed, one of the latter FtsZs, from *Mallomonas splendens*, has been localized to mitochondria and is thought to be involved in mitochondrial division (Beech et al. 2000). There is no obvious pattern regarding the number of FtsZ orthologues present in these organisms, although most possess one potential chloroplast FtsZ orthologue and one potential mitochondrial FtsZ orthologue. An exception is *Galdieria sulphuraria* (Takahara et al. 1999), which has two potential chloroplast FtsZ orthologues (which group with Cyanobacteria FtsZs and have potential chloroplast targeting sequences; data not shown) but no mitochondrial FtsZ orthologue. Further patterns may emerge once complete genome sequences are available for these species.

The two FtsZ homologues from the slime mold *Dictyostelium discoideum* (Gilson and Beech 2001) do not group strongly with the mitochondrial or chloroplastid clades described above. We find that one of the *D. discoideum* FtsZs (FtsZa; AAG37880) is predicted to contain a mitochondrial import sequence (TargetP score 0.606; see Materials and Methods). This FtsZ could therefore be involved in mitochondrial division. Maximum likelihood analysis predicts that both the *D. discoideum* FtsZs are monophyletic with the mitochondrial and α-proteobacterial FtsZs, but this topology is not strongly supported (Fig. 6). Significantly, a dynamin-like protein has been shown to be involved in *D. discoideum* mitochondrial morphology (Wienke et al. 1999). In addition, our searches of the incomplete *D. discoideum* genome project revealed clear homologues of all proteins identified as being involved in mitochondrial fission and fusion in eukaryotes which lack FtsZ (data not shown). Therefore *D. discoideum* mitochondrial division may involve a combination of FtsZ and dynamin-like proteins, or FtsZ may have been subsumed to a different role in this species. The idea that FtsZ functions along with proteins of eukaryotic origin during organelle division is supported by recent work into chloroplast division (Miyagishima et al. 2001; Fulgosi et al. 2002; Kuroiwa et al. 2002).

Genome projects for some organisms that possess nonphotosynthetic plastids, such as *Plasmodium falciparum*, *Toxoplasma gondii*, and *Leishmania major*, are under way or complete. Thus far we have not identified any FtsZ homologues from these organisms. Experimental evidence suggesting that basal bodies are involved in the division of plastids in *Toxoplasma gondii* (Striepen et al. 2000) may point to a very different mechanism to that driven by FtsZ.

The FtsZ C-Terminus

As mentioned previously, we define the FtsZ C-terminus region as the 9-amino acid C-terminus motif immediately followed by a variable (1–28) stretch of mainly basic and amide amino acids. The C-terminus of FtsZ is essential for cell division in *E. coli* (Ma and Margolin 1999), *Staphylococcus aureus* (Yan et al. 2000), and *Caulobacter crescentus* (Din et al. 1998) and a summary of published structural and functional data regarding the FtsZ C-terminus is presented in Fig. 7A.

The C-terminus motif has been described previously (Din et al. 1998; Ma and Margolin 1999; Yan et al. 2000; Erickson 2001), however, we have noted that, as more sequences become available, the consensus of this motif becomes less precise. We examined all 225 full-length FtsZ sequences for a recognizable match at the C-terminus and judged that 154 FtsZ sequences contained the motif (see Fig. 7B, upper panel, for examples). This purely qualitative analysis was then quantified using Multiple EM for Motif Elicitation (see Materials and Methods). Using this method only 1 of these 154 FtsZs (that from *Mycoplasma pulmonis*) was predicted not to possess this motif. We therefore find 72 FtsZ homologues that lack a conserved C-terminus motif (see Fig. 7B, lower panel, for examples) using our qualitative and quantitative methods. These include 22 from Bacteria and 16 from eukaryotic organisms, and, contrary to previous reports (Erickson 2001), we find no matching sequences at the C-terminus in any of the 34 Archaea FtsZs (see Materials and Methods for further details).

The C-terminus motif cannot be represented by a classical consensus sequence due to the variety of amino acid substitutions. We therefore derived a sequence logo (see Materials and Methods) from the 153 sequences (Fig. 7C). A completely conserved

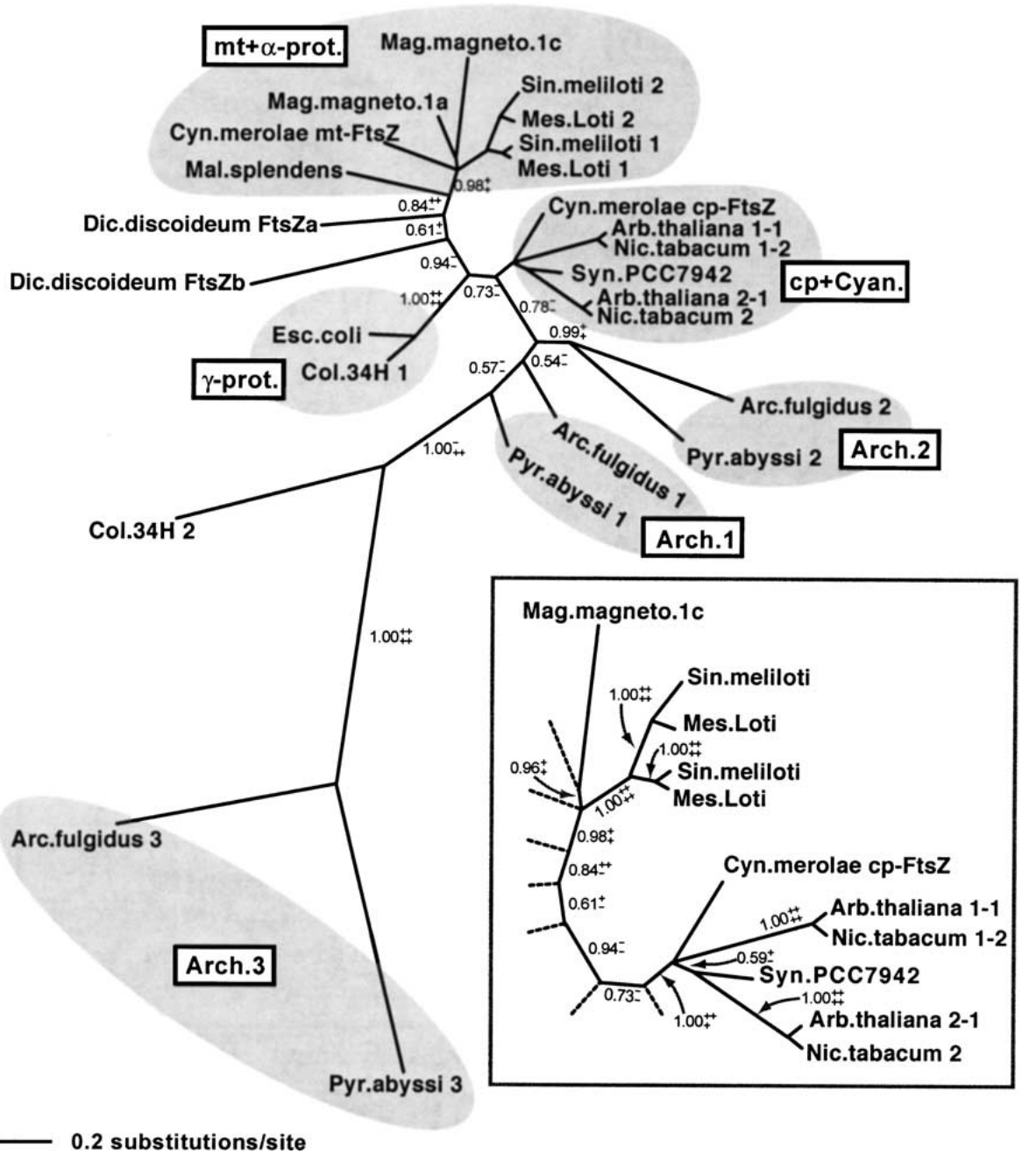


Fig. 6. Multiple independent FtsZ duplication events in Bacteria, Archaea, and Eukaryota. ML tree inferred as in Fig. 4. Bootstrap support for the topology from NJ (upper symbols) and MP (lower symbols) is shown next to posterior probabilities: ≥90% (++), ≥50% (+), or <50% (–). **Inset:** Topology support for short branches near center of tree. Scale bar indicates estimated number of changes per site used in analysis. Paralogous FtsZ1 and FtsZ2 sequences from different phyla or superkingdoms do not group together in separate clades. The following groups are separately highlighted by gray shading: Archaea FtsZ1s (Arch. 1); Archaea FtsZ2s (Arch. 2); Archaea FtsZ3s (Arch. 3); Viridiplantae FtsZ1s and FtsZ2s together with other plastid FtsZs and Cyanobacteria (cP + Cyan.); FtsZ1s and FtsZ2s of the Rhizobiales order together with other α-proteobacterial and mitochondrial FtsZs (mt + α-prot.); and *Colwellia psychrerythraea* 34H FtsZ1 together with other γ-proteobacterial FtsZs (γ-prot.). Note that *C. psychrerythraea* 34H FtsZ2 does not group with γ-proteobacteria and the two *D. discoideum* FtsZs do not group strongly with either Cyanobacteria or α-proteobacteria. Genus names: *Arabidopsis* (Arb), *Archaeoglobus* (Arc), *Colwellia* (Col), *Cyanidioschyson* (Cyn), *Dictyostelium* (Die), *Escherichia* (Esc), *Magnetospirillum* (Mag), *Mallomonas* (Mal), *Mesorhizobium* (Mes), *Nicotiania* (Nic), *Pyrococcus* (Pyr), *Sinorhizobium* (Sin), and *Synechococcus* (Syn).

proline lies at residue 6 and residues 4, 5, 8, and 9 are highly conserved, corresponding closely with functional data (Fig. 7A). The sequence logo allows us to describe the motif in general terms as follows: acidic/polar/amphoteric (positions 1 and 2), hydrophobic (position 3), acidic (position 4), hydrophobic (position 5), proline (position 6), polar/amphoteric/acidic (position 7), and hydrophobic (positions 8 and 9).

Table 1. Number and distribution of FtsZ homologues in Bacteria, Archaea, and Eukaryota

	No. of FtsZ sequences
Bacteria	
Phylum: Actinobacteria	15
Phylum: Aquificae	1
Phylum: Bacteroidetes	6
Phylum: Chlorobi	1
Phylum: Chloroflexi	1
Phylum: Cyanobacteria	8
Phylum: Deinococcus-Thermus	1
Phylum: Fibrobacteres	1
Phylum: Firmicutes	41*
Phylum: Fusobacteria	1
Phylum: Proteobacteria	82*
Phylum: Spirochaetes	3
Phylum: Thermotogae	1
Unclassified (*Dehalococcoides ethanogenes*)	1
Bacteria total	163
Archaea	
Phylum: Euryarchaeota	34*
Archaea total	34
Eukaryota	
Kingdom, Viridiplantae; phylum, Embryophta	16
Kingdom, Viridiplantae; phylum, Chlorophyta; class, Chlorophyceae	2
Class: Cryptophyta	1
Stramenopiles; class, Chrysophyceae	2
Mycetozoa; order, Dictyosteliida	2
Rhodophyta; class, Bangiophyceae	5
Eukaryota total	28
Total number of FtsZ homologues	225

Note: Asterisks indicate taxonomic divisions which also have FtsZ-like sequences (see text).

Immediately following the C-terminus motif is a short C-terminal tail which is dominated by basic or amide residues (Fig. 7B, upper panel, green underline). This tail varies in length between 1 and 13 amino acids (with the exception of *Dehalococcoides ethanogenes*, which has 28) and is well conserved within taxonomic divisions. The functional significance of this tail is not known. However, it is always present in sequences with a conserved C-terminus motif, indicating, perhaps, a combined function for these two portions of the C-terminus.

As previously noted (Din et al. 1998; Ma and Margolin 1999; Erickson 2001), it is significant that conservation of the FtsZ C-terminal motif does not necessarily correspond with the presence of an FtsA or ZipA homologue in the same genome. A good example of this is the absence of FtsA or ZipA homologues from two phyla—Cyanobacteria and Actinobacteria. In both phyla, we have found the C-terminal region to be present in every FtsZ homologue identified so far (Fig. 3). Therefore, as pointed out previously (Erickson 2001), the role of the C-terminal motif must not be exclusively to interact with FtsA and ZipA. More likely is that the FtsZ C-terminus is available to interact with accessory proteins which will differ between taxonomic divisions. The converse situation appears to exist in γ-proteobacteria of the Pasteurellales family and Bacteroidetes species. We find that these sequences lack a conserved C-terminus motif (Figs. 3 and 7B). However, when we performed BLAST searches on the genome sequences of these species, using *E. coli* FtsA and ZipA as queries, we found that all have FtsA homologues and most of the Pasteurellales have ZipA homologues (some genomes are incomplete; data not shown). It is possible therefore that neither FtsA nor ZipA interacts with FtsZ in these organisms. We consider it more likely, at least in the cases of two of the Pasteurellales (*H. influenzae* and *P. multocida*), that interaction does occur through a loosely similar stretch of amino acids at the FtsZ C-terminus. We find that both the *H. influenzae* and the *P. multocida* ZipA sequences have a number of nonconservative amino acid substitutions (data not shown) at positions identified as being in complex with the FtsZ C-terminus in *E. coli* (Mosyak et al. 2000). This may be evidence of complementary changes between the two interacting domains in these species.

A situation in which FtsA homologues are present alongside FtsZs lacking the C-terminus motif arises in Bacteria with multiple FtsZs. As discussed above, in Phyllobacteriaceae and Rhizobiaceae, FtsZ1s have the C-terminus motif and FtsZ2s do not. One might propose, therefore, that FtsZ1 function involves interaction with FtsA whereas FtsZ2 function does not.

In Viridiplantae, the C-terminus region is present in all of the FtsZ2 orthologues but not in the FtsZ1 orthologues. In this respect (and in terms of which paralogue is most similar to the original *E. coli* and *B. subtilis* sequences), the nomenclature is reversed from that of α-proteobacteria. Both FtsZ1 and FtsZ2 paralogues group with Cyanobacteria FtsZs (Fig. 6) and are implicated in chloroplast division (for review see Osteryoung and McAndrew 2001). In the other eukaryotic organisms, those that group with Cyanobacteria FtsZ homologues have the C-terminus motif, whereas those that group with α-proteobacteria (implicated in mitochondrial division) do not. This implies that at least one FtsZ which interacts with accessory proteins is required for chloroplast division but that this is not necessarily the case in mitochondria. Since no eukaryotes so far have FtsA or ZipA homologues, there are few candidates for binding to this motif during organelle division, but perhaps ARTEMIS is one such possibility (Fulgosi et al. 2002).

In summary, our analysis redefines the conserved C-terminus motif and illustrates the prevalence of this

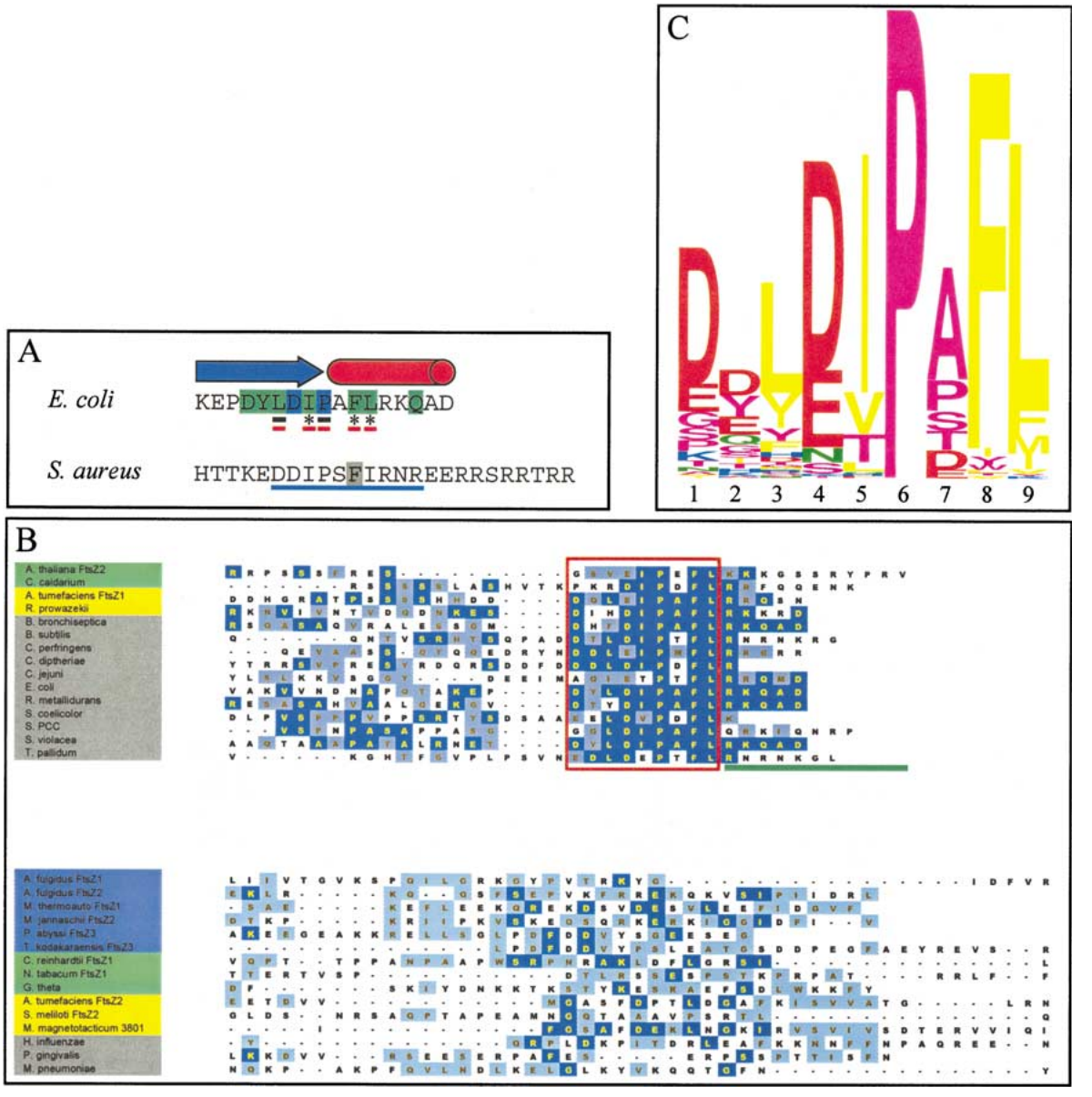


Fig. 7. The FtsZ C-terminus region. **A** In complex with ZipA, a 17-mer peptide containing the *E. coli* C-terminus motif, which forms a β-strand (blue arrow) followed by an α-helix (red cylinder), is buried in a shallow hydrophobic cavity within ZipA (Mosyak et al. 2000). Seven of the 17 residues contact ZipA (boxed in green), however, nearly all of the binding affinity for ZipA is provided by 3 residues (asterisks). Two other residues are thought to be important for the conformation of the peptide (boxed in blue) (Mosyak et al. 2000). In *E. coli*, five residues (underlined in red) were individually mutated to alanine; in each case the resulting FtsZ was unable to complement the *ftsZ84*(Ts) mutant. Yeast two-hybrid analysis of these FtsZs found two mutations which reduced binding to FtsA (underlined in black) (Haney et al. 2001). Deletion of a 10-residue stretch of amino acids overlapping the C-terminus motif from *Staphylococcus aureus* FtsZ (underlined in blue) rendered it unable to interact with FtsA using a yeast two-hybrid system, however, alanine scanning mutagenesis identified only one of these residues (boxed in gray) as being required for interaction with FtsA (Van et al. 2000). **B** The last 30 amino acids of FtsZ sequences from a range of organisms have been aligned to highlight the presence or absence of the C-terminal motif. Dark blue shading highlights identical residues and light blue shading highlights similar residues. Upper alignment: 15 FtsZ sequences with the C-terminus motif (orange box). The basic/amide-rich region is underlined in green (see text). Lower alignment: 15 FtsZ sequences without the C-terminus motif. **C** A sequence logo (see Materials and Methods) produced from an alignment of the C-terminus region of 153 FtsZ homologues that contain a conserved C-terminus motif. The height of the letters represents the frequency of the amino acid residue at that position.

motif among FtsZ homologues. The C-terminus motif may represent a protein–protein interaction domain, which has been adapted to bind to a variety of partners and lost when this function is no longer required.

FtsZ-like Sequences

In addition to those sequences already discussed, we found 10 sequences related to FtsZ which fall into two categories: (A) ORFs from 5 Bacteria and 2

Archaea that contain low sequence similarity to FtsZ and do not meet all our criteria to be classed as genuine FtsZs and (B) 1 ORF from Bacteria and 2 ORFs from Archaea that have moderate to high sequence similarity to FtsZ but have alterations in regions known to be important for FtsZ function. We consider it likely that all proteins described here do not bind and hydrolyze GTP in the same manner as FtsZ, therefore we choose to designate them "FtsZ-like." Except for the plasmid-borne FtsZ-like sequences from *Bacillus* species (see below), which are very closely related, FtsZ-like sequences do not group with each other between species in phylogenetic analyses (data not shown). They therefore do not constitute a distinct family of sequences. By documenting FtsZ-like sequences, we have encompassed in this paper all sequences, save from the tubulin superfamily (McKean et al. 2001), which show significant similarity to FtsZs.

FtsZ-like Sequences in Bacteria

Bacillus Species (Phylum, Firmicutes; Order, Bacillales; Family, Bacillaceae). Our searches identify two sequences related to FtsZ in the *Bacillus anthracis* whole-genome shotgun sequence. The genuine FtsZ groups with FtsZs from other Firmicutes in phylogenetic analysis (data not shown). The FtsZ-like ORF is encoded on the pXO1 virulence plasmid (Okinaka et al. 1999) and contains the tubulin signature motif. However, it shows only 17% sequence identity to the genuine, chromosomally encoded *B. anthracis* FtsZ and does not fit our criteria to be an FtsZ. Similarly, we find that the *B. cereus* genome (NC_003909) has a chromosomal *ftsZ* gene (99% identical to the chromosomal *B. anthracis* orthologue at the amino acid level) and a partial FtsZ-like ORF (85% identical to the equivalent portion of the *B. anthracis* FtsZ-like ORF) (Okinaka et al. 1999). Furthermore, DNA sequencing of plasmids from *B. thuringiensis* and *B. megaterium* reveals homologues of this FtsZ-like sequence. Cross-hybridization studies suggest the presence of the FtsZ-like ORF in *B. mycoides*, *B. amloliquifaciens*, and *Paenibacillus glucanolyticus* and, also, confirmed the presence of large extra-chromosomal DNA molecules in these species. However, the study did not confirm whether the FtsZ-like ORFs are encoded on these plasmids (Okinaka et al. 1999; Pannucci et al. 2002). The pX01 plasmid is not necessary for viability of *B. anthracis* in vitro but it is known to be essential for the manifestation of disease (Okinaka et al. 1999). It is therefore possible that the FtsZ-like ORF is required during host infection.

Clostridium acetobutylicum (*Phylum, Firmicutes; Class, Clostridia; Order, Clostridiales; Family, Clostridiaceae*). The completed genome sequence for *Clostridium acetobutylicum* contains two ORFs annotated as FtsZs. We found one to group with FtsZs from other *Clostridium* species and therefore to be a genuine FtsZ (data not shown). However, the second ORF, annotated as a "diverged homologue" of FtsZ (accession No. NP350048), encodes a protein with only 14% sequence identity to the *C. acetobutylicum* FtsZ. This sequence contains the tubulin signature motif but lacks some of the residues involved in GTP binding and hydrolysis. In addition, PRINTS does not predict this to be an FtsZ homologue and therefore we have designated this as an FtsZ-like sequence.

Magnetospirillum magnetotacticum (Phylum, Proteobacteria; Class, α-Proteobacteria; Order, Rhodospirillales; Family, Rhodospirillaceae). We designate one of the four FtsZ sequences found in this bacterium as FtsZ-like for reasons described above (see Bacteria with Multiple FtsZ Sequences).

FtsZ-like Sequences in the Archaea

An ORF from *Halobacterium nrc-1* (NP_280638) is over 40% identical to FtsZ3a from the same species but lacks important residues involved in GTP-hydrolysis at the N-terminal end of the core region. An ORF from *Methanosarcina acetivorans* (NP_615906) is most closely related to FtsZ3 sequences but lacks some important residues for GTP hydrolysis. Finally, two ORFs from *Methanopyrus kandleri*, which are each other's closest relatives in BLAST searches (NP_614409, NP_614146), show minimal similarity to other FtsZs. Both lack many of the residues involved in GTP hydrolysis and are very much shorter than any genuine FtsZ sequences. However, similarity to FtsZ does reside in and around the most conserved regions (e.g., the tubulin signature sequence and the synergy loop). Although both are annotated "FtsZ GTPase involved in cell division," these are the sequences least related to FtsZ which we present here.

Conclusions

We have comprehensively documented the FtsZ sequences present in freely available sequence databases (Table 1), and using multiple sequence alignments, we have summarized patterns of primary sequence structure diagrammatically by phylum (Fig. 3). This approach aids the identification of interesting exceptions (such as the *W. pipientis* and Pasteurellales FtsZs, which have insertions in their core domain) and should also prove useful for initial characterization of novel FtsZ sequences.

Phylogenetic analysis has highlighted extremely divergent FtsZ sequences such as those from *Mycoplasma pneumoniae*, *M. genitalium*, *Colwellia psychrerythraea 34H* (FtsZ2), *Dictyostelium discoideum*, and Euryarchaeota (FtsZ3s). Significantly, apart from those of the *Mycoplasma* genus, these all come from species with more than one FtsZ (Fig. 6) and may, therefore, be performing functions distinct from cell division. As yet, there is no functional data for any of these proteins. Nevertheless, they do have all the important residues for GTP hydrolysis in the same order and position as more conventional FtsZs (Fig. 2). It remains to be seen whether FtsZ-like sequences such as those from *Bacillus* and *Clostridium* species and *Methanopyrus kandleri* are simply even more divergent FtsZs which have mutated these important residues as they specialize in function.

In the cases of *Colwellia psychrerythraea* (FtsZ2) and *Dictyostelium discoidium* (FtsZ1a and 1b) we must consider the possibility that these divergent FtsZs were acquired by horizontal gene transfer; see Fig. 6. Indeed it is also possible that one of the *Chloroflexus aurantiacus* FtsZs was acquired in this way. The highest-scoring hit in BLAST searches using the full-length FtsZ (FtsZ2) from this organism comes from *Deinococcus radiodurans*, whereas for the truncated sequence (FtsZ1) the best hits come from Cyanobacteria. The two FtsZs in *C. aurantiacus* may therefore not be each other's closest relatives (data not shown).

Expression of heterologous FtsZ proteins in *E. coli* has often been used as a test for similar function, but it also indicates that copolymer formation occurs even between FtsZs of different species (e.g., Ma et al. 1997). Therefore it is reasonable to suggest that FtsZ in species with multiple FtsZs could be forming copolymers at various ratios, perhaps even through heterodimer formation (e.g., McAndrew et al. 2001). We have identified an interesting case in the α-proteobacteria *Magnetospirillium magnetotacticum* which has at least two FtsZs (probably three; see above) and an FtsZ-like protein. The latter is classed as such because of a nonconservative substitution in the synergy loop. Nevertheless, the FtsZ-like protein could act in polymer formation in a similar way to β-tubulin, which has a similar substitution and cannot catalyze hydrolysis of GTP bound to α-tubulin.

It has been suggested that rapid evolution of FtsZ, after gene duplication in Archaea, could represent the origin of eukaryotic tubulins (Faguy and Doolittle 1998). We find that FtsZ3s of Archaea group with tubulins when the latter are included as an outgroup for FtsZ phylogenetic analysis (data not shown), however, this is likely to be due to long branch attraction as previously suggested (Gilson and Beech 2001). It is significant that we can identify multiple independent instances of *ftsZ* gene duplication—in Viridiplantae (Osteryoung et al. 1998), Archaea (Faguy and Doolittle 1998), γ-proteobacteria (*Colwellia psychrerythraea*), *Chloroflexus auranticus*, and α-proteobacteria (Figs. 4–6). As around half of bacterial phyla still do not have a representative for which complete genome sequence is available, it seems likely that a clearer ancestor of tubulin remains to be found. In this context, a recent report of two tubulin genes in the bacterial genus *Prosthecobacter* (Jenkins et al. 2002) is intriguing. Although these tubulins are predicted to be monomeric, it will be interesting to see if they are required for cell division of *P. dejongeii*, which appears to be only the second free-living bacteria identified as lacking an FtsZ homologue (V. Vonstein and Y. Kogan, Integrated Genomics, Inc., personal communication).

Acknowledgments. This work was supported by grants to K.G. (Principal Fellow) and S.G.A. (Research Career Development Fellow) from the Wellcome Trust. We thank Professor T.K. Attwood and Dr. P. Scordis for early discussions of this work and Dr. T.M. Embley for help with phylogenetic analyses.

References

Adams MD, Celniker SE, Holt RA, et al. (2000) The genome sequence *Drosophila melanogaster*. Science 287:2185–2195

Addinall SG, Holland B (2002) The tubulin ancestor, FtsZ, draughtsman, designer and driving force for bacterial cytokinesis. J Mol Biol 318:219–236

Arimura S-i, Tsutsumi N (2002) A dynamin-like protein (ADL2b), rather than FtsZ, is involved in *Arabidopsis* mitochondrial division. PNAS 99:5727–5731

Attwood TK, Beck ME (1994) PRINTS—A protein motif fingerprint database. Protein Eng 7:841–848

Attwood TK, Beck ME, Bleasby AJ, Parry-Smith DJ (1994) PRINTS—A database of protein motif fingerprints. Nucleic Acids Res 22:3590–3596

Attwood TK, Beck ME, Flower DR, Scordis P, Selley JN (1998) The PRINTS protein fingerprint database in its fifth year. Nucleic Acids Res 26:304–308

Attwood TK, Blythe MJ, Flower DR, Gaulton A, Mabey JE, Maudling N, McGregor L, Mitchell AL, Moulton G, Paine K, Scordis P (2002) PRINTS and PRINTS-S shed light on protein ancestry. Nucleic Acids Res 30:239–241

Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc 2nd Int Conf Intelligent Syst Mol Biol, AAAI Press, Menlo Park, C, pp 28–36

Bailey TL, Gribskov (1997) Combining evidence using p-values: Application to sequence homology searches. Bioinf 14:48–54

Beech PL, Gilson PR (2000) FtsZ and organelle division in Protists. Protist 151:11–16

Beech PL, Nheu T, Schultz T, Herbert S, Lithgow T, Gilson PR, McFadden GI (2000) Mitochondrial FtsZ in a chromophyte alga. Science 287:1276–1279

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA, Wheeler DL (2000) GenBank. Nucleic Acids Res 28:15–18

Brown WJ, Rockey DD (2000) Identification of an antigen localized to an apparent septum within dividing chlamydiae. Infect Immun 68:708–715

Bucher P, Bairoch A (1994) A generalised profile syntax for biomolecular sequence motifs and its function in automatic sequence interpretation. In: Altman R, Brutlag D, Karp P, Lathrop R, Searls D (eds) 2nd International Conference on Intelligent Systems for Molecular Biology. AAAI Press, pp 53–61

Chen JC, Beckwith J (2001) FtsQ, FtsL and FtsI require FtsK, but not FtsN, for co-localization with FtsZ during *Escherichia coli* cell division. Mol Microbiol 42:395–413

Consortium (1998) Genome sequence of the nematode *C. elegans*: A platform for investigating biology. Science 282:2012–2018

Din N, Quardokus EM, Sackett MJ, Brun YV (1998) Dominant C-terminal deletions of FtsZ that affect its ability to localize in *Caulobacter* and its interaction with FtsA. Mol Microbiol 27:1051–1063

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300:1005–1016

Erickson HP (1995) FtsZ, a prokaryotic homolog of tubulin? Cell 80:367–370

Erickson HP (2000) Dynamin and FtsZ. Missing links in mitochondrial and bacterial division. J Cell Biol 148:1103–1105

Erickson HP (2001) The FtsZ protofilament and attachment of ZipA—Structural constraints on the FtsZ power stroke. Curr Opin Cell Biol 13:55–60

Faguy DM, Doolittle WF (1998) Cytoskeletal proteins: The evolution of cell division. Curr Biol 8:R338–R341

Fitz-Gibbon ST, Ladner H, Kim UJ, Stetter KO, Simon MI, Miller JH (2002) Genome sequence of the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. PNAS 99:984–989

Fujiwara M, Yoshida S (2001) Chloroplast targeting of chloroplast division Ftsz2 proteins in *Arabidopsis*. Biochem Biophys Res Commun 287:462–467

Fulgosi H, Gerdes L, Westphal S, Glockmann C, Soll J (2002) Cell and chloroplast division requires ARTEMIS. PNAS 99:11501–11506

Gage DJ, Margolin W (2000) Hanging by a thread: Invasion of legume plants by Rhizobia. Curr Opin Microbiol 3:613–617

Galibert F, Finan TM, Long SR, et al. (2001) The composite genome of the legume symbiont *Sinorhizobium meliloti*. Science 293:668–672

Garrity GM, Holt JG (2001) Taxonomic outline of the Archaea and Bacteria. In: Boone DR, Castenholz RW (eds) Bergey's manual of systematic bacteriology, 2nd ed, Vol 1. The Archaea and the deeply branching and phototrophic Bacteria. Springer-Verlag, New York, pp 155–166

Gilson PR, Beech PL (2001) Cell division protein FtsZ: Running rings around bacteria, chloroplasts and mitochondria. Res Microbiol 152:3–10

Glass JI, Lefkowitz EJ, Glass JS, Heiner CR, Chen EY, Cassell GH (2000) The complete sequence of the mucosal pathogen *Ureaplasma urealyticum*. Nature 407:757–762

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG (1996) Life with 6000 genes. Science 274:546, 563–547

Goodner B, Hinkle G, Gattung S, et al. (2001) Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. Science 294:2323–2328

Hale CA, de Boer PAJ (2002) ZipA is required for recruitment of FtsK, FtsQ, FtsL, and FtsN to the septal ring in *Escherichia coli*. J Bacteriol 184:2552–2556

Hale CA, Rhee AC, de Boer PA (2000) ZipA-induced bundling of FtsZ polymers mediated by an interaction between C-terminal domains. J Bacteriol 182:5153–5166

Haney SA, Glasfeld E, Hale C, Keeney D, He ZZ, de Boer P (2001) Genetic analysis of the Escherichia coli FtsZ · ZipA interaction in the yeast two-hybrid system—Characterization of FtsZ residues essential for the interactions with ZipA and with FtsA. J Biol Chem 276:11980–11987

Hogue CW (1997) Cn3D: A new generation of three-dimensional molecular structure viewer. Trends Biochem Sci 22:314–316

Huang J, Cao C, Lutkenhaus J (1996) Interaction between FtsZ and inhibitors of cell division. J Bacteriol 178:5080–5085

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinform 17:754–755

Jenkins C, Samudrala R, Anderson I, Hedlund BP, Petroni G, Michailova N, Pinel N, Overbeek R, Rosati G, Staley JT (2002) Genes for the cytoskeletal protein tubulin in the bacterial genus *Prosthecobacter*. PNAS 99:17049–17054

Kaneko T, Nakamura Y, Sato S, Asamizu E, Kato T, Sasamoto S, Watanabe A, Idesawa K, Ishikawa A, Kawashima K, Kimura T, Kishida Y, Kiyokawa C, Kohara M, Matsumoto M, Matsuno A, Mochizuki Y, Nakayama S, Nakazaki N, Shimpo S, Sugimoto M, Takeuchi C, Yamada M, Tabata S (2000a) Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. DNA Res 7:331–338

Kaneko T, Nakamura Y, Sato S, Asamizu E, Kato T, Sasamoto S, Watanabe A, Idesawa K, Ishikawa A, Kawashima K, Kimura T, Kishida Y, Kiyokawa C, Kohara M, Matsumoto M, Matsuno A, Mochizuki Y, Nakayama S, Nakazaki N, Shimpo S, Sugimoto M, Takeuchi C, Yamada M, Tabata S (2000b) Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti* (supplement). DNA Res 7:381–406

Kaneko T, Nakamura Y, Sato S, Minamisawa K, Uchiumi T, Sasamoto S, Watanabe A, Idesawa K, Iriguchi M, Kawashima K, Kohara M, Matsumoto M, Shimpo S, Tsuruoka H, Wada T, Yamada M, Tabata S (2002) Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. DNA Res 9:189–286

Kawarabayasi Y, Hino Y, Horikawa H, et al. (2001) Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain 7. DNA Res 8:123–140

Kuroiwa H, Mori T, Takahara M, Miyagishima SY, Kuroiwa T (2002) Chloroplast division machinery as revealed by immunofluorescence and electron microscopy. Planta 215:185–190

Liu Z, Mukherjee A, Lutkenhaus J (1999) Recruitment of ZipA to the division site by interaction with FtsZ. Mol Microbiol 31:1853–1861

Lowe J, Amos LA (1998) Crystal structure of the bacterial cell-division protein FtsZ. Nature 391:203–206

Lutkenhaus J, Addinall SG (1997) Bacterial cell division and the Z ring. Annu Rev Biochem 66:93–116

Lutkenhaus JF, Wolf-Watz H, Donachie WD (1980) Organization of genes in the ftsA-envA region of the Escherichia coli genetic map and identification of a new fts locus (ftsZ). J Bacteriol 142:615–620

Ma X, Ehrhardt DW, Margolin W (1996) Colocalization of cell division proteins FtsZ and FtsA to cytoskeletal structures in living *Escherichia coli* cells by using green fluorescent protein. PNAS 93:12998–13003

Ma X, Margolin W (1999) Genetic and functional analyses of the conserved C-terminal core domain of Escherichia coli FtsZ. J Bacteriol 181:7531–7544

Ma X, Sun Q, Wang R, Singh G, Jonietz EL, Margolin W (1997) Interactions between heterologous FtsA and FtsZ proteins at the FtsZ ring. J Bacteriol 179:6788–6797

Margolin W (2000a) Organelle division: Self-assembling GTPase caught in the middle. Curr Biol 10:R328–R330

Margolin W (2000b) Themes and variations in prokaryotic cell division. FEMS Microbiol Rev 24:531–548

Margolin W, Long SR (1994) *Rhizobium meliloti* contains a novel second homolog of the cell division gene ftsZ. J Bacteriol 176:2033–2043

Margolin W, Corbo JC, Long SR (1991) Cloning and characterization of a Rhizobium meliloti homolog of the *Escherichia coli* cell division gene ftsZ. J Bacteriol 173:5822–5830

Margolin W, Wang R, Kumar M (1996) Isolation of an ftsZ homolog from the archaebacterium *Halobacterium salinarium*: Implications for the evolution of FtsZ and tubulin. J Bacteriol 178:1320–1327

McAndrew RS, Froehlich JE, Vitha S, Stokes KD, Osteryoung KW (2001) Colocalization of plastid division proteins in the chloroplast stromal compartment establishes a new functional relationship between FtsZ1 and FtsZ2 in higher plants. Plant Physiol 127:1656–1666

McKean PG, Vaughan S, Gull K (2001) The extended tubulin superfamily. J Cell Sci 114:2723–2733

Miyagishima S, Takahara M, Mori T, Kuroiwa H, Higashiyama T, Kuroiwa T (2001) Plastid division is driven by a complex mechanism that involves differential transition of the bacterial and eukaryotic division rings. Plant Cell 13:2257–2268

Mori T, Kuroiwa H, Takahara M, Miyagishima S, Kuroiwa T (2001) Visualization of an FtsZ ring in chloroplasts of *Lilium longiflorum* leaves. Plant Cell Physiol 42:555–559

Mosyak L, Zhang Y, Glasfeld E, Haney S, Stahl M, Seehra J, Somers WS (2000) The bacterial cell-division protein ZipA and its interaction with an FtsZ fragment revealed by X-ray crystallography. EMBO J 19:3179–3191

Nogales E, Downing KH, Amos LA, Lowe J (1998a) Tubulin and FtsZ form a distinct family of GTPases. Nat Struct Biol 5:451–458

Nogales E, Wolf SG, Downing KH (1998b) Structure of the alpha beta tubulin dimer by electron crystallography. Nature 391:199–203

Okinaka RT, Cloud K, Hampton O, Hoffmaster AR, Hill KK, Keim P, Koehler TM, Lamke G, Kumano S, Mahillon J, Manter D, Martinez Y, Ricke D, Svensson R, Jackson PJ (1999) Sequence and organization of pXO1, the large *Bacillus anthracis* plasmid harboring the anthrax toxin genes. J Bacteriol 181:6509–6515

Osteryoung KW, McAndrew RS (2001) The plastid division machine. Ann Rev Plant Physiol Plant Mol Biol 52:315–333

Osteryoung KW, Stokes KD, Rutherford SM, Percival AL, Lee WY (1998) Chloroplast division in higher plants requires members of two functionally divergent gene families with homology to bacterial ftsZ. Plant Cell 10:1991–2004

Oyaizu H, Debrunner-Vossbrinck B, Mandelco L, Studier JA, Woese CR (1987) The green non-sulfur bacteria: A deep branching in the eubacterial line of descent. Syst Appl Micro biol 9:47–53

Pannucci J, Okinaka RT, Sabin R, Kuske CR (2002) *Bacillus anthracis* pX01 plamid sequence conservation among closely related bacterial species. J Bacteriol 184:134–141

RayChaudhuri D (1999) ZipA is a MAP-Tau homolog and is essential for structural integrity of the cytokinetic FtsZ ring during bacterial cell division. EMBO J 18:2372–2383

Read TD, Brunham RC, Shen C, Gill SR, Heidelberg JF, White O, Hickey EK, Peterson J, Utterback T, Berry K, Bass S, Linher K, Weidman J, Khouri H, Craven B, Bowman C, Dodson R, Gwinn M, Nelson W, DeBoy R, Kolonay J, McClarty G, Salzberg SL, Eisen J, Fraser CM (2000) Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. Nucleic Acids Res 28:1397–1406

Rothfield L, Justice S, Garcia-Lara J (1999) Bacterial cell division. Annu Rev Genet 33:423–448

Schachter J, Wyrick PB (1994) Culture and isolation of *Chlamydia trachomatis*. Methods Enzymol 236:377–390

Schneider TD, Stephens RM (1990) Sequence logos: A new way to display consensus sequences. Nucleic Acids Res 18:6097–6100

Shaw JM, Nunnari J (2002) Mitochondrial dynamics and division in budding yeast. Trends Cell Biol 12:178–184

She Q, Singh RK, Confalonieri F, et al. (2001) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. PNAS 98:7835–7840

Smirnova E, Griparic L, Shurland DL, van Der Bliek AM (2001) Dynamin-related protein drp1 is required for mitochondrial division in mammalian cells. Mol Biol Cell 12:2245–2256

Striepen B, Crawford MJ, Shaw MK, Tilney LG, Seeber F, Roos DS (2000) The plastid of *Toxoplasma gondii* is divided by association with the centrosomes. J Cell Biol 151:4123–1434

Swofford DL (1998) Phylogenetic analysis using parsimony (*and other methods). Sinauer, Sunderland, MA

Takahara M, Takahashi H, Matsunaga S, Sakai A, Kawano S, Kuroiwa T (1999) Two types of ftsZ genes isolated from the unicellular primitive red alga Galdieria sulphuraria. Plant Cell Physiol 40:784–791

Teng K, Li M, Yu W, Li H, Shen D, Liu D (1994) Comparison of PCR with culture for detection of *Ureaplasma urealyticum* in clinical samples from patients with urogenital infections. J Clin Microbiol 32:2232–2234

Trachtenberg S (1998) Mollicutes—Wall-less bacteria with internal cytoskeletons. J Struct Biol 124:244–256

Vitha S, McAndrew RS, Osteryoung KW (2001) FtsZ ring formation at the chloroplast division site in plants. J Cell Biol 153:111–120

Wang X, Lutkenhaus J (1996) FtsZ ring: The eubacterial division apparatus conserved in archaebacteria. Mol Microbiol 21:313–319

Wang X, Huang J, Mukherjee A, Cao C, Lutkenhaus J (1997) Analysis of the interaction of FtsZ with itself, GTP, and FtsA. J Bacteriol 179:5551–5559

Wheeler DL, Chappey C, Lash AE, Leipe DD, Madden TL, Schuler GD, Tatusova TA, Rapp BA (2000) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 28:10–14

Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum likelihood approach. Mol Biol Evol 18:691–699

Wienke DC, Knetsch ML, Neuhaus EM, Reedy MC, Manstein DJ (1999) Disruption of a dynamin homologue affects endocytosis, organelle morphology, and cytokinesis in *Dictyostelium discoideum*. Mol Biol Cell 10:225–243

Wood V, Gwilliam R, Rajandream MA, et al. (2002) The genome sequence of *Schizosaccharomyces pombe*. Nature 415:871–880

Yan K, Pearce KH, Payne DJ (2000) A conserved residue at the extreme C-terminus of FtsZ is critical for the FtsA-FtsZ interaction in *Staphylococcus aureus*. Biochem Biophys Res Commun 270:387–392

Yang Z (1993) Maximum likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Mol Biol Evol 10:1396–1401

Yaoi T, Laksanalamai P, Jiemjit A, Kagawa HK, Alton T, Trent JD (2000) Cloning and characterization of *ftsZ* and *pyrF* from the archaeon *Thermoplasma acidophilum*. Biochem Biophys Res Commun 275:936–945