Published online October 5, 2004

5260–5279 *Nucleic Acids Research, 2004, Vol. 32, No. 17*
doi:10.1093/nar/gkh828

Comparative genomics of the FtsK–HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging

Lakshminarayan M. Iyer, Kira S. Makarova, Eugene V. Koonin and L. Aravind*

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

Received May 25, 2004; Revised and Accepted August 27, 2004

ABSTRACT

Recently, it has been shown that a predicted P-loop ATPase (the HerA or MlaA protein), which is highly conserved in archaea and also present in many bacteria but absent in eukaryotes, has a bidirectional helicase activity and forms hexameric rings similar to those described for the TrwB ATPase. In this study, the FtsK–HerA superfamily of P-loop ATPases, in which the HerA clade comprises one of the major branches, is analyzed in detail. We show that, in addition to the FtsK and HerA clades, this superfamily includes several families of characterized or predicted ATPases which are predominantly involved in extrusion of DNA and peptides through membrane pores. The DNA-packaging ATPases of various bacteriophages and eukaryotic double-stranded DNA viruses also belong to the FtsK–HerA superfamily. The FtsK protein is the essential bacterial ATPase that is responsible for the correct segregation of daughter chromosomes during cell division. The structural and evolutionary relationship between HerA and FtsK and the nearly perfect complementarity of their phyletic distributions suggest that HerA similarly mediates DNA pumping into the progeny cells during archaeal cell division. It appears likely that the HerA and FtsK families diverged concomitantly with the archaeal–bacterial division and that the last universal common ancestor of modern life forms had an ancestral DNA-pumping ATPase that gave rise to these families. Furthermore, the relationship of these cellular proteins with the packaging ATPases of diverse DNA viruses suggests that a common DNA pumping mechanism might be operational in both cellular and viral genome segregation. The *herA* gene forms a highly conserved operon with the gene for the NurA nuclease and, in many archaea, also with the orthologs of eukaryotic double-strand break repair proteins MRE11 and Rad50. HerA is predicted to function in a complex with these proteins in DNA pumping and repair of double-stranded breaks introduced during this process and, possibly, also during DNA replication. Extensive comparative analysis of the 'genomic context' combined with in-depth sequence analysis led to the prediction of numerous previously unnoticed nucleases of the NurA superfamily, including a specific version that is likely to be the endonuclease component of a novel restriction-modification system. This analysis also led to the identification of previously uncharacterized nucleases, such as a novel predicted nuclease of the Sir2-type Rossmann fold, and phosphatases of the HAD superfamily that are likely to function as partners of the FtsK–HerA superfamily ATPases.

INTRODUCTION

Cell division in bacteria is mediated by several distinct protein complexes which are involved in chromosome segregation, choice of the division site and partitioning of the chromosomes between the daughter cells (1,2). In bacteria, unlike in eukaryotes, DNA replication, chromosome segregation and cell division are not temporally ordered by the phases of the cell cycle during which checkpoints ensure the proper progression of these events. The key event in bacterial cell division is the assembly of the oligomeric Z-ring formed by the tubulin-related GTPase, FtsZ (3,4). This ring typically forms near the center of the bacterial cell, where the DNA concentration is low. Aberrant formation of the Z-ring in regions closer to the poles of the cell is prevented by the action of the MinD ATPase and an associated protein complex (5). The FtsZ ring recruits another key cell division protein, FtsK, via interactions with its N-terminal region, and FtsK, in turn, recruits several additional cell division proteins to the ring complex (6,7). FtsK is a large protein that consists of an N-terminal transmembrane domain with four membrane-spanning helices, a central coiled-coil region and a C-terminal P-loop ATPase domain. Although disruption of the ATPase domain of FtsK is not lethal in *Escherichia coli*, the mutant cells are defective in

*To whom correspondence should be addressed. Tel: +1 301 594 2445; Fax: +1 301 435 7794; Email: aravind@ncbi.nlm.nih.gov

chromosome segregation as well as septation and exhibit asymmetrically positioned nucleoids and large anucleate regions (8). These observations suggest that the ATPase activity of FtsK is required for proper chromosome segregation (9,10). After the replication of bacterial circular chromosomes, homologous recombination can lead to the formation of dimeric circles (9,11–13). Recombinases XerC and XerD act in concert to resolve these dimers (14). The ATPase domain of FtsK tightly regulates the Xer recombinases and mediates a switch in the catalytic state of XerCD such that XerD initiates duplex recombination (9,10). Experiments in the *Bacillus subtilis* and the *E.coli* systems indicate that the FtsK protein translocates along DNA and mediates pumping of the chromosome across the closing septum (12,15). Furthermore, FtsK also interacts with the ParC subunit of topoisomerase IV and recruits it to regions close to the septum. Additionally, FtsK activates chromosome decatenation by topoisomerase IV (16). These observations indicate that FtsK plays a central role in chromosome segregation both by activating recombination and decatenation and by pumping the chromosomal DNA across the septum.

ATPases related to FtsK from Gram-positive bacteria and actinomycetes, with multiple ATPase domains, have been proposed to function as pumps for the extrusion of small polypeptides of the ESAT-6 superfamily (17). Sequence analysis has shown that FtsK belongs to a family of P-loop ATPases which also includes two proteins of the type IV secretion systems (T4SS), VirB4 and VirD4, and the TrwB-like proteins involved in the conjugal transfer of plasmids (18,19). The VirD4-like ATPases of agrobacteria and other conjugative plasmids are required for the coupling of plasmid DNA processing by the relaxosome to the mating bridges (20). VirB4 is involved in the transfer of agrobacterial T-DNA into the plant hosts (21,22). VirD4 and VirB4 proteins of other T4SS have been implicated in the extrusion of protein virulence factors or cell surface structures in an ATP-dependent manner (23,24).

The solution of the crystal structure of the TrwB protein from the conjugative plasmid R388 revealed that these proteins form a hexameric ring, which is similar to the tertiary structures of a number of other P-loop ATPases, such as those of the AAA^+ and the RecA/DnaB-like classes (25,26). A general model for the functioning of these ring ATPases has been suggested whereby the substrate (e.g. DNA) is threaded through the central pore of the ring and the ATPase activity facilitates pumping of the substrate (26,27) and/or (dis)assembly of other symmetric structures on the face of the ATPase ring.

Recently, we and others have shown that all archaea and some bacteria encode a highly conserved homolog of FtsK, TrwB and VirB4/VirD4 named HerA (the name used hereinafter) (18) or MlaA (28). The HerA protein from *Sulfolobus acidocaldarius* has been shown to have a bi-directional (3′–5′ and 5′–3′) DNA helicase activity (18). Electron microscopic studies have shown that MlaA from *Pyrococcus furiosus* forms hexameric rings similar to those formed by TrwB (28). The archaeal HerA proteins define a new family of FtsK-related ATPases, which includes additional divergent paralogs of HerA encoded in most archaeal genomes, as well as homologs from several phylogenetically distinct bacterial lineages (18). Examination of gene neighborhoods of the *herA* gene in archaeal genomes revealed strict co-occurrence in the same predicted operon with the *nurA* gene, which encodes an archaeal 5′→3′ nuclease (29). In addition, the *herA* and *nurA* genes often co-localize with the genes coding for components of the highly conserved DNA repair complex comprised of the archaeal orthologs of the eukaryotic Mre11 (a nuclease of the calcineurin-like phosphoesterase fold) and Rad50 (a P-loop ATPase of the ABC class) (18,28). Given that conserved gene neighborhoods in prokaryotic genomes are strong predictors of functional and physical interactions (30–33), it seems likely that these four proteins interact to form a DNA processing complex involved in DNA repair, replication and/or segregation during cell division.

While both prokaryotic superkingdoms, bacteria and archaea, share certain similarities in their cell division (34), little is known of the chromosome segregation process in archaea. The strict conservation of HerA in the archaea with sequenced genomes parallels the nearly ubiquitous presence of FtsK in bacteria, suggesting that HerA might have a biological role similar to that of FtsK in maintaining genome integrity and facilitating chromosomal separation during cell division. Furthermore, given the bacterial–archaeal split in the distribution of the HerA/FtsK ATPases, reconstruction of their evolutionary history might shed light on the origins of cell division and associated DNA processing, and the nature of these processes in the last universal common ancestor (LUCA) of cellular life forms. With this objective, we performed a detailed computational sequence analysis of the FtsK/HerA-related proteins, identified novel members and explored their evolutionary relationships as well as their relationships with other P-loop ATPases. In addition, we employed a comparative genomic approach to extract contextual information for the HerA/FtsK superfamily, which led to the identification of previously unrecognized probable functional partners of these ATPases, including nucleases and transmembrane proteins. These leads allow us to predict the structure of the chromosome separation and cell division apparatus of the archaea and several bacteria that lack FtsK. We propose that HerA and FtsK, along with several families of ATPases encoded by plasmids, conjugative transposons and viruses, constitute a superfamily of ATPases descending from an ancestral DNA-pumping enzyme of LUCA. Members of this superfamily appear to have been repeatedly used as ATP-dependent pumps, for the partitioning of DNA into the daughter cells during division, extruding proteins into the extracellular space and packaging DNA into viral capsids.

MATERIALS AND METHODS

The non-redundant (NR) database of protein sequences (National Center for Biotechnology Information, NIH, Bethesda, MD) was searched using the BLASTP program. Iterative database searches were conducted using the PSI-BLAST program (35) with either a single sequence or an alignment used as the query, with the position-specific scoring matrices (PSSM) inclusion expectation (*E*) value threshold of 0.01 (unless specified otherwise); the searches were iterated until convergence. For all searches with compositionally biased proteins, the statistical correction for this bias was employed. Multiple alignments were constructed using the

T_Coffee (36) or PCMA (37) programs, followed by manual correction based on the PSI-BLAST results. All large-scale sequence analysis procedures were carried out using the SEALS package (38). Transmembrane regions were predicted in individual proteins using the TMPRED, TMHMM2.0 and TOPRED1.0 programs with default parameters (39). For TOPRED1.0, the organism parameter was set to 'prokaryote' or 'eukaryote' depending on the source of the protein.

Protein structure manipulations were performed using the Swiss-PDB viewer program (40) and the ribbon diagrams were constructed using the MOLSCRIPT program (41). Protein secondary structure was predicted using a multiple alignment as the input for the PHD program (42). Similarity-based clustering of proteins was carried out using the BLASTCLUST program (ftp://ftp.ncbi.nih.gov/blast/documents/README.bcl).

Phylogenetic analysis was carried out using the maximum likelihood, neighbor-joining and minimum evolution (least squares) methods (43–45) Gene neighborhoods were determined by searching the NCBI PTT tables with a custom-written script. These tables can be accessed from the genomes division of the Entrez retrieval system.

RESULTS AND DISCUSSION

The FtsK–HerA superfamily

In order to identify all members of the FtsK–HerA superfamily, we performed PSI-BLAST searches (35) of the NR database with PSSMs for the bacterial FtsK orthologs and archaeal HerA orthologs. These searches were run with *E*-value thresholds in the range from 10^{-4} to 10^{-7} to avoid inclusion of P-loop ATPases of other families with highly conserved Walker A and B motifs into the PSSM. As a result of these controlled searches, we collected a divergent set of FtsK–HerA homologs which contain several specific sequence motifs defining this superfamily, to the exclusion of other groups of P-loop NTPases. Reciprocal searches with newly detected members of the superfamily employed as queries were carried out to eliminate false positives. Only those newly detected sequences that reciprocally recovered other HerA/FtsK-related proteins as the best hits and/or contained the conserved motifs characteristic of this superfamily were included in the PSSMs for subsequent iterations. Exhaustive, transitive searches with the newly detected superfamily members were also employed to identify more divergent homologs.

The searches initiated with the FtsK and HerA PSSMs readily detected each other as the best hits (*E*-value of 10^{-6}–10^{-8} at the point of first recovery in iterations 2–3) and also recovered the VirB4-, TrwB- and VirD4-like proteins (10^{-6}–10^{-9} in iterations 3–5), and numerous uncharacterized proteins from diverse bacteria. Interestingly, these searches also recovered, with *E*-values in the range of 10^{-4}–10^{-5}, the packaging ATPases of a variety of DNA viruses, including double-stranded DNA bacteriophages such as P9, large nucleocytoplasmic DNA viruses (NCLDV) (typified by the vaccinia virus A32 protein) and single-stranded DNA phages, such as F1 and M13 (gP1). All these proteins contain a unique set of conserved residues found only in the bona fide HerA and FtsK homologs, and none of them showed specific affinities to any other previously characterized group of NTPases (see below). Reciprocal searches with most viral proteins produced poor results due to their extreme sequence divergence. However, the A32 homologs from frog iridoviruses recovered the P9 packaging ATPase in iteration 2 (*E*-value = 10^{-6}) and the HerA proteins from the third iteration onwards. Reciprocal searches with the profiles for single-stranded DNA bacteriophages detected the HerA family as the best hits in the third iteration with *E*-value $\sim 10^{-2}$. Additionally, these searches recovered the *Zonula occludens* toxins (ZOT) from γ-proteobacteria, such as *Vibrio* and *Pseudomonas*, suggesting that these proteins are derivatives of the phage packaging enzymes. With the sole exception of a small orthologous set of HerA-like proteins from filamentous ascomycete fungi, none of these searches recovered any eukaryotic cellular members. Taken together, the results of these searches suggest that HerA/FtsK homologs form a large superfamily, which is distinct from all previously described groups of P-loop NTPases.

The sequence and structural signatures of the FtsK–HerA superfamily and relationships with other P-loop ATPases

Using the Gibbs sampling procedure (46), we identified three highly significant conserved motifs in the FtsK–HerA superfamily proteins; these motifs served as anchors for constructing a complete multiple alignment of the entire superfamily. The alignment was further refined by taking into account the secondary structure elements derived from the crystal structure of TrwB (PDB : 1e9r) (25). Superposition of the TrwB structure over the multiple alignment showed that the conserved core of the FtsK–HerA superfamily is a seven-stranded β-sheet with a 7615423 topology. The last strand in the sheet is antiparallel to the rest of the strands (Figure 1). The first conserved block includes the Walker A motif and encompasses the first strand, the P-loop and the following helix. In most FtsK–HerA superfamily members, the P-loop has the canonical form (GX4GK[TS]), but some of the phage packaging enzymes and the ZOT deviate from this pattern (Figure 2) (47). With the exception of the viral proteins, most members of this superfamily have a conserved histidine at the beginning of the Walker A-associated strand (Figure 2). In TrwB and, by inference, in other FtsK–HerA superfamily proteins, this histidine packs against a conserved hydrophobic residue at the C-terminus of the helix located immediately downstream of the Walker B motif (Figure 2). The Walker B motif defines the second conserved block, which has the consensus sequence hhhh[DE]E (where h is any hydrophobic residue). The first acidic residue, which is conserved in all P-loop NTPases, coordinates the Mg^{2+} cation involved in NTP hydrolysis; the second acidic residue that is present only in a subset of the P-loop enzymes primes a water molecule for the nucleophilic attack on the gamma-phosphate. The third conserved motif includes strand 4, which contains a polar residue, most often glutamine, at the C-terminus and the distinct helix with a highly conserved arginine that precedes this strand (Figures 1 and 2). The conservation pattern associated with this motif helps in distinguishing the FtsK–HerA superfamily from all other P-loop ATPase groups. In the three-dimensional (3D) structure of TrwB, strand 4 is positioned in between the Walker A and Walker B strands (Figure 1) within the core β-sheet. The C-terminal polar residue of strand 4 is structurally equivalent to the polar residue in the so-called sensor-1 motif

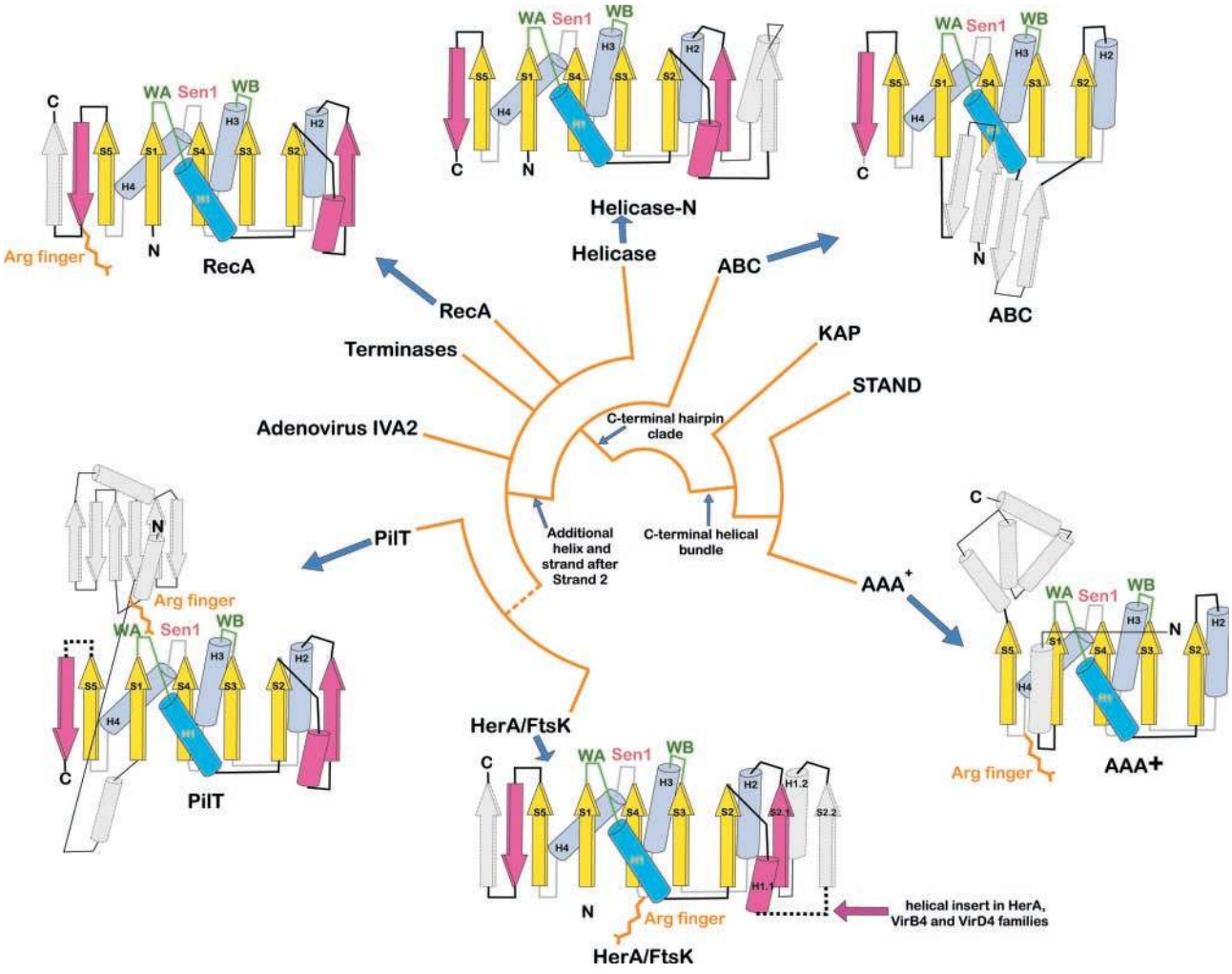


Figure 1. Topology diagram of the ASCE ATPases showing the putative higher order relationships of the FtsK–HerA superfamily. Strands are shown as arrows with the arrowhead at the C-terminus, helices are shown as cylinders. Strands and helices conserved across the ASCE group are numbered, and colored yellow and blue, respectively. The C-terminal β-hairpin synapomorphic to the RecA–ABC clade and the helix-strand unit synapomorphic to the RecA clade are colored pink. This hairpin is secondarily lost in most helicases. STAND is a large clade of NTPases that include the previously described AP-ATPases and NACHT NTPases, as well as several uncharacterized ATPase lineages predicted to participate in signal transduction (D. D. Leipe, Eugene V. Koonin and L. Aravind, unpublished data). Non-conserved secondary structural elements are colored white. Abbreviations: WA, Walker A; WB, Walker B and Sen1, sensor-1. The dotted connecting lines in the topology diagrams represent regions of the protein where insertions are observed. Broken lines in the cladogram reflect an uncertainty in relationship of the members within the clade supported by the broken line.

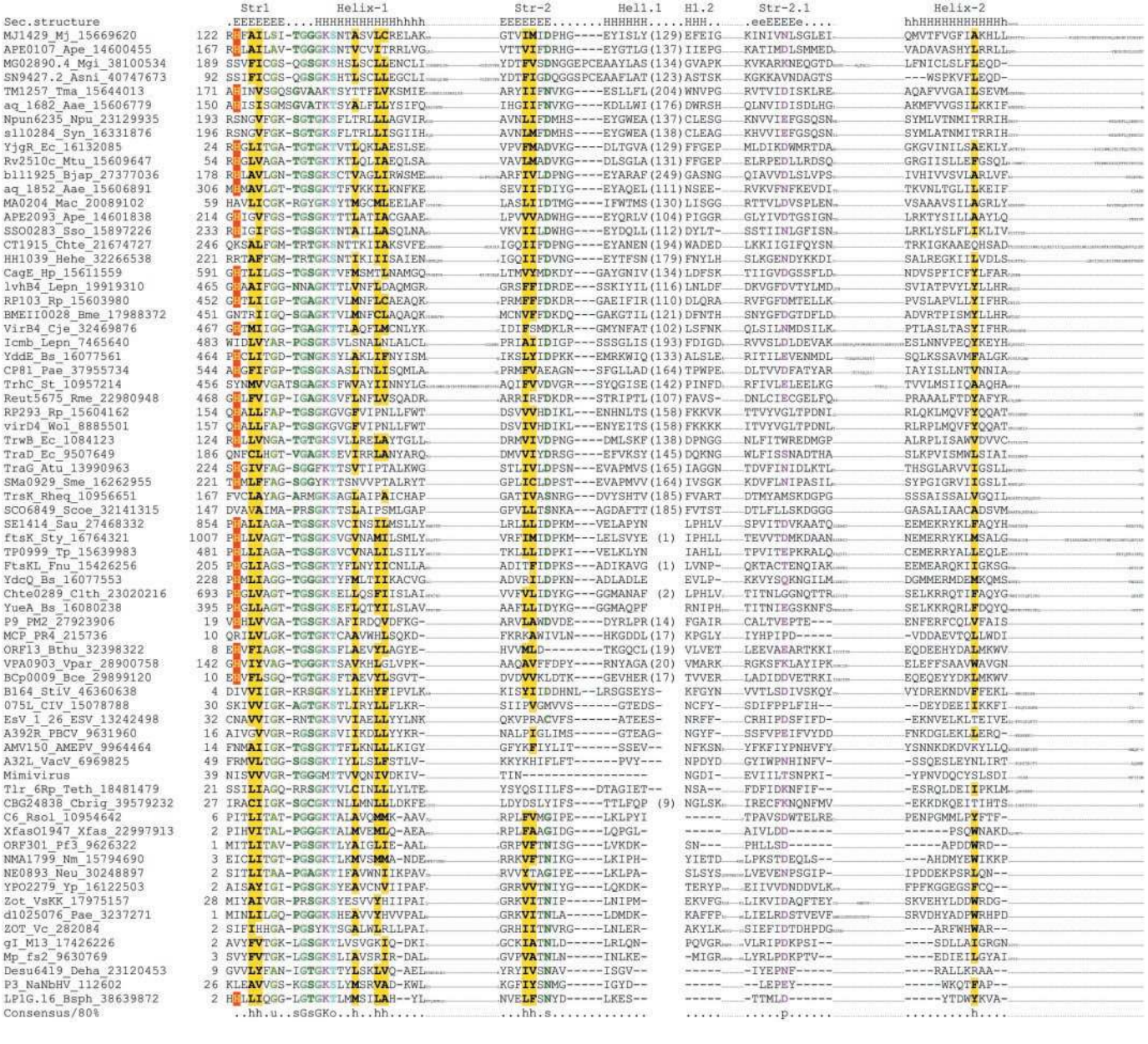
of the AAA^+ ATPases and the corresponding motif III of the SFI and SFII helicases (48–50). This residue appears to be required for sensing the triphosphate moiety of the bound nucleotide to trigger its hydrolysis.

Examination of the crystal structure of TrwB shows that the highly conserved arginine in the short helix upstream of strand 4 projects into the ATP-binding active site of the preceding protomer in the hexameric ring (25). Thus, this arginine is analogous to the arginine finger of the AAA^+ superfamily, which is located at the C-terminus of the helix following the sensor-1 strand of the AAA^+ ATPase domain (48,49,51). As in the AAA^+ ATPases, this conserved residue of the FtsK–HerA superfamily is likely to function as an arginine finger promoting inter-protomer cooperation in ATP-hydrolysis by binding the terminal phosphate of the substrate. Analogous arginine fingers supplied by adjacent subunits are implicated in cooperative ATP hydrolysis by ring ATPases of the RecA-like class, such as RecA, DnaB and ATP synthase (52–54). However, in this case, the arginine is located at the C-terminus of the ATPase domain, on a β-hairpin unique to the RecA-like ATPases (55) (Figure 1). Thus, the presence of a conserved arginine in P-loop NTPases is a good predictor of ring formation and inter-protomer cooperation. By this criterion, all FtsK–HerA superfamily ATPases are predicted to form hexameric rings. By contrast, in the PilT/VirB11 family, the arginine finger is supplied by a distinct domain, which is located in the same polypeptide, N-terminal of the ATPase domain (Figure 1). Thus, these proteins resemble the GTPases, in which the arginine finger is supplied by an external GTPase-activating protein (56–59).

The presence of the additional catalytic glutamate after the conserved acidic residue in the Walker B motif and an intervening strand between Walker A- and B-associated strands place the FtsK–HerA superfamily into the additional strand conserved E (ASCE) division of P-loop NTPases, which also includes the RecA-like, SFI/II helicase, ABC, AAA^+,

PilT/VirB11, KAP and STAND (a large clade of NTPases that include the previously described AP-ATPases and NACHT NTPases, as well as several uncharacterized ATPase lineages predicted to participate in signal transduction; D. D. Leipe, Eugene V. Koonin and L. Aravind, unpublished data) clades (49,60,61). Structural comparisons show that the FtsK–HerA superfamily shares a C-terminal hairpin, formed due to the presence of a terminal anti-parallel strand, with many other NTPases of the ASCE division, namely, the ABC, RecA-like and SFI/II helicase classes, and the PilT/VirB11 superfamily (Figure 1). Furthermore, the latter three groups have an additional common feature with the FtsK–HerA superfamily, namely, an additional strand 'to the right' of the core P-loop fold (Figure 1). Clustering based on DALI *Z*-scores (62) suggests grouping of the PilT/VirB11 and FtsK–HerA superfamilies into a single, higher order cluster. However, no definitive sequence or structural synapomorphies unifying these two groups were detected. Additional 3D structures from diverse representatives of each of these groups should help in assessing the validity of this higher order clustering.

The FtsK–HerA and the PilT/VirB11 superfamilies are both traceable to LUCA (see below) and appear to have diverged from each other at an even earlier stage of evolution. Additionally, two other superfamilies of the ASCE division, the adenoviral packaging ATPases and the terminases of diverse DNA viruses, such as the herpes viruses and Mu-like bacteriophages (63), also have an additional strand to the 'right' of the core P-loop domain, suggesting an evolutionary relationship with the FtsK–HerA and PilT/VirB11 superfamilies. However, the conserved sequence motifs characteristic of any of the latter superfamilies are not detectable in the viral ATPases.



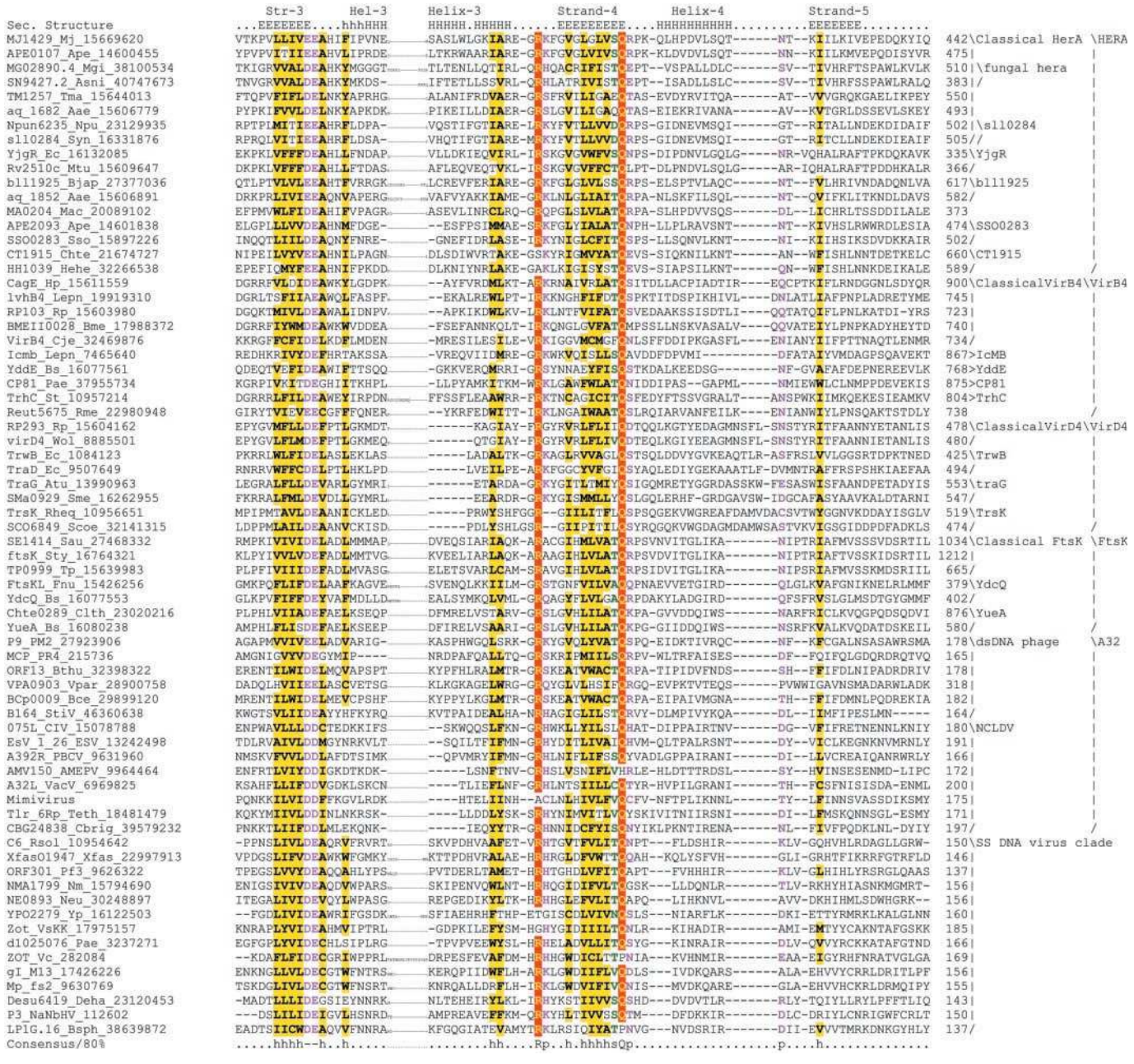


Figure 2. Multiple alignment of the FtsK–HerA superfamily. Proteins are denoted by their gene names, species abbreviations and gi numbers, separated by underscores. Amino acid residues are colored according to their side-chain properties and conservation in the multiple alignment. The coloring reflects 80% consensus and is shown underneath the alignment. The secondary structure shown above the alignment, is derived from the crystal structure of TrwB and secondary structure prediction programs. E and H represent a strand and helix, respectively. The consensus abbreviations and coloring scheme are as follows: h, hydrophobic residues (ACFILMVWY) shaded yellow; s, small residues (AGSVCDN) and u, tiny residues (GAS) colored green; o, alcohol group containing residues (ST) colored blue; p, polar residues (STEDKRNQHC) −, acidic residues (DE) and +, basic residues (HRK) colored purple. The conserved histidine in the Walker A strand, the arginine finger and the glutamine in sensor-1 are shaded red. Secondary structure elements that are conserved across the ASCE fold are numbered as integers. Species abbreviations are as follows: Aae, *A.aeolicus*; AMEPV, *Amsacta moorei* entomopoxvirus; Ape, *Aeropyrum pernix*; Asni, *Aspergillus nidulans*; Atu, *Agrobacterium tumefaciens*; Bce, *Bacillus cereus*; Bjap, *Bradyrhizobium japonicum*; Bme, *Brucella melitensis*; Bs, *B.subtilis*; Bsph, *Bacillus sphaericus*; Bthu, *B.thuringiensis*; CIV, Chilo iridescent virus; Cbrig, *Caenorhabditis briggsae*; Chte, *C.tepidum*; Cje, *Campylobacter jejuni*; Clth, *C.thermocellum*; Deha, *Desulfitobacterium hafniense*; ESV, *Ectocarpus siliculosus* virus; Ec, *E.coli*; Ec, Plasmid R100; Fnu, *Fusobacterium nucleatum*; fs2, *V.cholerae* filamentous bacteriophage fs-2; Hehe, *H.hepaticus*; Hp, *Helicobacter pylori*; Lepn, *Legionella pneumophila*; M13, Enterobacteria phage M13; Mac, *Methanosarcina acetivorans*; Mgi, *Magnaporthe grisea*; Mj, *Methanococcus jannaschii*; Mtu, *Mycobacterium tuberculosis*; NaNbHV, non-A, non-B hepatitis-associated virus; Neu, *Nitrosomonas europaea*; Nm, *Neisseria meningitidis*; Npu, *Nostoc punctiforme*; PBCV, *Paramecium bursaria* Chlorella virus 1; PM2, Alteromonas phage PM2; PR4, Bacteriophage PR4; Pae, *Pseudomonas aeruginosa*; Pf1, Pseudomonas phage Pf1; Pf3, Pseudomonas phage Pf3; Rheq, *Rhodococcus equi*; Rme, *Ralstonia metallidurans*; Rp, *Rickettsia prowazekii*; Rsol, *Ralstonia solanacearum*; Rsph, *Rhodobacter sphaeroides*; Scoe, *Streptomyces coelicolor*; Sep, *Staphylococcus epidermidis*; Sau, *Staphylococcus aureus*; Sme, *Sinorhizobium meliloti*; StiV, Sulfolobus turreted icosahedral virus; Sso, *Sulfolobus solfataricus*; St, *Salmonella typhi*, Sty, *Salmonella typhimurium*; Syn, *Synechocystis* sp.; Tel, *Thermosynechococcus elongatus*; Teth, *Tetrahymena thermophila*; Tma, *Thermotoga maritima*; Tp, *Treponema pallidum*; VacV, Vaccinia virus; Vc, *V.cholerae*; Vpar, *Vibrio parahaemolyticus*; VsKK, Bacteriophage VSKK; Vvul, *Vibrio vulnificus*; Wol, Wolbachia sp.; Xfas, *Xylella fastidiosa* and Yp, *Yersinia pestis*.

Evolutionary classification and phyletic spread of the FtsK–HerA superfamily

We analyzed the relationships between the members of the FtsK–HerA superfamily using a combination of approaches. To identify the major clades within the superfamily, the multiple alignment was examined for distinct sequence signatures characteristic of subsets of the superfamily members. Clustering by sequence similarity using the BLASTCLUST program was employed to identify subgroups and orthologous lineages. Finally, at the level of high sequence similarity, such as within an orthologous group or a closely related cluster of paralogs, conventional phylogenetic tree analysis using maximum-likelihood, neighbor-joining and minimum evolution methods was performed to decipher the evolutionary history of each such group. As a result of this analysis, we identified six major clades within the FtsK–HerA superfamily: (i) HerA, (ii) VirB4, (iii) VirD4/TrwB, (iv) FtsK, (v) ssDNA phage packaging ATPases and (vi) A32-like dsDNA viral packaging ATPases. Table 1 shows the classification of the FtsK–HerA superfamily.

The HerA clade is defined by several synapomorphies including a small residue (typically, glycine) after strand 2, a hydrophobic residue in the α-helix after strand 2, an aspartate in the α-helix immediately after strand 5. This family also contains a large helical insert immediately after the second strand-helix unit of the P-loop domain. The HerA family proper, which includes the experimentally characterized HerA protein of *Sulfolobus*, consists of a core orthologous group of archaeal proteins that are encoded in a conserved operon with the Mre11 and Rad50 orthologs, many additional, more diverged paralogs from each archaeal genome and numerous homologs scattered over a wide range of bacteria (Table 1). The simplest interpretation of this phyletic pattern is that the HerA family emerged in archaea, followed by horizontal gene transfers (HGTs) to and between bacteria.

Most HerA family members are encoded in a conserved operon with a gene for a NurA nuclease (18,28,29). The HerA family proteins contain a distinct N-terminal β-barrel domain which is homologous to the N-terminal domain of F1/F0 ATP synthases and is fused to the N-terminus of the P-loop domain (Figures 1 and 3); we named this domain the HAS-barrel (HerA-ATP Synthase barrel). The HAS-barrel is likely to form an independently folding toroidal structure stacked on one surface of the central ring formed by the P-loop domain of HerA. The presence of several shared residues between the HAS barrels of ATP synthases and those of the HerA family (Figure 3), and an analogous location at the N-terminus of the P-loop, suggest that these domains have similar functions. In ATP synthases, this domain is implicated in the assembly of the catalytic toroid and docking of accessory subunits, such as the δ subunit of the ATP synthase complex (64). Similar roles in docking of the functional partner, the NurA nuclease, and assembly of the HerA toroid complex appear likely for the HAS-barrel of the HerA family.

The HerA clade also includes several additional families with substantial differences in domain/operon organizations and phyletic patterns (Figures 4 and 5, Table 1). For example, a distinct family of HerA homologs, found primarily in proteobacteria (typified by bll1925 from *Bradyrhizobium*), has a specific form of the HAS barrel only weakly similar to that of the HerA family. The CT1915 family includes a divergent group of proteins with a distinct N-terminal domain that appears to be unrelated to any previously characterized domains. This family has an unusual phyletic pattern, with representatives from *Chlorobium tepidum*, *Helicobacter hepaticus*, *Chloroflexus aurantiacus* and *Methanosarcina mazei,* suggesting a high degree of lateral mobility.

The prototype of the VirB4 clade is the VirB4 ATPase which is a component of the T4SS in numerous bacteria (65,66).This clade is unified by several distinctive sequence signatures, which include conserved patterns in the α-helical insert located after the second β/α unit of the conserved core of the P-loop domain (Table 1). Most VirB4-type ATPases also have a long N-terminal extension that is less conserved than the ATPase portion but is likely to form a distinct globular domain. This large globular domain probably mediates interactions with other components of T4SS or the conjugative apparatus of transposons and plasmids. There are several distinct families within this clade, the best studied of which is the classical VirB4 family that is encoded by the mobile T4SS gene clusters of diverse proteobacteria. Other families are encoded by conjugative plasmids and conjugative transposons [e.g. Tn916 of Gram-positive bacteria (67)] from various bacterial taxa (Figure 4 and Table 1). Consistent with this, the genes coding for VirB4-type ATPases show little evidence of vertical inheritance across genomes and appear to have been disseminated widely by these mobile elements.

The VirD4 clade is typified by the T4SS component VirD4 which has a large insert in the ATPase domain in the same position as the HerA and VirB4 clades. This insert shows several unique sequence motifs characteristic of this clade. Additionally, most members of this family contain a small membrane-spanning domain N-terminal of the ATPase domain; this domain probably functions as a membrane anchor. In addition to the family typified by VirD4 proper, there are several smaller families within this clade, which function as DNA pumps of diverse conjugative plasmids from various bacterial lineages (19) (Figure 4; Table 1). Not unexpectedly, the evolutionary pattern of this clade seems to mirror that of the VirB4 clade.

The FtsK clade consists of proteins that have no inserts in the ATPase domain, unlike the HerA, VirB4 and VirD4 clades. Most of the FtsK-like proteins contain N-terminal membrane-spanning segments that probably function as anchors. The main family in this clade, the FtsK proteins proper, is represented by conserved orthologous ATPases involved in cell division in the great majority of bacteria. However, in spite of its essential function, FtsK is missing in several bacterial lineages, such as *Thermotoga*, *Aquifex*, *Chloroflexus* and cyanobacteria. Phylogenetic analysis of the FtsK family suggests a predominantly vertical pattern of inheritance, with the tree topology resembling those of other proteins with an apparent dominant vertical component (e.g., ribosomal proteins and RNA polymerase subunits) (data not shown; Supplementary Material). Certain plasmids and phages, especially those from actinomycetes and Gram-positive bacteria, encode divergent variants of FtsK, which probably function in *cis* as DNA pumps for transmission of the respective plasmids during cell division or packaging of the phage DNA. Additionally, this clade includes a few smaller distinct families which consist, principally, of proteins encoded in conjugative transposons from various bacteria (Figure 4; Table 1). One notable

Table 1. Classification of the FtsK–HerA superfamily

Post-strand 2 α-helical insert superclade

Large α helical domain inserted after strand 2

HerA clade

A small residue after strand 2; a hydrophobic residue in the helix after strand 2; a conserved aspartate in the helix immediately after strand 5.

- Classical Hera family (A > FF, Cya, Cau, Aq, Thth, Dr, Tma, Bha, Chth, Lme, Ruxy, Smu): Fused to a HAS barrel at the N-terminus; several members are in the neighborhood of NurA; fungi lack the HAS barrel
- bll1925 family (Pr, Fnu, Efae): Fused to a divergent HAS barrel at the N-terminus; several members are in the neighborhood of a Sir2-like predicted nuclease
- YjgR family (Pr, Spy, Chut, Act, Trde): Fused to a distinct α-helical domain at the C-terminus that has a conserved glutamate; glycine after conserved histidine before strand 1; PEhGD motif before Walker B strand; histidine in strand 5
- SSO0283 family (Sul, Ape, Pyae): Distinct N-terminal domain with two membrane-spanning helices
- CT1915 family (Chte, Hehe, Cau, Mma): Fused to a distinct N-terminal domain; Chte, Hehe and Mma are operonic with a methylase
- Ta1216 family (Tac, Sul): pNOB8 like conjugative plasmids

VirB4 clade

Distinct N-terminal domain with a conserved proline before the core ATPase domain; hxPh (x: any, h: hydrophobic) motif in post-strand 2 α-helical insert; alcohol or small residue prior to glutamine of sensor I strand

- Classical VirB4 family (Pr): Asparagine in strand 5; associated with type IV secretory system in proteobacteria
- CTnDOT-TraG family (Bdes): Present in conjugative transposons of Bacteroides and Porphyromonas
- CP81 family (Pr): EAG motif in strand 2; aspartate before helix 4; acidic residue after strand 5
- YddE family (LGC Gm +ve): PxxE motif after strand 2; encoded by Tn916-like conjugative transposons.
- TrhC family (Pr): aN (a: aromatic) motif before strand 1; lE (l: aliphatic) residue in helix 3; encoded by antibiotic resistant plasmids like R-27s
- R64-TraU family (Pr): Sporadically present in proteobacterial R64-like plasmids and genomes
- all8046 family (Cya): Distinct N-terminal domain with membrane-spanning helices that can exist as a solo domain; bN (b: big) motif before strand 1; DGosT (o: alcohol, s: small) motif after strand 2; hhGRI (h: hydrophobic) motif in strand 5; present on cyanobacterial plasmids
- TN5252 ORF26 family (LGC Gm +ve, Blon): Mostly present in conjugative plasmids and transposons

VirD4 clade

Proline in post-strand 2 helical insert; aromatic residue in post-strand 2 α-helical insert; glutamine N-terminal to the helix following strand 4; serine after strand 5; most members also have two to four N-terminal membrane-spanning domains

- Classical VirD4 family (Pr): Associated with type IV secretory system in proteobacteria
- CtnDOT-MobC (Bdes): Present in conjugative transposons of Bacteroides and Porphyromonas
- TrwB family (Pr, Pchm): RxW motif in strand 3; mainly present on proteobacterial plasmids
- TraG family (α Pr): Present on several plasmids of α proteobacteria, often in the vicinity of TraA
- TrsK family (Act): Serine in place of conserved aspartate in strand 2; conserved asparagine after strand 2; asparagine after Walker B strand; GS motif before strand 4; present on actinobacterial plasmids
- alr7539 family (Cya): Aspartate between helix 1 and strand 2; many members are in the neighborhood of a TrwC-like RCR superfamily nuclease; present on cyanobacterial plasmids
- TN5252 ORF21 family (LGC Gm +ve, Cgl, Smal): Mostly present in conjugative plasmids and transposons

Clades without insert after strand 2

FtsK clade

N-terminal membrane associated region; a variable coiled coil central region; C-terminal ATPase; glycine before core ATPase domain; Kh (h: hydrophobic) motif after strand 2; hxxR motif in helix 2; arginine or lysine at the beginning of strand 5; glycine after core ATPase domain

- Classical FtsK family (most bacteria): Histidine in helix before strand 2.1 and strand 4; RlsQ (l: aliphatic, s: small) motif in helix 3; DSR motif at the beginning of strand 6; SxhQR motif after core ATPase domain
- YueA family (LGC Gm +ve, Act, Cau): Three tandem ATPase domains often found in the vicinity of ESAT-6; some YueAs in Clostridium are fused to one to two FHA domains at their N-terminus; GGs (s: small) motif after strand 2 in first ATPase; WLPPL motif between the first two ATPase domains
- YdcQ family (LGC Gm +ve, Fnu): Alanine in Walker B strand; RD motif at the end of helix 4; present on Tn916-like conjugative transposons

A32 clade

- ds DNA phage packing ATPase family (V, prophages)
- A32 ATPase family (V, Nem, Teth): Aspartate in Walker B strand and in the C-terminal strand after the core ATPase domain
  - —Tlr transposon subfamily (Nem, Teth); transposons in Tetrahymena are often associated with a TrwC/TraA-like superfamily I helicase (Tlr 8Rp)
  - —A32 ATPases of NCLDV viruses

SS DNA virus clade (V)

Asparagine or glycine at the end of strand 2; C-terminal transmembrane region

A, Archaea; Act, actinobacteria; Ape, *Aeropyrum*; Aq, *Aquifex*; Bdes, Bacteroides; Bha, *Bacillus halodurans*; Blon, *Bifidobacterium longum*; Cau, *Chloroflexus*; Cgl, *Corynebacterium glutamicum*; Chte, *Chlorobium*; Chth, *Clostridium thermocellum*; Chut, *Cytophaga*; Cobu, *Coxiella*; Cya, cyanobacteria; Dr, *Deinococcus*; Efae, *Enterococcus*; FF, filamentous fungi; Fnu, *Fusobacterium*; Hehe, *Helicobacter hepaticus*; Lepn, *Legionella pneumophila*; LGC Gm +ve, low GC Gram positive bacteria; Lme, *Leuconostoc mesenteroides*; Nem, Nematodes; Pchm, Parachlyamydia; Teth, *Tetrahymena thermophila*; Mma, *Methanosarcina mazei*; Pr, proteobacteria; Pyae, *Pyrobaculum*; Ruxy, *Rubrobacter*; Smal, *Stenotrophomonas maltophilia*; Smu, *Streptococcus mutans*; Spy, *Streptococcus pyogenes*; Sul, *Sulfolobus*; Tac, *Thermoplasma*; Thth, *Thermus*; Tma, *Thermotoga*; Trde, *Treponema denticola*; V, viruses; > indicates lateral transfer.

family of the FtsK clade, typified by the YueA protein, is restricted to Gram-positive bacteria and actinomycetes and includes proteins with three tandem ATPase domains in the same polypeptide. These proteins are likely to dimerize and form toroidal structures with a total of 6 ATPase domains. The YueA-like proteins are implicated in the secretion of the unique extracellular peptides of Gram-positive bacteria and actinomycetes (17).

Packaging ATPases of single-stranded DNA bacteriophages comprise another distinct clade in the FtsK–HerA superfamily. These proteins, which are encoded by gene 1 of filamentous enterobacteriophages (e.g., F1 and M13) consist of an N-terminal, cytoplasmic ATPase domain, followed by a membrane-spanning region and an extracellular domain. Proteins with similar architectures are encoded by a variety of filamentous phages infecting several proteobacteria such as

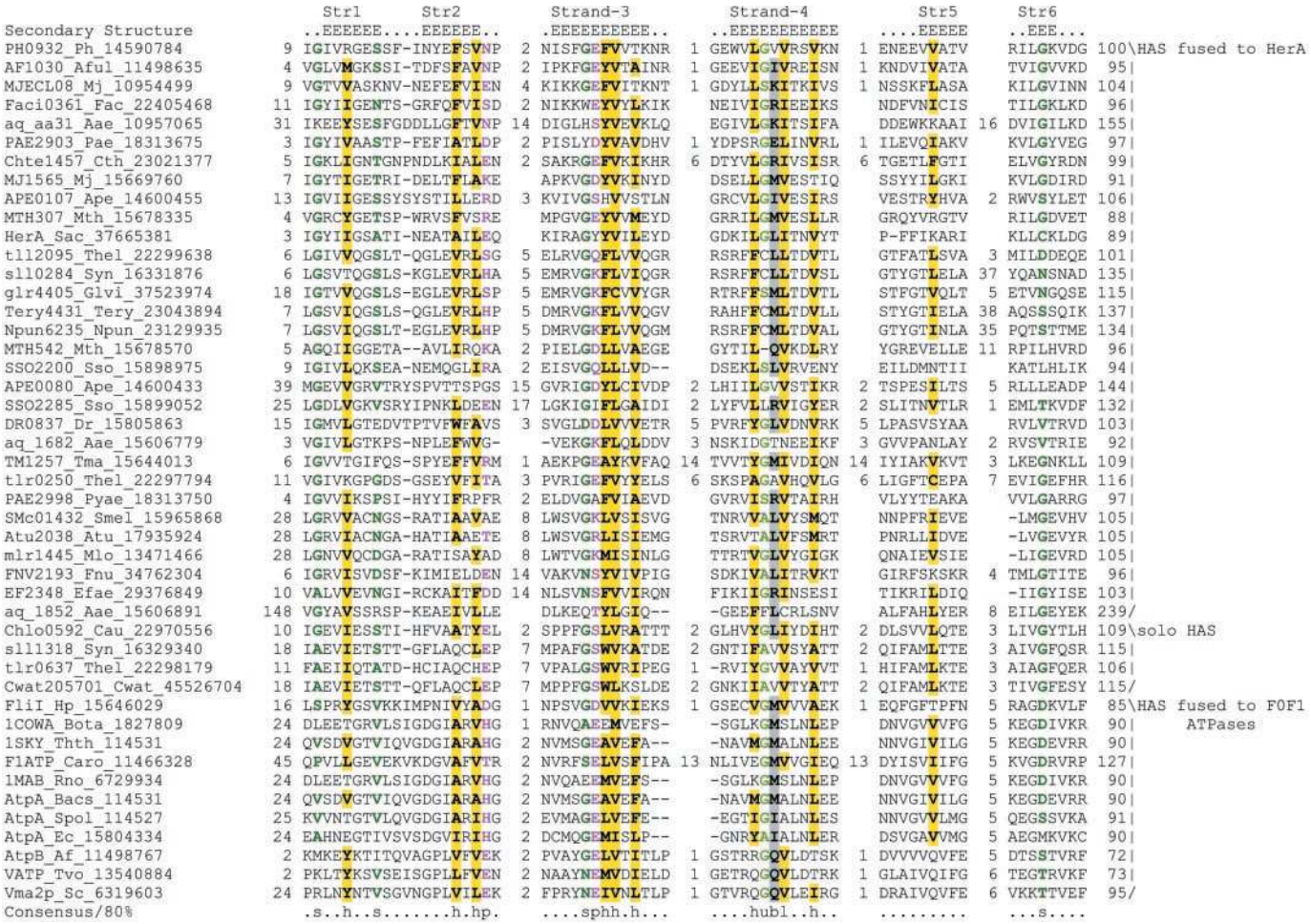


Figure 3. Multiple alignment of the HAS-barrel domain. The coloring reflects 80% consensus. The coloring scheme, consensus abbreviations and secondary structure representations are as in Figure 2. Additionally, big residues (LIYERFQKMW) are shaded gray. Species abbreviations are as follows: Af, *Archaeoglobus fulgidus*; Ape, *A.pernix*; Aae, *A.aeolicus*; Atu, *A.tumefaciens*; Bacs, *Bacillus* species; Bota, *Bos Taurus*; Caro, *Cafeteria roenbergensis*; Cau, *C.aurantiacus*; Cth, *C. thermocellum*; Cwat, *Crocosphaera watsonii*; Dr, *D.radiodurans*; Ec, *E.coli*; Efae, *Enterococcus faecalis*; Fac, *Ferroplasma acidarmanus*; Fnu, *F.nucleatum*; Glvi, *Gloeobacter violaceus*; Hp, *H.pylori*; Mj, *M.jannaschii*; Mlo, *Mesorhizobium loti*; Mth, *Methanothermobacter thermautotrophicus*; Npun, *N.punctiforme*; Ph, *Pyrococcus horikoshii*; Pyae, *Pyrobaculum aerophilum*; Rno, *Rattus norvegicus*; Sac, *S.acidocaldarius*; Sc, *Saccharomyces cerevisiae*; Smel, *S.meliloti*; Spol, *Spinacia oleracea*; Sso, *S.solfataricus*; Syn, *Synechocystis* sp.; Tery, *Trichodesmium erythraeum*; Thel, *T.elongatus*; Thth, *Thermus thermophilus*; Tma, *T.maritima* and Tvo, *Thermoplasma volcanium*.

Vibrio, *Pseudomonas*, *Neisseria*, *Nitrosomonas* and *Ralstonia*, as well as the actinomycete *Propionibacterium*. The ATPase domain does not contain any inserts and seems to correspond to the minimal conserved core of the FtsK–HerA superfamily. The mechanism of these ATPases has not been studied in detail but, by analogy to FtsK and TrwB, it seems likely that they associate with the bacterial membrane and act as ATP-dependent DNA pumps, which load the phage DNA into the capsids. The ZOT of *Vibrio cholerae* is the packaging ATPase of the integrated phage CTXΦ (47,68). ZOT has been shown to associate with the outer membrane through its single transmembrane region. The extracellular portion is cleaved off and binds to intestinal cells triggering a signaling cascade that leads to the disassembly of tight junctions (69). The potential role of the ATPase domain in the localization of the pro-toxin remains to be experimentally investigated.

Packaging ATPases of eukaryotic double-stranded DNA viruses, typified by the vaccinia virus A32R gene product, comprise a distinct clade of the FtsK–HerA superfamily; the similarity between the ATPase domains of these proteins and the ssDNA bacteriophage packaging enzymes has been noticed previously (70). The A32R-like ATPases comprise one of the several orthologous protein sets that unify poxviruses, asfarviruses, iridoviruses and phycodnaviruses into a monophyletic lineage of large NCLDV (71). Subsequently, an orthologous ATPase was also detected in the large mimivirus, an ameba virus (72), which also probably belongs to the NCLDV (Figure 2). Furthermore, homologous proteins are also encoded by the Tlr transposons of the ciliate *Tetrahymena* (73) and a distinct group of nematode transposons, which were discovered as part of this work (Table 1). In the present work, we also found that these proteins are also related to the packaging ATPases of the *Bacillus thuringiensis* phage Bam35c, enterobacteriophage PRD1 and the *Alteromonas* phage PM2. Notably, we found that a predicted ATPase of this family is also encoded in the recently sequenced genome of the

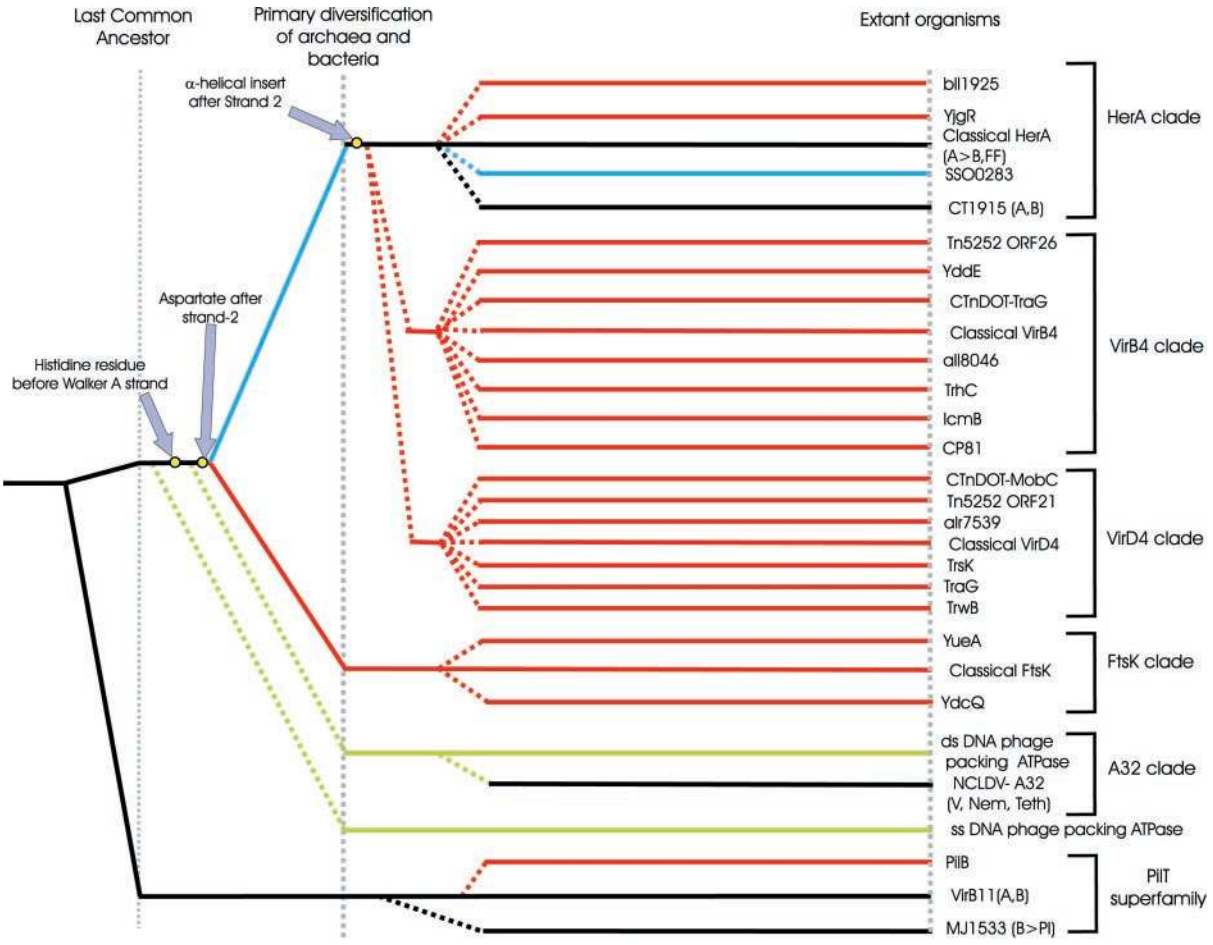


Figure 4. Major lineages of the FtsK–HerA superfamily. The horizontal lines show temporal epochs corresponding to two major transitions in evolution, namely, the LUCA and the divergence between the archaeo-eukaryotic lineage and the bacterial lineage. Solid lines indicate the maximum depth in time to which a particular lineage can be traced. The broken lines indicate uncertainty with respect to the exact point of origin of a lineage. Bacterial lineages are colored in red, archaeal in blue and viral in green. Black lines indicate lineages with representatives from more than one of the three major superkingdoms, bacteria, archaea or eukaryotes. In such mixed lineages the phyletic distribution is shown in brackets with A denoting archaea; B, bacteria; FF, filamentous fungi; Nem, Nematodes; Pl, plants; Teth, *T.thermophila* and > lateral transfer.

turreted icosahedral archaeal virus (74). This observation, taken together with the proposed common origin of the capsid proteins of several distinct DNA viruses (74), favor an early recruitment of these ATPases in viral DNA packaging. However, more recent dissemination of this family via HGT between different viral groups cannot be entirely ruled out either. The finding that viral packaging ATPases comprise a family of the FtsK–HerA superfamily suggests that they catalyze dsDNA pumping into viral capsids similarly to the function of FtsK, TrwB and other members of the superfamily in bacterial and plasmid DNA pumping. The function of the homologous ATPases in eukaryotic transposons is less obvious. They could have been recruited for an alternative function in DNA transposition but, given that these transposons have other uncharacterized ORFs, it cannot be ruled out that they are packaged into virus-like particles that are released from the cells.

Implications of the phyletic patterns and higher order relationships of the FtsK–HerA superfamily. Cladistic-type analysis provides for a reconstruction of the likely evolutionary history of the FtsK–HerA superfamily, even though the level of sequence conservation is insufficient for traditional phylogenetic analysis (Figure 4). The HerA clade is unified into a higher order lineage with the VirB4 and VirD4 clades on the basis of the presence of a shared, predominantly α-helical insert after the second conserved β/α unit of the ATPase domain. These three families, in turn, join the FtsK clade on the basis of several shared sequence features, such as the aspartate at the end of second core strand. The two viral clades, which lack these features, appear to lie outside of this assemblage of predominantly cellular proteins (Figure 4) with the packaging proteins of the double-stranded DNA viruses being closer to the cellular proteins as they share with the latter a conserved histidine at the N-terminus of the Walker A strand; however, it cannot be ruled out that this deep branching of the viral ATPases is an artifact of their extreme divergence.

The HerA clade, which includes a core, pan-archaeal orthologous set, appears to have originated in the common ancestor of the archaea, whereas the FtsK clade similarly can be inferred to have evolved in the ancestral bacterium. The clear-cut archaeo-bacterial complementarity in the distribution of the HerA and FtsK orthologs implies that LUCA encoded

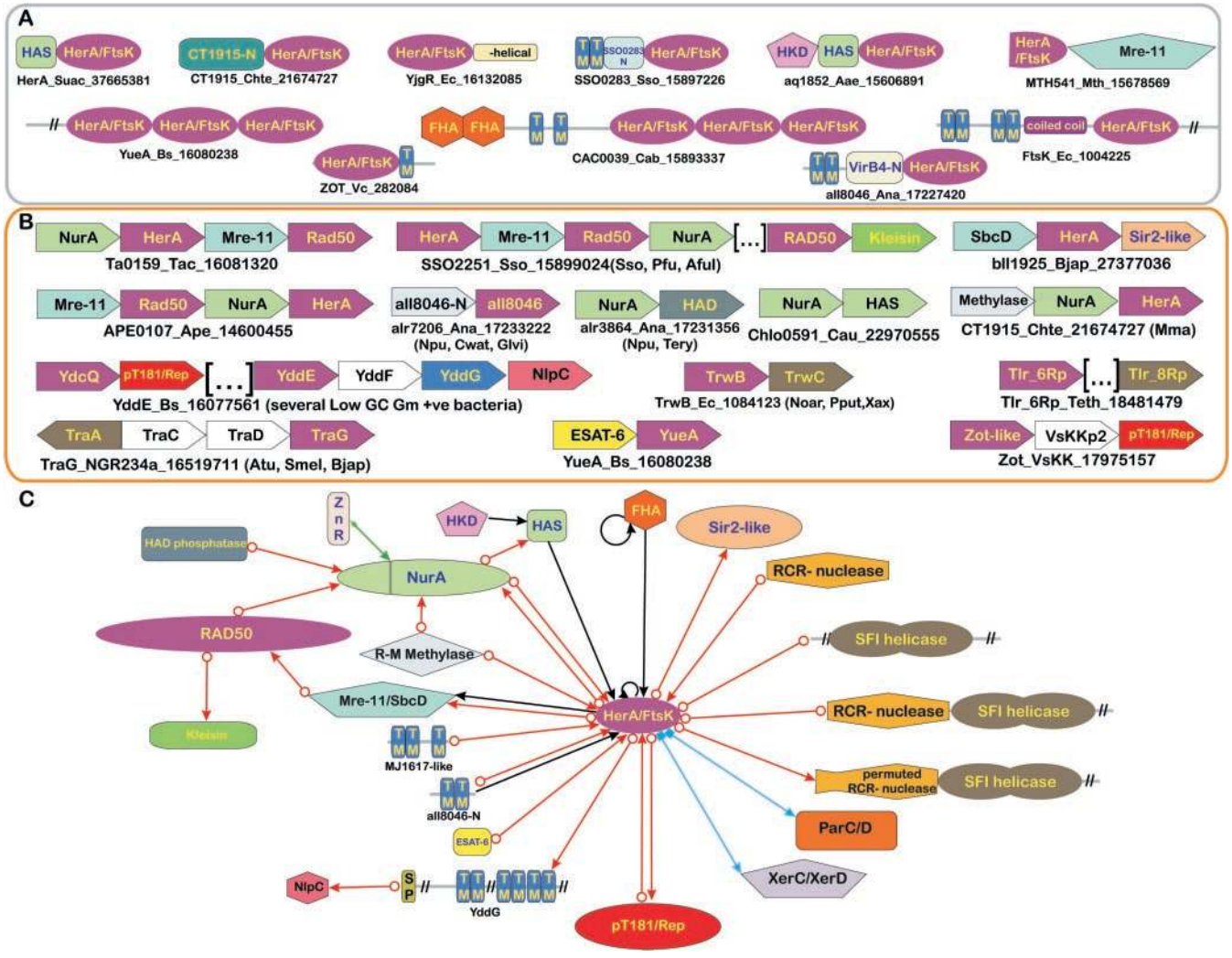


Figure 5. Domain architectures, conserved gene neighborhoods and contextual network graph for the FtsK–HerA superfamily. (**A**) Domain architectures of proteins containing a FtsK–HerA like ATPase. SSO0283-N, CT1915-N and VirB4-N are conserved N-terminal regions found in the SSO0283, CT1915 and VirB4 families respectively. Transmembrane regions are labeled TM. (**B**) Genes that have a conserved neighborhood are shown as boxed arrows. A representative gene, the species in which it is present and its gi number are shown below the boxes. The phyletic distribution of a particular gene context is shown in brackets. Species abbreviations are as in Figure 2. The dotted lines bounded by brackets indicate that the genes bounding the bracket are in the general neighborhood and do not show a close operonic association. Genes that are poorly characterized are repesented as white boxed arrows. (**C**) Contextual network graph for the FtsK–HerA family. Each vertex represents a domain and the edges represent a contextual association. Domain combinations are shown as black arrows, with the arrow pointing from the N-terminus to the C-terminus of the multi-domain protein. Circular arrows indicate multiple copies of the same domain. Operonic and neighborhood associations are shown as red arrows with an O at the tail and the direction of the arrows point from the 5′–3′ direction of the coding sequence. Lines with O at both ends indicate that the genes bounding the line are in not operonic but in close vicinity of each other. The blue arrows with the boxed tails represent experimentally observed functional associations. The green arrow with the feathered tail indicates an insertion of a Zn-ribbon with the arrow head pointing to the location of the insertion in NurA. Additional species abbreviations not in Figure 2. Aful, *A.fulgidus*; Ana, *Anabaena* sp.; Cab, *Clostridium acetobutylicum*; Cau, *C.aurantiacus*; Cwat, *C.watsonii*; Glvi, *G.violaceus*; Mth, *M.thermautotrophicus*; Pfu, *P.furiosus*; Suac, *S.acidocaldarius;* Tac, *Thermoplasma acidophilum* and Tery, *T.erythraeum*.

the common ancestor of these families, from which the HerA and FtsK clades diverged concomitantly with the split between the archaeal and bacterial lineages. This archaeo-bacterial dichotomy is similar to that in some families of proteins involved in DNA replication, such as PCNA/DNA polymerase III β subunit and ATP/NAD-dependent DNA ligase (75,76). In each of these cases, the fundamental separation among the conserved members of these families, which share only limited sequence similarity, corresponds to the split between the bacterial and archaeo-eukaryotic lineages. Moreover, those bacteria that lack FtsK always encode a HerA protein that belongs to the conserved core of this family, which is predominantly found in archaea. These bacterial genomes, without exception, also encode the nuclease partner of HerA, NurA (29) (Table 1 and Supplementary Material, Table S1). This complementarity in the phyletic distributions of the core orthologous set of HerA proteins and the FtsK family, even within the bacterial kingdom, along with the co-occurrence with NurA, suggests that HerA and FtsK are responsible for the same function, namely DNA pumping during cell division. It should be emphasized that, although there is no experimental data on the biological functions of various HerA paralogs, the strict conservation of the 'main' *herA* gene in archaea, in terms of both the ubiquitous presence and the sequence itself, implies that it is this gene that has an essential function in cell division rather than any of the extra *herA* paralogs present in some of

the archaea. The archaeal HerA–NurA system appears to have displaced FtsK in most bacterial extremophiles and cyanobacteria, which might have been facilitated by their ecological proximity with archaea. These observations imply that LUCA already had a DNA pumping system similar to those in the extant prokaryotes (Table 1 and Supplementary Material, Table S1).

Although not demonstrated experimentally, it seems likely that the pumping process could introduce double-strand breaks in the DNA. If this were the case, the bidirectional DNA helicase activity that has been detected *in vitro* in the purified *Sulfolobus* HerA protein (18) might be involved, together with MRE11 and Rad50, in double-strand break repair (see below). Alternatively, it cannot be ruled out that the helicase activity is unmasked in the *in vitro* assay as a result of the absence of other subunits which are associated with HerA *in vivo*.

The observed phyletic distributions suggest that VirB4 and VirD4 families might have been derived from DNA pumps of ancestral plasmids that were not part of the core cellular genomes. The relationship of the FtsK–HerA proteins with the viral packaging proteins suggests that DNA pumping activity in diverse systems, both cellular and viral, has an ancient common origin. The emergence of the FtsK–HerA ATPase might have marked the origin of structures in which copies of DNA were compartmentalized after replication. These primordial compartments, into which DNA was packaged by the ancient members of the FtsK–HerA superfamily, could have been the evolutionary precursors of cells and viral capsids. The absence of this superfamily in eukaryotes, with the exception of the apparent late HGT into filamentous ascomycetes, is consistent with the dramatic difference between the mechanisms of chromosome segregation in eukaryotes and prokaryotes. The emergence of eukaryotic cytoskeletal components facilitated segregation through the mitotic process, which involved chromosome translocation by ATPase motors, such as dynein and kinesin (77,78). This radically different segregation mechanism appears to have rendered the ancestral HerA-like DNA pump superfluous or even deleterious, thereby favoring its elimination through gene loss at an early stage of eukaryotic evolution.

Contextual information from gene fusions, domains architecture and conserved operons: functional implications for the FtsK–HerA superfamily

Conserved operons, gene fusions and domain architectures are useful in extracting functional information for otherwise uncharacterized proteins based on the principle of 'genomic context' or 'guilt by association'. Products of genes co-occurring in the same operon in multiple, sufficiently distant genomes (conserved gene neighborhoods) or undergoing gene fusions tend to interact physically and functionally (30–33). Accordingly, we systematically surveyed the genomic context information for the proteins of the FtsK–HerA superfamily. In Figure 5, this information is represented as domain architectures (Figure 5A), gene organizations (Figure 5B) and a graph where the nodes are the connected proteins and the edges denote different types of contextual connections (Figure 5C).

The archaeal herA operons and the widespread herA–nurA gene pairs. The largest conserved operons including genes for FtsK–HerA superfamily ATPases are those that contain the highly conserved archaeal HerA proteins of the core orthologous set. These operons typically encode four proteins, namely, HerA, orthologs of Mre11 and Rad50, and the NurA nuclease; the same gene order (HerA–Mre11–Rad50–NurA) is conserved in six genera of euryarchaea and crenarchaea (Figure 5B). Variants of this order are seen in *Aeropyrum* (Mre11–Rad50–NurA–HerA) and *Thermoplasma* (NurA–HerA–Mre11–RAD50). In *Methanothermobacter,* the *herA* gene in this operon is apparently split into two genes and the gene encoding the C-terminal half is fused to the gene for Mre11. In *Methanosarcina*, *Halobacterium* and *Methanococcus,* the operon is split into separate HerA–NurA and Mre11–Rad50 (predicted) operons. In the genome of *Nanoarchaeum,* the operon is completely disrupted, although all four genes are present. A parsimonious evolutionary scenario places the complete operon consisting of the four genes in the typical order into the genome of the common ancestor of archaea, with partial disruption of the operon during subsequent evolution of individual lineages. There is no trace of this conserved operon in any of the currently available bacterial genomes, with the sole exception of *Bradyrhizobium*, which shows a linkage of a HerA protein of the bll1925 family with SbcD, the bacterial ortholog of Mre11 (Figure 5B).

In addition to this core orthologous lineage, archaea have several additional paralogs of HerA, most of which are encoded next to a conserved ORF of approximately the same size as NurA. Comparison of these sequences to the NurA PSSM showed that the HerA-associated ORFs were divergent paralogs of NurA. Further iterations of these searches, with PSSMs that included the newly detected NurA homologs resulted in the detection of numerous additional divergent members of the NurA family from archaea and bacteria. Remarkably, these searches showed that the phyletic distribution of the NurA family is nearly identical to that of the HerA family (Supplementary Material; Table S1). In most of the archaeal genomes, the newly detected NurA proteins are encoded next to *herA* genes. Among bacteria, *nurA* genes of *Bacillus halodurans*, *Clostridium thermocellum*, *Deinococcus radiodurans* and *Aquifex aeolicus* are adjacent to genes for HerA homologs whereas, in the rest of bacteria, the *nurA* and *herA* genes are located in different parts of the chromosome. In *Chloroflexus,* the NurA homolog is encoded next to a gene encoding a stand-alone HAS-barrel (Figure 5B). These observations suggest that the nuclease NurA and the ATPase HerA are not only functionally linked, but also tend to be horizontally transferred among archaea and bacteria as a gene pair. Furthermore, the situation in *Chloroflexus* supports the prediction that interaction between NurA and HerA is mediated by the HAS-barrel. The tight functional association between HerA and NurA mirrors the functional connections between FtsK and the ParCD topoisomerase or the Xer recombinases, suggesting that NurA has a function similar to the functions of these bacterial enzymes in DNA processing during chromosome segregation (14,16).

Detection of the new, diverged NurA homologs provided for a better characterization of the conserved structural elements and the active site of the NurA family nucleases (Figure 6). Secondary structure prediction combined with examination of the alignment suggests that NurA has an $\alpha + \beta$-fold with a central, conserved β-sheet formed by at

least eight strands. NurA has at least five conserved α-helices, with the last three forming a characteristic triple helical unit. These patterns do not bear any obvious resemblance to previously characterized folds found in nucleases or other proteins. The predicted active site is comprised of six charged/polar residues, which include two characteristic aspartates at the ends of core strands 1 and 5, a conserved glutamate in the first core helix, a basic residue and acidic residue after strand 8, and a polar residue (usually histidine or aspartate) in the C-terminal helical unit (Figure 6). These residues might coordinate a metal cation as observed for the restriction endonuclease fold enzymes. The NurA family appears to be a rapidly diverging group, with a low level of sequence similarity between paralogs and even within orthologous groups. There are several inserts of variable size in different members including a small Zn-cluster in the cyanobacterial and *Chloroflexus* NurA homologs, and a Zn-ribbon in the NurAs associated with the CT1915-like proteins of the HerA clade. The extreme sequence divergence in the NurA family is reminiscent of restriction endonucleases, suggesting the possibility that, similar to restriction enzymes, different NurA family members recognize specific target sequences in DNA. The analogy could extend even further in that the NurA–HerA pairs might form mobile elements similar to restriction-modification system operons. Consistently with this hypothesis, a distinct subfamily of *nurA* genes (typified by HH1040 of *H.hepaticus* that associate with the distinctive CT1915 family HerA) comprises a predicted operon with a gene for a DNA methylase, which is related to methylases encoded by several restriction-modification operons (Figure 5B) (79). The organization of this predicted operon closely resembles that of several restriction-modification systems with the *NurA* gene taking the place of the endonuclease, and the *HerA* gene that of the accessory helicase or ATPase subunit (79). Thus, this family of NurA family is predicted to function as a bona fide restriction endonuclease. In addition, in cyanobacteria, the nurA genes co-occur with a gene for a distinct predicted hydrolase of the HAD (haloacid dehalogenase) superfamily (80); this particular orthologous set of predicted HAD hydrolases was detected only in cyanobacteria and plants (Figure 5B, Supplementary Material). Many proteins of the HAD superfamily have phosphatase activity, and some of them, such as the DNA 3′ phosphatase, Tpp1p, cooperate with endonucleases in strand break repair (81). Hence we speculate that, in cyanobacteria, this particular HAD protein cooperates with NurA in DNA repair, probably as a polynucleotide phosphatase.

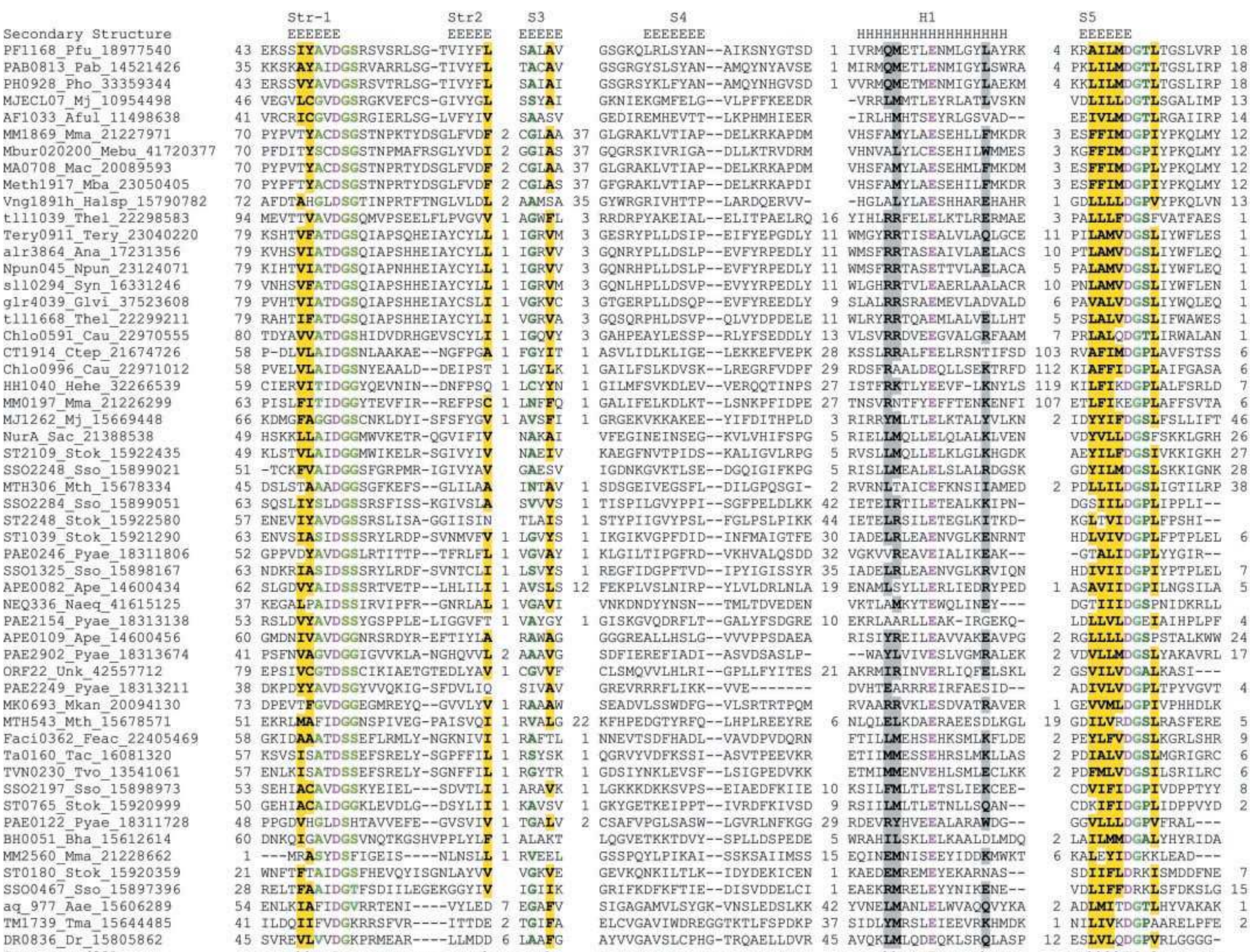
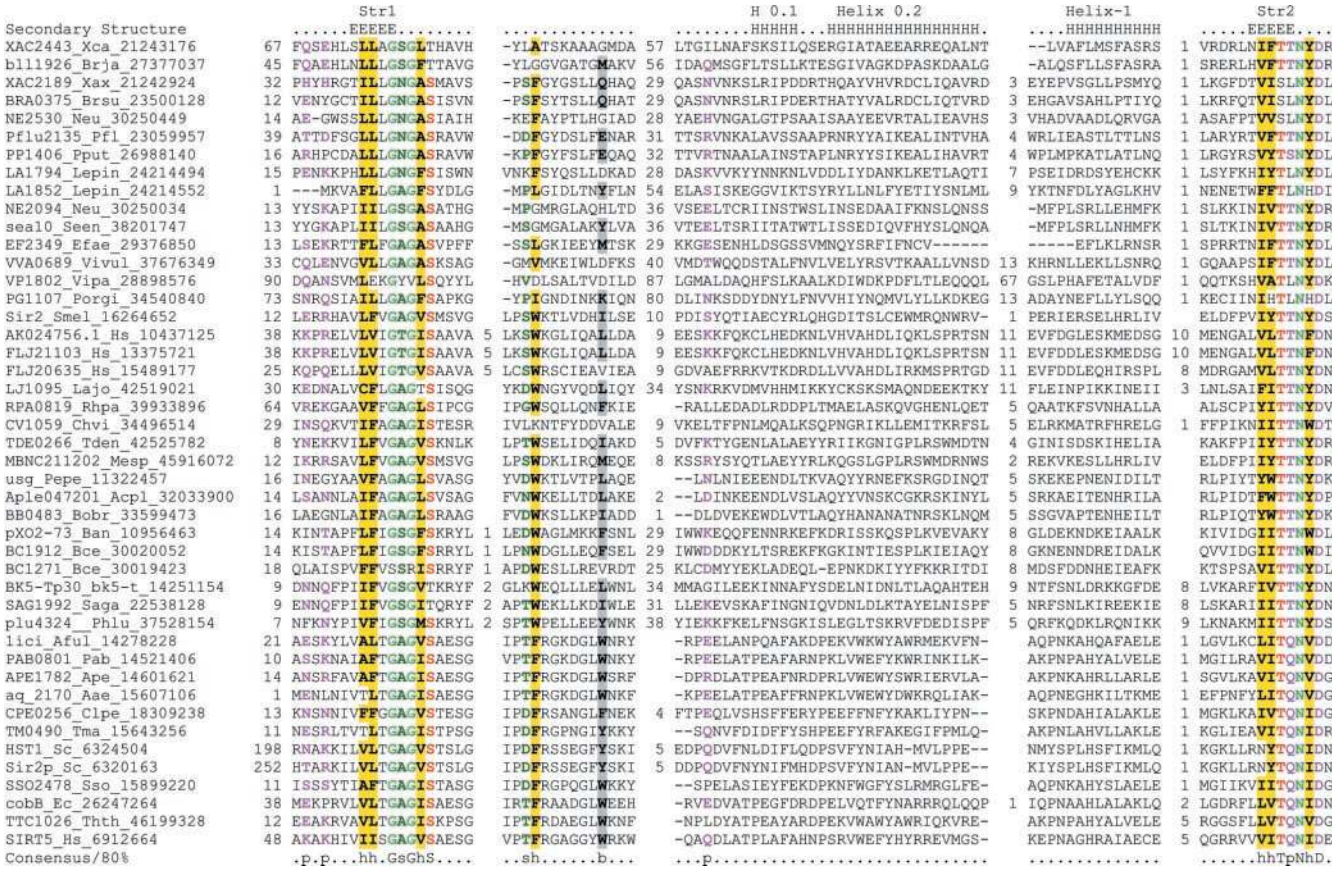

Str-1 Str2 S3 S4 H1 S5
Secondary Structure EEEEEE EEEEE EEEEE EEEEEEE HHHHHHHHHHHHHHHHHH EEEEEE
PF1168_Pfu_18977540 43 EKSSIYAVDGSRSVSRLSG-TVIYFL SALAV GSGKQLRLSYAN--AIKSNYGTSD 1 IVRMQMETLENMLGYLAYRK 4 KRAILMDGTLTGSLVRP 18
PAB0813_Pab_14521426 35 KKSKAYAIDGSRVARRLSG-TIVYFL TACAV GSGRGYSLSYAN--AMQYNYAVSE 1 MIRMQMETLENMIGYLSWRA 4 PKLILMDGTLTGSLIRP 18
PH0928_Pho_33359344 43 ERSSVYAVDGSRSVTRLSG-TIVYFL SAIAI GSGRSYKLFYAN--AMQYNHGVSD 1 VVRMQMETMENMIGYLAEKM 4 KKLILMDGTLTGSLIRP 18
MJECL07_Mj_10954498 46 VEGVLCGVDGSRGKVEFCS-GIVYGL SSYAI GKNIEKGMFELG--VLPFFKEEDR -VRRLMMTLEYRLATLVSKN VDLILLDGTLSGALIMP 13
AF1033_Aful_11498638 41 VRCRICGVDGSRGIERLSG-LVFYIV SAASV GEDIREMHEVTT--LKPHMHIEER -IRLHMHTSEYRLGSVAD-- EEIVLMDGTLRGAIIRP 14
MM1869_Mma_21227971 70 PYPVTYACDSGSTNPKTYDSGLFVDF 2 CGLAA 37 GLGRAKLVTIAP--DELKRKAPDM VHSFAMYLAESEHLLFMKDR 3 ESFFIMDGPIYPKQLMY 12
Mbur020200_Mebu_41720377 70 PFDITYSCDSGSTNPMAFRSGLYVDI 2 GGIAS 37 GQGRSKIVRIGA--DLLKTRVDRM VHNVALYLCESEHILWMMES 3 KGFFIMDGPIYPKQLMY 12
MA0708_Mac_20089593 70 PYPVTYACDSGSTNPRTYDSGLFVDF 2 CGLAA 37 GLGRAKLVTIAP--DELKRKAPDM VHSFAMYLAESEHMLFMKDM 3 ESFFIMDGPLYPKQLMY 12
Meth1917_Mba_23050405 70 PYPFTYACDSGSTNPRTYDSGLFVDF 2 CGLAS 37 GFGRAKLVTIAP--DELKRKAPDI VHSFAMYLAESEHILFMKDR 3 ESFFIMDGPIYPKQLMY 12
Vng1891h_Halsp_15790782 72 AFDTAHGLDSGTINPRTFTNGLVLDL 2 AAMSA 35 GYWRGRIVHTTP--LARDQERVV- -HGLALYLAESHHAREHAHR 1 GDLLLLDGPVYPKQLVN 13
tll1039_Thel_22298583 94 MEVTTVAVDGSQMVPSEELFLPVGVV 1 AGWFL 3 RRDRPYAKEIAL--ELITPAELRQ 16 YIHLRRFELELKTLRERMAE 3 PALLLFDGSFVATFAES 1
Tery0911_Tery_23040220 79 KSHTVFATDGSQIAPSQHEIAYCYLL 1 IGRVM 3 GESRYPLLDSIP--EIFYEPGDLY 11 WMGYRRTISEALVLAQLGCE 11 PILAMVDGSLIYWFLES 1
alr3864_Ana_17231356 79 KVHSVIATDGSQIAPSHHEIAYCYLL 1 IGRVV 3 GQNRYPLLDSLP--EVFYRPEDLY 11 WMSFRRTASEAIVLAELACS 10 PTLAMVDGSLIYWFLEQ 1
Npun045_Npun_23124071 79 KIHTVIATDGSQIAPNHHEIAYCYLL 1 IGRVV 3 GQNRHPLLDSLP--EVFYRPEDLY 11 WMSFRRTASETTVLAELACA 5 PALAMVDGSLIYWFLEQ 1
sll0294_Syn_16331246 79 VNHSVFATDGSQIAPSHHEIAYCYLL 1 IGRVM 3 GQNLHPLLDSVP--EVYYRPEDLY 11 WLGHRRTVLEAERLAALACR 10 PNLAMVDGSLIYWFLEN 1
glr4039_Glvi_37523608 79 PVHTVIATDGSQIAPSHHEIAYCSLI 1 VGKVC 3 GTGERPLLDSQP--EVFYREEDLY 9 SLALRRSRAEMEVLADVALD 6 PAVALVDGSLIYWQLEQ 1
tll1668_Thel_22299211 79 RAHTIFATDGSQIAPSHHEIAYCYLI 1 VGRVA 3 GQSQRPHLDSVP--QLVYDPDELE 11 WLRYRRTQAEMLALVELLHT 5 PSLALVDGSLIFWAWES 1
Chlo0591_Cau_22970555 80 TDYAVVATDGSHIDVDRHGEVSCYLI 1 IGQVY 3 GAHPEAYLESSP--RLYFSEDDLY 13 VLSVRRDVEEGVALGRFAAM 7 PRLALQDGTLIRWALAN 1
CT1914_Ctep_21674726 58 P-DLVLAIDGSNLAAKAE--NGFPGA 1 FGYIT 1 ASVLIDLKLIGE--LEKKEFVEPK 28 KSSLRRALFEELRSNTIFSD 103 RVAFIMDGPLAVFSTSS 6
Chlo0996_Cau_22971012 58 PVELVLAIDGSNYEAALD--DEIPST 1 LGYLK 1 GAILFSLKDVSK--LREGRFVDPF 29 RDSFRAALDEQLLSEKTRFD 112 KIAFFIDGPLAIFGASA 6
HH1040_Hehe_32266539 59 CIERVITIDGGYQEVNIN--DNFPSQ 1 LCYYN 1 GILMFSVKDLEV--VERQQTINPS 27 ISTFRKTLYEEVF-LKNYLS 119 KILFIKDGPLALFSRLD 7
MM0197_Mma_21226299 63 PISLFITIDGGYTEVFIR--REFPSC 1 LNFFQ 1 GALIFELKDLKT--LSNKPFIDPE 27 TNSVRNTFYEFFTENKENFI 107 ETLFIKEGPLAFFSVTA 6
MJ1262_Mj_15669448 66 KDMGFAGGDGSCNKLDYI-SFSFYGV 1 AVSFI 1 GRGEKVKKAKEE--YIFDITHPLD 3 RIRRYMLTLELKTALYVLKN 2 IDYYIFDGSLFSLLIFT 46
NurA_Sac_21388538 49 HSKKLLAIDGGMWVKETR-QGVIFIV NAKAI VFEGINEINSEG--KVLVHIFSPG 5 RIELLMQLLELQLALKLVEN VDYVLLDGSFSKKLGRH 26
ST2109_Stok_15922435 49 KLSTVLAIDGGMWIKELR-SGIVYIV NAEIV KAEGFNVTPIDS--KALIGVLRPG 5 RVSLLMQLLELKLGLKHGDK AEYILFDGSIVKKIGKH 27
SSO2248_Sso_15899021 51 -TCKFVAIDGGSFGRPMR-IGIVYAV GAESV IGDNKGVKTLSE--DGQIGIFKPG 5 RISLLMEALELSLALRDGSK GDYILMDGSLSKKIGNK 28
MTH306_Mth_15678334 45 DSLSTAAADGGSGFKEFS--GLILAA INTAV 1 SDSGEIVEGSFL--DILGPQSGI- 2 RVRNLTAICEFKNSIIAMED 2 PDLLILDGSLIGTILRP 38
SSO2284_Sso_15899051 63 SQSLIYSLDGSSRSFISS-KGIVSLA SVVVS 1 TISPILGVYPPI--SGFPELDLKK 42 IETEIRTILETEALKKIPN- DGSIILDGPLIPPLI--
ST2248_Stok_15922580 57 ENEVIYAVDGSSRSLISA-GGIISIN TLAIS 1 STYPIIGVYPSL--FGLPSLPIKK 44 IETELRSILETEGLKITKD- KGLTVIDGPLFPSHI--
ST1039_Stok_15921290 63 ENVSIASIDSSSRYLRDP-SVNMVFV 1 LGVYS 1 IKGIKVGPFDID--INFMAIGTFE 30 IADELRLEAENVGLKENRNT HDLVIVDGPLFPTPLEL 6
PAE0246_Pyae_18311806 52 GPPVDYAVDGSLRTITTP--TFRLFL 1 VGVAY 1 KLGILTIPGFRD--VKHVALQSDD 32 VGKVVREAVEIALIKEAK-- -GTALIDGPLYYGIR--
SSO1325_Sso_15898167 63 NDKRIASIDSSSRYLRDF-SVNTCLI 1 LSVYS 1 REGFIDGPFTVD--IPYIGISSYR 35 IADELRLEAENVGLKRVIQN HDIVIIDGPIYPTPLEL 7
APE0082_Ape_14600434 62 SLGDVYAIDSSSRTVETP--LHLILI 1 AVSLS 12 FEKPLVSLNIRP--YLVLDRLNLA 19 ENAMLSYLLERLIEDRYPED 1 ASAVIIDGPILNGSILA 5
NEQ336_Naeq_41615125 37 KEGALPAIDSSIRVIPFR--GNRLAL 1 VGAVI VNKDNDYYNSN---TMLTDVEDEN VKTLAMKYTEWQLINEY--- DGTIIIDGSPNIDKRLL
PAE2154_Pyae_18313138 53 RSLDVYAVDSSYGSPPLE-LIGGVFT 1 VAYGY 1 GISKGVQDRFLT--GALYFSDGRE 10 EKRLAARLLEAK-IRGEKQ- LDLLVLDGEIAIHPLPF 4
APE0109_Ape_14600456 60 GMDNIVAVDGGNRSRDYR-EFTIYLA RAWAG GGGREALLHSLG--VVVPPSDAEA RISIYREILEAVVAKEAVPG 2 RGLLLLDGSPSTALKWW 24
PAE2902_Pyae_18313674 41 PSFNVAGVDGGIGVVKLA-NGHQVVL 2 AAAVG SDFIEREFIADI--ASVDSASLP- --WAYLVIVESLVGMRALEK 2 VDVLLMDGSLYAKAVRL 17
ORF22_Unk_42557712 79 EPSIVCGTDSSCIKIAETGTEDLYAV 1 CGVVF CLSMQVVLHLRI--GPLLFYITES 21 AKRMIRINVERLIQFELSKL 2 GSVILVDGALKASI---
PAE2249_Pyae_18313211 38 DKPDYYAVDSGYVVQKIG-SFDVLIQ SIVAV GREVRRRFLIKK--VVE------- DVHTEARRREIRFAESID-- ADIVLVDGPLTPYVGVT 4
MK0693_Mkan_20094130 73 DPEVTFGVDGGEGMREYQ--GVVLYV 1 RAAAW SEADVLSSWDFG--VLSRTRTPQM RVAARRVKLESDVATRAVER 1 GEVVMLDGPIVPHHDLK
MTH543_Mth_15678571 51 EKRLMAFIDGGNSPIVEG-PAISVQI 1 RVALG 22 KFHPEDGTYRFQ--LHPLREEYRE 6 NLQLELKDAERAEESDLKGL 19 GDILVRDGSLRASFERE 5
Faci0362_Feac_22405469 58 GKIDAAATDSSEFLRMLY-NGKNIVI 1 RAFTL 1 NNEVTSDFHADL--VAVDPVDQRN FTILLMEHSEHKSMLKFLDE 2 PEYLFVDGSLKGRLSHR 9
Ta0160_Tac_16081320 57 KSVSISATDSSEFSRELY-SGPFFIL 1 RSYSK 1 QGRVYVDFKSSI--ASVTPEEVKR ETIIMMESSEHRSLMKLLAS 2 PDIALVDGSLMGRIGRC 6
TVN0230_Tvo_13541061 57 ENLKISATDSSEFSRELY-SGNFFIL 1 RGYTR 1 GDSIYNKLEVSF--LSIGPEDVKK ETMIMMENVEHLSMLECLKK 2 PDFMLVDGSILSRILRC 6
SSO2197_Sso_15898973 53 SEHIACAVDGSKYEIEL---SDVTLI 1 ARAVK 1 LGKKKDKKSVPS--EIAEDFKIIE 10 KSILFMLTLETSLIEKCEE- CDVIFIDGPIVDPPTYY 8
ST0765_Stok_15920999 50 GEHIACAIDGGKLEVDLG--DSYLII 1 KAVSV 1 GKYGETKEIPPT--IVRDFKIVSD 9 RSIILMLTLETNLLSQAN-- CDKIFIDGPLIDPPVYD 2
PAE0122_Pyae_18311728 48 PPGDVHGLDSHTAVVEFE--GVSVIV 1 TGALV 2 CSAFVPGLSASW--LGVRLNFKGG 29 RDEVRYHVEEALARAWDG-- GGVLLLDGPVFRAL---
BH0051_Bha_15612614 60 DNKQIGAVDGSVNQTKGSHVPPLYLF 1 ALAKT LQGVETKKTDVY--SPLLDSPEDE 5 WRAHILSKLELKAALDLMDQ 2 LAILMMDGALYHYRIDA
MM2560_Mma_21228662 1 ---MRASYDSFIGEIS----NLNSLL 1 RVEEL GSSPQYLPIKAI--SSKSAIIMSS 15 EQINEMNISEEYIDDKMWKT 6 KALEYIDGKKLEAD---
ST0180_Stok_15920359 21 WNFTFTAIDGSFHEVQYISGNLAYVV VGKVE GEVKQNKILTLK--IDYDEKICEN 1 KAEDEMREMEYEKARNAS-- SDIIFLDRKISMDDFNE 7
SSO0467_Sso_15897396 28 RELTFAAIDGTFSDIILEGEKGGYIV IGIIK GRIFKDFKFTIE--DISVDDELCI 1 EAEKRMRELEYYNIKENE-- VDLIFFDRKLSFDKSLG 15
aq_977_Aae_15606289 54 ENLKIAFIDGVRRTENI----VYLED 7 EGAFV SIGAGAMVLSYGK-VNSLEDSLKK 42 YVNELMANLELWVAQQVYKA 2 ADLMITDGTLHYVAKAK 1
TM1739_Tma_15644485 41 ILDQIIFVDGKRRSFVR----ITTDE 2 TGIFA ELCVGAVIWDREGGTKTLFSPDKP 37 SIDLYMRSLEIEEVRKHMDK 1 NILIVKDGPAARELPFE 2
DR0836_Dr_15805862 45 SVREVLVVDGKPRMEAR----LLMDD 6 LAAFG AYVVGAVSLCPHG-TRQAELLDVR 45 AVQKLMLQDEQKLSRQLASP 12 ESLVLQDGPVRLGGGG-
Consensus/80% ....hhs.Dus..............h....s.h. ....bb...E.....b.... ..hhhhDGsl.......

Figure 6. Multiple alignment of the NurA superfamily. The coloring reflects 80% consensus and the coloring scheme, consensus abbreviations and secondary structure representations are as in Figures 2 and 3. Species abbreviations are as follows. Aae, *A.aeolicus*; Aful, *A.fulgidus*; Ana, *Anabaena* sp.; Ape, *A.pernix*; Bha, *B.halodurans*; Cau, *C.aurantiacus*; Ctep, *C.tepidum*; Dr, *D.radiodurans*; Feac, *F.acidarmanus*; Glvi, *G.violaceus*; Halsp, *Halobacterium* sp.; Hehe, *H.hepaticus*; Mac, *M.acetivorans*; Mba, *Methanosarcina c*; Mebu, *Methanococcoides burtonii*; Mj, *M.jannaschii*; Mkan, *Methanopyrus kandleri*; Mma, *M.mazei*; Mth, *M.thermautotrophicus*; Naeq, *Nanoarchaeum equitans*; Npun, *N.punctiforme*; Pab, *Pyrococcus abyssi*; Pfu, *P.furiosus*; Pho, *P.horikoshii*; Pyae, *P.aerophilum*; Sac, *S.acidocaldarius*; Sso, *S.solfataricus*; Stok, *Sulfolobus tokodaii*; Syn, *Synechocystis* sp.; Tac, *T.acidophilum*; Tery, *T.erythraeum*; Thel, *T.elongatus*; Tma, *T.maritima*; Tvo, *T.volcanium* and Unk, Uncultured crenarchaeote.

The bacterial orthologs of Mre11 and Rad50 are, respectively, SbcD and SbcC proteins, which typically are encoded in a conserved operon present in most major bacterial lineages. Most likely, the bacterial SbcD–SbcC operon and the orthologous archaeal Mre11–Rad50 operon descended from an ancestral nuclease–ATPase operon of LUCA. Since both HerA–NurA and Mre11–Rad50 operons are much more common that the complete four-gene operon, it appears likely that the latter evolved in the common ancestor of crenarchaea and euryarchaea as a result of fusion of the two gene pairs. The available information on the functions of the eukaryotic Mre11 and Rad50 proteins provide hints regarding the possible functional significance of the genomic linkage of these four genes. The ABC ATPases of the SMC-family, which includes Rad50, are involved in chromatin dynamics associated with chromosome condensation and segregation (82,83). In particular, Rad50 bridges the double-strand breaks in DNA and facilitates end processing by the Mre11 nuclease (84,85). Therefore, in archaea, the Rad50 and Mre11 orthologs could function in a complex with HerA to repair double-strand breaks, which could potentially arise during the process of chromosomal segregation. Furthermore, Rad50 could also function in reorganizing the higher order chromatin structure during segregation. Archaeal kleisins, which are predicted to be functional partners of Rad50 proteins (86), are also likely to participate in this process. The predicted HerA–Rad50–Mre11–kleisin repair system might also function in double-strand break repair during archaeal DNA replication. However, in view of the structural, functional and evolutionary relationships between HerA and FtsK discussed above, it seems most likely that the principal, essential role of this system is linked to chromosomal segregation.



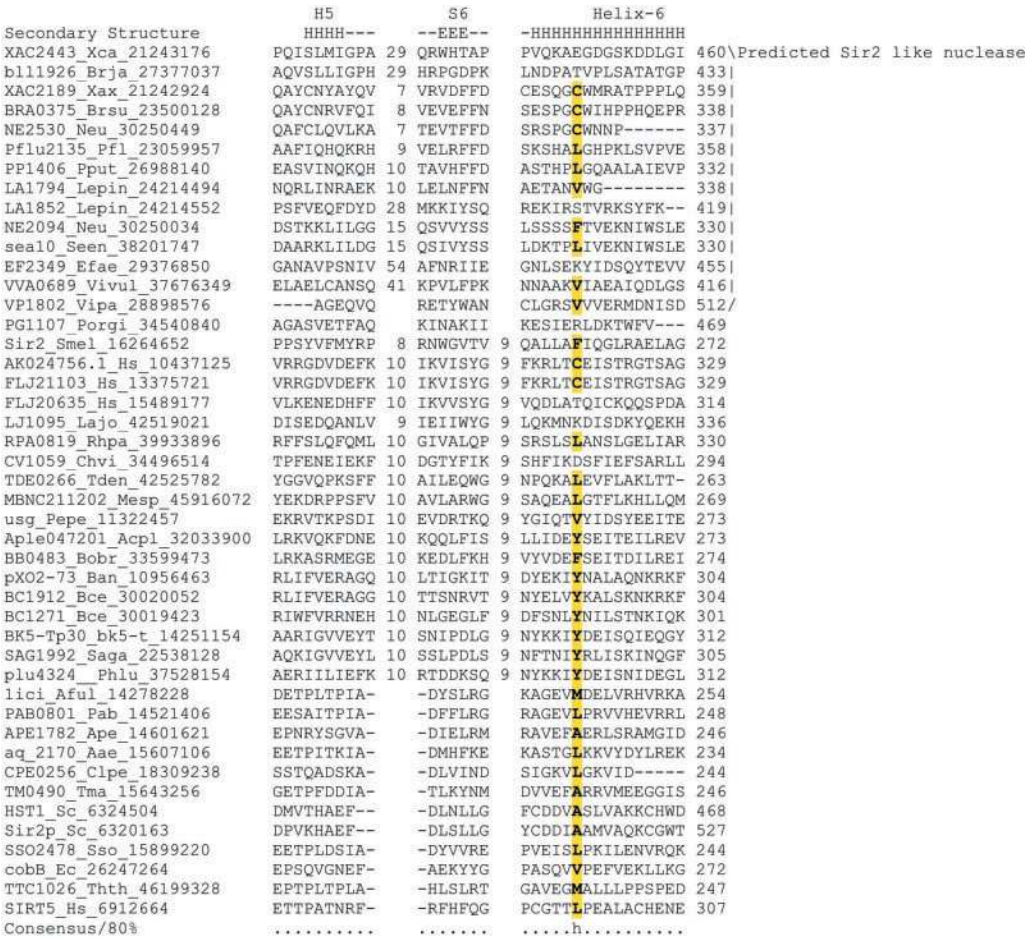


Figure 7. Multiple alignment of the predicted Sir2-like nuclease. The coloring reflects 80% consensus and the consensus abbreviations, coloring scheme and secondary structure designations are as in Figures 2 and 3. The histidine and aspartate residue conserved in the predicted nucleases are shaded red. Secondary structure elements are numbered according to their position in the core Rossmann fold. Helix 0.1 and 0.2 reflect helices that are synapomorphic to the Sir2-clade. Species abbreviations are as follows: Aae, *A.aeolicus*; Acpl, *Actinobacillus pleuropneumoniae*; Aful, *A.fulgidus*; Ape, *A.pernix*; Ban, *Bacillus anthracis*; Bce, *B.cereus*; bk5-t, Lactococcus phage bk5-t; Bobr, *Bordetella bronchiseptica*; Brja, *B.japonicum*; Brsu, *Brucella suis*; Chvi, *Chromobacterium violaceum*; Clpe, *Clostridium perfringens*; Ec, *E.coli*; Efae, *E.faecalis*; Hs, *Homo sapiens*; Lajo, *Lactobacillus johnsonii*; Lepin, *Leptospira c*; Mesp, *Mesorhizobium* sp.; Neu, *N.europaea*; Pab, *P.abyssi*; Pepe, *Pediococcus pentosaceus*; Pfl, *Pseudomonas fluorescens*; Phlu, *Photorhabdus luminescens*; Porgi, *Porphyromonas gingivalis*; Pput, *Pseudomonas putida*; Rhpa, *Rhodopseudomonas palustris*; Saga, *Streptococcus agalactiae*; Sc, *S.cerevisiae*; Seen, *Serratia entomophila*; Smel, *S.meliloti*; Sso, *Sulfolobus solfataricus*; Tden, *Treponema denticola*; Thth, *T.thermophilus*; Tma, *T.maritima*; Vipa, *V.parahaemolyticus*; Vivul, *V.vulnificus*; Xax, *Xanthomonas axonopodis* and Xca, *Xanthomonas campesteris*.

Many members of the FtsK–HerA superfamily contain membrane-spanning regions that probably anchor them to the cell membrane during DNA pumping. No such membrane-spanning regions are present in the core orthologous set of archaeal HerA proteins. The contextual association (albeit weak) of HerA and the highly conserved small membrane proteins typified by MJ1617 (COG2034) implicates these proteins as potential candidates for the role of a membrane tether for HerA. In the case of other HerA ATPases, additional, poorly conserved membrane proteins might function as their partners. However, the absence of membrane-spanning regions in HerA proteins themselves or conserved genes for membrane proteins in the predicted *herA* operons raises the possibility of fundamental functional differences between HerA proper and the rest of the FtsK–HerA superfamily ATPases.

Additional nuclease connections of the FtsK–HerA superfamily and prediction of a novel nuclease with the Sir2 fold. Several previously described conserved gene neighborhoods of VirB4, VirD4 and FtsK encode components of the T4SS of proteobacteria or the ESAT-6 system of Gram-positive bacteria and actinomycetes (17,87). However, in other conserved gene neighborhoods, FtsK–HerA superfamily ATPases are encoded together with nucleases involved in DNA processing. In particular, genes for ATPases of the TrwB/TrsK and the TraG families of the VirD4 clade are found in operons that also contain genes for conjugative relaxases of the TrwC and TraA families, respectively (Figure 5B) (65,88). These relaxases have an N-terminal nuclease domain of the rolling circle replication (RCR) fold combined with a C-terminal SF-I DNA helicase domain. The TraA relaxases belong to the RCR superfamily proper, with a HXH active site motif and a catalytic tyrosine (89,90), whereas the TrwC relaxase domain shows an evolutionarily distant, circularly permuted version of the fold [(91); L. Aravind, unpublished data]. Thus, at least on two independent occasions, VirD-like ATPases appear to have been combined with distinct members of the RCR nuclease fold in conserved operons.

The VirB4-like ATPases of the YddE family encoded by conjugative TN916 transposons often co-occur with genes for a large membrane protein with six transmembrane regions (YddG), a hydrolase of the NlpC/P60 superfamily (92), a smaller membrane protein with a single transmembrane region and a catalytic tyrosine containing relaxase of the pT181–Rep domain superfamily, which is unrelated to the RCR superfamily relaxases (Figure 5B). Using iterative sequence database searches with the PSSM for this relaxase family, we showed that they are homologous to the nicking enzyme of the filamentous bacteriophages, such as M13 and f1. In these phages, the nicking enzyme functions in conjunction with the packaging ATPase that also belongs to the FtsK–HerA superfamily. Thus, as in the case of the RCRs with the HXH motif, the pT181–Rep relaxases have also formed multiple, independent functional associations with ATPases of the FtsK–HerA superfamily. NlpC family hydrolases encoded by the adjacent ORFs in these transposons are likely to facilitate local degradation of the cell wall, whereas the transmembrane proteins are likely to be components of the conjugation tube through which the DNA of the conjugative transposons is pumped by YddE after it is processed by the associated relaxase (Figure 5; Supplementary Material). Thus, persistent operonic associations with several unrelated nucleases are prevalent in different clades of the FtsK–HerA superfamily.

This contextual theme was exploited to predict previously uncharacterized nucleases with probable functional links to FtsK–HerA ATPases. Several members of the proteobacterial bll1925 family of the HerA clade, which contains a divergent version of the HAS barrel (Table 1, Figure 5), are encoded next to a conserved, co-directional ORF. This ORF, bll1926, is unrelated to NurA, but iterated database searches using PSI-BLAST showed that it defines a previously undetected protein family, which is distantly related to the Sir2 proteins. Despite their high sequence divergence, the bll1926-like proteins contained all the hallmarks of the Sir2 fold (a variant Rossmann fold), such as the glycine-rich loop at the N-terminus, the central NhD motif (where h is any hydrophobic residue) and the C-terminal HG motif. However, the bll1926 family proteins lack the Zn-ribbon insert characteristic of the Sir2 family and contain a distinct, C-terminal DXH motif which is absent in Sir2 (Figure 7). Members of the Sir2 family deacetylate acetyl-lysines in a variety of protein substrates, a reaction that utilizes NAD and produces 2′-O-acetyl-ADP-ribose (93–96). A superposition of the conservation pattern of the bll1926 family onto the crystal structure of the Sir2 catalytic domain (94–97) suggests that, despite the conservation of the active site residues, the surface involved in peptide interaction in Sir2 is not conserved between the two proteins (Figure 7). Furthermore, the additional DXH motif of the bll1926 family, together with the conserved histidine of the HG motif, forms a potential di-histidine active site configuration similar to those in nucleases or phosphoesterases of the RNAse A and 2H superfamilies (98,99). Given the persistent linkage of the FtsK–HerA ATPases with nucleases, we predict that the bll1926 family proteins are nucleases rather than deacetylases like the Sir-2 proteins. Hence, two very different catalytic activities appear to have emerged within the same fold as a result of recruitment of partially different sets of conserved residues for the active center of Sir2 and bll1926. The apparent horizontal mobility of the bll1925–bll1926 pair in proteobacteria mirrors that of the HerA–NurA gene pair. This observation suggests a close functional parallel between these systems and supports the prediction of the nuclease function for the bll1926 family of Sir2 homologs.

Despite the close functional association with various nucleases, as indicated by the presence of conserved operons, HerA ATPases do not form fused genes with any of these nucleases. There might be a single exception to this trend. A protein from *A.aeolicus*, aq_1852, consists of a HerA domain and an N-terminal HKD domain, the catalytic module of numerous phosphohydrolases, such as phospholipase D, eukaryotic tyrosine–DNA phosphodiesterases and certain DNAses, such as Nuc (100–102). Thus, the HKD domain fused to the HerA ATPase in aq_1852 could be a DNAse, polynucleotide phosphatase or a tyrosine–DNA phosphodiesterase. The (near) absence of HerA-nuclease fusions is somewhat unexpected because, on many occasions, genes that are part of the same operon in some genomes are fused others (103). The absence of such fusions suggests that FtsK–HerA superfamily ATPases and the associated nucleases might be present in the respective functional complexes in non-stoichiometric amounts.

GENERAL EVOLUTIONARY CONSIDERATIONS AND CONCLUSIONS

The majority of experimentally characterized members of the FtsK–HerA ATPase superfamily are involved in pumping substrates, particularly DNA, through membrane-spanning pores. The two primary clades of this superfamily, HerA and FtsK, show nearly perfect complementarity in their phyletic patterns: predominantly archaeal HerA (the core orthologous set) versus mostly bacterial FtsK. Together with the evolutionary relationship between these proteins discussed here, this suggests that HerA and FtsK perform analogous functions in DNA pumping during cell division. The operonic organization of HerA, NurA, MRE11 and Rad50 that is conserved in most archaea suggests additional players in this process and points to the potential importance in it of double-strand break repair. The bacterial orthologs of MRE11 and Rad50 do not form operons with FtsK and so far have been implicated only in recombinational repair pathways (104,105). It appears likely that functional association of HerA–NurA with Rad50–MRE11 is an archaeal innovation.

While at least one representative of the FtsK–HerA families is present in each prokaryotic genome, they are practically absent in eukaryotes except for some fungal forms which probably were acquired via relatively late HGT. Given that eukaryotes evolved a mechanism of chromosome segregation that is radically different from the prokaryotic one, this observation lends further support to the conjecture that FtsK and HerA are functionally equivalent enzymes, which are ancestral in the bacterial and archaeal lineages, respectively. Eukaryotes probably lost HerA and its nuclease partner, NurA, concomitantly with the advent of the new segregation mechanism, whereas their functional partners, MRE11 and Rad50, have been retained as essential repair enzymes.

Under the most parsimonious evolutionary scenario, FtsK and HerA descended from a single ancestral ATPase pump that was present in LUCA, along with several other P-loop ATPases. It seems likely that separation of the FtsK–HerA

lineage from other related ASCE ATPases coincided with a critical early stage in the evolution of life, the origin of a specialized, active mechanism for segregation of daughter genomes during cell division. It is also notable that viral packaging ATPases comprise two of the early branching lineages of the FtsK–HerA superfamily. Thus, DNA packaging into capsid-like structures might have evolved roughly synchronously with chromosomal segregation.

Other conserved bacterial cell division proteins, such as FtsA, MreB and FtsZ (106,107), are not universally represented in all prokaryotic lineages. To date, FtsA is absent in almost all archaea, MreB is absent in all crenarchaea and several euryarchaea and FtsZ is absent in all crenarchaea [V. Anantharaman and L. Aravind, unpublished data; (108)]. These phyletic patterns raise the possibility that the cell septation apparatus in LUCA lacked some of the key extant components. At least part of the septation apparatus, along with the cell wall, might have evolved later than the putative DNA-pumping complex that included the prototype FtsK–HerA ATPase. Hence, the proto-cells, prior to and including LUCA, probably were relatively simple structures that did not possess a complex apparatus for septation that is seen in extant cells and principally depended on DNA-pumping for daughter genome segregation. Despite functionally similar associations of the FtsK–HerA superfamily ATPases with nucleases, none of the nuclease partners of the FtsK–HerA superfamily ATPases can be traced to LUCA. Given that even the smallest plasmids, conjugative transposons and phages have a nuclease or topoisomerase that functions along with the pumping ATPase of the FtsK–HerA superfamily, the ancestral nuclease might have been displaced during evolution of large cellular genomes. The mechanisms for decatenation of replication products of the larger chromosomes appear to have evolved independently in the archaeal and bacterial lineages, resulting in the independent recruitment of NurA and Xer/ParCD enzymes, respectively.

Thus, using computational analysis of proteins sequences and structures along with genome context analysis, we predict the central components and the possible mechanism for chromosomal segregation in archaea. The observations described here may help in designing further experiments aimed at dissection of two of the most fundamental biological processes, chromosomal segregation and cell division.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We thank C. Elie and F. Constalinesco for their contributions in the early stages of this work and thank P. Forterre and D. D. Leipe for helpful discussions.

REFERENCES

1. Pogliano,K., Pogliano,J. and Becker,E. (2003) Chromosome segregation in eubacteria. *Curr. Opin. Microbiol.*, **6**, 586–593.
2. Gerdes,K., Moller-Jensen,J., Ebersbach,G., Kruse,T. and Nordstrom,K. (2004) Bacterial mitotic machineries. *Cell*, **116**, 359–366.
3. Romberg,L. and Levin,P.A. (2003) Assembly dynamics of the bacterial cell division protein FTSZ: poised at the edge of stability. *Annu. Rev. Microbiol.*, **57**, 125–154.
4. Margolin,W. (2003) Bacterial division: the fellowship of the ring. *Curr. Biol.*, **13**, R16–18.
5. Margolin,W. (2001) Spatial regulation of cytokinesis in bacteria. *Curr. Opin. Microbiol.*, **4**, 647–652.
6. Chen,J.C. and 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.
7. Errington,J., Daniel,R.A. and Scheffers,D.J. (2003) Cytokinesis in bacteria. *Microbiol. Mol. Biol. Rev.*, **67**, 52–65.
8. Yu,X.C., Weihe,E.K. and Margolin,W. (1998) Role of the C terminus of FtsK in *Escherichia coli* chromosome segregation. *J. Bacteriol.*, **180**, 6424–6428.
9. Aussel,L., Barre,F.X., Aroyo,M., Stasiak,A., Stasiak,A.Z. and Sherratt,D. (2002) FtsK is a DNA motor protein that activates chromosome dimer resolution by switching the catalytic state of the XerC and XerD recombinases. *Cell*, **108**, 195–205.
10. Sherratt,D.J., Soballe,B., Barre,F.X., Filipe,S., Lau,I., Massey,T. and Yates,J. (2004) Recombination and chromosome segregation. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, **359**, 61–69.
11. Barre,F.X., Soballe,B., Michel,B., Aroyo,M., Robertson,M. and Sherratt,D. (2001) Circles: the replication-recombination-chromosome segregation connection. *Proc. Natl Acad. Sci. USA*, **98**, 8189–8195.
12. Donachie,W.D. (2002) FtsK: Maxwell's demon? *Mol. Cell*, **9**, 206–207.
13. Errington,J., Bath,J. and Wu,L.J. (2001) DNA transport in bacteria. *Nature Rev. Mol. Cell. Biol.*, **2**, 538–545.
14. Ip,S.C., Bregu,M., Barre,F.X. and Sherratt,D.J. (2003) Decatenation of DNA circles by FtsK-dependent Xer site-specific recombination. *EMBO J.*, **22**, 6399–6407.
15. Massey,T.H., Aussel,L., Barre,F.X. and Sherratt,D.J. (2004) Asymmetric activation of Xer site-specific recombination by FtsK. *EMBO Rep.*, **5**, 399–404.
16. Espeli,O., Lee,C. and Marians,K.J. (2003) A physical and functional interaction between *Escherichia coli* FtsK and topoisomerase IV. *J. Biol. Chem.*, **278**, 44639–44644.
17. Pallen,M.J. (2002) The ESAT-6/WXG100 superfamily—and a new Gram-positive secretion system? *Trends Microbiol.*, **10**, 209–212.
18. Constantinesco,F., Forterre,P., Koonin,E.V., Aravind,L. and Elie,C. (2004) A bipolar DNA helicase gene, herA, clusters with rad50, mre11 and nurA genes in thermophilic archaea. *Nucleic Acids Res.*, **32**, 1439–1447.
19. Moncalian,G., Cabezon,E., Alkorta,I., Valle,M., Moro,F., Valpuesta,J.M., Goni,F.M. and de La Cruz,F. (1999) Characterization of ATP and DNA binding activities of TrwB, the coupling protein essential in plasmid R388 conjugation. *J. Biol. Chem.*, **274**, 36117–36124.
20. Hamilton,C.M., Lee,H., Li,P.L., Cook,D.M., Piper,K.R., von Bodman,S.B., Lanka,E., Ream,W. and Farrand,S.K. (2000) TraG from RP4 and TraG and VirD4 from Ti plasmids confer relaxosome specificity to the conjugal transfer system of pTiC58. *J. Bacteriol.*, **182**, 1541–1548.
21. Dang,T.A., Zhou,X.R., Graf,B. and Christie,P.J. (1999) Dimerization of the *Agrobacterium tumefaciens* VirB4 ATPase and the effect of ATP-binding cassette mutations on the assembly and function of the T-DNA transporter. *Mol. Microbiol.*, **32**, 1239–1253.
22. Fullner,K.J., Stephens,K.M. and Nester,E.W. (1994) An essential virulence protein of *Agrobacterium tumefaciens*, VirB4, requires an intact mononucleotide binding domain to function in transfer of T-DNA. *Mol. Gen. Genet.*, **245**, 704–715.
23. Berger,B.R. and Christie,P.J. (1993) The *Agrobacterium tumefaciens* virB4 gene product is an essential virulence protein requiring an intact nucleoside triphosphate-binding domain. *J. Bacteriol.*, **175**, 1723–1734.
24. Kado,C.I. (1994) Promiscuous DNA transfer system of *Agrobacterium tumefaciens*: role of the virB operon in sex pilus assembly and synthesis. *Mol. Microbiol.*, **12**, 17–22.
25. Gomis-Ruth,F.X., Moncalian,G., Perez-Luque,R., Gonzalez,A., Cabezon,E., de la Cruz,F. and Coll,M. (2001) The bacterial conjugation protein TrwB resembles ring helicases and F1-ATPase. *Nature*, **409**, 637–641.
26. Egelman,E.H. (2001) Structural biology. Pumping DNA. *Nature*, **409**, 573–575.
27. Laskey,R.A. and Madine,M.A. (2003) A rotary pumping model for helicase function of MCM proteins at a distance from replication forks. *EMBO Rep.*, **4**, 26–30.

28. Manzan,A., Pfeiffer,G., Hefferin,M.L., Lang,C.E., Carney,J.P. and Hopfner,K.P. (2004) MlaA, a hexameric ATPase linked to the Mre11 complex in archaeal genomes. *EMBO Rep.*, **5**, 54–59.
29. Constantinesco,F., Forterre,P. and Elie,C. (2002) NurA, a novel 5′–3′ nuclease gene linked to rad50 and mre11 homologs of thermophilic *Archaea*. *EMBO Rep.*, **3**, 537–542.
30. Aravind,L. (2000) Guilt by association: contextual information in genome analysis. *Genome Res.*, **10**, 1074–1077.
31. Galperin,M.Y. and Koonin,E.V. (2000) Who's your neighbor? New computational approaches for functional genomics. *Nat. Biotechnol.*, **18**, 609–613.
32. Huynen,M., Snel,B., Lathe,W. and Bork,P. (2000) Exploitation of gene context. *Curr. Opin. Struct. Biol.*, **10**, 366–370.
33. Huynen,M.J. and Snel,B. (2000) Gene and context: integrative approaches to genome analysis. *Adv. Prot. Chem.*, **54**, 345–379.
34. Margolin,W. (2000) Themes and variations in prokaryotic cell division. *FEMS Microbiol. Rev.*, **24**, 531–548.
35. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
36. Notredame,C., Higgins,D.G. and Heringa,J. (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.*, **302**, 205–217.
37. Pei,J., Sadreyev,R. and Grishin,N.V. (2003) PCMA: fast and accurate multiple sequence alignment based on profile consistency. *Bioinformatics*, **19**, 427–428.
38. Walker,D.R. and Koonin,E.V. (1997) SEALS: a system for easy analysis of lots of sequences. *Proceedings of the Fifth International Conference on Intelligent Systems for Molecular Biology (ISMB-97)*, Halkidiki, Greece, pp. 333–339.
39. Ikeda,M., Arai,M., Lao,D.M. and Shimizu,T. (2002) Transmembrane topology prediction methods: a re-assessment and improvement by a consensus method using a dataset of experimentally-characterized transmembrane topologies. *In Silico Biol.*, **2**, 19–33.
40. Peitsch,M.C. (1996) ProMod and Swiss-model: internet-based tools for automated comparative protein modelling. *Biochem. Soc. Trans.*, **24**, 274–279.
41. Kraulis,P. (1991) A program to produce both detailed and schematic plots of proteins. *J. Appl. Crystallogr.*, **24**, 946–950.
42. Rost,B., Sander,C. and Schneider,R. (1994) PHD—an automatic mail server for protein secondary structure prediction. *Comput. Appl. Biosci.*, **10**, 53–60.
43. Felsenstein,J. (1996) Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol.*, **266**, 418–427.
44. Hasegawa,M., Kishino,H. and Saitou,N. (1991) On the maximum likelihood method in molecular phylogenetics. *J. Mol. Evol.*, **32**, 443–445.
45. Wolf,Y.I., Rogozin,I.B., Grishin,N.V., Tatusov,R.L. and Koonin,E.V. (2001) Genome trees constructed using five different approaches suggest new major bacterial clades. *BMC Evol. Biol.*, **1**, 8.
46. Neuwald,A.F., Liu,J.S., Lipman,D.J. and Lawrence,C.E. (1997) Extracting protein alignment models from the sequence database. *Nucleic Acids Res.*, **25**, 1665–1677.
47. Koonin,E.V. (1992) The second cholera toxin, Zot, and its plasmid-encoded and phage-encoded homologues constitute a group of putative ATPases with an altered purine NTP-binding motif. *FEBS Lett.*, **312**, 3–6.
48. Neuwald,A.F., Aravind,L., Spouge,J.L. and Koonin,E.V. (1999) AAA^+: a class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.*, **9**, 27–43.
49. Iyer,L.M., Leipe,D.D., Koonin,E.V. and Aravind,L. (2004) Evolutionary history and higher order classification of AAA^+ ATPases. *J. Struct. Biol.*, **146**, 11–31.
50. Gorbalenya,A.E. and Koonin,E.V. (1993) Helicases: amino acid sequence comparisons and structure-function relationships. *Curr. Opin. Struct. Biol.*, **3**, 419–429.
51. Putnam,C.D., Clancy,S.B., Tsuruta,H., Gonzalez,S., Wetmur,J.G. and Tainer,J.A. (2001) Structure and mechanism of the RuvB Holliday junction branch migration motor. *J. Mol. Biol.*, **311**, 297–310.
52. Sawaya,M.R., Guo,S., Tabor,S., Richardson,C.C. and Ellenberger,T. (1999) Crystal structure of the helicase domain from the replicative helicase-primase of bacteriophage T7. *Cell*, **99**, 167–177.
53. Singleton,M.R., Sawaya,M.R., Ellenberger,T. and Wigley,D.B. (2000) Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. *Cell*, **101**, 589–600.
54. Nadanaciva,S., Weber,J., Wilke-Mounts,S. and Senior,A.E. (1999) Importance of F1-ATPase residue alpha-Arg-376 for catalytic transition state stabilization. *Biochemistry*, **38**, 15493–15499.
55. Leipe,D.D., Aravind,L., Grishin,N.V. and Koonin,E.V. (2000) The bacterial replicative helicase DnaB evolved from a RecA duplication. *Genome Res.*, **10**, 5–16.
56. Hillig,R.C., Renault,L., Vetter,I.R., Drell,T.,4h, Wittinghofer,A. and Becker,J. (1999) The crystal structure of rna1p: a new fold for a GTPase-activating protein. *Mol. Cell*, **3**, 781–791.
57. Bourne,H.R. (1997) G proteins. The arginine finger strikes again. *Nature*, **389**, 673–674.
58. Ahmadian,M.R., Stege,P., Scheffzek,K. and Wittinghofer,A. (1997) Confirmation of the arginine-finger hypothesis for the GAP-stimulated GTP-hydrolysis reaction of Ras. *Nature Struct. Biol.*, **4**, 686–689.
59. Leipe,D.D., Wolf,Y.I., Koonin,E.V. and Aravind,L. (2002) Classification and evolution of P-loop GTPases and related ATPases. *J. Mol. Biol.*, **317**, 41–72.
60. Leipe,D.D., Koonin,E.V. and Aravind,L. (2003) Evolution and classification of P-loop kinases and related proteins. *J. Mol. Biol.*, **333**, 781–815.
61. Lupas,A.N. and Martin,J. (2002) AAA proteins. *Curr. Opin. Struct. Biol.*, **12**, 746–753.
62. Holm,L. and Sander,C. (1995) Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.*, **20**, 478–480.
63. Mitchell,M.S., Matsuzaki,S., Imai,S. and Rao,V.B. (2002) Sequence analysis of bacteriophage T4 DNA packaging/terminase genes 16 and 17 reveals a common ATPase center in the large subunit of viral terminases. *Nucleic Acids Res.*, **30**, 4009–4021.
64. Bakhtiari,N., Lai-Zhang,J., Yao,B. and Mueller,D.M. (1999) Structure/function of the beta-barrel domain of F1-ATPase in the yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **274**, 16363–16369.
65. Llosa,M., Gomis-Ruth,F.X., Coll,M. and de la Cruz Fd,F. (2002) Bacterial conjugation: a two-step mechanism for DNA transport. *Mol. Microbiol.*, **45**, 1–8.
66. Ding,Z., Atmakuri,K. and Christie,P.J. (2003) The outs and ins of bacterial type IV secretion substrates. *Trends Microbiol.*, **11**, 527–535.
67. Burrus,V., Pavlovic,G., Decaris,B. and Guedon,G. (2002) Conjugative transposons: the tip of the iceberg. *Mol. Microbiol.*, **46**, 601–610.
68. Fasano,A., Fiorentini,C., Donelli,G., Uzzau,S., Kaper,J.B., Margaretten,K., Ding,X., Guandalini,S., Comstock,L. and Goldblum,S.E. (1995) *Zonula occludens* toxin modulates tight junctions through protein kinase C-dependent actin reorganization, *in vitro*. *J. Clin. Invest.*, **96**, 710–720.
69. Di Pierro,M., Lu,R., Uzzau,S., Wang,W., Margaretten,K., Pazzani,C., Maimone,F. and Fasano,A. (2001) *Zonula occludens* toxin structure-function analysis. Identification of the fragment biologically active on tight junctions and of the zonulin receptor binding domain. *J. Biol. Chem.*, **276**, 19160–19165.
70. Koonin,E.V., Senkevich,T.G. and Chernos,V.I. (1993) Gene A32 product of vaccinia virus may be an ATPase involved in viral DNA packaging as indicated by sequence comparisons with other putative viral ATPases. *Virus Genes*, **7**, 89–94.
71. Iyer,L.M., Aravind,L. and Koonin,E.V. (2001) Common origin of four diverse families of large eukaryotic DNA viruses. *J. Virol.*, **75**, 11720–11734.
72. La Scola,B., Audic,S., Robert,C., Jungang,L., de Lamballerie,X., Drancourt,M., Birtles,R., Claverie,J.M. and Raoult,D. (2003) A giant virus in amoebae. *Science*, **299**, 2033.
73. Wuitschick,J.D., Gershan,J.A., Lochowicz,A.J., Li,S. and Karrer,K.M. (2002) A novel family of mobile genetic elements is limited to the germline genome in *Tetrahymena thermophila*. *Nucleic Acids Res.*, **30**, 2524–2537.
74. Rice,G., Tang,L., Stedman,K., Roberto,F., Spuhler,J., Gillitzer,E., Johnson,J.E., Douglas,T. and Young,M. (2004) The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. *Proc. Natl Acad. Sci. USA*, **3**, 3.

75. Edgell,D.R. and Doolittle,W.F. (1997) Archaea and the origin(s) of DNA replication proteins. *Cell*, **89**, 995–998.
76. Leipe,D.D., Aravind,L. and Koonin,E.V. (1999) Did DNA replication evolve twice independently? *Nucleic Acids Res.*, **27**, 3389–3401.
77. Andrews,P.D., Knatko,E., Moore,W.J. and Swedlow,J.R. (2003) Mitotic mechanics: the auroras come into view. *Curr. Opin. Cell Biol.*, **15**, 672–683.
78. Heald,R.W. (2004) Cell division: burning the spindle at both ends. *Nature*, **427**, 300–301.
79. Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2003) REBASE: restriction enzymes and methyltransferases. *Nucleic Acids Res.*, **31**, 418–420.
80. Aravind,L., Galperin,M.Y. and Koonin,E.V. (1998) The catalytic domain of the P-type ATPase has the haloacid dehalogenase fold. *Trends Biochem. Sci.*, **23**, 127–129.
81. Karumbati,A.S., Deshpande,R.A., Jilani,A., Vance,J.R., Ramotar,D. and Wilson,T.E. (2003) The role of yeast DNA 3′-phosphatase Tpp1 and rad1/Rad10 endonuclease in processing spontaneous and induced base lesions. *J. Biol. Chem.*, **278**, 31434–31443.
82. Strunnikov,A.V. and Jessberger,R. (1999) Structural maintenance of chromosomes (SMC) proteins: conserved molecular properties for multiple biological functions. *Eur. J. Biochem.*, **263**, 6–13.
83. Jessberger,R. (2002) The many functions of SMC proteins in chromosome dynamics. *Nature Rev. Mol. Cell Biol.*, **3**, 767–778.
84. van den Bosch,M., Bree,R.T. and Lowndes,N.F. (2003) The MRN complex: coordinating and mediating the response to broken chromosomes. *EMBO Rep.*, **4**, 844–849.
85. Borde,V., Lin,W., Novikov,E., Petrini,J.H., Lichten,M. and Nicolas,A. (2004) Association of Mre11p with double-strand break sites during yeast meiosis. *Mol. Cell*, **13**, 389–401.
86. Schleiffer,A., Kaitna,S., Maurer-Stroh,S., Glotzer,M., Nasmyth,K. and Eisenhaber,F. (2003) Kleisins: a superfamily of bacterial and eukaryotic SMC protein partners. *Mol. Cell*, **11**, 571–575.
87. Christie,P.J. (2001) Type IV secretion: intercellular transfer of macromolecules by systems ancestrally related to conjugation machines. *Mol. Microbiol.*, **40**, 294–305.
88. Lanka,E. and Wilkins,B.M. (1995) DNA processing reactions in bacterial conjugation. *Annu. Rev. Biochem.*, **64**, 141–169.
89. Ilyina,T.V. and Koonin,E.V. (1992) Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. *Nucleic Acids Res.*, **20**, 3279–3285.
90. Datta,S., Larkin,C. and Schildbach,J.F. (2003) Structural insights into single-stranded DNA binding and cleavage by F factor TraI. *Structure (Camb.)*, **11**, 1369–1379.
91. Guasch,A., Lucas,M., Moncalian,G., Cabezas,M., Perez-Luque,R., Gomis-Ruth,F.X., de la Cruz,F. and Coll,M. (2003) Recognition and processing of the origin of transfer DNA by conjugative relaxase TrwC. *Nature Struct. Biol.*, **10**, 1002–1010.
92. Anantharaman,V. and Aravind,L. (2003) Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. *Genome Biol.*, **4**, R11.
93. Moazed,D. (2001) Enzymatic activities of Sir2 and chromatin silencing. *Curr. Opin. Cell Biol.*, **13**, 232–238.
94. Min,J., Landry,J., Sternglanz,R. and Xu,R.M. (2001) Crystal structure of a SIR2 homolog-NAD complex. *Cell*, **105**, 269–279.
95. Zhao,K., Chai,X. and Marmorstein,R. (2004) Structure and substrate binding properties of cobB, a Sir2 homolog protein deacetylase from *Escherichia coli*. *J. Mol. Biol.*, **337**, 731–741.
96. Zhao,K., Chai,X. and Marmorstein,R. (2003) Structure of the yeast Hst2 protein deacetylase in ternary complex with 2′-O-acetyl ADP ribose and histone peptide. *Structure (Camb.)*, **11**, 1403–1411.
97. Finnin,M.S., Donigian,J.R. and Pavletich,N.P. (2001) Structure of the histone deacetylase SIRT2. *Nature Struct. Biol.*, **8**, 621–625.
98. Deshpande,R.A. and Shankar,V. (2002) Ribonucleases from T2 family. *Crit. Rev. Microbiol.*, **28**, 79–122.
99. Mazumder,R., Iyer,L.M., Vasudevan,S. and Aravind,L. (2002) Detection of novel members, structure-function analysis and evolutionary classification of the 2H phosphoesterase superfamily. *Nucleic Acids Res.*, **30**, 5229–5243.
100. Koonin,E.V. (1996) A duplicated catalytic motif in a new superfamily of phosphohydrolases and phospholipid synthases that includes poxvirus envelope proteins. *Trends Biochem. Sci.*, **21**, 242–243.
101. Zaremba,M., Urbanke,C., Halford,S.E. and Siksnys,V. (2004) Generation of the BfiI restriction endonuclease from the fusion of a DNA recognition domain to a non-specific nuclease from the phospholipase D superfamily. *J. Mol. Biol.*, **336**, 81–92.
102. Pouliot,J.J., Yao,K.C., Robertson,C.A. and Nash,H.A. (1999) Yeast gene for a Tyr-DNA phosphodiesterase that repairs topoisomerase I complexes. *Science*, **286**, 552–555.
103. Yanai,I., Wolf,Y.I. and Koonin,E.V. (2002) Evolution of gene fusions: horizontal transfer versus independent events. *Genome Biol.*, **3**, research0024.
104. Leach,D.R., Okely,E.A. and Pinder,D.J. (1997) Repair by recombination of DNA containing a palindromic sequence. *Mol. Microbiol.*, **26**, 597–606.
105. Pan,X. and Leach,D.R. (2000) The roles of mutS, sbcCD and recA in the propagation of TGG repeats in *Escherichia coli*. *Nucleic Acids Res.*, **28**, 3178–3184.
106. van den Ent,F., Amos,L. and Lowe,J. (2001) Bacterial ancestry of actin and tubulin. *Curr. Opin. Microbiol.*, **4**, 634–638.
107. Amos,L.A., van den Ent,F. and Lowe,J. (2004) Structural/functional homology between the bacterial and eukaryotic cytoskeletons. *Curr. Opin. Cell Biol.*, **16**, 24–31.
108. Anantharaman,V. and Aravind,L. (2004) The SHS2 module is a common structural theme in functionally diverse protein groups, like Rpb7p, FtsA, GyrI, and MTH1598/TM1083 superfamilies. *Proteins*, in press.