ABCF ATPases involved in protein synthesis, ribosome assembly and antibiotic resistance: structural and functional diversification across the tree of life

Victoria Murina*[1,2], Marje Kasari*[1], Michael Reith[1], Vasili Hauryliuk[1,2,3] and Gemma Catherine Atkinson[†1]

[1] Department of Molecular Biology, Umeå University, 901 87, Umeå, Sweden

[2] Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå University, 901 87, Umeå, Sweden

[3] University of Tartu, Institute of Technology, Nooruse 1, 50411 Tartu, Estonia

* These authors contributed equally

[†] To whom correspondence should be addressed

Abstract

Within the larger ABC superfamily of ATPases, ABCF family members eEF3 in *Saccharomyces cerevisiae* and EttA in *Escherichia coli* have been characterised as ribosomal translation factors. Several other ABCFs including VgaA and LsaA confer resistance to MLS-type ribosome-targeting antibiotics. However, the diversity of ABCF subfamilies, the relationships among subfamilies and the evolution of antibiotic resistance factors from other ABCFs have not been explored. To address this, we analysed the presence of ABCFs and their domain architectures in 4505 genomes across the tree of life. We find that there are 45 distinct subfamilies of ABCFs, which are widespread across bacterial and eukaryotic phyla, suggesting they were present in the last common ancestor of both. Surprisingly, currently known antibiotic resistance (ARE) ABCFs are not confined to a distinct lineage of the ABCF family tree. This suggests that either antibiotic resistance is a general feature of bacterial ABCFs, or it is relatively easy to evolve antibiotic resistance from other ABCF functions. While eEF3 was thought to be fungi-specific, we have found eEF3-like factors in a range of single celled eukaryotes, suggesting an ancient origin in this domain of life. Finally, we address ribosome association of the four *E. coli* ABCFs EttA, YbiT, YheS and Uup through polysome profiling combined with Western blot detection, as well as ^{35}S-methionine pulse labeling assays upon expression of the ATPase deficient ABCF mutants.

Introduction

Protein biosynthesis – translation – is the reading and deciphering of information coded in genes to produce proteins. It is one of the most ancient and central cellular processes, and control of the various stages of translation is achieved via an intricate interplay of multiple molecular interactions. For many years, enzymatic control of the ribosomal cycle was thought to be orchestrated mainly by translational GTPases (trGTPases), and as such those proteins have had the close attention of the ribosome community for decades. Recently that view has been expanded by the identification of multiple ATPases in the ABC superfamily that have important roles in translational regulation on the ribosome. The ABC protein eEF3 (eukaryotic Elongation Factor 3) has long been known as a third factor required, in addition to the GTPases eEF1A and eEF2, for polypeptide elongation in *Saccharomyces cerevisiae* (Skogerson and Wakatama 1976). However, its specific molecular role as a ribosome E-site tRNA release factor was established only relatively recently (Triana-Alonso, et al. 1995; Andersen, et al. 2006). In addition to its role in elongation, eEF3 also participates in ribosomal recycling, i.e. disassembly of the ribosome upon completion of the polypeptide (Kurata, et al. 2010). This fungi-specific translational ABC ATPase seemed to be an exception to the general rule that trGTPases rule the ribosome, until ABCE1 (also known as Rli1), a highly conserved protein in eukaryotes and archaea was identified as another ribosome recycling factor promoting the dissociation of the ribosome into subunits (Pisarev, et al. 2010; Becker, et al. 2012; Young, et al. 2015). The ribosome-binding domain of ABCE1 is unique to this protein, suggesting that ribosome interactions of different ABC ATPases have evolved convergently.

The ABC ATPases together comprise one of the most ancient superfamilies of proteins, evolving well before the last common ancestor of life on earth (Caetano-Anolles, et al. 2011). The superfamily contains members with wide varieties of functions, but is probably best known for its many transporters, some of which are important for antibiotic resistance (Davidson, et al. 2008). Families of proteins within the ABC superfamily are named alphabetically ABCA to ABCH, following the nomenclature of the human proteins (Dean, et al. 2001). While most ABCs carry membrane-spanning domains (MSDs), these are lacking in ABCE and ABCF families (Kerr 2004). ABCE1 is the sole orthologue of the ABCE family of ABCs, but the ABCF family contains a number of subfamilies found in bacterial and eukaryotes (Kerr 2004). The ABCF family includes eEF3, and also other ribosome-associated proteins: Gcn20 is involved in sensing starvation by the presence of uncharged tRNAs on the eukaryotic ribosome (Vazquez de Aldana, et al. 1995); ABC50 (ABCF1) promotes translation initiation in eukaryotes (Paytubi, et al. 2009); and both Arb1 (ABCF2) and New1p have been proposed to be involved in biogenesis of the eukaryotic ribosome (Dong, et al. 2005; Li, et al. 2009). The adaptation of ABCE and F proteins for ribosome-associated activities seemed to be a eukaryotic and archaeal invention until characterisation of ABCF member EttA (energy-dependent translational throttle) found in diverse bacteria. *Escherichia coli* EttA was

found to bind to the E site of the ribosome where it was proposed to 'throttle' the elongation stage of translation in response to change in the intercellular ATP/ADP ratio (Boël, et al. 2014; Chen, Boel, et al. 2014). EttA is one of four *E. coli* ABCFs, the others being YheS, Uup and YbiT (Boël, et al. 2014). There is no functional information available for YheS, and only very limited data for Uup and YbiT; YbiT plays a role in infective fitness of plant pathogen *Erwinia chrysanthemi* (Llama-Palacios, et al. 2002), while Uup appears to be a DNA binding protein involved in transposon excisions (Murat, et al. 2006; Burgos Zepeda, et al. 2010) and has a role in competitive fitness and contact-dependent inhibition (CDI) (Murat, et al. 2008). DNA binding by Uup is mediated by its C terminal domain (CTD), which forms a coiled coil structure (Carlier, et al. 2012).

ABCF family members have been found to confer resistance to ribosome-inhibiting antibiotics widely used in clinical practise, such as ketolides (Wolter, et al. 2008), lincosamides (Singh, et al. 2002; Ohki, et al. 2005; Novotna and Janata 2006; Hot, et al. 2014), macrolides (Ross, et al. 1990; Buche, et al. 1997), oxylidiones (Wang, et al. 2015), phenicols (Wang, et al. 2015), pleuromutilins (Hot, et al. 2014) and streptogramins A (Kitani, et al. 2010; Hot, et al. 2014) and B (Ross, et al. 1990). These antibiotic resistance (AREs) ABCFs have been identified in antibiotic-resistant clinical isolates of *Staphylococcus*, *Streptomyces* and *Enterococcus* among others (Kerr, et al. 2005). This includes the so-called ESKAPE pathogens *Enterococcus faecium* and *Staphylococcus aureus* that contribute to a substantial proportion of hospital-acquired drug-resistant infections (Rice 2010). As some efflux pumps carry the ABC ATPase domain, it was originally thought that ARE ABCFs confer resistance by pumping out antibiotics. However, as they do not carry the necessary transmembrane domains, they are unlikely to be efflux pumps (Wilson 2016). In support of this, it was recently shown that *Staphylococcus aureus* ARE VgaA protects translation in lysates from inhibition by MLS antibiotics and that *Enterococcus faecalis* ARE LsaA displaces radioactive lincomycin from vacant 70S S. aureus ribosomes (Sharkey, et al. 2016). The ARE ABCFs therefore appear to act in an analogous way to the trGTPase Tet proteins that dislodge tetracycline from the ribosome (Donhofer, et al. 2012; Li, et al. 2013; Atkinson 2015). However, how ABCF AREs function on the ribosome to confer antibiotic resistance is still unclear. The ARE may physically interact with the drug and displace it from the ribosome, or may allosterically induce a change in conformation of the ribosome that leads to drug drop-off.

With the exception of ABCE1, all the translational ATPases identified to date belong to the ABCF family within the ABC superfamily. Thus, this large but previously overlooked family is likely to be a rich source of ribosome interacting proteins; both translation and antibiotic resistance factors. Here, we carry out an in depth survey of the diversity of ABCFs across all species with sequenced genomes. We find 45 groups (15 in eukaryotes and 30 in bacteria), including 7 groups of AREs. There is no clear sequence signature for antibiotic resistance ABCFs, suggesting that there maybe

be multiple ways of evolving antibiotic resistance in the ABCF family, or that modulating the conformation of the ribosome to remove a co-factor is a general function of ABCFs. So-called EQ_2 mutations, double glutamic acid to glutamine substitutions in the two ATPase active sites of EttA lock the enzyme in an ATP-bound conformation, inhibiting protein synthesis and cellular growth (Boël, et al. 2014; Chen, Boel, et al. 2014). We have tested the effect of equivalent mutations in the other three *E. coli* ABCFs, YbiT, Yhes and Uup in addition to EttA. We find that all four mutants inhibit growth and protein synthesis and are found in ribosomal fractions of sucrose gradient centrifugations, together indicating their association with translating ribosomes.

Results

The bioinformatic protocol was an iterative process of sequence searching and phylogenetic analysis to identify candidate subfamilies of ABCFs and then refine the classifications. Sequence searching was carried out against a local database of 4505 genomes from across the tree of life. To first get an overview of the breadth of diversity of ABCFs across life, sequence searching began with a local BlastP search against a translated coding sequence database limited by taxonomy to one representative per class, or order if there was no information on class for that species in the NCBI taxonomy database. From phylogenetic analysis of the hits, preliminary groups were identified and extracted to make Hidden Markov Models (HMMs) for further sequence searching and classification. Additional sequences from known ABCF AREs were included in phylogenetic analyses in order to identify groups of ARE-like ABCFs. HMM searching was carried out at the genus level followed by phylogenetic analysis to refine subfamily identification, with final predictions made at the species level. The resulting classification of 16848 homologous sequences comprises 45 subfamilies, 15 in eukaryotes and 30 in bacteria. Phylogenetic analysis of representatives across the diversity of ABCFs shows a roughly bipartite structure, with most eukaryotic sequences being excluded from those of bacteria with strong support (fig. 1). Five eukaryotic groups that fall in the bacterial part of the tree are likely to be endosymbiotic in origin (see the section *Bacteria-like eukaryotic ABCFs*, below).

Distribution of ABCFs

ABCFs are widespread among bacteria and eukaryotes; there are on average four ABCFs per bacterial genome, and five per eukaryotic genome. However, there is a large variation in how widespread each subfamily is (table 1). The presence of all subfamilies in each genome considered here is shown in table S1, with the full set of sequence IDs and domain composition recorded in table S2. Actinobacteria, and Firmicutes are the phyla with the largest numbers of ABCFs (up to 11 per genome), due to expansions in ARE and potential novel ARE ABCF subfamilies. In eukaryotes, it

is plants and algae that have the most subfamilies, probably due to gene acquisition from endosymbiosis, and – in some cases – secondary endosymbiosis events. The diatom *Fragilariopsis cylindrus* has 30 ABCFs and Haptophyte *Emiliania huxleyi* has 26. Bacterial contamination can sometimes inflate the number of genes in eukaryotic genomes as noted previously for trGTPases (Atkinson 2015). However, as all the *Fragilariopsis* and *Emiliania* sequences belong to typically eukaryotic subgroups, they do not appear to be the result of bacterial contamination. The Tibetan antelope *Pantholops hodgsonii*, on the other hand has 25 ABCFs, 20 of which belong to bacterial subgroups. The genome is known to be contaminated by *Bradyrhizobium* sequences (Laurence, et al. 2014), and thus the bacteria-like hits from *P. hodgsonii* are most likely artifacts.

The wide distribution and multi-copy nature of ABCFs suggests an importance of these proteins. However they are not completely universal, and are absent in almost all Archaea. The euryarcheaotes *Candidatus Methanomassiliicoccus intestinalis*, *Methanomethylophilus alvus*, *Methanomassiliicoccus luminyensis*, and *Thermoplasmatales archaeon BRNA1* are the only archaea found to encode ABCFs, in each case YdiF. ABCFs are also lacking in 214 bacterial species from various phyla, including many endosymbionts. However, the only phylum that is totally lacking ABCFs is Aquificae. ABCFs are almost universal in Eukaryotes; the only genomes where they were not detected were those of Basidomycete *Postia placenta*, Microsporidium *Enterocytozoon bieneusi*, and apicomplexan genera *Theileria, Babesia* and *Eimeria*.

Many of the phylogenetic relationships among subfamilies of ABCFs are poorly resolved (fig. 1). This is not surprising, since these represent very deep bacterial relationships that predate the diversification of major phyla and include gene duplication, and differential diversification and loss, and likely combine both vertical inheritance and horizontal gene transfer. Although relationships can not be resolved among all subfamilies, some deep relationships do have significant support; EttA and Uup share a common ancestor to the exclusion of other ABCFs with full support (fig. 1, fig. S1) and YheS is the closest bacterial group to the eukaryotic ABCFs with strong support (91% maximum likelihood bootstrap percentage (MLBP) and 1.0 Bayesian inference posterior probability (BIPP); (fig. 1, fig. S1). This latter observation suggests that eukaryotic-like ABCFs evolved from within the diversity of bacterial ABCFs. However, this depends on the root of the ABCF family tree. To address this, phylogenetic analysis was carried out of all ABCFs from *E. coli*, *Homo sapiens*, *S. cerevisiae* and *Bacillus subtilis*, along with ABCE family sequences from the UniProt database (The UniProt Consortium 2017). Rooting with ABCE does not provide statistical support for a particular group at the base of the tree, but does support the eukaryotic subgroups being nested within bacteria, with YheS as the closest bacterial group to the eukaryotic types (fig. 2). It also shows eEF3 and New1 as nested within the rest of the ABCF family, thus confirming its identity as an ABCF, despite its unusual domain structure.

Domain architectures in the ABCF family

The most common domain and subdomain structure is an N terminal ABC1 nucleotide binding domain (NBD) containing an internal Arm subdomain, followed by the Linker region joining to the ABC2 NBD (fig. 3A). Variations on this basic structure include deletions in the Arm and Linker regions, insertion of a Chromo subdomain in the ABC2 NBD, and extension of N and C termini by sequence extensions (fig. 3A). The domain structures of eukaryotic ABCFs are more diverse than those of bacteria, with greater capacity for extensions of the N terminal regions to create new domains (fig. 3A). An increased propensity to evolve extensions, especially at the N terminus is also seen with eukaryotic members of the trGTPase family (Atkinson 2015). In bacteria, terminal extensions of ABCFs tend to be at the C terminus (fig. 3A).

Cryo-electron microscopy structures show that eEF3 and EttA have very different binding sites on the ribosome (Andersen, et al. 2006; Chen, Boel, et al. 2014), and this is reflected in their domain architectures. eEF3 does not have the Arm subdomain of EttA that binds the L1 stalk and ribosomal protein L1, but carries a Chromo (Chromatin Organization Modifier) subdomain in ABC2 that – from a different orientation to the Arm domain of EttA - interacts with, and stabilizes the conformation of the L1 stalk (Andersen, et al. 2006). The Arm subdomain is also missing in eEF3-like close relatives New1 and eEF3L. Although the Arm is widespread in bacterial ABCFs, it is not universal; it is greatly reduced in a number of subfamilies, with the most drastic loss seen in ARE1 and 2 (figs. 3A and see *Putative ARE* section, below). Subdomain composition can even vary within subfamilies; Uup has lost its arm independently in multiple lineages (fig. 1, fig. S1).

Arms, linkers and CTD extensions are poorly conserved at the primary sequence level, but are similar in terms of composition, all being rich in charged amino acids, particularly arginine and lysine. Their variable presence and length suggests they can be readily extended or reduced during evolution. The CTD of Uup forms a coiled coil structure that is capable of binding DNA (Carlier, et al. 2012). However, whether DNA binding is its primary function is unclear. The CTD of YheS has significant sequence similarity (E value 3.58e-03) to the tRNA binding CTD of Valine-tRNA synthetase (NCBI conserved domains database accession cl11104). *In silico* coiled coil prediction suggests there is a propensity of all these regions to form coiled coil structures (fig. 3B). Extensions and truncations of the arms, linkers and CTD extensions possibly modulate the length of coiled coil protrusions that extend from the globular mass of the protein.

Eukaryotic ABCFs

eEF3, New1 and eEF3L

Together, eEF3, eEF3L and New1p are particularly distinct members of the ABCF family. eEF3 represents the classical fungal protein, while eEF3L is a more divergent group found mainly in protists (see below). eEF3 has a recent paralogue in *S. cerevisiae* (Hef3 (fig. 2), YEF3B) that apparently arose as a result of the whole genome duplication in yeast (Maurice, et al. 1998; Sarthy, et al. 1998). The conservation of domain structure in these proteins suggests they bind the ribosome similarly (fig. 3A). Ribosome binding by eEF3 involves the Chromo subdomain, and the HEAT domain (Andersen, et al. 2006), which in addition to New1 and eEF3L, is also found in the protein Gcn1. eEF3 has a distinct C-terminal extension (fig. 3A), through which it interacts with eEF1A (Anand, et al. 2003). It has been suggested that this leads to the recruitment of eEF1A to *S. cerevisiae* ribosomes (Anand, et al. 2003; Anand, et al. 2006). The New1 CTD contains a region of sequence similarity to eEF3; both contain a polylysine/arginine-rich tract of 20-25 amino acids (fig. S2). In *S. cerevisiae* eEF3, this is at positions 1009-1031, which falls within the eEF1A-binding site. Thus eEF1A binding may be a common feature of eEF3, New1, and also eEF3L, which commonly includes an eEF3-like C-terminal extension (table S2).

While eEF3 was thought to be fungi-specific (Belfield and Tuite 1993), we show eEF3-like (eEF3L) factors are found in a range of eukaryotes including choanoflagellates, haptophytes, heterokonts, dinoflagellates, cryptophytes and red and green algae. This suggests the progenitor of eEF3 was an ancient protein within eukaryotes, and has been lost in a number of taxonomic lineages. Alternatively, eEF3L may have been horizontally transferred in eukaryotes. eEF3 is found in *Chlorella* viruses (Yamada, et al. 1993) and *Phaeocystis* viruses (fig. S3), suggesting this may be a medium of transfer. There is no "smoking gun" for viral-mediated transfer in the phylogenies, in that eukaryotic eEF3L sequences do not nest within viral sequence clades (fig. S3), however given the close association of viral and protist eEF3L, this still remains a possibility. Curiously, the taxonomic distribution of eEF3L in diverse and distantly related protists is similar to that of the unusual elongation factor 1 (eEF1A) paralogue EFL (Atkinson, et al. 2014). EFL is also found in viruses such as *Aureococcus anophagefferens* virus (NCBI accession YP_009052194.1). As eEF3 interacts with eEF1A (Anand, et al. 2006), the equivalents eEF3L and EFL may also interact in the organisms that encode them.

New1 in Saccharomycetale yeasts has a prion-like Y/N/Q/G repetitive region in the N terminal region before the HEAT domain (fig. 3A and fig. S2). It is unclear whether New1 is a true prion itself, i.e. that it forms infectious aggregates. However *S. cerevisiae* New1p can break Sup35 (the prionogenic eukaryotic translation termination factor eRF3) amyloid fibrils into fragments in an ATP-dependent manner (Inoue, et al. 2011). This suggests that whatever the main function of New1, in Saccharomycetales it may attenuate the stop codon readthrough that is seen in [PSI+]

(Sup35 prion active) strains of yeast. In *Schizosaccharomyces pombe*, a non-prionogenic yeast, the New1 orthologue is named Elf1, and it was suggested to have a role in mRNA export from the nucleus (Kubinski, et al. 2006).

ABCF1-7

ABCF1-7 comprise the "ancestral-type" eukaryotic ABCFs, in that they have the typical ABC domain structure that is seen in bacterial ABCFs, and they lack the Chromo and HEAT domains that are found in eEF3, New1 and eEF3L (fig. 3A). ABCF1, also called ABCF50, is essential for mammalian development (Wilcox, et al. 2017). It is involved in start site selection during translation initiation in humans (Paytubi, et al. 2009), interacting with initiation factor eIF2 via its N terminal domain (Paytubi, et al. 2008) (fig. 3A), and facilitating 5' cap-independent translation in the presence of m(6)A mRNA methylation (Coots, et al. 2017). ABC50 can be present as multiple isoforms that differ in the length of the NTD (Tyzack, et al. 2000). All of the terminal extensions found in ABCF subfamilies are biased towards charged amino acids, often present as repeated motifs. The ABCF1 NTD is the most striking however, due to its length and number of repeats (fig. S2). It contains multiple tracts of poly-lysine/arginine, poly–glutamic acid/aspartic acid and – in animals - poly-glutamine. ABCF1 and ABCF2 have moderate support as sister groups (fig. 1), and both have representatives in all eukaryotic superphyla, but are not universal. Notable absences of ABCF1 are fungi, amoebozoa and most Aves (birds) (table 1, table S1).

ABCF2 (Arb1) is essential in yeast and its disruption leads to abnormal ribosome assembly (Dong, et al. 2005). Human ABCF2 can complement an Arb1 deletion, suggesting conservation of function (Kachroo, et al. 2015). The gene is upregulated in human cancers, including tumour cell lines that are resistant to cisplatin (Bao, et al. 2017), a chemotherapy drug that binds and modifies many nucleotides, including ribosomal RNAs (Melnikov, et al. 2016). A general importance of ABCF2 in eukaryotic cells is suggested by its essentiality in yeast, but also by its broad distribution across eukaryotes. It is, however absent in the *Alveolata* superphylum (Table S1). Like other ABCF terminal extensions, the ABCF2 N terminal domain is rich in lysine, in this case lysine and alanine repeats.

Like ABCF2, human ABCF3 is associated with cancer, promoting proliferation of liver cancer cells (Zhou, et al. 2013). Mouse ABCF3 has antiviral activity against flaviviruses, interacting with the flavivirus resistance protein Oas1b (Courtney, et al. 2012). In yeast, where it is known as Gcn20, ABCF3 is a component of the general amino acid control (GAAC) response to amino acid starvation (Castilho, et al. 2014). Amino acid starvation in eukaryotes is sensed by the presence of uncharged (deacylated) tRNA in the ribosomal A-site in an analogous (but not homologous) mechanism to the stringent response of bacteria (Atkinson, et al. 2011). Unlike the stringent response, which uses the

nucleotide (p)ppGpp as a signal for downregulation of translation and upregulation of amino acid biosynthesis, starvation in eukaryotes is signaled through phosphorylation of initiation factor eIF2-alpha by the kinase Gcn2 (Castilho, et al. 2014). Gcn2 is activated by a physical interaction with the Gcn20-Gcn1 complex, which perhaps facilitates the interaction of Gcn2 with deacylated tRNA (Garcia-Barrio, et al. 2000; Castilho, et al. 2014). However, Gcn20's exact role in the process, and its ribosome and tRNA binding capabilities are unclear. Gcn20 binds to Gcn1 via the latter's HEAT-containing N terminal domain (Marton, et al. 1997), an interaction that is conserved in ABCF3 and Gcn1 of *Caenorhabditis elegans* (Hirose and Horvitz 2014). As the HEAT domain is also found in eEF3, this raises the possibility that Gcn20/ABCF3 and eEF3 interact in encoding organisms (fig. 3A). Possible support for this comes from the observation that eEF3 overexpression impairs Gcn2 activation (Visweswaraiah, et al. 2012). ABCF3 is widespread in eukaryotes, but absent in heterkont algae and archaeplastida. However these taxa carry ABCF4 and ABCF7 of unknown function, which have homologous N terminal domains with ABCF3. Thus, ABCF4 and ABCF7 may be the functional equivalents of ABCF3 in these taxa, potentially interacting with HEAT domain-containing eEF3L in organisms that encode the latter (fig. 3A). ABCF3 from four fungi (*Setosphaeria turcica,* NCBI accession XP_008030281.1; *Cochliobolus sativus,* XP_007703000.1; *Bipolaris oryzae,* XP_007692076.1; and *Pyrenophora teres,* XP_003306113.1) are fused to a protein with sequence similarity to WHI2, an activator of the yeast general stress response (Kaida, et al. 2002).

ABCF5 is a monophyletic group limited to fungi and green algae (*Volvox* and *Chlamydomonas*), with a specific NTD and CTD (fig. 1 and 2A). ABCF5 is found in a variety of Ascomycete and Basidiomycete fungi, including the yeast *Debaryomyces hansenii*, but is absent in yeasts *S. pombe* and *S. cerevisiae*. ABCF4 is a polyphyletic group of various algal and amoeba protists that can not be assigned to the ABCF5, or eEF3/eEF3L/New1 clades. In eukaryotic ABCF-specific phylogenetic analysis, ABCF4/ABCF5/eEF3/eEF3L/New1 are separated from all other eukaryotic ABCFs with moderate support (85% MLBP fig. S3). ABCF6 represents a collection of algal and amoebal sequences that associate with the ABCF1+ABCF2 clade with mixed support in phylogenetic analysis (fig. 1).

<u>*Bacteria-like eukaryotic ABCFs*</u>

Five eukaryotic subfamilies are found in the bacteria-like subtree: algAF1, algUup, mEttA, algEttA and cpYdiF (fig. 1). Given their affiliation with bacterial groups, they may have entered the cell with an endosymbiotic ancestor. Indeed, chloroplast targeting peptides are predicted at the N termini of the majority of cpYdiF sequences, and mitochondrial localization peptides at most mEttA N termini. The situation is less clear for the three remaining groups, with a mix of signal peptides, or none at all being predicted across the group members (table S3).

Bacterial ABCFs:

There are 30 groups of bacterial ABCFs, the most broadly distributed being YdiF (the subfamily is given the name of the *Bacillus subtilis* protein as it is not present in *E. coli*) (Table 1). This subfamily is a paraphyletic grouping comprising ABCF sequences that can not confidently be classified into any of the other subgroups (Figure 1). The next most broadly distributed group is Uup (*B. subtilis* protein name YfmR), which itself is paraphyletic to EttA. *B. subtilis* does not encode EttA, but does encode YbiT (*B. subtilis* name YpkA). It also encodes two ABCFs not present in E. coli: YfmM and VmlR (also known as ExpZ). VmlR is in the ARE2 subfamily, and confers resistance to virginiamycin M1 and lincomycin (Ohki, et al. 2005). Insertional disruptants of all the chromosomal ABCF genes in *B subtilis* strain 168 have been examined for resistance to a panel of nine MLS class antibiotics, and only VmlR showed any hypersensitivity (Ohki, et al. 2005). With the exception of EttA (Boël, et al. 2014; Chen, Boel, et al. 2014), and the seven antibiotic resistance ARE ABCFs, the biological roles of the other 22 bacterial ABCFs are largely obscure. As more is discovered about the function of individual ABCFs, the more it will be possible to predict functions of uncharacterised members, using the ABCF family tree as a framework.

Ribosomal interactions of E. coli *ABCFs: effects of ATPase-deficient EQ_2 mutants of EttA, YbiT, YheS and Uup on growth and protein synthesis*

E. coli carries four ABCFs, but only EttA has been shown to bind the ribosome (Boël, et al. 2014; Chen, Boel, et al. 2014), with published data on this capability lacking for the other three. Ribosome binding by EttA was determined using mutations of conserved catalytic glutamate residues following the Walker B motif in the two ABC domains (Boël, et al. 2014; Chen, Boel, et al. 2014). Simultaneous mutation of both glutamate residues for glutamine (EQ_2) renders ABC ATPases deficient in ATP hydrolysis (Orelle, et al. 2003), locking the enzymes in an ATP-bound active conformation (Smith, et al. 2002; Chen, Böel, et al. 2014). In the case of EttA, expression of the EQ_2 mutant results in a dominant-negative phenotype as without ATP hydrolysis, EttA is locked on the ribosome, inhibiting protein synthesis and, consequently, bacterial growth (Boël, et al. 2014). We used inducible expression of EQ_2 mutants of *E. coli* ABCFs to probe and compare their role in translation. Wild type and EQ_2 mutants of EttA, YbiT, YheS and Uup with N-terminal FLAG TEV 6His (FTH)-tags were expressed from a low copy pBad plasmid under an arabinose-inducible araBAD (pBAD) promoter in an *E. coli* BW25113 strain that can not metabolise arabinose (*Δ(araD-araB)567* genotype) (Grenier, et al. 2014). The effect on growth was followed using optical density at 600 nm (OD_{600}), and the effect of ABCF-EQ_2 expression on protein was probed by pulse labeling with radioactive ^{35}S-methionine. Induction by L-arabinose results in growth inhibition in all four mutant strains, but not the corresponding wild types (fig. 4 A-D). This, however, does not mean

that all of the ABCFs interact with the ribosome – locking on any essential target would cause a growth arrest. The ^{35}S-methionine pulse-labelling assay provides a more direct readout of translational inhibition. In the case of all of the ABCFs, the methionine incorporation decreases showing protein synthesis is clearly inhibited. The strongest effect is observed for YbiT-EQ_2 and EttA-EQ_2, and the weakest in the case of Uup-EQ_2 (fig. 4A and C). Finally, we directly probed the ribosomal association of the four FTH-tagged EQ_2 ABCF mutants using polysome profiling in a sucrose gradient combined with Western blotting with anti-FLAG antibodies. Similarly to ribosome-associated EttA (see Supplementary Information of Boël *et al.* (2014)), all the four ABCF-EQ_2 proteins are detectable by Western blotting well into the polysomal fractions (fig. S4 A to F).

Putative AREs

Through phylogenetic analysis of predicted proteins and previously documented AREs with sequences available in UniProt (The UniProt Consortium 2017) and the Comprehensive Antibiotic Resistance Database (CARD) (Jia, et al. 2017), we have identified seven groups of AREs (fig. 1, table 1). Surprisingly, these can be quite variable in their subdomain architecture (fig. 5A and B). Some AREs (ARE1-5) have experienced extension of the Linker by on average around 30 amino acids, which can potentially bring the Linker into close contact with the bound antibiotic (fig. 5A). However, linker extension is not the rule for AREs; ARE7 (OptrA) has a linker of comparable length to the *E. coli* ABCFs (fig. 5B). Similarly, the Arm subdomain that in EttA interacts with the L1 ribosomal protein and the L1 stalk rRNA (fig. 5A) varies in length, and the CTD extension may or not be present (fig. 5B). Surprisingly, the Uup protein from cave bacterium *Paenibacillus sp. LC231*, a sequence that is unremarkable among Uups, confers resistance to the pleuromutilin antibiotic tiamulin (Pawlowski, et al. 2016) (fig. S1).

Actinobacteria are the source of many ribosome-targeting antibiotics (Mahajan and Balachandran 2012), and they have evolved measures to protect their own ribosomes, including ARE4 (OleB) and ARE5 (VarM). In addition to these known AREs, Actinobacteria encode a number of other ABCFs specific to this phylum (AAF1-6; table 1). It is possible that some – or all – of these groups are in fact AREs. Other subfamilies may also be unidentified AREs, but two particularly strong candidates are BAF2 and BAF3 that have strong support for association with ARE5 (fig. 1) and are found in a wide range of bacteria (table 1). PAF1 is also worthy of investigation; it is found in the genomes of several pathogens in the proteobacterial genera *Vibrio*, *Enterobacter*, *Klebsiella*, *Serratia* and *Citrobacter*, but not *Escherichia* (table S1).

Discussion

Towards a general model for non-eEF3 ABCF function

The conserved core domain structure that non-eEF3 ABCFs share with EttA suggests they may all bind to the same site of the ribosome, the E site, analogously to most trGTPase family members binding the A site (Atkinson 2015). Like the trGTPases, ABCFs can act during translation (Paytubi, et al. 2009; Boël, et al. 2014; Chen, Boel, et al. 2014) and during ribosome biogenesis (Dong, et al. 2005). In the case of ABCFs that act on the assembled ribosome during translation, they would presumably bind most effectively when the E site is vacant. Specifically, this would be when the E site has not yet received a tRNA (during initiation, where ABCF50 (ABCF1) functions (Paytubi, et al. 2009)) or when an empty tRNA has dissociated and not been replaced (during slow or stalled translation such as in the presence of an MLS antibiotic). EttA has been reported to promote the first peptide bond after initiation, via a proposed model of modulating the conformation of the PTC (Boël, et al. 2014; Chen, Boel, et al. 2014). This structural modulation or stabilisation could conceivably be a general function of ABCFs, with the specific ribosomal substrate differing, depending on the stage of translation, assembly or cellular conditions. Differences in subdomains would determine both what is sensed and what is the resulting signal. For instance the presence of the Arm would affect signal transmission between the PTC and the L1 protein and/or the L1 stalk.

Evolution of AREs

In order to determine antibiotic resistance capabilities and track the transfer routes of resistance, it is critical to be able to annotate antibiotic resistance genes in genomes. This requires discrimination of antibiotic genes from homologous genes from which resistance functions evolved. At present this is not straightforward for ABCF AREs, as the distinction between potential translation and antibiotic resistance factors is ambiguous. ABCF AREs have been compared to the Tet family of antibiotic resistance proteins that evolved from trGTPase EF-G to remove tetracycline from the ribosome (Wilson 2016). However the distinction between Tet proteins and EF-G is much more clear-cut, with Tet comprising a distinct lineage in the evolutionary history of trGTPases (Atkinson 2015). The surprising lack of a clear sequence signature for antibiotic resistance in the ARE ABCFs suggests that antibiotic resistance functions may evolve in multiple ways in ABCFs, and that ABCFs closely related to AREs may have similar functions to the AREs, while not conferring resistance. For example, there are multiple small molecules that bind the PTC and exit tunnel (Seip and Innis 2016), and conceivably ABCFs may be involved in sensing such cases, removing the small molecule, or allowing translation of a subset of mRNAs to continue in its presence. Even macrolide antibiotics that target the exit tunnel do not kill protein synthesis entirely, but rather reshape the translational landscape (Kannan, et al. 2012; Almutairi, et al. 2017).

Polyprolines and translatability of ABCFs

Curiously, ABCFs from both bacteria and eukaryotes are often rich in polyproline sequences, which are known to cause ribosome stalls during translation. 56% of all the sequences in the ABCF

database contain at least two consecutive prolines, compared to an overall 37% of all the proteins in the predicted proteomes considered here. There is a particularly proline-rich hotspot in the C terminal part of the linker (fig. 5), which can be up to nine consecutive prolines long in the case of Uup from *Novosphingobium aromaticivorans* (WP_028641352.1). Arms and CTD extensions are also polyproline hotspots; PPP is a common motif in the Arm of EttA proteins, and polyprolines are frequent in YdiF, Uup and YheS CTD extensions. *Phenylobacterium zucineum* YheS (YP_002130661.1) contains eight consecutive prolines in its CTD region. Prolines are rigid amino acids, and conceivably their presence may support the tertiary structures of ABCFs, particularly the orientation of the subdomain coiled coils. However, while there are hotpots, polyprolines can be distributed across the full lengths of the proteins; *Deinococcus murrayi* YdiF (WP_027459336.1) has seven distinct regions of at least two consecutive prolines, four of them in the CTD extension, with one region containing eight consecutive prolines. Proline-rich regions are found in many different types of proteins, and are often involved in protein-protein interactions (Williamson 1994). The structure of EttA (Boël, et al. 2014) suggests these proline-rich regions often occur in exposed loops, supporting a role in intermolecular interactions.

Polyprolines cause ribosome slow-down or stalling due to proline being a poor donor and acceptor for peptide-bond formation. This stalling is alleviated by the elongation factor EF-P in bacteria (Ude, et al. 2013) and eIF5A in eukaryotes (Gutierrez, et al. 2013). EttA carries two PPX motifs, PPG and PPK, and its expression is decreased if EF-P is deleted (Elgamal, et al. 2014). Translatability of the hyperpolyproline ABCFs is potentially low, and strictly dependent on EF-P/eIF5A. Paradoxically, while many ABCFs may need EF-P to translate efficiently, EF-P in the E site would actually compete with the E site-binding ABCFs. Thus there seems to be some intricate interplay between EF-P and ABCFs that for now is unresolved.

Conclusion

ABCFs are stepping into the limelight as important translation, ribosome assembly and antibiotic resistance factors. Using hydrolysis-incompetent EQ_2 mutants of *E. coli* ABCFs, we have detected ribosomal association of all ABCF-EQ_2 proteins as well as inhibition of protein synthesis. Along with known ribosome association of antibiotic resistance ABCFs, and eukaryotic ABCFs, this suggests ribosome-binding is a general - perhaps ancestral - feature of ABCFs. However, the ABCF family is diverse, and even within subfamilies, there can be differences in subdomain architecture. As more structural and functional information comes to light for these proteins, further sequence analyses will allow more precise resolution of sequence-structure-function relationships and prediction of function across the family. We have also identified clusters of antibiotic resistance ARE ABCFs, and predicted likely new AREs. Strikingly, the AREs do not form a clear monophyletic

group, and either antibiotic resistance has evolved multiple times independently from the ABCF diversity, or this is an innate ability of ABCFs, raising the possibility of a general role of ABCFs in ribosome-binding small molecule sensing and signaling. These questions can be resolved by targeted biochemical investigations of ABCFs and their resistance capabilities.

Methods

Sequence searching and classification

Predicted proteomes were downloaded from the NCBI genome FTP site (2nd December 2014). One representative was downloaded per species of bacteria (i.e. not every strain). Seven additional proteomes were downloaded from JGI (*Aplanochytrium kerguelense, Aurantiochytrium limacinum, Fragilariopsis cylindrus, Phytophthora capsici, Phytophthora cinnamomi, Pseudo-nitzschia multiseries,* and *Schizochytrium aggregatum*). Previously documented AREs were retrieved from UniProt (The UniProt Consortium 2017) and the Comprehensive Antibiotic Resistance Database (CARD) (Jia, et al. 2017). Taxonomy was retrieved from NCBI, and curated manually where some ranks were not available.

An initial local BlastP search was carried out locally with BLAST+ v 2.2.3 (Camacho, et al. 2009) against a proteome database limited by taxonomy to one representative per class or (order if there was no information on class from the NCBI taxonomy database), using EttA as the query. Subsequent sequence searching against the proteome collections used hmmsearch from HMMER 3.1b1, with HMMs made from multiple sequence alignments of subfamilies, as identified below. The E value threshold for hmmsearch was set to $1e^{-70}$, a value at which the subfamily models hit outside of the eukaryotic-like or bacterial-like ABCF bipartitions, ensuring complete coverage while not picking up sequences outside of the ABCF family.

Sequences were aligned with MAFFT v7.164b (default settings) and Maximum Likelihood phylogenetic analyses were carried out with RAxML-HPC v.8 (Stamatakis 2014) on the CIPRES Science Gateway v3 (Miller, et al. 2010) using the LG model of substitution. Alignment positions with >50% gaps were excluded from the analysis. Trees were visualized with FigTree v. 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/) and visually inspected to identify putative subfamilies that preferably satisfied the criteria of 1) containing mostly orthologues, and 2) had at least moderate (>60% bootstrap support). These subfamilies were then isolated, aligned separately and used to make HMM models. Models were refined with subsequent rounds of searching and classification into subfamilies first by comparisons of the E values of HMM hits, then by curating with phylogenetic analysis. After the final classification of all ABCF types in all predicted protein sequences using HMMER, some manual correction was still required. For example, cyanobacterial

sequences always hit the chloroplast HMM with a more significant E value than the bacterial model, and eEF3/New1/eEF3L sequences could not be reliably discriminated between using E value comparisons. Therefore, the final classification is a manually curated version of that generated from automatic predictions (table S2). All sequence handling was carried out with bespoke Python scripts, and data was stored in a MySQL database (exported to Excel files for the supplementary material tables).

Domain prediction

Domain HMMs were made from subalignments extracted from subfamily alignments. Partial and poorly aligned sequences were excluded from the alignments. All significant domain hits (<E value $1e^{-3}$) for each ABCF sequence were stored in the MySQL database. Sequence logos of domains were created with WebLogo with default algorithm settings (Crooks, et al. 2004).

Putative transit peptides for mitochondrial and plastid subcellular localization were predicted with the TargetP web server hosted at the Technical University of Denmark (Emanuelsson, et al. 2007).

Phylogenetic analysis of representatives

For the representative tree of the ABCF family, taxa were selected from the ABCF database to sample broadly across the tree of life, including eukaryotic protistan phyla, while also covering all subfamilies of ABCFs. Sequences were aligned with MAFFT with the L-ins-i strategy (Katoh and Standley 2013) and positions with >50% gaps, and several ambiguously aligned positions at the termini were removed. The resulting 249 sequences, and 533 positions were subject to RAxML Maximum Likelihood phylogenetic analysis as above, using the LG substitution matrix, as favoured by ProtTest 2.4 (Abascal, et al. 2005).

Bayesian inference phylogenetic analysis was carried out with MrBayes v3.2.6, also on the CIPRES gateway. The analysis was run for 1 million generations, after which the standard deviation of split frequencies (SDSF) was 0.08. The mixed model setting was used for determining the amino acid substitution matrix, which converged on WAG. RAxML analysis with the WAG model showed no difference in topology for well-supported branches compared to the RAxML tree with the LG model.

For the rooted tree with ABCE as the outgroup, all ABCFs were selected from *Arabidopsis thaliana, Homo sapiens, Escherichia coli, Bacillus subtilis, Saccharomyces cerevisae, Schizosaccharomyces pombe* and *Methanococcus maripaludis*. Ambiguously aligned sites were identified and removed manually. RAxML and MrBayes were carried out as above on the resulting 344 positions from 35

sequences. The MrBayes analysis stopped automatically when the SDSF got down to the 0.009 threshold, which was at 235,000 generations.

For the tree of eukaryotic and viral ABCFs rooted with YheS, sequences were extracted from the ABCF database, and viral sequences were found in the NCBI protein database using BlastP. The resulting 658 sequences were aligned with MAFFT and a RAxML analysis of 645 positions was carried out as above. The alignments used to build the phylogenies presented in the main text as supplementary information are available in text S1.

Structural analyses

Homology modeling was carried out using Swiss Model (Biasini, et al. 2014) with EttA (PDB ID 3J5S) as the template structure. Because the linkers were lacking secondary structure, QUARK (Xu and Zhang 2012) was used for *ab initio* structure modeling of these regions, and the resulting coils were aligned back to the homology model using the structural alignment method of MacPyMOL (Schrodinger 2015). The presence of coiled coil regions was predicted with the COILS program hosted at the ExPASy Bioinformatics Research Portal (Lupas, et al. 1991).

Experimental methods

For descriptions of plasmid construction, cell growth, ^{35}S-methionine pulse labeling, polysome profiling and Western blotting, see Supplementary Methods.

Acknowledgements

This work was supported by the Swedish Research council (2015-04746 to GCA, grant 2013-4680 to VH); Ragnar Söderberg foundation (VH); Carl Tryggers grants CTS 14:34 & CTS 15:35 (GCA); Kempe stiftelse grant JCK-1627 (GCA); Jeanssons Styrelserna grant (GCA); Molecular Infection Medicine Sweden (MIMS) (VH); Umeå Universitet Insamlingsstiftelsen för medicinsk forskning (GCA and VH) and the European Regional Development Fund through the Centre of Excellence for Molecular Cell Technology (VH).

Author contributions:

GCA conceived the study and carried out bioinformatic analyses. VH coordinated the experimental analyses, which were carried out by VM and MK with assistance from MR. GCA drafted the manuscript with input from VH, VM and MK.

References

Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21:2104-2105.

Almutairi MM, Svetlov MS, Hansen DA, Khabibullina NF, Klepacki D, Kang HY, Sherman DH, Vazquez-Laslop N, Polikanov YS, Mankin AS. 2017. Co-produced natural ketolides methymycin and pikromycin inhibit bacterial growth by preventing synthesis of a limited number of proteins. Nucleic Acids Res 45:9573-9582.

Anand M, Balar B, Ulloque R, Gross SR, Kinzy TG. 2006. Domain and nucleotide dependence of the interaction between Saccharomyces cerevisiae translation elongation factors 3 and 1A. J Biol Chem 281:32318-32326.

Anand M, Chakraburtty K, Marton MJ, Hinnebusch AG, Kinzy TG. 2003. Functional interactions between yeast translation eukaryotic elongation factor (eEF) 1A and eEF3. J Biol Chem 278:6985-6991.

Andersen CB, Becker T, Blau M, Anand M, Halic M, Balar B, Mielke T, Boesen T, Pedersen JS, Spahn CM, et al. 2006. Structure of eEF3 and the mechanism of transfer RNA release from the E-site. Nature 443:663-668.

Atkinson GC. 2015. The evolutionary and functional diversity of classical and lesser-known cytoplasmic and organellar translational GTPases across the tree of life. BMC Genomics 16:78.

Atkinson GC, Kuzmenko A, Chicherin I, Soosaar A, Tenson T, Carr M, Kamenski P, Hauryliuk V. 2014. An evolutionary ratchet leading to loss of elongation factors in eukaryotes. BMC Evol Biol 14:35.

Atkinson GC, Tenson T, Hauryliuk V. 2011. The RelA/SpoT homolog (RSH) superfamily: distribution and functional evolution of ppGpp synthetases and hydrolases across the tree of life. PLoS ONE 6:e23479.

Bao L, Wu J, Dodson M, Rojo de la Vega EM, Ning Y, Zhang Z, Yao M, Zhang DD, Xu C, Yi X. 2017. ABCF2, an Nrf2 target gene, contributes to cisplatin resistance in ovarian cancer cells. Mol Carcinog 56:1543-1553.

Becker T, Franckenberg S, Wickles S, Shoemaker CJ, Anger AM, Armache JP, Sieber H, Ungewickell C, Berninghausen O, Daberkow I, et al. 2012. Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. Nature 482:501-506.

Belfield GP, Tuite MF. 1993. Translation elongation factor 3: a fungus-specific translation factor? Mol Microbiol 9:411-418.

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Gallo Cassarino T, Bertoni M, Bordoli L, et al. 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 42:W252-258.

Boël G, Smith PC, Ning W, Englander MT, Chen B, Hashem Y, Testa AJ, Fischer JJ, Wieden HJ, Frank J, et al. 2014. The ABC-F protein EttA gates ribosome entry into the translation elongation cycle. Nat Struct Mol Biol.

Buche A, Mendez C, Salas JA. 1997. Interaction between ATP, oleandomycin and the OleB ATP-binding cassette transporter of Streptomyces antibioticus involved in oleandomycin secretion. Biochem J 321 (Pt 1):139-144.

Burgos Zepeda MY, Alessandri K, Murat D, El Amri C, Dassa E. 2010. C-terminal domain of the Uup ATP-binding cassette ATPase is an essential folding domain that binds to DNA. Biochim Biophys Acta 1804:755-761.

Caetano-Anolles D, Kim KM, Mittenthal JE, Caetano-Anolles G. 2011. Proteome evolution and the metabolic origins of translation and cellular life. J Mol Evol 72:14-33.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC bioinformatics 10:421.

Carlier L, Haase AS, Burgos Zepeda MY, Dassa E, Lequin O. 2012. The C-terminal domain of the Uup protein is a DNA-binding coiled coil motif. J Struct Biol 180:577-584.

Castilho BA, Shanmugam R, Silva RC, Ramesh R, Himme BM, Sattlegger E. 2014. Keeping the eIF2 alpha kinase Gcn2 in check. Biochim Biophys Acta 1843:1948-1968.

Chen B, Boel G, Hashem Y, Ning W, Fei J, Wang C, Gonzalez RL, Jr., Hunt JF, Frank J. 2014. EttA regulates translation by binding the ribosomal E site and restricting ribosome-tRNA dynamics. Nat Struct Mol Biol.

Chen B, Böel G, Hashem Y, Ning W, Fei J, Wang C, Gonzalez RL, Jr., Hunt JF, Frank J. 2014. EttA regulates translation by binding the ribosomal E site and restricting ribosome-tRNA dynamics. Nat Struct Mol Biol 21:152-159.

Coots RA, Liu XM, Mao Y, Dong L, Zhou J, Wan J, Zhang X, Qian SB. 2017. m6A Facilitates eIF4F-Independent mRNA Translation. Mol Cell.

Courtney SC, Di H, Stockman BM, Liu H, Scherbik SV, Brinton MA. 2012. Identification of novel host cell binding partners of Oas1b, the protein conferring resistance to flavivirus-induced disease in mice. J Virol 86:7953-7963.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res 14:1188-1190.

Davidson AL, Dassa E, Orelle C, Chen J. 2008. Structure, function, and evolution of bacterial ATP-binding cassette systems. Microbiol Mol Biol Rev 72:317-364, table of contents.

Dean M, Rzhetsky A, Allikmets R. 2001. The human ATP-binding cassette (ABC) transporter superfamily. Genome Res 11:1156-1166.

Dong J, Lai R, Jennings JL, Link AJ, Hinnebusch AG. 2005. The novel ATP-binding cassette protein ARB1 is a shuttling factor that stimulates 40S and 60S ribosome biogenesis. Mol Cell Biol 25:9859-9873.

1 Donhofer A, Franckenberg S, Wickles S, Berninghausen O, Beckmann R, Wilson DN. 2012.
2 Structural basis for TetM-mediated tetracycline resistance. Proc Natl Acad Sci U S A 109:16900-
3 16905.
4 Elgamal S, Katz A, Hersch SJ, Newsom D, White P, Navarre WW, Ibba M. 2014. EF-P dependent
5 pauses integrate proximal and distal signals during translation. PLoS Genet 10:e1004553.
6 Emanuelsson O, Brunak S, von Heijne G, Nielsen H. 2007. Locating proteins in the cell using
7 TargetP, SignalP and related tools. Nat Protoc 2:953-971.
8 Garcia-Barrio M, Dong J, Ufano S, Hinnebusch AG. 2000. Association of GCN1-GCN20 regulatory
9 complex with the N-terminus of eIF2alpha kinase GCN2 is required for GCN2 activation. EMBO J
10 19:1887-1899.
11 Grenier F, Matteau D, Baby V, Rodrigue S. 2014. Complete Genome Sequence of Escherichia coli
12 BW25113. Genome Announc 2.
13 Gutierrez E, Shin BS, Woolstenhulme CJ, Kim JR, Saini P, Buskirk AR, Dever TE. 2013. eIF5A
14 promotes translation of polyproline motifs. Mol Cell 51:35-45.
15 Hirose T, Horvitz HR. 2014. The translational regulators GCN-1 and ABCF-3 act together to
16 promote apoptosis in C. elegans. PLoS Genet 10:e1004512.
17 Hot C, Berthet N, Chesneau O. 2014. Characterization of sal(A), a novel gene responsible for
18 lincosamide and streptogramin A resistance in Staphylococcus sciuri. Antimicrob Agents
19 Chemother 58:3335-3341.
20 Inoue Y, Kawai-Noma S, Koike-Takeshita A, Taguchi H, Yoshida M. 2011. Yeast prion protein New1
21 can break Sup35 amyloid fibrils into fragments in an ATP-dependent manner. Genes Cells 16:545-
22 556.
23 Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S, Sharma
24 AN, et al. 2017. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic
25 resistance database. Nucleic Acids Res 45:D566-D573.
26 Kachroo AH, Laurent JM, Yellman CM, Meyer AG, Wilke CO, Marcotte EM. 2015. Evolution.
27 Systematic humanization of yeast genes reveals conserved functions and genetic modularity.
28 Science 348:921-925.
29 Kaida D, Yashiroda H, Toh-e A, Kikuchi Y. 2002. Yeast Whi2 and Psr1-phosphatase form a complex
30 and regulate STRE-mediated gene expression. Genes Cells 7:543-552.
31 Kannan K, Vazquez-Laslop N, Mankin AS. 2012. Selective protein synthesis by ribosomes with a
32 drug-obstructed exit tunnel. Cell 151:508-520.
33 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
34 improvements in performance and usability. Mol Biol Evol 30:772-780.
35 Kerr ID. 2004. Sequence analysis of twin ATP binding cassette proteins involved in translational
36 control, antibiotic resistance, and ribonuclease L inhibition. Biochem Biophys Res Commun
37 315:166-173.

Kerr ID, Reynolds ED, Cove JH. 2005. ABC proteins and antibiotic drug resistance: is it all about transport? Biochem Soc Trans 33:1000-1002.

Kitani S, Yamauchi T, Fukushima E, Lee CK, Ningsih F, Kinoshita H, Nihira T. 2010. Characterization of varM Encoding Type II ABC Transporter in Streptomyces virginiae, a Virginiamycin M1 Producer. Actinomycetologica 24:51-57.

Kubinski K, Zielinski R, Hellman U, Mazur E, Szyszka R. 2006. Yeast elf1 factor is phosphorylated and interacts with protein kinase CK2. J Biochem Mol Biol 39:311-318.

Kurata S, Nielsen KH, Mitchell SF, Lorsch JR, Kaji A, Kaji H. 2010. Ribosome recycling step in yeast cytoplasmic protein synthesis is catalyzed by eEF3 and ATP. Proc Natl Acad Sci U S A 107:10854-10859.

Laurence M, Hatzis C, Brash DE. 2014. Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. PLoS ONE 9:e97876.

Li W, Atkinson GC, Thakor NS, Allas U, Lu CC, Chan KY, Tenson T, Schulten K, Wilson KS, Hauryliuk V, et al. 2013. Mechanism of tetracycline resistance by ribosomal protection protein Tet(O). Nat Commun 4:1477.

Li Z, Lee I, Moradi E, Hung NJ, Johnson AW, Marcotte EM. 2009. Rational extension of the ribosome biogenesis pathway using network-guided genetics. PLoS Biol 7:e1000213.

Llama-Palacios A, Lopez-Solanilla E, Rodriguez-Palenzuela P. 2002. The ybiT gene of Erwinia chrysanthemi codes for a putative ABC transporter and is involved in competitiveness against endophytic bacteria during infection. Appl Environ Microbiol 68:1624-1630.

Lupas A, Van Dyke M, Stock J. 1991. Predicting coiled coils from protein sequences. Science 252:1162-1164.

Mahajan GB, Balachandran L. 2012. Antibacterial agents from actinomycetes - a review. Front Biosci (Elite Ed) 4:240-253.

Marton MJ, Vazquez de Aldana CR, Qiu H, Chakraburtty K, Hinnebusch AG. 1997. Evidence that GCN1 and GCN20, translational regulators of GCN4, function on elongating ribosomes in activation of eIF2alpha kinase GCN2. Mol Cell Biol 17:4474-4489.

Maurice TC, Mazzucco CE, Ramanathan CS, Ryan BM, Warr GA, Puziss JW. 1998. A highly conserved intraspecies homolog of the Saccharomyces cerevisiae elongation factor-3 encoded by the HEF3 gene. Yeast 14:1105-1113.

Melnikov SV, Soll D, Steitz TA, Polikanov YS. 2016. Insights into RNA binding by the anticancer drug cisplatin from the crystal structure of cisplatin-modified ribosome. Nucleic Acids Res 44:4978-4987.

Miller MA, Pfeiffer W, Schwartz T editors. Gateway Computing Environments Workshop (GCE). 2010 New Orleans, LA.

Murat D, Bance P, Callebaut I, Dassa E. 2006. ATP hydrolysis is essential for the function of the Uup ATP-binding cassette ATPase in precise excision of transposons. J Biol Chem 281:6850-6859.

Murat D, Goncalves L, Dassa E. 2008. Deletion of the Escherichia coli uup gene encoding a protein of the ATP binding cassette superfamily affects bacterial competitiveness. Res Microbiol 159:671-677.

Novotna G, Janata J. 2006. A new evolutionary variant of the streptogramin A resistance protein, Vga(A)LC, from Staphylococcus haemolyticus with shifted substrate specificity towards lincosamides. Antimicrob Agents Chemother 50:4070-4076.

Ohki R, Tateno K, Takizawa T, Aiso T, Murata M. 2005. Transcriptional termination control of a novel ABC transporter gene involved in antibiotic resistance in Bacillus subtilis. J Bacteriol 187:5946-5954.

Orelle C, Dalmas O, Gros P, Di Pietro A, Jault JM. 2003. The conserved glutamate residue adjacent to the Walker-B motif is the catalytic base for ATP hydrolysis in the ATP-binding cassette transporter BmrA. J Biol Chem 278:47002-47008.

Pawlowski AC, Wang W, Koteva K, Barton HA, McArthur AG, Wright GD. 2016. A diverse intrinsic antibiotic resistome from a cave bacterium. Nat Commun 7:13803.

Paytubi S, Morrice NA, Boudeau J, Proud CG. 2008. The N-terminal region of ABC50 interacts with eukaryotic initiation factor eIF2 and is a target for regulatory phosphorylation by CK2. Biochem J 409:223-231.

Paytubi S, Wang X, Lam YW, Izquierdo L, Hunter MJ, Jan E, Hundal HS, Proud CG. 2009. ABC50 promotes translation initiation in mammalian cells. J Biol Chem 284:24061-24073.

Pisarev AV, Skabkin MA, Pisareva VP, Skabkina OV, Rakotondrafara AM, Hentze MW, Hellen CU, Pestova TV. 2010. The role of ABCE1 in eukaryotic posttermination ribosomal recycling. Mol Cell 37:196-210.

Rice LB. 2010. Progress and challenges in implementing the research on ESKAPE pathogens. Infect Control Hosp Epidemiol 31 Suppl 1:S7-10.

Ross JI, Eady EA, Cove JH, Cunliffe WJ, Baumberg S, Wootton JC. 1990. Inducible erythromycin resistance in staphylococci is encoded by a member of the ATP-binding transport super-gene family. Mol Microbiol 4:1207-1214.

Sarthy AV, McGonigal T, Capobianco JO, Schmidt M, Green SR, Moehle CM, Goldman RC. 1998. Identification and kinetic analysis of a functional homolog of elongation factor 3, YEF3 in Saccharomyces cerevisiae. Yeast 14:239-253.

Schrodinger L. (PyMOL co-authors). 2015. The PyMOL Molecular Graphics System, Version 1.8.

Seip B, Innis CA. 2016. How Widespread is Metabolite Sensing by Ribosome-Arresting Nascent Peptides? J Mol Biol 428:2217-2227.

Sharkey LK, Edwards TA, O'Neill AJ. 2016. ABC-F Proteins Mediate Antibiotic Resistance through Ribosomal Protection. MBio 7.

Singh KV, Weinstock GM, Murray BE. 2002. An Enterococcus faecalis ABC homologue (Lsa) is required for the resistance of this species to clindamycin and quinupristin-dalfopristin. Antimicrob Agents Chemother 46:1845-1850.

Skogerson L, Wakatama E. 1976. A ribosome-dependent GTPase from yeast distinct from elongation factor 2. Proc Natl Acad Sci U S A 73:73-76.

Smith PC, Karpowich N, Millen L, Moody JE, Rosen J, Thomas PJ, Hunt JF. 2002. ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer. Mol Cell 10:139-149.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-1313.

The UniProt Consortium. 2017. UniProt: the universal protein knowledgebase. Nucleic Acids Res 45:D158-D169.

Triana-Alonso FJ, Chakraburtty K, Nierhaus KH. 1995. The elongation factor 3 unique in higher fungi and essential for protein biosynthesis is an E site factor. J Biol Chem 270:20473-20478.

Tyzack JK, Wang X, Belsham GJ, Proud CG. 2000. ABC50 interacts with eukaryotic initiation factor 2 and associates with the ribosome in an ATP-dependent manner. J Biol Chem 275:34131-34139.

Ude S, Lassak J, Starosta AL, Kraxenberger T, Wilson DN, Jung K. 2013. Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. Science 339:82-85.

Vazquez de Aldana CR, Marton MJ, Hinnebusch AG. 1995. GCN20, a novel ATP binding cassette protein, and GCN1 reside in a complex that mediates activation of the eIF-2 alpha kinase GCN2 in amino acid-starved cells. EMBO J 14:3184-3199.

Visweswaraiah J, Lee SJ, Hinnebusch AG, Sattlegger E. 2012. Overexpression of eukaryotic translation Elongation Factor 3 (eEF3) impairs GCN2 activation. J Biol Chem.

Wang Y, Lv Y, Cai J, Schwarz S, Cui L, Hu Z, Zhang R, Li J, Zhao Q, He T, et al. 2015. A novel gene, optrA, that confers transferable resistance to oxazolidinones and phenicols and its presence in Enterococcus faecalis and Enterococcus faecium of human and animal origin. J Antimicrob Chemother 70:2182-2190.

Wilcox SM, Arora H, Munro L, Xin J, Fenninger F, Johnson LA, Pfeifer CG, Choi KB, Hou J, Hoodless PA, et al. 2017. The role of the innate immune response regulatory gene ABCF1 in mammalian embryogenesis and development. PLoS ONE 12:e0175918.

Williamson MP. 1994. The structure and function of proline-rich regions in proteins. Biochem J 297 (Pt 2):249-260.

Wilson DN. 2016. The ABC of Ribosome-Related Antibiotic Resistance. MBio 7.

Wolter N, Smith AM, Farrell DJ, Northwood JB, Douthwaite S, Klugman KP. 2008. Telithromycin resistance in Streptococcus pneumoniae is conferred by a deletion in the leader sequence of erm(B) that increases rRNA methylation. Antimicrob Agents Chemother 52:435-440.

1 Xu D, Zhang Y. 2012. Ab initio protein structure assembly using continuous structure fragments
2 and optimized knowledge-based force field. Proteins 80:1715-1735.
3 Yamada T, Fukuda T, Tamura K, Furukawa S, Songsri P. 1993. Expression of the gene encoding a
4 translational elongation factor 3 homolog of Chlorella virus CVK2. Virology 197:742-750.
5 Young DJ, Guydosh NR, Zhang F, Hinnebusch AG, Green R. 2015. Rli1/ABCE1 Recycles Terminating
6 Ribosomes and Controls Translation Reinitiation in 3'UTRs In Vivo. Cell 162:872-884.
7 Zhou J, Lin Y, Shi H, Huo K, Li Y. 2013. hABCF3, a TPD52L2 interacting partner, enhances the
8 proliferation of human liver cancer cell lines in vitro. Mol Biol Rep 40:5759-5767.
9

Table 1 | The subfamilies of the ABCF family, and the numbers (N) of phyla and species in which they are encoded

Subfamily	N Phyla	N Species	Notes on function, relationships and taxonomic distribution
YdiF	42	1852	Broad distribution in bacteria; polyphyletic
Uup [a]	24	3104	Broad distribution in bacteria; paraphyletic to EttA. Inc. P resistance TaeA
EttA	18	2337	Broad distribution in bacteria; translation factor
YbiT	15	1874	Broad distribution in bacteria; potential translation factor
BAF2	7	305	Proteobacteria, Planctomycetes, Spirochaetes, Actinobacteria, Bacteroidetes, Gemmatimonadetes, Cyanobacteria
ARE1[a]	6	269	M,L,S,P,K resistance, inc. VgaA, MsrA; Firmicutes, Actinobacteria, Spirochaetes, Bacteroidetes, Proteobacteria, Tenericutes
ARE3[a]	5	261	L resistance inc Lsa; Firmicutes, Spirochaetes, Proteobacteria, Fusobacteria, Actinobacteria
DAF1	5	58	Spirochaetes, Proteobacteria, Deferribacteres, Fibrobacteres, Chlamydiae
YfmM	5	587	Firmicutes, Tenericutes, Proteobacteria, Fusobacteria, Bacteroidetes
YheS	5	1234	Proteobacteria, Bacteroidetes, Cyanobacteria, Arthropoda, Elusimicrobia
BAF3	4	111	Bacteroidetes, Proteobacteria, Cyanobacteria, Elusimicrobia
DAF2	4	24	Proteobacteria, Planctomycetes, Spirochaetes, Omnitrophica
ARE2[a]	2	54	Antibiotic resistance inc. VmlR; Firmicutes, Tenericutes
ARE4[a]	2	173	M,M16 resistance inc. CarA, srmB, tlrC; Actinobacteria, Chloroflexi
ARE5[a]	2	408	L,S resistance inc. VarM, LmrC; Actinobacteria, Proteobacteria
BAF1	2	20	Firmicutes, Actinobacteria
PAF1	1	138	Proteobacteria
AAF1	1	313	Actinobacteria
AAF2	1	301	Actinobacteria
AAF4	1	219	Actinobacteria
AAF6	1	649	Actinobacteria
AAF3	1	8	Actinobacteria
AAF5	1	25	Actinobacteria
ARE6[a]	1	8	L,S resistance SalA; Firmicutes
ARE7[a]	1	35	Oxazolidinone resistance OptrA, Firmicutes
BdAF1	1	176	Bacteroidetes
DAF3	1	36	Proteobacteria
FAF1	1	37	Firmicutes
FAF2	1	42	Firmicutes
SAF1	1	66	Firmicutes
ABCF2	27	560	Arb1 ribosome biogenesis factor; broadly distributed but lacking in Apicomplexa and Microsporidia
ABCF1	23	376	ABC50 translation initiation factor; found in plants, diverse algae, and opisthokonts excluding fungi
ABCF7	16	131	Found in plants, diverse algae, Alveolata, Excavata and Microsporidia
ABCF3	13	382	Gcn20 starvation response; Opisthokonts
eEF3L	9	83	Diverse algae, Chytridiomycota and choanoflagellates
ABCF4	7	37	Diverse algae and Filozoa
ABCF5	7	105	Diverse algae and fungi
cpYdiF	7	85	Diverse algae
algAF1	6	25	Diverse algae
cpEttA	6	18	Chloroplast targeting peptides predicted; plants and diverse algae
algUup	5	17	Found in diverse algae
ABCF6	5	17	Found in diverse algae and Amoebozoa
mEttA	3	14	mitochondrial targeting peptides predicted; diverse algae, Amoebozoa
New1	3	127	Fungi
eEF3	2	138	Translation factor; fungi

Notes - [a] Subfamilies containing known AREs; resistance to antibiotic class in notes column is as follows: M, 14- and 15-membered ring macrolides; M16, 16-membered ring macrolides; L, lincosamides; S, streptogramins; K, ketolides; P, Pleuromutilins

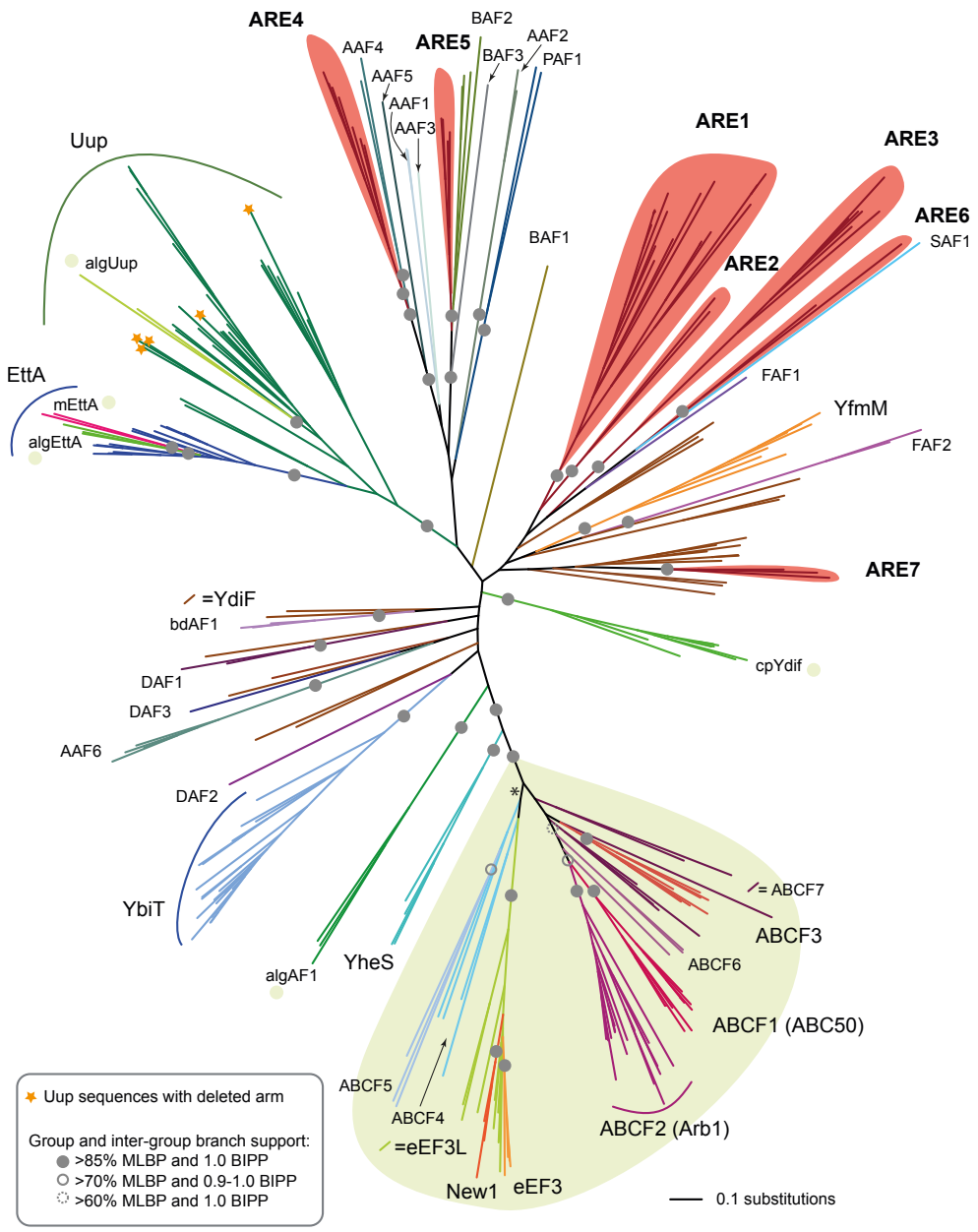

Figure 1 | The family tree of ABCFs has a bipartite structure corresponding to eukaryotic-like and bacterial (and organellar)-like sequences

The tree is a maximum likelihood phylogeny of representatives across the ABCF family with branch support values from 100 bootstrap replicates (maximum likelihood bootstrap percentage; MLBP) and Bayesian inference posterior probability (BIPP). The inset box shows the legend for subfamily and intersubfamily support; support values within subfamilies and that are less that 60% MLBP are not shown. Species were chosen that sample broadly across the tree of ABCF-encoding life, sampling at least one representative from each subfamily. Green shading shows the eukaryotic type ABCFs; other subgroups are bacterial unless marked with a green shaded circle to indicate eukaryotic groups with potentially endosymbiotic origin. The full tree with taxon names and sequence IDs is shown in fig. S1. Branch lengths are proportional to amino acid substitutions as per the scale bar in the lower right. The asterisked branch was not supported by this data set, however it was supported at 85% MLBP in phylogenetic analysis of the eukaryotic subgroup and its viral relatives, rooted with YheS (fig. S3).

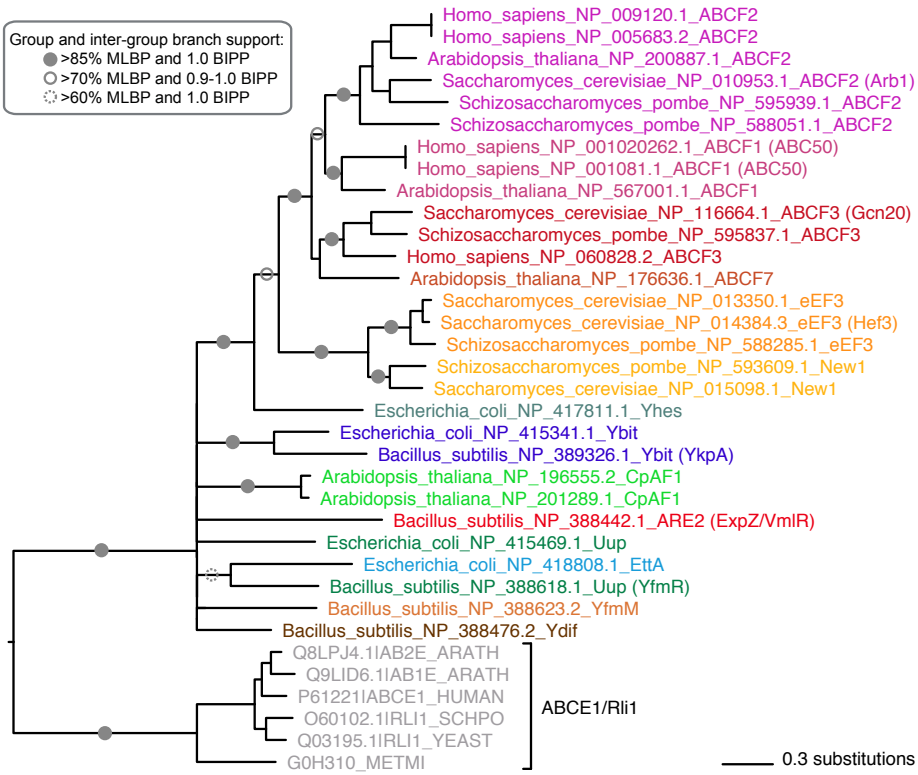

Figure 2 | Rooting with ABCE shows eukaryotic-like ABCFs nesting within bacterial-like ABCFs, with YheS as the sister group to the eukaryotic-like clade

Maximum likelihood phylogeny of representatives across the ABCF family, and ABCE sequences from the UniProt database. Branch support from 200 bootstrap replicates and Bayesian inference posterior probability is indicated with the key in the inset box. Branch lengths are proportional to amino acid substitutions as per the scale bar in the lower right.

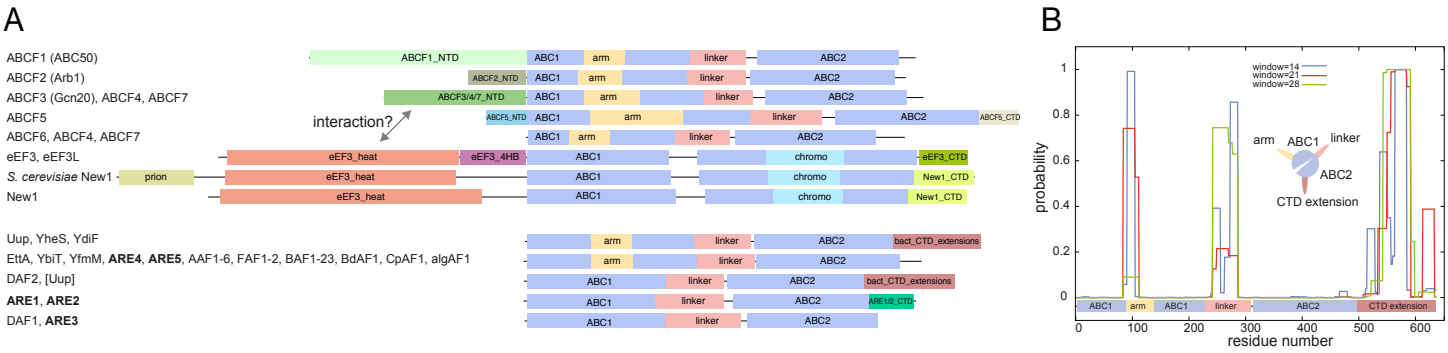

Figure 3 | Typical domain and subdomain architectures of ABCFs

A) Boxes show domains as predicted by HMMs. Dotted lines shows possible interactions between the HEAT domain of eEF3/New1/eEF3L and the N terminal domain of ABCF3,4 and 7. B) Predicted coiled coil regions of *E. coli* YheS along the protein length. Inset: cartoon representation of the coiled coil subdomains protruding from the core ABC domains.

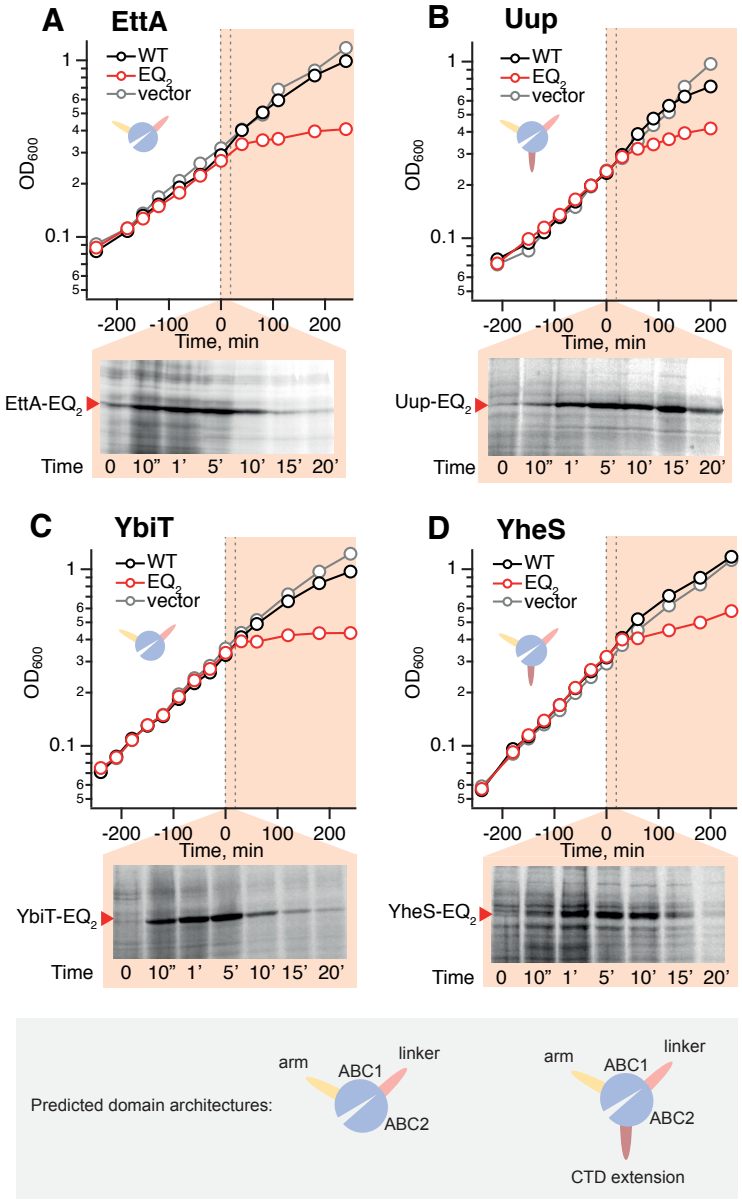

Figure 4 | Expression of *E. coli* ABCF-EQ$_2$ mutants inhibits growth and protein synthesis
Growth of wildtype *E. coli* BW2513 (WT, black line) and EQ$_2$ mutants (red line) of EttA (A), Uup (B), YbiT (C), and YheS (D) as measured by optical density at 600 nm (OD_{600}). The grey line is the empty vector control. Gels below the graphs show the effect of EQ$_2$ ABCF expression on protein synthesis, as probed by pulse labeling with radioactive ^{35}S-methionine. Expression was induced by L-arabinose at time point 0. The inset cartoons are a representation of ABCF domains and sub-domains, as per the legend in the lower box.

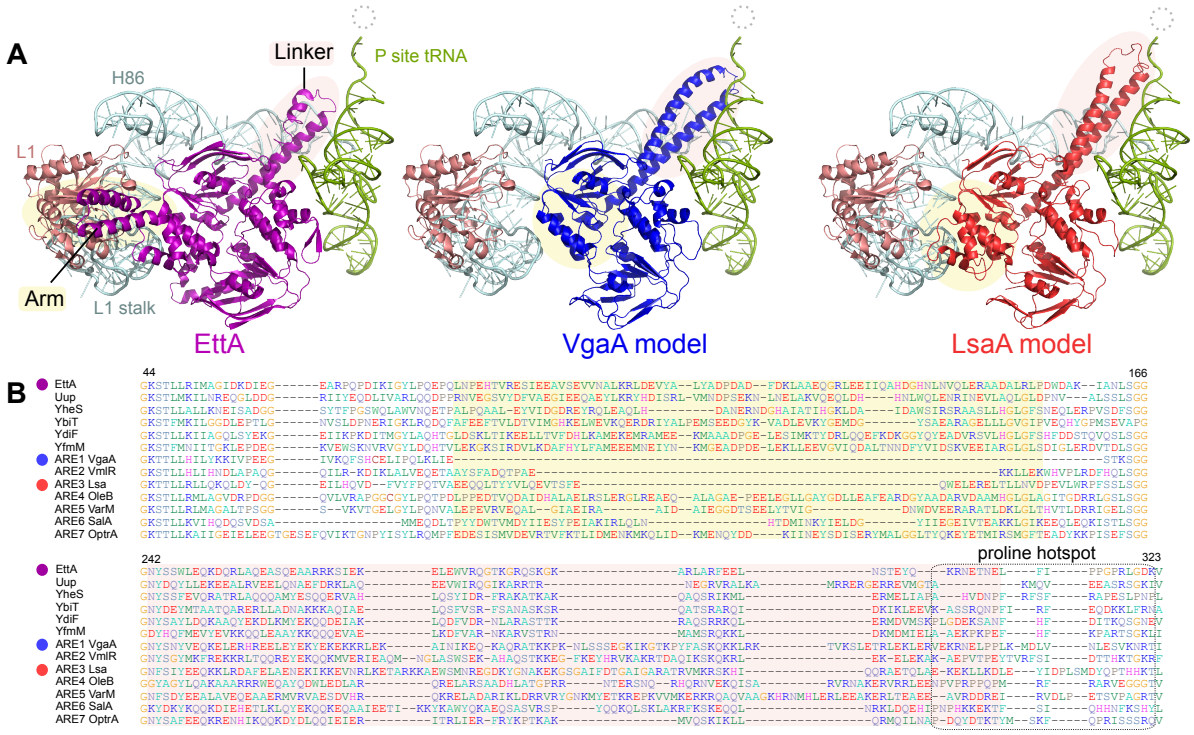

Figure 5 | AREs tend to have relatively long linker regions that potentially extend towards the ribosome bound antibiotics

A) The structure of EttA and its interacting ribosomal components from PDB 3J5S (Chen, Boel, et al. 2014) is shown alongside homology models of *Staphylococcus aureus* VgaA and *Entercoccus faecalis* LsaA, using 3J5S as the template, with *de novo* modeling of the linker regions. The dotted circle shows the relative location of PTC-inhibiting antibiotics. Arm and linker regions are shaded in yellow and turquoise respectively. B) Extracts from the multiple sequence alignment of *E. coli* and *B. subtilis* ABCFs, and representative AREs, containing the Arm (yellow shading) and Linker (turquoise shading) subdomains. Alignment numbering is according to the EttA sequence. A boxed region shows a region that is particularly rich in proline and polyproline in various ABCF family members.