Homology between the flagellar export apparatus and ATP synthetase: evidence from synteny predating the Last Universal Common Ancestor

Nicholas J. Matzke[1], Angela Lin[2], Micaella Stone[1], Matthew A. B. Baker[2,3*]

[1]School of Biological Sciences, University of Auckland, Auckland 1010, New Zealand

[2]School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, Australia

[3]CSIRO Synthetic Biology Future Science Platform, Brisbane, Australia.

[*]correspondence: matthew.baker@unsw.edu.au

Abstract

Evidence of homology between proteins in the ATP synthetase and the bacterial flagellar motor (BFM) has been accumulating since the 1980s. Specifically, the BFM's Type 3 Secretion System (T3SS) export apparatus FliH, FliI, and FliJ are considered homologous to F_O-b + F_1-δ, F_1-α/β, and F_1-γ, and have similar structure and interactions. We review the discoveries that advanced the homology hypothesis and then conduct a further test by examining gene order in the two systems and their relatives. Conservation of gene order, or synteny, is often observed between closely related prokaryote species, but usually degrades with phylogenetic distance. As a result, observed conservation of synteny over vast phylogenetic distances can be evidence of shared ancestral coexpression, interaction, and function. We constructed a gene order dataset by examining the order of *fliH*, *fliI*, and *fliJ* genes across the phylogenetic breadth of flagellar and nonflagellar T3SS. We compared this to published surveys of gene order in the F_1F_O-ATP synthetase, its N-ATPase relatives, and the bacterial/archaeal V- and A-type ATPases. Strikingly, the *fliHIJ* gene order was deeply conserved, with the few exceptions appearing derived, and exactly matching the widely conserved F-ATPase gene order *atpFHAG*, coding for subunits *b-δ-α-γ*. The V/A-type ATPases have a similar conserved gene order shared for homologous components. Our results further strengthen the argument for homology between these systems, and suggest a rare case of synteny conserved over billions of years, dating back to well before the Last Universal Common Ancestor (LUCA).

Gene order and homology hypotheses

Synteny is often observed between close relatives but tends to decay rapidly with phylogenetic distance (Lathe et al. 2000; Gómez et al. 2004; Zhulin 2017). While widely conserved operons are rare, when found, they may indicate conserved functional relationships. Co-expression, stoichiometry, and order of expression may provide preliminary evidence for the assembly and function of cooperating protein products (Dandekar et al. 1998; Gómez et al. 2004).

The observation of synteny can also be useful for proposing, or strengthening, hypotheses of homology between proteins or protein complexes. It is well-known that DNA sequence

similarity between a pair of genes can eventually decay to statistical undetectability (saturation; Philippe et al. 2011); amino acid sequences, although more conserved, are subject to the same problem. When amino acid sequence similarity has decayed into the "twilight zone" (Doolittle 1986; Rost 1999), homology arguments can still be made based on conserved protein structure, function, or protein-protein interactions, but confidence will be reduced unless several pieces of evidence line up in support of the hypothesis. The discovery of synteny can support other lines of evidence in favor of a hypothesized homology and can have practical benefits as well. For example, if synteny is strongly conserved between a set of genes, this might assist identification and naming of ORFs that have been missed by automated BLAST (Lathe et al. 2000).

Furthermore, confirmation of homology between distantly-related protein complexes can inspire productive research avenues. This occurs because experimental information gathered about a well-studied protein complex – protein structure, protein interactions, functional mechanisms, etc. – provides starting points for experimental research into a less-studied, but evolutionarily related, complex. A process of "reciprocal illumination" (Hennig 1966) can ensue, where experimental work on either system helps inform work on the other, as similarities and differences are elucidated.

Here, we review a remarkable case where a sequence of homology discoveries has reciprocally illuminated research on the structure and function of two molecular machines: the F_1F_O-ATP synthetase (or F-ATPase) and the flagellar protein export apparatus (and their respective relatives). We add to the evidence for homology by describing shared gene order between the two systems, suggesting a case of synteny that pre-dates the Last Universal Common Ancestor (LUCA).

The rotary motors and their relatives

For simplicity, we refer to "the" flagellum and F_1F_O-ATP synthetase, as they are exemplar systems that are well-studied, particularly in model systems, and because our argument can be made with reference to these. However, we acknowledge the major related systems, which we summarize below (Fig. 1).

While there is no "the" flagellum, and considerable structural diversity exists across bacterial flagella (Snyder et al. 2009; Beeby et al. 2020), about 40 proteins are well-conserved across most flagella (Fig. 1; Pallen and Matzke 2006). About 11 of the proteins form a flagellar Type 3 Secretion System (F-T3SS) that exports proteins through a central channel (Milne-Davies et al 2020). The exported proteins grow a rod, hook, and flagellar filament atop the basal body. The whole structure is surrounded by an outer ring of stator/motor protein complexes ($MotA_5B_2$; Deme et al. 2020; Santiveri et al. 2020) which use the flow of H^+ (or other cations; Ishida et al. 2019; Lai et al. 2020) to rotate the basal body (reviewed in detail by Lai et al. 2020).

The flagellar T3SS (F-T3SS) is homologous to the nonflagellar T3SSs (NF-T3SS) which form needle-like structures that secrete effector proteins into eukaryote cells; NF-T3SS are common in plant and animal pathogens. The most recent major analysis suggests that NF-T3SS are phylogenetically derived from the flagellum (Abby and Rocha 2012; Denise et al. 2020),

although it appears to be a deep branch that did not originate within any particular extant bacterial phylum.

The bacterial F_1F_O-ATP synthetase (von Ballmoos et al. 2009; Fillingame and Steed 2014; Sobti et al. 2019; Sobti et al. 2020; Sobti et al. 2020), in contrast, is much smaller (Fig. 1). The canonical system of *E. coli* consists of 8 well-conserved proteins all localized to the inner membrane. Subunits of F_O-c form a rotating ring in the inner membrane, and F_O-a acts as a stator; F_O-a and F_O-c together form a channel allowing the flow of H^+. The soluble F_1-subunit consists of a heterohexameric ATPase with alternating α (noncatalytic) and β (catalytic) subunits. These are attached to the F_O complex with an external stalk formed by a dimer of F_O-b, and an internal stalk formed by F_1-γ and F_1-ε (refs). The stalks allow ATP hydrolysis to exert force on the F_O-c-ring, driving rotation with respect to F_O-a and causing proton export, or the reverse reaction (proton flow driving ATP synthesis).

Two subcategories of bacterial F_1F_O-ATPases are the mitochondrial (He et al. 2018) and chloroplast (Hahn et al. 2018) versions (MF_1F_O and CF_1F_O, respectively). They each have their own experimental literature, varying nomenclature, and additional accessory subunits (Chaban 2005; Supp. Mat.), but are of relatively recent origin, being related to the F-ATPases of Rickettsiales and cyanobacteria, respectively (although many of the genes for these systems have relocated to the nuclear genome; Stoebe and Kowallik 1999; Johnston and Williams 2016). However, another subcategory, the N-ATPases, which pump Na^+ instead of H^+, are potentially sister to the rest of the F-ATPase clade. More broadly, the F- and N-ATPases are homologous to the archaeal A-ATPases and the V-ATPases of eukaryotic vacuoles (however, V-ATPases have since been discovered in bacterial as well, sometimes called V/A-ATPases or bacterial V-ATPases; ref). These other ATPases carry homologs to most or all of the F-ATPase subunits, although homologies are sometimes difficult to confirm due to lack of sequence similarity (Mulkidjanian et al. 2007); they also contain a variety of less-conserved accessory proteins.

Flagellum-ATP synthetase homology: a history of discovery

The bacterial flagellum and the F_1F_O-ATPase are both famous for using protonmotive force to power rotary motors to do useful work. These generic similarities inspired some early scenarios that attempted to derive a flagellum directly from a synthetase (Goodenough 1998; Rizzotti 2000), but any detailed comparison of the systems reveals these as vague and fanciful (Matzke 2003).

Homology of the ATPase proteins. The strong case for homology between the flagellum and ATP synthetase does not involve ion-powered rotary motion, but instead concerns the ATPase proteins of both systems. Early protein sequence comparisons quickly recognized highly significant amino acid sequence similarity between FliI of the flagellum, and the F_1-α and F_1-β subunits of the F_1F_O-ATP synthetase (Gogarten et al. 1989; Iwabe et al. 1989; Gogarten et al. 1992). Early discussions focused on the fact that the F_1-ATPase's heterohexameric structure of alternating catalytic (β) and non-catalytic (α) subunits proteins is also found in the V- and A-ATPases, suggesting that the heterohexameric ATPase was found in the Last Universal Common Ancestor (LUCA). The catalytic and non-catalytic subunit were derived by gene duplication from

a common ancestor that was a catalytic homohexamer (as in more distant relatives like the transcription termination factor Rho, and the P-loop ATPase RecA; Mulkidjanian et al. 2007). Initial interest focused on the suggestion that these ancient paralogs could thus be used to provide outgroups for each other, allowing rooting of the phylogenetic Tree of Life (ToL) between the bacteria and the archaea/eukaryote clade (Gogarten and Kibak 1992; Shih and Matzke 2013; Shih et al. 2017). While conceptually appealing, later work has questioned the idea the small collection of pre-LUCA paralogs retains sufficient signal to estimate the ToL root or the premise that different genes would record the same organismal history in the first place (Matzke et al. 2014; Gouy et al. 2015).

Homology of the outer stalk. Surprisingly, although the homology between FliI and the F_1-α/β was recognized early, it was not until Jackson & Plano (2000) that it was suggested some flagellar proteins might form a multiprotein complex similar in structure to F1 and V1, inspired by BLAST hits between the FliH homolog YscL and the external stalk protein V1-e (Ge et al. 1996 also noticed sequence similarity between *Borrelia* FliH and F_O-b). Only in 2003 was homology deployed to inform models of flagellum structure. Prior to this date, cartoons of flagellum structure represent FliI as a monomer proximal to the inner membrane. However, in Blocker et al. (2003) hypothesized, on the basis of F_1-ATPase homology, that the *Shigella* NF-T3SS Spa47 (a FliI homolog) formed a homohexamer. Claret *et al.* (2003) showed hexamerization of FliI with electron microscopy.

Further observations of FliH/F_O-b sequence similarity, and additional similarities (both FliH and F0-b dimerize, both associate with ATPases (FliI, F_1-α/β), and both associate with membrane proteins via N-terminal α-helices) inspired suggestions that a FliH dimer might serve an external stalk role similar to F_O-b (Matzke 2003; Pallen et al. 2005; Lane et al. 2006). Lane et al. (2006) also found analogies between F_1-δ and the globular, C-terminal domain of FliH; F_1-δ forms the membrane-distal part of the F_1F_O-ATPase external stalk, and both FliH-C and F_1F_O-δ bind to amphipathic helices at the N-terminus of FliI and F_1-β, respectively. They summed up the similar interactions between FliHI and F_O-b, F_1-δ, and F_1-α/β as having "uncanny similarity," although they termed them analogous, rather than homologous.

Pallen et al. (2006) used PSI-BLAST to conduct a systematic search for distant homologs of FliH/YscL and found remote, but significant, sequence similarity (~18-20%) to F_O-b and the external stalk E subunit of V-ATPase. These matches occurred with just the N-terminal ~115 residues of FliH/YscL, so Pallen et al. conducted a separate search with the C-terminal region (YscL positions 115-223). Variations on these searches yielded repeated, although nonsignificant (expected number of hits E-value ranging from 0.14-8.5) hits to the F_1-δ subunit (known as the OSCP in the MF_1Fo-ATPase). They buttressed the argument by presenting highly suggestive multiple sequence alignments, and by noting that the genes for F_O-b and F_1-δ (atpF and atpH in *E. coli*) are often adjacent, and that natural fusions of F_O-b and F_1-δ are observed in some bacterial genomes.

The homology of three F-ATPase proteins to two flagellar export proteins inspired exploration of possible implications, leading Pallen & Matzke (2006) to note "the existence of ancient structural and functional similarities between these two types of molecular rotary motor, which we predict will become ever more apparent as work on these systems progresses."

Mulkidjanian, et al. (2007) reviewed the similarities among V/F/A-ATPases, flagellar and nonflagellar T3SS, but found no homology between the central stalks proteins (FliJ/YscO in T3SS; F_1-γ in F-ATPases, and V_1-D/F in V-type ATPases). However, they note the membership of FliI/F_1-α/β in the vast group of P-loop NTPases (Lupas and Martin 2002), specifically within the RecA-like motor ATPases (Ye et al. 2004), and even more specifically their relationship to Rho, the transcription-termination factor that forms a homohexamer ring to act as an ATP-powered RNA helicase. They propose that a cytoplasmic RNA helicase could associate with a transmembrane channel (the proto-F_O-c ring) to become an RNA translocase, and later transition into a protein translocase with the addition of an external stalk. From here, they propose the divergence of the VFA-ATPases and the T3SS. On the VFA-ATPase line, ion translocation (Na+ or H+; Mulkidjanian 2009 argue for Na+ as the easier starting point) is added for unclear reasons. Then, the central stalks (F_1-γ, V1-D) evolved independently from translocating proteins that became temporarily trapped and instead rotated the proto-c-ring, driving ion translocation in reverse, coupling ATP synthesis/hydrolysis with an ion gradient. Permanent recruitment of these proto-stalk proteins and specialisation on the ion-pumping/ATP synthase function results in the modern systems. The further evolution of the proto-T3SS is not detailed, although it is suggested that FliI still acts directly in the protein translocase activity, threading polypeptides through the middle of the hexamer.

Homology of the inner stalk. Additional speculation about the descent of T3SS and V/F/A-ATPases from a common ancestral system accumulated (Imada et al. 2007; Mulkidjanian et al. 2009) until the next major advance homologizing the systems. This was the crystallization of FliJ (Ibuki et al. 2009), leading to a solved 3D structure that was immediately recognized as highly similar to F_1-γ and V1-D (Ibuki et al. 2011). Both proteins consist of two long antiparallel α-helices. In spite of no significant sequence similarity, structure-guided sequence alignment indicated conserved residues in regions involved in interaction with other proteins. Strikingly, cryo-EM images showed FliJ localizing off-center within the FliI hexamer ring; in other words, "[t]he FliI6FliJ complex looks just like F_1-ATPase" (Ibuki et al. 2011). Similar results were soon found for the NF-T3SS homolog of FliJ, EscO/YscO (Romo-Castillo et al. 2014). Additional similarities accumulated. Ibuki et al. (2011) noticed a conserved binding site between FliJ and F_1-γ corresponding to where F_1-γ interacts with F_1-ε. Ibuki et al. (2013) showed that the corresponding residues of FliJ are involved in interactions with FlhA-L, a possibly α-helical linker region between the N-terminal transmembrane helices of FlhA and the cytoplasmic C-terminal domain. A FlhA nonamer ring forms the core of the membrane-bound portion of the T3SS, suggesting at least a general similarity to the F_O subunit, and demonstration that FliJ stretched from inside the FliI ring to a membrane-bound protein finally clinched the case that the FliI hexamer took the same orientation with respect to the inner membrane as F_1-$\alpha_3\beta_3$ (that is, FliI C-terminal region faces the membrane, and N-terminal faces away; (Ibuki et al. 2011).

Finally, Kishikawa et al. (2013) tested whether FliJ could rotate like F_1-γ, constructing a chimeric complex of *Salmonella* FliJ with *Thermus aquaticus* V_1-A_3B_3, finding that FliJ localized to the center of the ring and rotated bidirectionally. Diepold & Armitage (2015) suggested that "the FliJ/SctO stalk might rotate," and incredibly, a chimera of FliJ with the 21 C-terminal residues of V_1-D was shown to rotate unidirectionally inside the V1A3B3-ATPase in the presence of ATP at a speed comparable to V_1-D (Baba et al. 2016). Rotation, though weaker, was also shown for a F_1-

FliJ chimera (Baba et al. 2016), suggesting torque generation through a coarse-grained, non-sequence-specific interaction that must date back before the LCA (Noji et al. 2017).

The discoveries mentioned above indicate that the homology hypothesis has served as a productive research avenue (as suggested by e.g. Pallen et al. 2006; Pallen and Matzke 2006), and homology between the F_1-ATPase and FliHIJ is now well-accepted and routinely considered in studies of both structural/functional studies of both systems. However, it should be kept in mind that, apart from the highly significant sequence similarity between FliI and F_1-α/β (and many more distant relatives; Iyer et al. 2004; Ye et al. 2004), which are large proteins with conserved sites of catalysis and conserved structure at all levels, the other proteins are small and structurally simple, formed primarily of extended α-helices. The sequence similarity evidence, judged by e-values, is weak (FliH/F_O-b+F_1-δ, particularly the latter) or non-existent (FliJ/F_1-γ). And, while the case for homology is reasonably strong *in toto*, as a large number of similarities, individually only suggestive, point in the same direction, it is somewhat conceivable that they might be explained away by convergence. Furthermore, experimental work suggests the possibility of some substantial dissimilarities. For example, while experiments and homology have suggested that $FliH_2$ dimerizes to form the outer stalk in a fashion similar to F_O-b, cryoelectron tomography of the NF-T3SS of *Shigella flexneri* indicates that the FliH homolog (MxiN) forms six "spokes" that link the N-terminal regions of the FliI hexamer subunits to FliN subunits, of the C-ring, capping the T3SS structure (Hu et al. 2015; Gao et al. 2018). The functions of F_1-γ and FliJ also appear quite different; F_1-γ rotates with the F_O-c ring, coupling ATP synthetase activity to ion flow, whereas FliJ seems to play a role in making substrate proteins (axial proteins, including flagellin and rod/hook/linker/cap proteins and NF-T3SS equivalents) export-competent by removing chaperones (chaperones FlgN, FliS, FliT are all structurally similar; Khanra et al. 2016) and delivering substrate to the ion/protein antiporter via interactions with FlhA (Gao et al. 2018; Minamino et al. 2020).

Another source of evidence: Gene order. We investigated whether gene order evidence might strengthen or weaken the hypothesis of homology between FliHIJ and F_1-ATPase. Gene order has occasionally been used to identify homologous proteins in T3SS. For example, FliJ homologs in NF-T3SS (Romo-Castillo et al. 2014) as the gene for the FliJ homolog is typically downstream from the gene for the FliI homolog (Evans et al. 2006). Similarly, Pallen et al. (2006) noted that the genes for F_O-b and F_1-δ are typically adjacent (Wilkens and Capaldi 1998) and sometimes fused (Féthière et al. 2004; Pallen et al. 2006). While *fliH*, *fliI*, and *fliJ* were named alphabetically in order in the unified flagellar gene nomenclature established by Iino et al. (1988), this scheme was based on *E. coli* & *Salmonella* sp. Liu & Ochman (2007) surveyed operon structure and gene order across many flagellar systems and found that the alphabetical ordering broke down at larger phylogenetic scales. They did note that certain small clusters were usually adjacent, e.g. *flgBC*, *flgKL*, *flhBA*, *fliMN*, *fliPQR*, but they did not mention *fliHIJ*. This was probably because the *fliHIJ* cluster was missing from 4 of the 10 groups they summarized in their Figure 3. The *fliJ* gene was thought to encode one of several flagellar chaperons, and Liu and Ochman stated that this "auxiliary" gene had a "sporadic" distribution in their dataset. This was likely due to failure to detect FliJ homologs due to its less-conserved sequence.

The study of gene order has been more systematic in F- and V/A-type ATPases, so we are able to rely on previously-published reviews for these systems. Gene order in the F_1F_O-ATPase and

relatives is highly, but not universally, conserved across bacterial phyla and divergent systems (Koumandou and Kossida 2014; Niu et al. 2017). The same is true for V/A-type ATPases in bacteria and archaea (Lolkema et al. 2003).

Methods

We surveyed the order of the *fliH*, *fliI*, and *fliJ* genes across bacteria. We chose species representing each major flagellated clade/phylum (Fig. 2), using Snyder et al. (2009) for guidance. Our strategy was to identify *fliI* homologs, as FliI is well-conserved and usually identified by automated annotations, and then identify the flanking genes. For each species, we searched Ensembl's bacterial genome browser (Howe et al. 2020). We started by searching by name for FliI/YscN. If that failed, we searched for FliF, which is also highly conserved and sometimes annotated when FliI is not, and then searched neighboring genes for a FliI homolog. If both of those approaches failed, we BLASTed the genome with the *E. coli* FliI amino acid sequence.

Once *fliI* was located, we identified the flanking genes using the following approach. If the protein was automatically identified, we used that label. If it was not, we checked the protein family and then conserved domain information, using the first identification found. If these approaches revealed nothing, we downloaded the amino acid sequence and used PSI-BLAST (Altschul et al. 1997) on the RefSeq Select proteins database (refseq_select; Li et al. 2020). If all of those approaches failed, we recorded the protein as a hypothetical protein, or a DUF (conserved Domain of Unknown Function), if so identified in the database. To provide genomic context, we used the same approach on the 2 genes upstream and downstream of the *fliHIJ* cluster, recording the flanking 2 genes on each side of *fliHIJ* if they were flagellar, or 1 if the further genes were nonflagellar.

Gene identifications were recorded in transcription order from upstream to downstream, ordered with respect to the *fliI* homolog. Cases where coding switched to the opposite strand from *fliI* were indicated with "<". In cases where *fliH* or *fliJ* was not found near *fliI*, we BLASTed the genome with the FliH or FliJ amino acid sequence from a relative. Once the homolog was identified, we identified the flanking 2 genes on each side if they were flagellar, or 1 if the further genes were nonflagellar. The non-neighbor *fliH/J* genes are displayed in Fig. 2 in the same order as the canonical *fliHIJ* ordering, but with "|" indicating a discontinuity meaning that the genes are not in the same neighborhood.

-

F-ATPase and prokaryote V/A-ATPase gene orders were taken from Fig. 4 of Koumandou & Kossida (2014) and Table 2 of Lolkema et al. (2003), using their clade names and gene names. The Koumandou & Kossida (2014) dataset represents summaries of the most common gene arrangements in each clade according to their analyses; Lolkema et al. (2003) uses representative species. We did not investigate gene order in eukaryotic V-ATPases as eukaryotic genome organization is vastly more complex and presumably derived with respect to the A-type systems.

Results

Figure 2 depicts gene order across highly divergent representatives of F-T3SS, NF-T3SS, F-type ATPases, and V/A-type ATPases from bacteria and archaea. A high degree of conservation of gene order of the *fliHIJ* homologs can be observed across the T3SS without sophisticated analysis. While gene order breaks down for some Alpha- and Epsilon-proteobacterial systems, as noted by Liu & Ochman (2007), it is otherwise conserved across proteobacteria as well as deep splits in the bacterial tree, including Chloroflexi, Spirochetes, Firmicutes, Nitrospirae, Thermotogae, and Verrucomicrobia. The gene order is also found in the presumed "flagellar remnant" (Betts-Hampikian and Fields 2010) system found in Chlamydiae in addition to their classic NF-T3SS; some marine Chlamydiae are now known to have flagella (Collingro et al. 2017; Dharamshi et al. 2020).

While the conserved gene order is disrupted in *E. coli* and *Vibrio* NF-T3SS by the relocation of *escL*/*yscL* (*fliH* homologs), the standard gene order is retained across the deeper branches of NF-T3SS, including *Myxococcus* and *Chlamydia* lineages. Overall, whatever exact phylogeny is assumed to represent the relationships between the different bacterial phyla (a difficult question; Ishida et al. 2019) and the NF-T3SS (Abby and Rocha 2012), it is clear that *fliHIJ* represents the ancestral gene order.

Turning to the F-type ATPases, the gene order found in the model *E. coli* system is conserved with rare exceptions across phyla; Koumandou & Kossida (2014) conclude it is the ancestral gene order. Strikingly, the genes for F_O-b, F_1-δ, F_1-α, and F_1-γ are arranged identically with their *fliHIJ* homologs, taking into account that F_O-b and F_1-δ are considered to be homologous to the N-terminal and C-terminal regions of FliH, respectively.

The prokaryotic V/A-ATPases also show strong conservation of gene order, with I-K-E-C-G-A-B-D representing the likely ancestral arrangement (Lolkema et al. 2003), with the poorly-conserved F (and the associated nomenclatural nightmare; Lolkema et al. 2003) appearing in various locations. In 2003, Lolkema et al. assumed that subunit E formed the central stalk and that D was the peripheral stalk, so the similarity to the F-type ATPase gene order was not apparent. However, it is now recognized that E and G are external stalks (Vasanthakumar and Rubinstein 2020; see Supp. Mat. for nomenclatural conversions between prokaryotic and eukaryotic systems), where E corresponds to F_O-b and F_1-δ, and G is a shorter external stalk corresponding to F_O-b (Pallen et al. 2006). Subunit D is the central stalk, homologous to F_1-γ (reviewed above). The major exceptions to the conserved gene order are the variable positions of the central stalk components F and C. The C-subunit is named "d" in the eukaryotic systems (Féthière et al. 2004; Chaban 2005) and forms the interface between the central stalk and the ion-transducing membrane ring (V_O-K, corresponding to F_O-c).

Examination of the order of the genes for the ATPase proteins with respect to the central stalk gene shows that the V_1-B subunit (noncatalytic and considered homologous to F_1-α) is almost always directly upstream of the gene for V_1-D. This corresponds with the arrangement of F_1-α and F_1-γ. However, the order of the catalytic and noncatalytic subunits is reversed between the V_1 (AB) and F_1 (αβ, with γ in-between).

Discussion

The gene orders of the *fliHIJ* homologs are observed to be highly, although not universally, conserved, across flagellar and nonflagellar T3SS and the prokaryotic F- and V/A-type ATPase systems. As the duplication of an ancestral FliI-like ATPase to produce the paralogous catalytic F_1-β/V_1-A and noncatalytic F_1-α/V_1-B clades predates the LUCA (Iwabe et al. 1989; Shih and Matzke 2013), and the divergence of FliI must precede that, this constitutes the most ancient case of synteny involving more than 2 genes of which we are aware.

Our results suggest a parsimonious scenario to explain the ancestral gene orders in the studied systems. The proto-F_1-like system was encoded by an operon with genes coding for the equivalents of F_O-b, F_1-δ, FliI, and FliJ. Duplication of the FliI homolog produced 2 catalytic subunits in tandem. The second of these (representing proto-F_1-α/V_1-B) lost catalytic ability, and then after the divergence of the F- and V-type systems, the ancestor of F_1-β was transposed to be just downstream of F_1-γ.

In the same scenario, the external stalk subunits originated by (1) fusion of the F_O-b and F_1-δ subunits to produced proto-FliH in the F-T3SS ancestral lineage; (2) retention of the ancestral arrangement in the F-ATPase lineage, with occasional later duplications of F_O-b and fusions with F_1-δ; and (3) a similar process of duplication and fusion producing the E and G external stalk subunits in the V-ATPase lineage, along with the addition of other subunits associated with the central stalk.

Several implications follow from our results. First, conserved synteny further strengthens the case for homology between the two systems, and the specific homologized proteins. Although there was little remaining doubt in the literature concerning these homologies, additional evidence helps to confirm research is on the right track, and adds genomic context to inferences about ancestral complexes based on shared sequence, structure, and functional detail.

Secondly, our results confirm earlier conclusions, based only on operon structure, that the F-type ATPases are made up of evolutionary submodules (Niu et al. 2017); indeed, these submodules appear to sit even deeper into evolutionary history than Niu et al. suggest.

Thirdly, they constitute a dramatic illustration of the hypothesis that widely-conserved gene order in prokaryotes indicates conserved protein-protein interactions (Dandekar et al. 1998; Lathe et al. 2000). In addition, the fact that exceptions to the conserved gene order are known and thus are possible, suggests that systems with alternative arrangements have somehow compensated for what is otherwise a common, conserved mechanism for regulating gene expression and stoichiometry across these systems.

Fourthly, cases of highly-conserved gene order suggest an obvious way to identify gene products in prokaryote genomes, when ORFs are unidentified by automated approaches: if *fliI* is identified, then *fliH* and *fliJ* are likely the flanking genes, even if the sequence similarity in these short genes has decayed beyond detectability. For example, we expect that the hypothetical gene upstream of *fliI* in *Planctomyces* sp. SH PL14 (Fig. 2) will turn out to be a *fliH* homolog,

even though PSI-BLAST searches using the amino acid sequence turned up only a conserved, but unidentified, protein family.

Finally, discoveries of homology suggest avenues for further homology searches. It is tempting to postulate that the conserved F_1-like complex suggests an ancestral association with a membrane-bound system (Mulkidjanian et al. 2007; Mulkidjanian et al. 2009), and that therefore at least parts of the membrane-bound components of the T3SS export apparatus should be ancient homologs of the F_O subunits. However, despite the fact that experimental discoveries have indicated certain unexpected similarities between the systems – notably, the T3SS uses ion-motive force to power protein export – Beeby et al. (2020) report no homology with the FliPQR proteins. These proteins were long strongly predicted to be transmembrane proteins, but solved structures indicate they form a gate that is withdrawn above the inner membrane to bind with FliE, the beginning of the axial filament. Beeby et al. thus conclude that there is no homology between T3SSs and F_O, and the F_1-like complex was a late addition to the flagellar export system. They support this by noting that the FliHIJ complex is not absolutely required for protein export, which can be driven by ion motive force alone, although ATPase activity it makes it highly efficient.

Our results indicate a few points for consideration. First, it is a mistake to assume that either the extant systems in either the T3SS or V/F/A-ATPase branch represent "the ancestral" configuration. Rather, both systems retain a mixture of ancestral and derived features of gene order. This principle will likely apply to other homologies between the systems; for instance, it is possible that the F_O subunit is substantially simplified and specialized from a larger and more complex protein-translocating ancestor (Mulkidjanian et al. 2007). In addition, the fact that the ancestral F_1-like subunit contained both external and internal stalks suggests a multi-subunit interactive function that might not be easily transposed into a new system. We suggest caution when evaluating possible indications of homology between F_O and parts of the T3SS. It is possible that homology might appear in retained interactions (e.g., FliJ interacts in part with FlhA-L, the linker region between the N-terminal transmembrane and C-terminal cytoplasmic domains of FlhA; Ibuki et al. 2013) and molecular mechanisms (e.g., the conversion of ion-motive force into conformational changes), even if the three-dimensional structures have been highly modified.

Acknowledgements / Funding

MABB is supported by the UNSW Scientia Research Fellowship, the CSIRO Synthetic Biology Future Science Platform 2018 Project Grant and ARC Discovery Project DP190100497. NJM is supported by the University of Auckland Faculty Research Development Fund (FRDF) Strategic Initiative - Early Career Researcher Fund, Project #3722433, and New Zealand Marsden Grants 16-UOA-277 and 18-UOA-034. MS was a summer student volunteer at the University of Auckland.

Figure 1. Cartoon depicting location of core structural components of F_1F_O-ATP synthetase, V/A-type ATPases, and the bacterial flagellum. Colored components match the colors in Fig. 2. For the V-type ATPases, naming follows the bacterial V-ATPases (Kolkema et al. 2003); see Supp. Material for corresponding names across systems. Stoichiometry is indicated with subscripts where important in this discussion, but most flagellar proteins form large oligomers (rings or filaments) with many subunits. Cartoon after Macnab (2003), updated with Minamino & Imada (2015), Minamino et al. (2017), Beeby (2020). FliH interacts with both FlhA and FliN so both positions are depicted, after Minamino & Imada (2015).

Figure 2. Conserved gene order between *fliHIJ* genes and V/F/A-ATPase genes. Homologs of *fliH*, *fliI*, and *fliJ* are colored orange, red, and dark pink, respectively. Homologs of V/F/A-ATPase a-, c-, and ε-subunits are dark green, light green, and light pink. Other genes are white. Genes are ordered from upstream to downstream with respect to the *fliI* exon. Genes coded on the opposite strand are indicated by left arrows (<); return to the *fliI* strand by right arrows (>). Large breaks are indicated by vertical lines (|) as in Lolkema et al. (2003). For T3SS gene order, see Methods and Supplemental Material; Supp. Mat. also includes the full names of strains. F-ATPase genes are named by the protein subunit, with the order taken from Koumandou & Kossida (2014). V/A-ATPase gene names & order from Kolkema et al. (2003). *E. coli* F_1F_O-ATPase after Nielsen et al. (1984). Chloroplast (*Arabidopsis*) gene order from *https://www.ncbi.nlm.nih.gov/gene/844790* ; chloroplast *atpG* (b', II), *atpC* (γ), and *atpD* (δ), are encoded in the plant nuclear genome (Hermans et al. 1988; Stoebe & Kowallik 1999). This is a schematic depiction to highlight gene order; the gene lengths and spaces between genes are not intended to be exact. *Notes:* N-ATPase "b" is a b+δ fusion (Dibrova et al. 2010), as is V1-E (Pallen et al. 2006). The V_O I-subunit is homologous to F_O–a and the subunit named "a" in yeast V-ATPases (Chaban et al. 2005). The F_O–I subunit (and the homologous sI) are Ca^{2+}/Mg^{2+} transporters, unique to F-ATPases (Koumandou & Kossida 2014). DUF: conserved Domain of Unknown Function; hyp. = hypothetical protein in database, not identified as homologous to T3SS genes by the database or manual searches.

References

Abby, S. S. and E. P. C. Rocha (2012). "The Non-Flagellar Type III Secretion System Evolved from the Bacterial Flagellum and Diversified into Host-Cell Adapted Systems." *PLOS Genetics* **8**(9), e1002983.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." *Nucleic Acids Research* **25**(17), 3389-3402.

Beeby, M., J. L. Ferreira, P. Tripp, S.-V. Albers and D. R. Mitchell (2020). "Propulsive nanomachines: the convergent evolution of archaella, flagella and cilia." *FEMS Microbiology Reviews* **44**(3), 253-304.

Betts-Hampikian, H. and K. Fields (2010). "The Chlamydial Type III Secretion Mechanism: Revealing Cracks in a Tough Nut." *Frontiers in Microbiology* **1**, 114.

Blocker, A., K. Komoriya and S. I. Aizawa (2003). "Type III secretion systems and bacterial flagella: Insights into their function from structural similarities." *Proc Natl Acad Sci U S A* **100**(6), 3027-3030.

Chaban, Y. (2005). Subunit topology in the V type ATPase and related enzymes. *University of Groningen*. Groningen, Germany, University of Groningen. **Ph.D.**

Claret, L., S. R. Calder, M. Higgins and C. Hughes (2003). "Oligomerization and activation of the FliI ATPase central to bacterial flagellum assembly." *Mol Microbiol* **48**(5), 1349-1355.

Collingro, A., S. Köstlbacher, M. Mussmann, R. Stepanauskas, S. J. Hallam and M. Horn (2017). "Unexpected genomic features in widespread intracellular bacteria: evidence for motility of marine chlamydiae." *The ISME Journal* **11**(10), 2334-2344.

Dandekar, T., B. Snel, M. Huynen and P. Bork (1998). "Conservation of gene order: a fingerprint of proteins that physically interact." *Trends in Biochemical Sciences* **23**(9), 324-328.

Deme, J. C., S. Johnson, O. Vickery, A. Aron, H. Monkhouse, T. Griffiths, R. H. James, B. C. Berks, J. W. Coulton, P. J. Stansfeld and S. M. Lea (2020). "Structures of the stator complex that drives rotation of the bacterial flagellum." *Nature Microbiology* **5**(12), 1553-1564.

Denise, R., S. S. Abby and E. P. C. Rocha (2020). "The Evolution of Protein Secretion Systems by Co-option and Tinkering of Cellular Machineries." *Trends in Microbiology* **28**(5), 372-386.

Dharamshi, J. E., D. Tamarit, L. Eme, C. W. Stairs, J. Martijn, F. Homa, S. L. Jørgensen, A. Spang and T. J. G. Ettema (2020). "Marine Sediments Illuminate Chlamydiae Diversity and Evolution." *Current Biology* **30**(6), 1032-1048.e7.

Diepold, A. and J. P. Armitage (2015). "Type III secretion systems: the bacterial flagellum and the injectisome." *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**(1679), 20150020.

Doolittle, R. F. (1986). *Of URFs and ORFs: a primer on how to analyze derived amino acid sequences*. Mill Valley, CA, University Science Books.

Evans, L. D. B., G. P. Stafford, S. Ahmed, G. M. Fraser and C. Hughes (2006). "An escort mechanism for cycling of export chaperones during flagellum assembly." *Proceedings of the National Academy of Sciences* **103**(46), 17474.

Féthière, J., D. Venzke, M. Diepholz, A. Seybert, A. Geerlof, M. Gentzel, M. Wilm and B. Böttcher (2004). "Building the Stator of the Yeast Vacuolar-ATPase: Specific Interaction Between Subunits E and G." *Journal of Biological Chemistry* **279**(39), 40670-40676.

Fillingame, R. H. and P. R. Steed (2014). "Half channels mediating H+ transport and the mechanism of gating in the Fo sector of Escherichia coli F1Fo ATP synthase." *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1837**(7), 1063-1068.

Gao, X., Z. Mu, X. Yu, B. Qin, J. Wojdyla, M. Wang and S. Cui (2018). "Structural Insight Into Conformational Changes Induced by ATP Binding in a Type III Secretion-Associated ATPase From Shigella flexneri." *Frontiers in Microbiology* **9**, 1468.

Ge, Y., I. Old, I. Saint Girons, D. B. Yelton and N. W. Charon (1996). "FliH and FliI of Borrelia burgdorferi are similar to flagellar and virulence factor export proteins of other bacteria." *Gene* **168**(1), 73-75.

Gogarten, J. P. and H. Kibak (1992). "The bioenergetics of the last common ancestor and the origin of the eukaryotic endomembrane system." *Biosystems* **28**(1-3), 131-153.

Gogarten, J. P., H. Kibak, P. Dittrich, L. Taiz, E. J. Bowman, B. J. Bowman, M. F. Manolson, R. J. Poole, T. Date and T. Oshima (1989). "Evolution of the vacuolar H^+-ATPase: implications for the origin of eukaryotes." *Proc Natl Acad Sci U S A* **86**(17), 6661-6665.

Gogarten, J. P., T. Starke, H. Kibak, J. Fishman and L. Taiz (1992). "Evolution and isoforms of V-ATPase subunits." *J Exp Biol* **172**, 137-47.

Gómez, M. J., I. Cases and A. Valencia (2004). "Gene order in Prokaryotes: conservation and implications." *Molecules in Time and Space: Bacterial Shape, Division and Phylogeny*. M. Vicente, J. Tamames, A. Valencia and J. Mingorance. Boston, MA, Springer US, 209-237.

Goodenough, U. (1998). *The Sacred Depths of Nature*. New York, Oxford University Press.

Gouy, R., D. Baurain and H. Philippe (2015). "Rooting the tree of life: the phylogenetic jury is still out." *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**(1678), 20140329.

Hahn, A., J. Vonck, D. J. Mills, T. Meier and W. Kühlbrandt (2018). "Structure, mechanism, and regulation of the chloroplast ATP synthase." *Science* **360**(6389), eaat4318.

He, J., H. C. Ford, J. Carroll, C. Douglas, E. Gonzales, S. Ding, I. M. Fearnley and J. E. Walker (2018). "Assembly of the membrane domain of ATP synthase in human mitochondria." *Proceedings of the National Academy of Sciences* **115**(12), 2988.

Hennig, W. (1966). *Phylogenetic Systematics*. Illinois, USA, University of Illinois Press.

Howe, K. L., B. Contreras-Moreira, N. De Silva, G. Maslen, W. Akanni, J. Allen, J. Alvarez-Jarreta, M. Barba, D. M. Bolser, L. Cambell, M. Carbajo, M. Chakiachvili, M. Christensen, C. Cummins, A. Cuzick, P. Davis, S. Fexova, A. Gall, N. George, L. Gil, P. Gupta, K. E. Hammond-Kosack, E. Haskell, S. E. Hunt, P. Jaiswal, S. H. Janacek, P. J. Kersey, N. Langridge, U. Maheswari, T. Maurel, M. D. McDowall, B. Moore, M. Muffato, G. Naamati, S. Naithani, A. Olson, I. Papatheodorou, M. Patricio, M. Paulini, H. Pedro, E. Perry, J. Preece, M. Rosello, M. Russell, V. Sitnik, D. M. Staines, J. Stein, M. K. Tello-Ruiz, S. J. Trevanion, M. Urban, S. Wei, D. Ware, G. Williams, A. D. Yates and P. Flicek (2020). "Ensembl Genomes 2020—enabling non-vertebrate genomic research." *Nucleic Acids Research* **48**(D1), D689-D695.

Hu, B., D. R. Morado, W. Margolin, J. R. Rohde, O. Arizmendi, W. L. Picking, W. D. Picking and J. Liu (2015). "Visualization of the type III secretion sorting platform of <em>Shigella flexneri</em&gt." *Proceedings of the National Academy of Sciences* **112**(4), 1047.

Ibuki, T., K. Imada, T. Minamino, T. Kato, T. Miyata and K. Namba (2011). "Common architecture of the flagellar type III protein export apparatus and F- and V-type ATPases." *Nature Structural & Molecular Biology* **18**(3), 277-282.

Ibuki, T., M. Shimada, T. Minamino, K. Namba and K. Imada (2009). "Crystallization and preliminary X-ray analysis of FliJ, a cytoplasmic component of the flagellar type III protein-export apparatus from Salmonella sp." *Acta Crystallographica Section F* **65**(1), 47-50.

Ibuki, T., Y. Uchida, Y. Hironaka, K. Namba, K. Imada and T. Minamino (2013). "Interaction between FliJ and FlhA, Components of the Bacterial Flagellar Type III Export Apparatus." *Journal of Bacteriology* **195**(3), 466.

Iino, T., Y. Komeda, K. Kutsukake, R. M. Macnab, P. Matsumura, J. S. Parkinson, M. I. Simon and S. Yamaguchi (1988). "New unified nomenclature for the flagellar genes of Escherichia coli and Salmonella typhimurium." *Microbiological Reviews* **52**(4), 533.

Imada, K., T. Minamino, A. Tahara and K. Namba (2007). "Structural similarity between the flagellar type III ATPase FliI and F₁-ATPase subunits." *Proceedings of the National Academy of Sciences* **104**(2), 485.

Ishida, T., R. Ito, J. Clark, N. J. Matzke, Y. Sowa and M. A. B. Baker (2019). "Sodium-powered stators of the bacterial flagellar motor can generate torque in the presence of phenamil with mutations near the peptidoglycan-binding region." *Molecular Microbiology* **111**(6), 1689-1699.

Iwabe, N., K.-I. Kuma, M. Hasegawa, S. Osawa and T. Miyata (1989). "Evolutionary Relationship of Archaebacteria, Eubacteria, and Eukaryotes Inferred from Phylogenetic Trees of Duplicated Genes." *Proceedings of the National Academy of Sciences of the United States of America* **86**(23), 9355-9359.

Iyer, L. M., D. D. Leipe, E. V. Koonin and L. Aravind (2004). "Evolutionary history and higher order classification of AAA+ ATPases." *Journal of Structural Biology* **146**(1), 11-31.

Jackson, M. W. and G. V. Plano (2000). "Interactions between type III secretion apparatus components from *Yersinia pestis* detected using the yeast two-hybrid system." *FEMS Microbiol Lett* **186**(1), 85-90.

Johnston, I. G. and B. P. Williams (2016). "Evolutionary Inference across Eukaryotes Identifies Specific Pressures Favoring Mitochondrial Gene Retention." *Cell Systems* **2**(2), 101-111.

Khanra, N., P. Rossi, A. Economou and C. G. Kalodimos (2016). "Recognition and targeting mechanisms by chaperones in flagellum assembly and operation." *Proceedings of the National Academy of Sciences* **113**(35), 9798.

Kishikawa, J.-i., T. Ibuki, S. Nakamura, A. Nakanishi, T. Minamino, T. Miyata, K. Namba, H. Konno, H. Ueno, K. Imada and K. Yokoyama (2013). "Common Evolutionary Origin for the Rotor Domain of Rotary Atpases and Flagellar Protein Export Apparatus." *PLOS ONE* **8**(5), e64695.

Koumandou, V. L. and S. Kossida (2014). "Evolution of the F0F1 ATP Synthase Complex in Light of the Patchy Distribution of Different Bioenergetic Pathways across Prokaryotes." *PLOS Computational Biology* **10**(9), e1003821.

Lai, Y.-W., P. Ridone, G. Peralta, M. M. Tanaka and M. A. B. Baker (2020). "Evolution of the Stator Elements of Rotary Prokaryote Motors." *Journal of Bacteriology* **202**(3), e00557-19.

Lane, M. C., P. W. O'Toole and S. A. Moore (2006). "Molecular Basis of the Interaction between the Flagellar Export Proteins FliI and FliH from Helicobacter pylori." *Journal of Biological Chemistry* **281**(1), 508-517.

Lathe, W. C., B. Snel and P. Bork (2000). "Gene context conservation of a higher order than operons." *Trends in Biochemical Sciences* **25**(10), 474-479.

Li, W., K. R. O'Neill, D. H. Haft, M. DiCuccio, V. Chetvernin, A. Badretdin, G. Coulouris, F. Chitsaz, Myra K. Derbyshire, A. S. Durkin, N. R. Gonzales, M. Gwadz, Christopher J. Lanczycki, J. S. Song, N. Thanki, J. Wang, Roxanne A. Yamashita, M. Yang, C. Zheng, A. Marchler-Bauer and F. Thibaud-Nissen (2020). "RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation." *Nucleic Acids Research*.

Liu, R. and H. Ochman (2007). "Origins of Flagellar Gene Operons and Secondary Flagellar Systems." *Journal of Bacteriology* **189**(19), 7098.

Lolkema, J. S., Y. Chaban and E. J. Boekema (2003). "Subunit Composition, Structure, and Distribution of Bacterial V-Type ATPases." *Journal of Bioenergetics and Biomembranes* **35**(4), 323-335.

Lupas, A. N. and J. Martin (2002). "AAA proteins." *Current Opinion in Structural Biology* **12**(6), 746-753.

Matzke, N. J. (2003). "Evolution in (Brownian) space: a model for the origin of the bacterial flagellum." 2020, from http://www.talkdesign.org/faqs/flagellum.html.

Matzke, N. J., P. M. Shih and C. A. Kerfeld (2014). "Bayesian analysis of congruence of core genes in *Prochlorococcus* and *Synechococcus* and implications on horizontal gene transfer." *PLoS ONE* **9**(1), e85103.

Minamino, T., Y. V. Morimoto, M. Kinoshita and K. Namba (2020). "Membrane voltage-dependent activation of the flagellar protein export engine." *bioRxiv*, 2020.07.18.210377.

Mulkidjanian, A. Y., M. Y. Galperin and E. V. Koonin (2009). "Co-evolution of primordial membranes and membrane proteins." *Trends in Biochemical Sciences* **34**(4), 206-215.

Mulkidjanian, A. Y., K. S. Makarova, M. Y. Galperin and E. V. Koonin (2007). "Inventing the dynamo machine: the evolution of the F-type and V-type ATPases." *Nature Reviews Microbiology* **5**(11), 892-899.

Niu, Y., S. Moghimyfiroozabad, S. Safaie, Y. Yang, E. A. Jonas and K. N. Alavian (2017). "Phylogenetic Profiling of Mitochondrial Proteins and Integration Analysis of Bacterial Transcription Units Suggest Evolution of F1Fo ATP Synthase from Multiple Modules." *Journal of Molecular Evolution* **85**(5), 219-233.

Noji, H., H. Ueno and D. G. G. McMillan (2017). "Catalytic robustness and torque generation of the F1-ATPase." *Biophysical Reviews* **9**(2), 103-118.

Pallen, M. J., C. M. Bailey and S. A. Beatson (2006). "Evolutionary links between FliH/YscL-like proteins from bacterial type III secretion systems and second-stalk components of the FoF1 and vacuolar ATPases." *Protein Science* **15**(4), 935-941.

Pallen, M. J., S. A. Beatson and C. M. Bailey (2005). "Bioinformatics analysis of the locus for enterocyte effacement provides novel insights into type-III secretion." *BMC Microbiology* **5**(1), 9.

Pallen, M. J. and N. J. Matzke (2006). "From The Origin of Species to the origin of bacterial flagella." *Nature Reviews Microbiology* **4**(10), 784-790.

Philippe, H., H. Brinkmann, D. V. Lavrov, D. T. J. Littlewood, M. Manuel, G. Wörheide and D. Baurain (2011). "Resolving Difficult Phylogenetic Questions: Why More Sequences Are Not Enough." *PLOS Biology* **9**(3), e1000602.

Rizzotti, M. (2000). *Early Evolution: From the appearance of the first cell to the first modern organisms*. Boston, Birkhäuser Verlag.

Romo-Castillo, M., A. Andrade, N. Espinosa, J. Monjarás Feria, E. Soto, M. Díaz-Guerrero and B. González-Pedrajo (2014). "EscO, a Functional and Structural Analog of the Flagellar FliJ Protein, Is a Positive Regulator of EscN ATPase Activity of the Enteropathogenic Escherichia coli Injectisome." *Journal of Bacteriology* **196**(12), 2227.

Rost, B. (1999). "Twilight zone of protein sequence alignments." *Protein Engineering, Design and Selection* **12**(2), 85-94.

Santiveri, M., A. Roa-Eguiara, C. Kühne, N. Wadhwa, H. Hu, H. C. Berg, M. Erhardt and N. M. I. Taylor (2020). "Structure and Function of Stator Units of the Bacterial Flagellar Motor." *Cell* **183**(1), 244-257.e16.

Shih, P. M., J. Hemp, L. M. Ward, N. J. Matzke and W. W. Fischer (2017). "Crown group Oxyphotobacteria postdate the rise of oxygen." *Geobiology* **15**(1), 19-29.

Shih, P. M. and N. J. Matzke (2013). "Primary endosymbiosis events date to the later Proterozoic with cross-calibrated phylogenetic dating of duplicated ATPase proteins." *Proceedings of the National Academy of Sciences* **110**(30), 12355-12360.

Snyder, L. A. S., N. J. Loman, K. Fütterer and M. J. Pallen (2009). "Bacterial flagellar diversity and evolution: seek simplicity and distrust it?" *Trends in Microbiology* **17**(1), 1-5.

Sobti, M., R. Ishmukhametov, J. C. Bouwer, A. Ayer, C. Suarna, N. J. Smith, M. Christie, R. Stocker, T. M. Duncan and A. G. Stewart (2019). "Cryo-EM reveals distinct conformations of E. coli ATP synthase on exposure to ATP." *eLife* **8**, e43864.

Sobti, M., H. Ueno, H. Noji and A. G. Stewart (2020). "The six steps of the F₁-ATPase rotary catalytic cycle." *bioRxiv*, 2020.12.15.422962.

Sobti, M., J. L. Walshe, D. Wu, R. Ishmukhametov, Y. C. Zeng, C. V. Robinson, R. M. Berry and A. G. Stewart (2020). "Cryo-EM structures provide insight into how E. coli F1Fo ATP synthase accommodates symmetry mismatch." *Nature Communications* **11**(1), 2615.

Stoebe, B. and K. V. Kowallik (1999). "Gene-cluster analysis in chloroplast genomics." *Trends in Genetics* **15**(9), 344-347.

Vasanthakumar, T. and J. L. Rubinstein (2020). "Structure and Roles of V-type ATPases." *Trends in Biochemical Sciences* **45**(4), 295-307.

von Ballmoos, C., A. Wiedenmann and P. Dimroth (2009). "Essentials for ATP Synthesis by F1F0 ATP Synthases." *Annual Review of Biochemistry* **78**(1), 649-672.

Wilkens, S. and R. A. Capaldi (1998). "ATP synthase's second stalk comes into focus." *Nature* **393**(6680), 29-29.

Ye, J., A. R. Osborne, M. Groll and T. A. Rapoport (2004). "RecA-like motor ATPases—lessons from structures." *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1659**(1), 1-18.

Zhulin, I. B. (2017). "By Staying Together, Two Genes Keep the Motor Running." *Structure* **25**(2), 214-215.

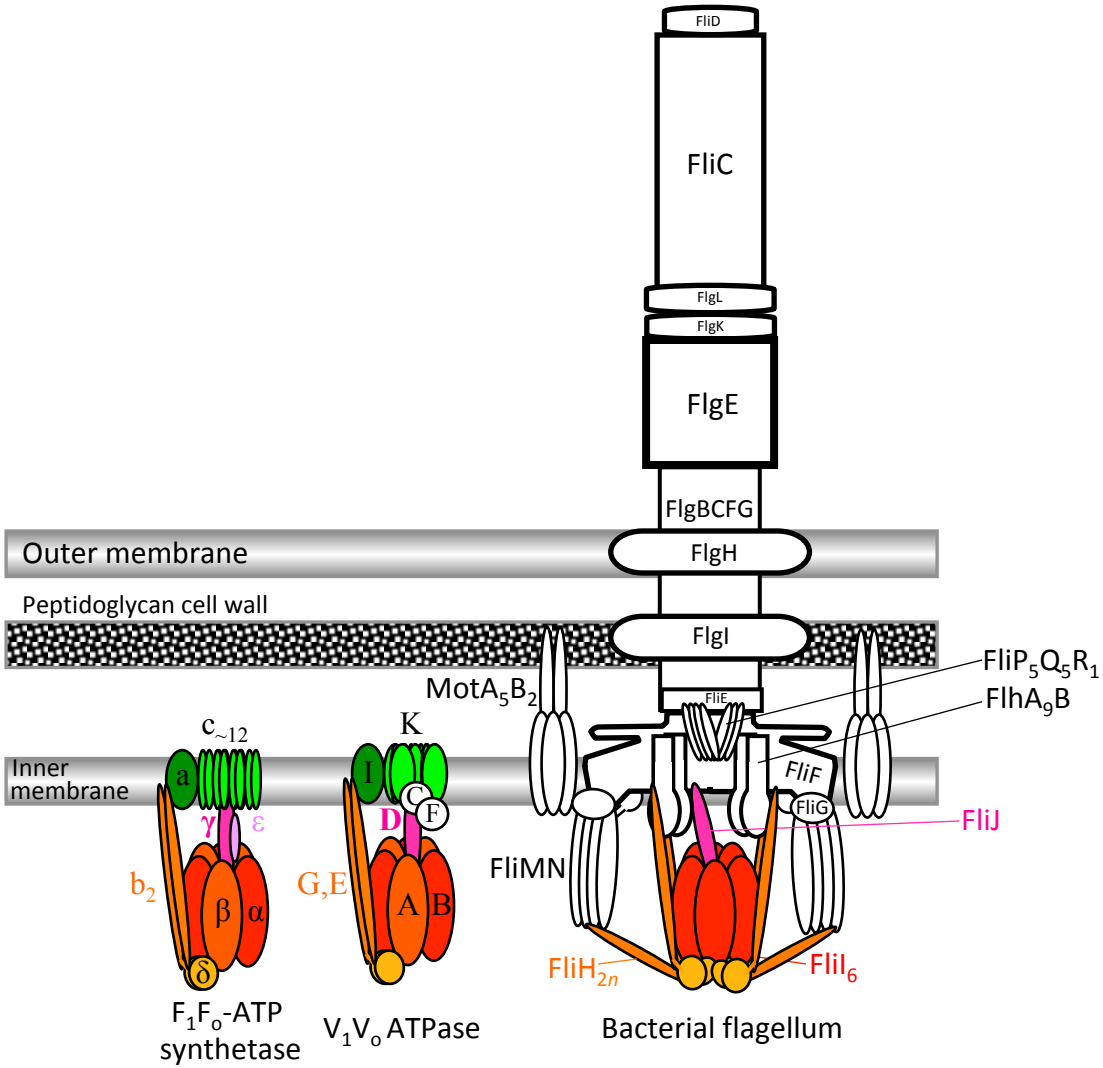

Figure 1. Cartoon depicting location of core structural components of F_1F_o-ATP synthetase, V/A-type ATPases, and the bacterial flagellum. Colored components match the colors in Fig. 2. For the V-type ATPases, naming follows the bacterial V-ATPases (Kolkema et al. 2003); see Supp. Material for corresponding names across systems. Stoichiometry is indicated with subscripts where important in this discussion, but most flagellar proteins form large oligomers (rings or filaments) with many subunits. Cartoon after Macnab (2003), updated with Minamino & Imada (2015), Minamino et al. (2017), Beeby (2020). FliH interacts with both FlhA and FliN so both are depicted, after Minamino & Imada (2015).

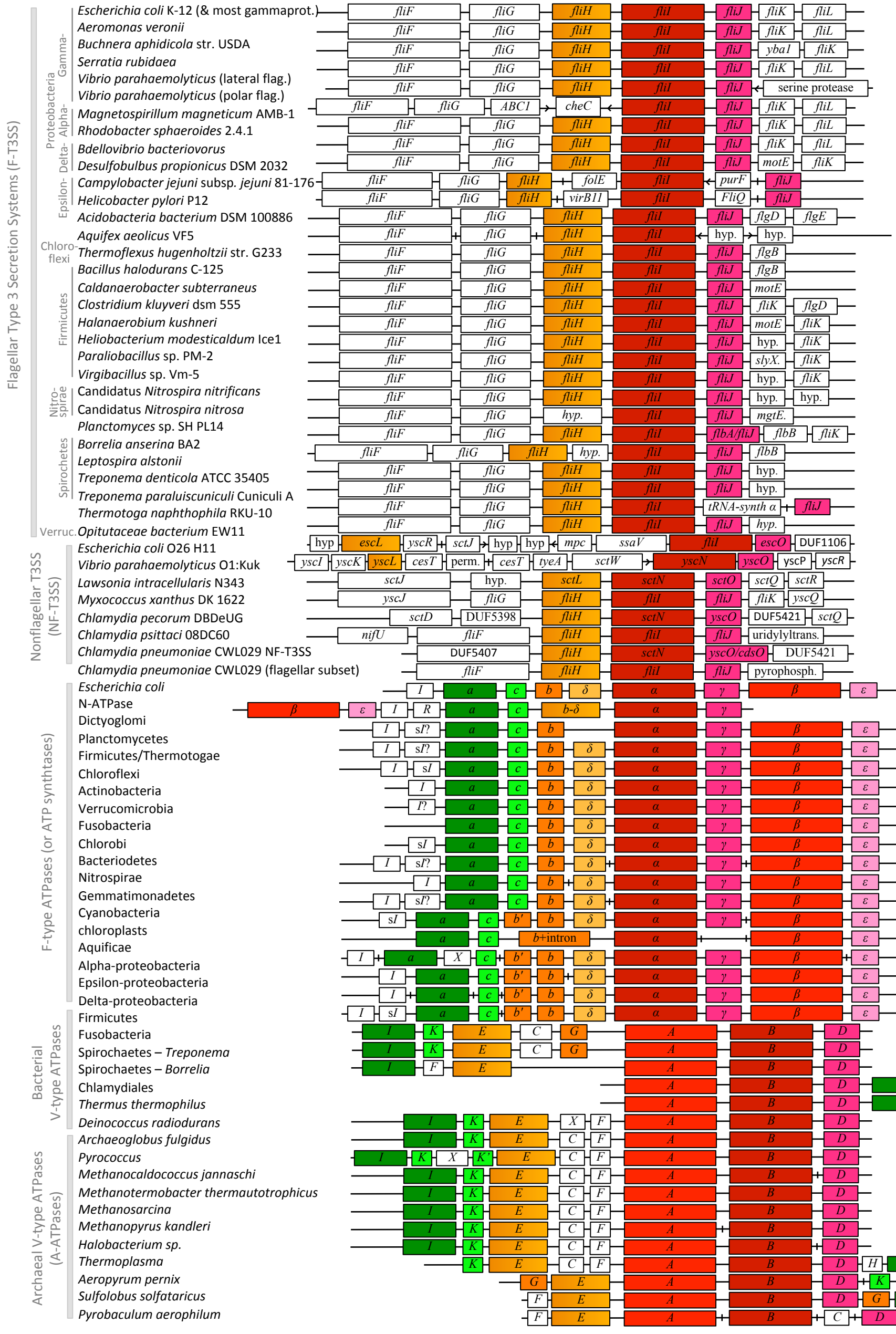
Flagellar Type 3 Secretion Systems (F-T3SS)
Proteobacteria
Gamma-
Escherichia coli K-12 (& most gammaprot.)
Aeromonas veronii
Buchnera aphidicola str. USDA
Serratia rubidaea
Vibrio parahaemolyticus (lateral flag.)
Vibrio parahaemolyticus (polar flag.)
Alpha-
Magnetospirillum magneticum AMB-1
Rhodobacter sphaeroides 2.4.1
Delta-
Bdellovibrio bacteriovorus
Desulfobulbus propionicus DSM 2032
Epsilon-
Campylobacter jejuni subsp. jejuni 81-176
Helicobacter pylori P12
Acidobacteria bacterium DSM 100886
Chloro-flexi
Aquifex aeolicus VF5
Thermoflexus hugenholtzii str. G233
Firmicutes
Bacillus halodurans C-125
Caldanaerobacter subterraneus
Clostridium kluyveri dsm 555
Halanaerobium kushneri
Heliobacterium modesticaldum Ice1
Paraliobacillus sp. PM-2
Virgibacillus sp. Vm-5
Nitro-spirae
Candidatus Nitrospira nitrificans
Candidatus Nitrospira nitrosa
Planctomyces sp. SH PL14
Spirochetes
Borrelia anserina BA2
Leptospira alstonii
Treponema denticola ATCC 35405
Treponema paraluiscuniculi Cuniculi A
Thermotoga naphthophila RKU-10
Verruc.
Opitutaceae bacterium EW11
Nonflagellar T3SS (NF-T3SS)
Escherichia coli O26 H11
Vibrio parahaemolyticus O1:Kuk
Lawsonia intracellularis N343
Myxococcus xanthus DK 1622
Chlamydia pecorum DBDeUG
Chlamydia psittaci 08DC60
Chlamydia pneumoniae CWL029 NF-T3SS
Chlamydia pneumoniae CWL029 (flagellar subset)
F-type ATPases (or ATP synthtases)
Escherichia coli
N-ATPase
Dictyoglomi
Planctomycetes
Firmicutes/Thermotogae
Chloroflexi
Actinobacteria
Verrucomicrobia
Fusobacteria
Chlorobi
Bacteriodetes
Nitrospirae
Gemmatimonadetes
Cyanobacteria
chloroplasts
Aquificae
Alpha-proteobacteria
Epsilon-proteobacteria
Delta-proteobacteria
Bacterial V-type ATPases
Firmicutes
Fusobacteria
Spirochaetes – Treponema
Spirochaetes – Borrelia
Chlamydiales
Thermus thermophilus
Deinococcus radiodurans
Archaeal V-type ATPases (A-ATPases)
Archaeoglobus fulgidus
Pyrococcus
Methanocaldococcus jannaschi
Methanotermobacter thermautotrophicus
Methanosarcina
Methanopyrus kandleri
Halobacterium sp.
Thermoplasma
Aeropyrum pernix
Sulfolobus solfataricus
Pyrobaculum aerophilum
fliF fliG fliH fliI fliJ fliK fliL
yba1
serine protease
ABC1 cheC
motE
folE purF
virB11 FliQ
flgD flgE
hyp.
flgB
slyX.
mgtE.
flbA/fliJ flbB
tRNA-synth α
escL yscR sctJ mpc ssaV escO DUF1106
yscI yscK yscL cesT perm. tyeA sctW yscN yscO yscP
sctL sctN sctO sctQ sctR
yscJ yscQ
sctD DUF5398 DUF5421
nifU uridylyltrans.
DUF5407 yscO/cdsO
pyrophosph.
I a c b δ α γ β ε
β ε I R a c b-δ α γ
I sI? a c b α γ β ε
I sI a c b δ α γ β ε
I? a c b δ α γ β ε
sI a c b δ α γ β ε
I sI? a c b δ α γ β ε
sI a c b' b δ α γ β ε
a c b+intron α β ε
I a X c b' b δ α γ β ε
I a c b' b δ α γ β ε
I a c b' b δ α γ β ε
I sI a c b' b δ α γ β ε
I K E C G A B D
I K E C G A B D
I F E A B D
A B D I K
A B D I K
I K E X F A B D
I K E C F A B D
I K X K' E C F A B D
I K E C F A B D
I K E C F A B D
I K E C F A B D
I K E C F A B D
K E C F A B D H I
G E A B D K
F E A B D G K
F E A B C D

Figure 2. Conserved gene order between *fliHIJ* genes and V/F/A-ATPase genes. Homologs of *fliH*, *fliI*, and *fliJ* are colored orange, red, and dark pink, respectively. Homologs of V/F/A-ATPase a-, c-, and ε-subunits are dark green, light green, and light pink. Other genes are white. Genes are ordered from upstream to downstream with respect to the *fliI* exon. Genes coded on the opposite strand are indicated by left arrows (<); return to the *fliI* strand by right arrows (>). Large breaks are indicated by vertical lines (|) as in Lolkema et al. (2003). For T3SS gene order, see Methods and Supplemental Material; Supp. Mat. also includes the full names of strains. F-ATPase genes are named by the protein subunit, with the order taken from Koumandou & Kossida (2014). V/A-ATPase gene names & order from Kolkema et al. (2003). *E. coli* F_1F_O-ATPase after Nielsen et al. (1984). Chloroplast (*Arabidopsis*) gene order from *https://www.ncbi.nlm.nih.gov/gene/844790* ; chloroplast *atpG* (b', II), *atpC* (γ), and *atpD* (δ), are encoded in the plant nuclear genome (Hermans et al. 1988; Stoebe & Kowallik 1999). This is a schematic depiction to highlight gene order; the gene lengths and spaces between genes are not intended to be exact. *Notes:* N-ATPase "b" is a b+δ fusion (Dibrova et al. 2010), as is V1-E (Pallen et al. 2006). The V_O I-subunit is homologous to F_O−a and the subunit named "a" in yeast V-ATPases (Chaban et al. 2005). The F_O−I subunit (and the homologous sI) are Ca^{2+}/Mg^{2+} transporters, unique to F-ATPases (Koumandou & Kossida 2014). DUF: conserved Domain of Unknown Function; hyp. = hypothetical protein in database, not identified as homologous to T3SS genes by the database or manual searches.