Gene 893 (2024) 147895

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

FicD genes in invertebrates: A tale of transposons, pathogenic and integrated viruses

Umberto Rosani [a,*], Sofia De Felice [a], Riccardo Frizzo [a], Satoshi Kawato [b], K. Mathias Wegner [c]

[a] Department of Biology, University of Padova, 35121 Padova, Italy
[b] Laboratory of Genome Science, Tokyo University of Marine Science and Technology, 108-8477 Tokyo, Japan
[c] Alfred Wegener Institute - Helmholtz Centre for Polar and Marine Research, Waddensea Station Sylt, 25992 List, Germany

ARTICLE INFO

Edited by: Justin Blumenstiel

Keywords:
AMPylation
OsHV-1
WSSV
Integrated viruses

ABSTRACT

Many gene families are shared across the tree of life between distantly related species because of horizontal gene transfers (HGTs). However, the frequency of HGTs varies strongly between gene families and biotic realms suggesting differential selection pressures and functional bias. One gene family with a wide distribution are FIC-domain containing enzymes (FicDs). FicDs catalyze AMPylation, a post-translational protein modification consisting in the addition of adenosine monophosphate to accessible residues of target proteins. Beside the well-known conservation of FicDs in deuterostomes, we report the presence of a conserved FicD gene ortholog in a large number of protostomes and microbial eukaryotes. We also reported additional FicD gene copies in the genomes of some rotifers, parasitic worms and bivalves. A few dsDNA viruses of these invertebrates, including White spot syndrome virus, Cherax quadricarinatus iridovirus, Ostreid herpesvirus-1 and the beetle nudivirus, carry copies of FicDs, with phylogenetic analysis suggesting a common origin of these FicD copies and the duplicated FicDs of their invertebrate hosts. HGTs and gene duplications possibly mediated by endogenous viruses or genetic mobile elements seem to have contributed to the transfer of AMPylation ability from bacteria and eukaryotes to pathogenic viruses, where this pathway could have been hijacked to promote viral infection.

1. Introduction

Horizontal gene transfer (HGT) can transfer novel functions to recipient organisms (Oliveira et al., 2017; Irwin et al., 2022). In other cases HGT might simply transfer "junk DNA", making the understanding of the biological significance of this phenomenon puzzling (Vogan and Higgs, 2011; van Dijk et al., 2020). Indeed, the functional roles of widespread transferred genes have remained poorly investigated. One function that has been transferred between biological domains is post-translational protein modification by AMPylation (Khater and Mohanty, 2015), which is gaining attention for its role in neurodegenerative diseases, physiology, and host-bacteria interactions (Woolery et al., 2010; Chambers and Scheck, 2020; Yarbrough et al., 2009; Truttmann et al., 2018). AMPylation consists in the covalent addition of 5'-Adenosine monophosphate (AMP) to the hydroxyl group of an accessible Ser, Thr or Tyr residue located on target proteins (Woolery et al., 2010). The Filamentation-induced-by-cyclic-AMP enzyme (FicD) was firstly discovered in *Escherichia coli* as a protein associated with cell filamentation at elevated temperatures (Utsumi et al., 1982). Bioinformatic characterization of FicD homologs revealed a conserved motif (HxFx(D/E)GNGRxxR) associated with AMPylation for, among others, *Vibrio outer protein S* (Vops) and *Histophilus somni Immunoglobin-binding protein A* (IbpA) (Roy and Mukherjee, 2009). Bacterial FicD proteins catalyze a broad range of reactions besides AMPylation, such as phosphorylation, phosphocholination and UMPylation (Castro-Roa et al., 2013; Campanacci et al., 2013), able to affect translation regulation, host defense suppression, and regulation of eukaryotic cell processes (Garcia-Pino, Zenkin, and Loris, 2014). Particularly, bacterial-promoted AMPylation has been shown to induce cytotoxicity in the infected cells by targeting host proteins of the Rho GTPase family (Chatterjee and Truttmann, 2021; Harms et al., 2016; Chambers and Scheck, 2020; Yarbrough et al., 2009).

Metazoan FicDs encode a slightly different FIC motif (HxFx(D/E)GN(G/K)RxxR) (Roy and Cherfils, 2015), and can also show adenylyltransferase activity (Sanyal et al., 2015). In contrast to bacterial FicDs, knowledge about the FicD range of reactions and functions is scarce

Abbreviations: HGT, Horizontal Gene Transfer; OsHV-1, Ostreid herpesvirus-1; WSSV, White Spot Syndrome Virus.

* Corresponding author at: v. U. Bassi 58/b, 35121 Padova, Italy.
E-mail address: umberto.rosani@unipd.it (U. Rosani).

https://doi.org/10.1016/j.gene.2023.147895
Received 4 March 2023; Received in revised form 5 October 2023; Accepted 10 October 2023
Available online 11 October 2023

among metazoans. Among animals, FicDs are generally recognized as AMP transferases, and were found predominantly involved in the regulation of the unfolded protein response (UPR) and in the suppression of Transforming Growth Factors (TGF). In *Drosophila*, FicD (dFic) regulates UPR in the endoplasmic reticulum (ER) by reversibly modifying the constitutively-expressed heat shock protein 70 BiP through the covalent attachment of AMP (Ham et al., 2014). Moreover, *C. elegans* FIC-1 was shown to influence pathogen avoidance behavior by suppressing the production of TGF-β ligands (Hernandez-Lima et al., 2022), and to affect the dynamics of protein aggregation acting on Heat Shock Protein families, as well as neuro-development and neuro-degeneration (Truttmann et al., 2018). Similar functions are hypothesized for HYPE, the human ortholog of *Drosophila* and *C. elegans* FicDs (Moehlman et al., 2018; Truttmann et al., 2016; Hernandez-Lima et al., 2022). I*n vitro* studies further added core histones as AMPylation targets of FIC-1 (Truttmann et al., 2016), and provided proof of alternative post-translational reactions linked to FicD activity, such as GMPylation, UMPylation and phosphorylation (Cruz et al., 2014; Veyron et al., 2018; Preissler et al., 2017; Perera et al., 2019).

FicDs are widespread in the domains of life, including a few dsDNA viruses (Mistry et al., 2020; Gulen and Itzen, 2021). Phylogenetic analyses suggest that metazoan and archaea FicDs have originated from multiple HGTs from bacteria (Khater and Mohanty, 2015). Although the presence of the FIC domain among viruses has been reported for a few bacteriophages and other DNA viruses, the origin of viral FicDs, their conservation, and biological function are currently unknown (Gulen and Itzen, 2021). Beside phages, the FIC domain has been reported in white spot syndrome virus (WSSV) and Ostreid herpesvirus-1 (OsHV-1) (Rosani and Venier, 2017; Yarbrough et al., 2009), with the OsHV-1 FIC encoded within the small ribonucleotide reductase (RDR2) gene (Yarbrough et al., et al., 2009). Both WSSV and OsHV-1 cause diseases in crustaceans or bivalves, with detrimental consequences for aquaculture (FAO, 2020). Although global marine metagenomic projects (e.g. TARA (Zhang and Ning, 2015; Sunagawa et al., 2020)) are currently expanding the known gene repertoire of prokaryotes, viruses and protists, together with their coding genes and phylogenetic connections (Delmont et al., 2022; Dorrell et al., 2021; Salazar et al., 2019; Schulz et al., 2020; Cheng et al., 2021), the phylogenetic origin of these viruses remains enigmatic like in OsHV-1 (Rosani et al., 2023).

Here, we investigated the taxonomic distribution of FicD genes in eukaryotes and related viruses, highlighting the presence of a few gene expansion events and HGTs. We also investigated the structural conservation of WSSV and OsHV-1 FicDs, as emblematic examples of functional novelties in highly pathogenic DNA viruses likely introduced by HGTs.

2. Materials and methods

2.1. Sequence data retrieving and preliminary analysis

Annotated genome assemblies of protostome metazoans (N = 556), OsHV-1 (N = 7) and WSSV (N = 29) were retrieved from the NCBI databases (Supplementary Table 1). The collection of 47 million (M) gene sequences were obtained from the assemblies of 370 metagenomic samples produced by the TARA ocean consortium (https://doi.org/10.5281/zenodo.3473199). Similarly, eukaryotic metagenomic-assembled genomes (MAGs) provided with taxonomic classification were downloaded from the TARA dataset (Delmont et al., 2022). Predicted proteomes were obtained by translating the open reading frame (ORF) of the genomic-annotated features using CLC Genomic Workbench v.21 (Qiagen, US). Genomes, metagenomic-derived genes and proteomes were transformed into nucleotidic and protein *blast* databases for gene and protein identification, respectively.

2.2. Identification of FicD and sequence analysis

HMMer v3.3.2 (Eddy, 2011) was used in combination with the FIC Hidden Markov Model (Pfam HMM ID: PF02661) to identify putative FicD hits among the previously described proteomes. FicD genomic loci were identified in selected species by *tblastn* searches using the corresponding FicD proteins as queries. To identify FIC-associated conserved domains, *hmmsearch* was used in combination with the Pfam-A HMM dataset v.34 (Mistry et al., 2020). The conserved FIC and inhibitory motifs were identified using Java regular expression syntax. All the retrieved FicD sequences were taxonomically classified using DIAMOND *blastp* searches (Buchfink et al., 2015) against the NCBI *nr* database (downloaded in September 2021). The presence of the FIC domain in OsHV-1 relatives or in WSSV- and OsHV-1-like sequences putatively integrated into host genomes (crustacean species,for WSSV, *Branchiostoma* spp. and *Capitella teleta*, for OsHV-1, that have been previously identified (Rosani et al., 2023; Mushegian et al., 2018; Kawato et al., 2019), was investigated as described above after obtaining the relevant genome sequences from NCBI. Additionally, 210 deuterostome FicD orthologs were retrieved from the Ensembl genome browser (release 109) (https://www.ensembl.org/). The complete set of analyzed sequences is provided as Supplementary Data 1.

2.3. Phylogenetic analysis

A subset of the FIC domains were used to build a phylogenetic tree. Redundant sequences (>99 % identity) were removed using *CD-HIT v4.7* (Fu et al., 2012) and the remaining hits were aligned with the G-INS-i algorithm of *MAFFT v7.490* (Katoh and Standley, 2013). Sites with more than 80 % gap and sequences containing more than 50 % of gaps were trimmed using *Goalign v0.3.1* (Lemoine and Gascuel, 2021). Tree reconstruction was done with *IQ-TREE v1.6.12* (Nguyen et al., 2015) with the LG + R10 substitution model, as determined to be most suitable for the analysed data using *ModelFinder* (Kalyaanamoorthy et al., 2017). The branch supports were computed using the SH-aLRT test (Guindon et al., 2010, 0) and ultrafast bootstrap estimation (Hoang et al., 2018, 2). The alignment in fasta format and the phylogenetic tree are available as Supplementary Data 2 and Supplementary Data 3, respectively. To integrate our analysis with a previous report of the phylogenesis of FicD proteins (Khater and Mohanty, 2015), 1,345 hits were downloaded from the supporting files associated to that study, combined with the ones described in this work, and used for a second phylogenetic analysis.

2.4. Genomic context of FicD genes

To investigate the presence of conserved genes flanking FicDs in selected Bivalvia, Gastropoda and Crustacea species (reported in Table 1), the 60 kb genomic regions upstream and downstream of FicD genes were considered. The conserved domains were identified on the 6-frame translations using *hmmsearch*. To investigate the nucleotidic conservation of FicD among strains of OsHV-1 and WSSV, 29 complete genome sequences of WSSV (NCBI IDs: NC_075105.1, NC_003225.3, AP027278.1, AP027279.1, AP027280.1, AP027281.1, AP027282.1, AP027283.1, AP027284.1, AP027285.1, AP027286.1, AP027287.1, AP027288.1, AP027289.1, AP027290.1, MN840357.1, MH090824.1, KX686117.1, MG702567.1, MF768985.1, KY827813.1, KU216744.2, KT995470.1, KT995471.1, KT995472.1, JX515788.1, AF440570.1, AF332093.3, AF369029.2) and 30 of OsHV-1 (NCBI IDs: NC_005881.2, OQ101585.1, OM811577.1, OM811578.1, OM811579.1, OM811580.1, OM811581.1, OM811582.1, OM811583.1, OM811584.1, OM811585.1, M811586.1, OM811587.1, OM811588.1, OM811589.1, OM811590.1, OM811591.1, OM811592.1, OM811593.1, OM811594.1, OM811595.1, OM811596.1, OM811597.1, MW412420.1, MF509813.1, MG561751.2, KY242785.1, KY271630.1, KP412538.1, AY509253.2) were retrieved from NCBI and the FicD loci were extracted and aligned using MUSCLE. The resulting two alignments were inspected for the presence of single-

Table 1
Summary of the FicD genes of bivalves, crustaceans and dsDNA viruses infecting these species. Subphylum, species name and family and the number of FicD genes divided between multi-exons and single exon genes are reported.

Subphylum	Species name	Family	No. of FicD genes	
			multi-exon	single exon
Conchifera	*Scapharca broughtonii*	*Arcidae*	1	
	Panopea generosa	*Hiatellidae*		1
	Lutraria rhynchaena	*Mactridae*	1	1
	Bathymodiolus platifons	*Mytilidae*	1	
	Limnoperna fortunei	*Mytilidae*	1	2
	Modiolus philippinarum	*Mytilidae*	1	
	Mytilus coruscus	*Mytilidae*	1	6
	Mytilus edulis	*Mytilidae*	1	5
	Mytilus galloprovincialis	*Mytilidae*	1	6
	Crassostrea gigas	*Ostreidae*	1	
	Crassostrea virginica	*Ostreidae*	1	
	Ostrea edulis	*Ostreidae*	1	
	Saccostrea glomerata	*Ostreidae*	1	
	Mizuhopecten yessoensis	*Pectinidae*	1	1
	Pecten maximus	*Pectinidae*	1	
	Sinonovacula constricta	*Solecurtidae*	1	
	Cyclina sinensis	*Veneridae*	1	2
	Mercenaria mercenaria	*Veneridae*	1	1
	Ruditapes philippinarum	*Veneridae*	1	2
	Archivesica marissinica	*Vesicomyidae*	1	
	Pomacea caniculata	*Ampullariidae*	1	
	Gigantopelta aegis	*Peltospiridae*	1	
	Haliotis rufescens	*Haliotidae*	1	
	Haliotis diversicolor	*Haliotidae*	1	
Crustacea	*Armadillidium vulgare*	*Armadillidiidae*	1	8
	Armadillidium nasutum	*Armadillidiidae*	1	
	Panaeus japonicus	*Penaeidae*	1	
	Panaeus vannamei	*Penaeidae*	1	
	Panaeus monodon	*Penaeidae*	1	
	Lepeophtheirus salmonis	*Caligidae*	1	
	Homarus americanus	*Nephropidae*	1	
	Hyalella azteca	*Hyalellidae*	1	
dsDNA viruses	Ostreid herpesvirus-1	*Malacoherpesviridae*		1
	White Spot Syndrome Virus	*Nimaviridae*		1
	Cherax quadricarinatus iridovirus	*Iridoviridae*		1
	Oryctes rhinoceros nudivirus	*Nudiviridae*		1

nucleotide variations.

2.5. Structural analysis

The OsHV-1 and WSSV FIC protein structures were modeled by AlphaFold2, using YP_024565.1 and YP_009220606.1 sequences as queries, respectively, and restricting the homology search to the 70 % of the PDB. The AlphaFold model structure of human protein adenylyltransferase FICD was downloaded from Uniprot (ID A0A024RBM8). The structural comparison was performed using UCSF Chimera, limited to the two catalytic motifs of the FIC domains.

3. Results

3.1. Tracing FicD genes revealed conservation in eukaryotes and scarcity among viruses

The FIC domain associated to a transmembrane region and to tetratricopeptide helical domains (Uniprot ID: IPR011990) is a protein architecture described as *FIC-domain containing protein* (FicD) in vertebrates (Chatterjee and Truttmann, 2021). Screening the genome annotations of 556 protostomes we identified at least one FicD gene in 501 genomes, for a total of 752 hits. Seventy-five percent of these protostome species encoded a single-copy FicD gene, as previously described for deuterostome metazoans, and 16 % possessed two gene-copies (Fig. 1a). Multiple FicD gene copies are present in the bivalves *Mya arenaria* and *Mytilus californianus* (17 and 9 copies, respectively), in the spiders *Nephila pilipes* and *Trichonephila clavata*, in the flatworm *Macrostomum lignano*, in the rotifer *Rotaria_sp. Silwood 1* and in some fruit flies (Fig. 1b). Redundancy removal revealed that some of these expanded FicD genes are identical, as for *M. arenaria* or *Drosophila spp*. (Fig. 1c).

We then searched the FIC domain in genes predicted from 370 globally distributed ocean metagenomic assemblies and in the paired eukaryotic Metagenomic Assembled Genomes (MAGs) (Salazar et al., 2019; Delmont et al., 2022). Proteomes derived from metagenomic assemblies included 1,360 FicDs, collapsed into 387 non-redundant proteins and taxonomic analysis indicated that most of them correctly referred to microbes, namely bacteria (76 %) or archaea (19 %), whereas a minority of them is likely part of eukaryotic genomes (3 %) or remained unclassified. By mining 130 eukaryotic MAGs we obtained 264 FicDs, with a maximum of 6 genes per MAG (Fig. 1c). To update the distribution of the FIC domain among viruses, we interrogated the NCBI *nt* database using the 68 viral hits available from the PFAM database. Only two additional hits were identified, referring to slightly divergent sequences of Cherax quadricarinatus iridovirus (CQIV). As result, the FIC domain was only detectable in phages, in the bivalve *Malacoherpesviridae* (OsHV-1, but not in its nearest relative, the abalone herpesvirus), in two crustacean viruses of the *Nimaviridae* (WSSV) and *Iridoviridae* (CQIV) families and in the beetle Oryctes rhinoceros nudivirus (OrNV, *Nudiviridae*). Viral FicD proteins mostly encoded the FIC domain alone or associated to the N-terminal FIC (FIC_N) or the Helix-turn-helix (HTH) domains among phages. A unique exception is the FIC of OsHV-1, since the domain is encoded within the ribonucleotide reductase small subunit gene (RNR2). Overall, we considered 1,649 FicD sequences (Supplementary Data 1).

3.2. Phylogenetic analysis of the FIC domains revealed a limited distribution of HGTs

We performed multiple sequence alignment and phylogenetic analysis of FIC domains obtained from protostome, deuterostome, viral, *Wolbachia* and marine microbial sequences (derived from TARA database), for a total of 1,269 non-redundant (>99 %) hits. The refined alignment included 102 informative positions referring to 1,251 sequences (Supplementary Data 2) and was used to generate a consensus phylogenetic tree (Fig. 2a, Supplementary Data 3 and https://itol.embl.de/shared/umbe1984). The tree was split into two halves. The first half formed a more derived clade including most of the FicDs of metazoan species and had a similar topology to the phylogenetic relationships among taxa, with a subclade including deuterostome FicDs (green taxonomy labels in Fig. 2a) and one including protostome ones (orange labels). Among the latter group, we found a separate clade referring to FicD sequences from *Hexanauplia* arthropods obtained from the eukaryotic MAGs (violet taxonomy labels, indicated as I in Fig. 2a). As expected, these hits clustered together with the few copepod FicDs obtained from genomic datasets. A second group of eukaryotic MAG-derived hits clustered together with bacterial and tardigrada hits (indicated as II in Fig. 2a). At the base of the whole clade, we found two

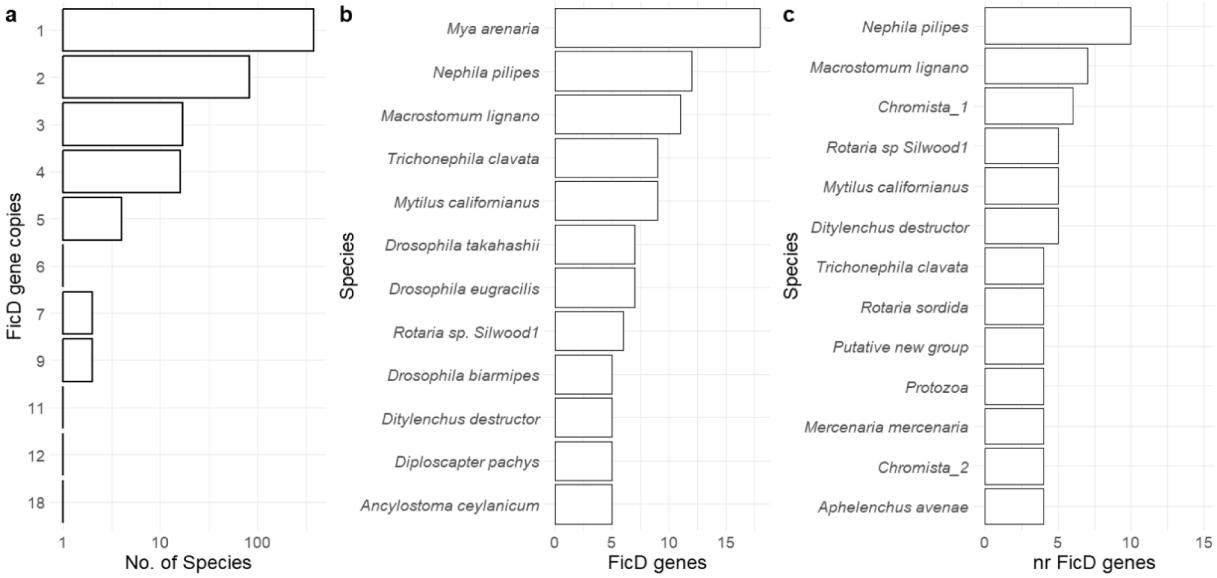

Fig. 1. Identification of FicD genes. a. The abundance of FicD genes in the analyzed species is reported. b. The species with the highest number of FicD gene copies is reported, (c) as well as the species with the highest number of non-redundant FicD gene copies.

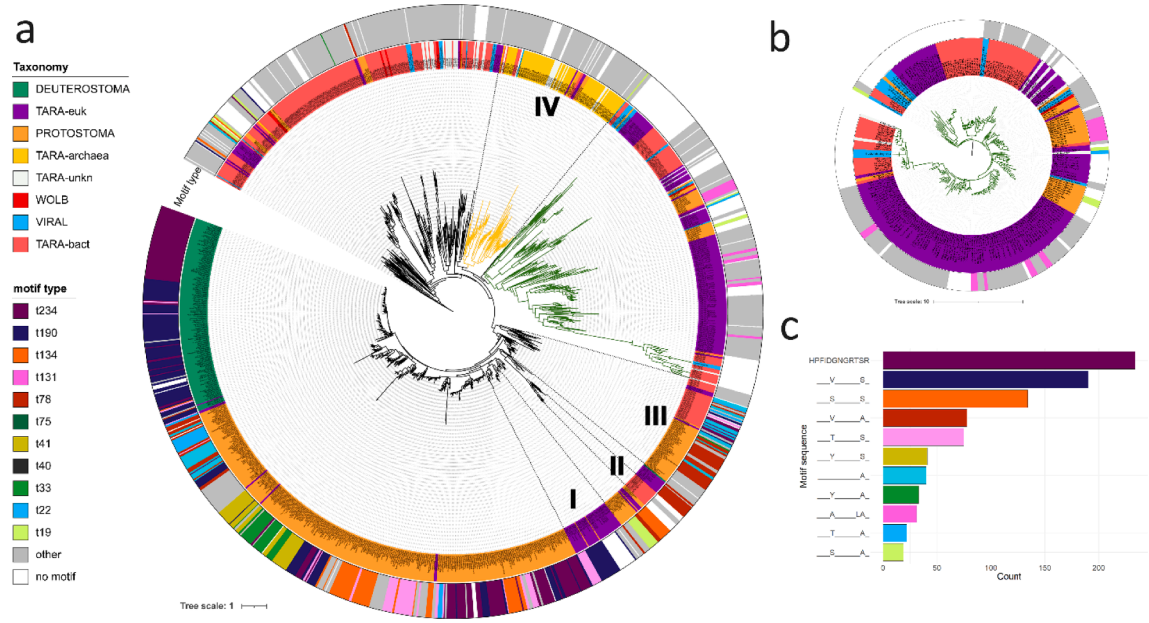

Fig. 2. Phylogenetic tree of the FIC domain. a. The phylogenetic tree was obtained from the 1,251 aligned FIC domain. The circular annotations referred, from the most inner to the outer circle, to the taxonomic annotation of the hits (green, deuterostome; orange, protostomes; violet, eukaryotic MAGs; red, bacteria; yellow, archaea; light blue, viruses; red, Wolbachia and to the conservation of the FIC enzymatic motifs. The 11 most conserved motifs are reported in order of abundance using a color-code (referred both to the legend on the left and to panel c). A few clades mentioned in the results have been highlighted (I, II, III and IV). b. Magnification of the clade referring to eukaryotic MAGs (highlighted with green branches in panel a). The circular annotations are maintained consistent with panel a. c. The sequences of the 11 most-represented enzymatic motifs have been reported together with their abundance. Lines in the consensus motif sequences referred to conserved residues. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

groups of arthropod and bacterial hits (orange and red labels, indicated as III in Fig. 2a). The good match between FIC and species phylogenies along with the short branch lengths indicate that this part of the tree did not see frequent HGT. The second half of the tree included the majority of FicDs found in bacteria (light red labels), *Wolbachia* hits (dark red labels), archaea (yellow labels, indicated as IV in Fig. 2a), all the viral hits (blue labels), and a clade containing the micro-eukaryotic Chromista and Protozoa MAG-derived hits (violet labels) clustering together with the extra copies of FicDs described for some protostome metazoans (orange labels). This clade denoted by green branches in Fig. 2a (magnified in Fig. 2b), included one first group of FicDs of platyhelminthes (*Macrostomum lignano*) and coleopteran hits (*Holotrichia oblita, Tenebrio molitor, Zophobas morio* and *Ignelater luminosus*), together with the Oryctes rhinoceros nudivirus FicD and bacterial hits. A second group included bivalve and crustacean hits (e.g. *Armadillidium vulgare* OJH30677), the FicD of the annelid *Owenia fusiformis*, and WSSV and CQIV FicDs. As an outgroup of this clade, we found a group of Chromista FicDs clustered together with bacterial hits, phages and with the OsHV-1 FicD (Fig. 2b). Among the archaea (yellow labels) and bacteria (red labels) we also found sequences classified as rotifers, as well as the

spider *Trichonephila spp*. associated with *Wolbachia spp*. FIC sequences. The high number of discrepancies between gene and species phylogenies in this part of the tree and especially the close association of hosts and their virus and associated bacteria could indicate frequent HGTs.

3.3. The FIC enzymatic motif showed taxa-specific patterns of conservation

Next, we searched for the conserved enzymatic and inhibitory motifs in the 1,649 FicD sequences. The enzymatic FIC motif (HxFx(D/E)GN (G/K)RxxR) is present in 84 % of FicD proteins, whereas the motif responsible for the inhibitory activity of the alpha helix ((S/T)xxxE(G/N), α-inh) was detectable in 77.5 % of the sequences. Considering metazoan FicDs, the α-inh motif was detectable in at least one FicD gene copy per species, but it was absent in some duplicated FicD genes. We identified 214 different combinations of this 12-residues long FIC motif, but only 11 of them were found more than 20 times (Fig. 2c). These frequent motifs only showed variation in three positions with the 4th position being the most variable, whereas in the 11th position we found an Alanine with high frequency (Fig. 2c). Except for sporadic cases (21 hits), we always identified the conserved enzymatic motif on metazoan FicDs (Fig. 2a, outer circle with white labels). The cases of missing motifs were distributed randomly across the tree and were only more frequent in a MAG-derived FicDs clustering within arthropods and for the bacterial hits clustering with tardigrada (indicated as 1 and 2 in Fig. 2a). Considering the clades of metazoan hits clustering in the second main clade, we noticed that most of the coleoptera FicDs did not encode a conserved enzymatic motif and should likely be considered as inactive proteins. The enzymatic motif is conserved in WSSV, CQIV and OsHV-1 FicDs.

3.4. Genomic landscape of selected mollusk and crustacean FicDs

The bivalve, gastropod (N = 53) and crustacean FicDs (N = 16) reported in Table 1 were selected to investigate the genomic landscape around FicD genes. Most of the bivalve and crustacean FicDs, which clustered within the main metazoan clade, were made up from multiple exons, whereas the extra numeral FicD genes identified in bivalve species are characterized by a single exon, suggesting a bacterial origin. Conserved flanking genes were identified only for the multi-exon FicD genes in some species, e.g. bivalves, *Haliotis* spp. and the malacostracan crustaceans considered in this study (decapods and *A. vulgare*; Supplementary Fig. 1). In detail, at the 3′-ends of FicD genes, a helicase domain sometimes associated with a DEAD domain is present in all the tested mollusk species (except for *Lottia gigantea*). In contrast, at the 5′-ends of FicD genes, we observed a higher variability, with clear family-specific patterns among bivalves. In fact, in species of the *Pectinidae* and *Arcidae* there is a conserved *Transcription initiation factor TFIID subunit 8* (TAF8), whereas species of the *Heteroconchia* clade displayed a *Myelin regulatory factor* gene (Supplementary Fig. 1). Among gastropods, *Haliotis* spp. and *Gigantopelta aegis* conserved the 5′ TAF8 gene, with the latter species encoding a helix-turn-helix DNA binding domains (HTH_Tnp_Tc3_2) at both sides of FicD. Among crustaceans, the decapods and the isopod *A. vulgare* shared a 5′-end neighboring gene (BAH), whereas no other conserved flanking genes were evident (Supplementary Fig. 1). The presence of 3 nearly identical *M. coruscus* FicDs on the same scaffold and, similarly, two *M. edulis* FicDs likely indicating recent duplications, possibly mediated by transposable elements. Indeed, transposable elements are found at the flanks of the FicDs of *C. sculpturatus* (RNAse T), *M. edulis* (integrase, HTH_Tnp_Tc3_2 and reverse-transcriptase RNAseH) and *L. fortune*i (DUF1759, gag-polyprotein). One hit identified as *Wolbachia* endosymbiont of *A. vulgare* (OJH30677.1) is likely part of a *Nimaviridae-like* integrated virus, due to the presence of multiple flanking proteins showing a moderate similarity with WSSV hits.

3.5. OsHV-1 and WSSV FicDs are conserved between strains and structurally related to human FicDs

We investigated the nucleotidic conservation of OsHV-1 and WSSV FicDs between viral strains. By aligning available genomic scaffolds of 29 WSSV and, separately, of 30 OsHV-1 isolates, we could show that the FicD gene regions appeared conserved at the nucleotide level for each virus (100 % of ANI). We could identify only one deletion involving the third-last coding triplet of WSSV FicD, and only two variations impacting the RDR part of the OsHV-1 gene. We further searched for WSSV and OsHV-1 FicD loci in genomic assemblies of host/reservoir species, which included partial viral scaffolds referring to viral integrations or co-sequencing, but no match was found.

Subsequently, we used AlphaFold2 to model the structure of the 579-residues long OsHV-1 RDR2 (Supplementary Data 4). The secondary structure is composed of 33 helices and one antiparallel beta sheet. The protein comprises the RDR2 domain (Ala72- Asn339) and the FIC domain that extends from Val435 to Ile545 and it includes the catalytic motif (His527-Arg538), responsible for the AMPylation activity (Fig. 3a). The OsHV-1 protein sequence retains the conserved α-inh motif from Ser427 to Gly432, located between helix 25 and 26. After we assured the presence of the conserved enzymatic FIC motif in the protein sequence, we investigated the structural conservation of the catalytic domain by comparing it with the human FicD (Supplementary Fig. 2a). The catalytic motif of OsHV-1 FicD, located from residue His527 to Arg538, resulted highly conserved with the human one (His363 to Arg374, with a RMSD value of 0.262, Supplementary Fig. 2b). Using the same approach, we modeled the 259 residues long WSSV FicD protein (Supplementary Data 5), which was composed of 12 helices and one antiparallel beta sheet. The protein comprises the FIC domain, extended from Met104 to Leu213, and including the catalytic motif from His195 to Arg206. Similar to the OsHV-1 protein, the WSSV FicD protein contains the α-inh motif in helix 3 from Thr46 to Asn51 (Fig. 3b). Both the OsHV-1 and the WSSV protein cover the FIC catalytic motif and were highly conserved between both species (RMSD = 0.193, OsHV-1 from His527 to Arg538 and WSSV from His195 to Arg206, Fig. 3c,).

4. Discussion

FicDs have been reported as highly mobile genes, associated with transposable elements and frequently exchanged through HGTs (Khater and Mohanty, 2015; Veyron et al., 2018). Bacterial FicDs are often associated with pathogenic islands, also representing rapidly mobilizing elements, supporting the virulence of bacterial strains (Veyron et al., 2018; Ma et al., 2013).

In this study, we updated the phylogenetic distribution of FicDs beyond bacteria and deuterostomes by considering protostome and metagenomic-derived eukaryotic genomes. A remarkable conservation of a single-copy FicD gene is present in metazoans, with a relevant exception consisting in multiple FicDs found in parasitic worms (e.g., *Trichinella* spp.). Here, the FIC domains are organized in different protein architectures but, nevertheless, these FicDs showed high domain similarities to the main metazoan cluster of FicDs. This suggests that the extra numeral FicD genes in *Trichinella* probably originated from gene duplications and neofunctionalization, which is often reported during radiations of parasitic lineages to explore new host exploitation strategies (Zajac et al., 2021; Lespinet et al., 2002). The conservation of the enzymatic and inhibitory motifs of FicDs as well as the good correlation to the species phylogeny possibly suggested a preserved mode-of-action among metazoans, although experimental support is largely lacking and only available for a few species, like for *C. elegans*, *Drosophila spp*. and humans (Truttmann et al., 2016; Ham et al., 2014; Broncel et al., 2016). Notably, the cluster of single-copy metazoan FicDs appeared not impacted by HGTs, since co-clustered bacterial hits are absent. Indeed, the presence of a cluster of bacterial FicDs in a basal position of the protostome clade with poorly conserved enzymatic motifs, suggested

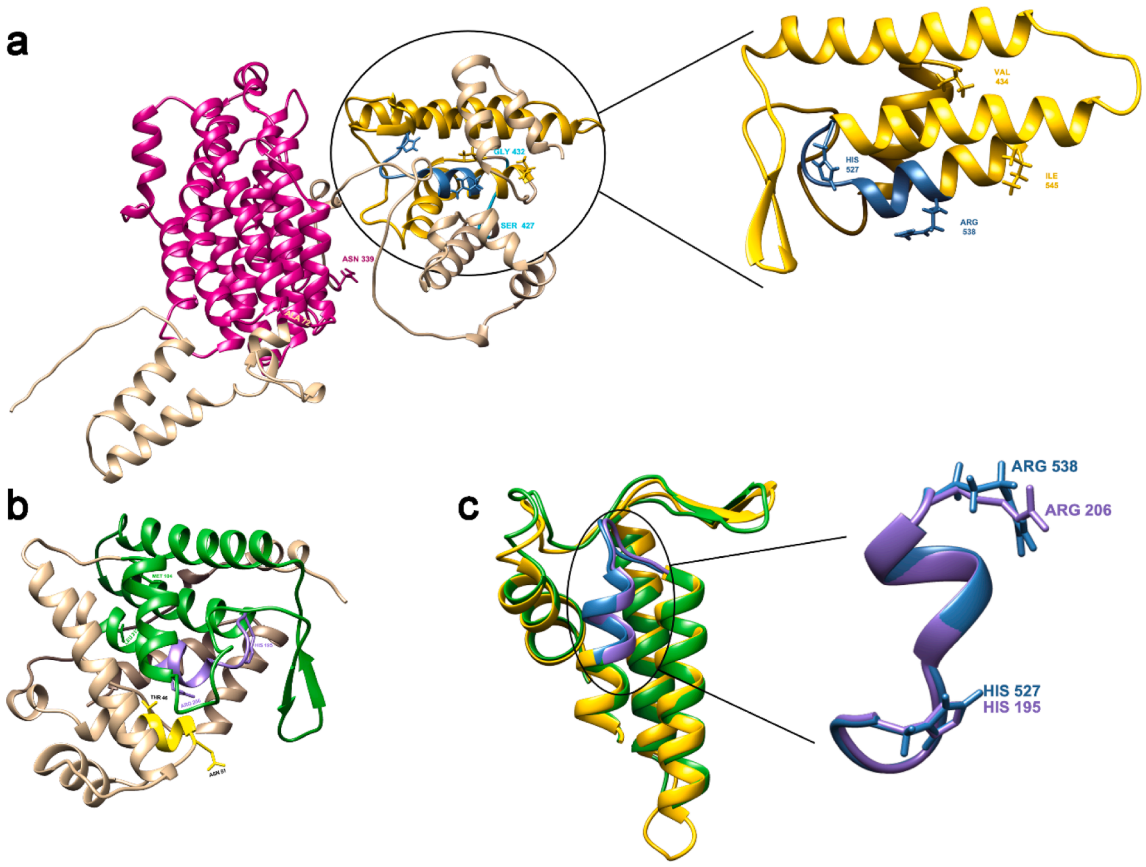

Fig. 3. AlphaFold models of OsHV-1 and WSSV FicDs. a. The OsHV-1 FIC protein structure modelled by Alphafold. the RDR2 domain (Ala 72- Asn 339) is depicted in violet-red, the FIC one (Val435- Ile 545) in gold, the catalytic site (His 527-Arg 538) in cornflower blue and the inhibitory helix (Ser427- Gly432) in cyan. b. The WSSV FIC protein structure modelled by Alphafold. The FIC domain is depicted in green (Met104- Leu213), the catalytic motif (His195 Arg206) in violet and the inhibitory one (Thr46-Asn51) in yellow. c. Structural comparison between the OsHV-1 and the WSSV FIC domain. The color code is the same as in the panel A and B. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

that, if HGT has occurred, this likely was from eukaryotes to bacteria, leading to non-conserved and likely non-functional FicDs.

In other clades we highlighted some possible evidence of HGTs, as previously reported (Khater and Mohanty, 2015). Beside a well-defined clade mostly composed by MAG-derived FicDs of Protozoa and Chromista, we found another small clade combining eukaryotic, bacterial and viral FicDs. The metazoan hits clustering in these clades were multicopy FicDs described for a few species, and sometimes encoded less-conserved or rare enzymatic motifs. A different situation characterized the cluster of bivalve, crustacean, WSSV and CQIV FicDs. Notably, one Wolbachia/*A. vulgare* FicD present in this group contains a part of a *Nimaviridae-like* integrated virus, and at the base of that clade we also found OsHV-1 FicD. All these hits maintained the enzymatic motif, likely suggesting a recent mobilization of the FicD gene possibly facilitated by viruses and other mobile genetic elements. However, the enzymatic motif found in OsHV-1 appeared identical to the one found in some arthropods, while WSSV and CQIV motifs included some mutations. Despite huge phylogenetic divergence between these viruses, their FIC domains are somewhat similar, suggesting recent HGT. Moreover, the absence of FIC domains in genomic scaffolds of integrated WSSV-like viruses found in crabs or in OsHV-1-like viruses found in *Capitella teleta* and *Branchiostoma floridae* (Davison et al., 2005; Bao et al., 2020)), as well as in the nearest OsHV-1 relative (Haliotid herpesvirus-1) and in other malacoherpes-like viruses recently identified (Rosani et al., 2023), further suggested a recent transfer. A direct virus-to-virus HGT is unlikely to occur, although we cannot exclude this possibility for WSSV and CQIV (*Nimaviridae-Iridoviridae*) since co-infection in decapod crustaceans has been reported (Bateman and Stentiford, 2017). In addition to symbiont/bacteria/host interactions, in the marine environment the predatory interaction between crustaceans and bivalves might have further contributed to connecting their microbiomes and viromes.

More interestingly, the FicD genes of invertebrate viruses showed conserved enzymatic motifs and structural features, suggesting a preserved AMPylation ability, which raises the question about its involvement in viral pathogenesis mechanisms. Subversion of host AMPylation by pathogenic bacteria has been shown as a strategy to promote infection (Yarbrough et al., 2009; Roy and Mukherjee, 2009). Recently, a FIC gene has been identified in a putative virulence plasmid of an *Arcobacter* strain isolated from New Zealand mussels (On et al., 2019), postulating a role of AMPylation during pathogenesis also in marine invertebrates. Arguably, WSSV and OsHV-1 FicDs seem to become firmly integrated into the viruses' replication cycle, suggesting an evolutionary advantage in maintaining this gene/domain. The peculiar protein construction of OsHV-1 FicD, which is associated with the RDR2 gene, is of particular interest in this context. The small and big subunits of RDR enzymatically convert ribonucleotide diphosphates to corresponding deoxyribonucleotides, allowing virus replication in non-dividing cells (Goldstein and Weller, 1988), and have been reported as "early genes" in several herpesviruses infections (Dufour et al., 2011). One could speculate that the association of RDR2 with the FIC enzyme can represent a trojan horse strategy to hijack host defenses at an early stage of infection. However, experimental evidence is required to test this hypothesis.

5. Conclusions and future perspectives

Complex interactions occurring between animals and their

microbiomes and viromes have likely shaped the extant distribution of FicDs, with the effect of HGTs further amplified by intra-species duplication possibly facilitated by mobile genetic elements and integrated viruses. In this context, the shift of AMPylation capability from bacteria to viruses can provide an example of pathogen evolution, highlighted through the lens of evolutionary genomics. Disrupting AMPylation in the early stage of viral infection could strategically help a productive infection, whereas, later in the infection, the same mechanism can promote apoptosis and viral spread. To evaluate this hypothesis, dedicated studies to determine the enzymatic activity through recombinant protein production and the identification of AMPylation targets in invertebrates are necessary to reveal evolutionary trajectories leading to post-translational protein modification in these viruses.

Funding

UR is supported by the Italian National Project PRIN2017 (Viral diversity and impacts on deep-sea biodiversity and ecosystem functioning, VIRIDE).

CRediT authorship contribution statement

Umberto Rosani: Formal analysis, Conceptualization, Data curation, Writing - original draft, Writing - review & editing. **Sofia De Felice:** Formal analysis, Writing - review & editing. **Riccardo Frizzo:** Formal analysis, Writing - review & editing. **Satoshi Kawato:** Formal analysis, Data curation. **K. Mathias Wegner:** Conceptualization, Supervision , Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All the accession codes of the analyzed data and the produced analyses are provided within the Supplementary materials. The phylogenetic tree has been uploaded to iTol for an easier visualization (https://itol.embl.de/shared/umbe1984).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2023.147895.

References

Bao, W., Tang, K.F.J., Alcivar-Warren, A., 2020. The complete genome of an endogenous nimavirus (Nimav-1_LVa) from the pacific whiteleg *Shrimp penaeus* (Litopenaeus) Vannamei. Genes 11 (1), 94. https://doi.org/10.3390/genes11010094.

Bateman, K.S., Stentiford, G.D., 2017. A taxonomic review of viruses infecting crustaceans with an emphasis on wild hosts. J. Invertebr. Pathol., Invertebr. Viruses Food Chain 147 (July), 86–110. https://doi.org/10.1016/j.jip.2017.01.010.

Broncel, M., Serwa, R.A., Bunney, T.D., Katan, M., Tate, E.W., 2016. Global profiling of huntingtin-associated protein E (HYPE)-mediated AMPylation through a chemical proteomic approach *. Mol. Cell. Proteomics 15 (2), 715–725. https://doi.org/10.1074/mcp.O115.054429.

Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12 (1), 59–60. https://doi.org/10.1038/nmeth.3176.

Campanacci, V., Mukherjee, S., Roy, C.R., Cherfils, J., 2013. Structure of the legionella effector AnkX reveals the mechanism of phosphocholine transfer by the FIC domain. EMBO J. 32 (10), 1469–1477. https://doi.org/10.1038/emboj.2013.82.

Castro-Roa, D., Garcia-Pino, A., De Gieter, S., van Nuland, N.A.J., Loris, R., Zenkin, N., 2013. The Fic protein doc uses an inverted substrate to phosphorylate and inactivate EF-Tu. Nat. Chem. Biol. 9 (12), 811–887. https://doi.org/10.1038/nchembio.1364.

Chambers, K.A., Scheck, R.A., 2020. Bacterial virulence mediated by orthogonal post-translational modification. Nat. Chem. Biol. 16 (10), 1043–1051. https://doi.org/10.1038/s41589-020-0638-2.

Chatterjee, B.K., Truttmann, M.C., 2021. Fic and non-Fic AMPylases: protein AMPylation in metazoans. Open Biol. 11 (5), 210009 https://doi.org/10.1098/rsob.210009.

Cheng, S., Wong, G.-S., Melkonian, M., 2021. Giant DNA viruses make big strides in eukaryote evolution. Cell Host Microbe 29 (2), 152–214. https://doi.org/10.1016/j.chom.2021.01.008.

Cruz, J.W., Rothenbacher, F.P., Maehigashi, T., Lane, W.S., Dunham, C.M., Woychik, N. A., 2014. Doc toxin is a kinase that inactivates elongation factor Tu. J. Biol. Chem. 289 (11), 7788–7798. https://doi.org/10.1074/jbc.M113.544429.

Davison, A.J., Trus, B.L., Cheng, N., Steven, A.C., Watson, M.S., Cunningham, C., Le Deuff, R.-M., Renault, T., 2005. A novel class of herpesvirus with bivalve hosts. J. Gen. Virol. 86 (Pt 1), 41–53. https://doi.org/10.1099/vir.0.80382-0.

Delmont, T.O., Gaia, M., Hinsinger, D.D., Frémont, P., Vanni, C., Antonio Fernandez-Guerra, A., Eren, M., et al., 2022. Functional repertoire convergence of distantly related eukaryotic plankton lineages abundant in the sunlit ocean. Cell Genom. 2 (5), 100123 https://doi.org/10.1016/j.xgen.2022.100123.

Dorrell, Richard G., Adrien Villain, Benoît Perez-Lamarque, Guillemette Audren de Kerdrel, Giselle McCallum, Andrew K. Watson, Ouardia Ait-Mohamed, et al. 2021. Phylogenomic fingerprinting of tempo and functions of horizontal gene transfer within ochrophytes. Proc. Natl. Acad. Sci. 118(4), e2009974118. 10.1073/pnas.2009974118.

Dufour, F., Marie-Josée Sasseville, A., Chabaud, S., Massie, B., Siegel, R.M., Langelier, Y., 2011. The ribonucleotide reductase R1 subunits of herpes simplex virus types 1 and 2 protect cells against TNFα- and FasL-induced apoptosis by interacting with caspase-8. Apoptosis: Int. J. Program. Cell Death 16 (3), 256–271. https://doi.org/10.1007/s10495-010-0560-2.

Eddy, S.R., 2011. Accelerated profile HMM searches. PLoS Comput. Biol. 7 (10), e1002195.

FAO, 2020. The State of World Fisheries and Aquaculture 2020: Sustainability in action. The State of World Fisheries and Aquaculture (SOFIA) 2020-SOFIA 2020. FAO, Rome, Italy. 10.4060/ca9229en (Also Available in: Chinese Spanish Arabic French Russian).

Fu, L., Niu, B., Zhu, Z., Sitao, W.u., Li, W., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics (Oxford, England) 28 (23), 3150–3312. https://doi.org/10.1093/bioinformatics/bts565.

Garcia-Pino, A., Zenkin, N., Loris, R., 2014. The many faces of fic: structural and functional aspects of fic enzymes. Trends Biochem. Sci 39 (3), 121–219. https://doi.org/10.1016/j.tibs.2014.01.001.

Goldstein, D.J., Weller, S.K., 1988. Factor(s) present in herpes simplex virus type 1-infected cells can compensate for the loss of the large subunit of the viral ribonucleotide reductase: characterization of an ICP6 deletion mutant. Virology 166 (1), 41–51. https://doi.org/10.1016/0042-6822(88)90144-4.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59 (3), 307–321. https://doi.org/10.1093/sysbio/syq010.

Gulen, Burak, Itzen, Aymelt, 2021. Revisiting AMPylation through the Lens of Fic Enzymes. Trends Microbiol, September, S0966-842X(21)00188-8. 10.1016/j.tim.2021.08.003.

Ham, H., Woolery, A.R., Tracy, C., Stenesen, D., Krämer, H., Orth, K., 2014. Unfolded protein response-regulated drosophila fic (dFic) protein reversibly AMPylates BiP chaperone during endoplasmic reticulum homeostasis. J. Biol. Chem. 289 (52), 36059–36069. https://doi.org/10.1074/jbc.M114.612515.

Harms, A., Stanger, F.V., Dehio, C., 2016. Biological diversity and molecular plasticity of FIC domain proteins. Annu. Rev. Microbiol. 70 (September), 341–360. https://doi.org/10.1146/annurev-micro-102215-095245.

Hernandez-Lima, M.A., Champion, M., Mattiola, Z., Truttmann, M.C., 2022. The AMPylase FIC-1 modulates TGF-β signaling in caenorhabditis elegans. Front. Mol. Neurosci. 15, 912734 https://doi.org/10.3389/fnmol.2022.912734.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35 (2), 518–522. https://doi.org/10.1093/molbev/msx281.

Irwin, N.A.T., Pittis, A.A., Richards, T.A., Keeling, P.J., 2022. Systematic evaluation of horizontal gene transfer between eukaryotes and viruses. Nat. Microbiol. 7 (2), 327–336. https://doi.org/10.1038/s41564-021-01026-3.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14 (6), 587–659. https://doi.org/10.1038/nmeth.4285.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30 (4), 772–780. https://doi.org/10.1093/molbev/mst010.

Kawato, S., Shitara, A., Wang, Y., Nozaki, R., Kondo, H., Hirono, I., 2019. Crustacean genome exploration reveals the evolutionary origin of white spot syndrome virus. J. Virol. 93 (3), e01144-18 https://doi.org/10.1128/JVI.01144-18.

Khater, S., Mohanty, D., 2015. In silico identification of AMPylating enzymes and study of their divergent evolution. Sci. Rep. 5 (June), 10804. https://doi.org/10.1038/srep10804.

Lemoine, Frédéric, Gascuel, Olivier, 2021. "Gotree/Goalign: Toolkit and Go API to facilitate the development of phylogenetic workflows. NAR Genom. Bioinform. 3(3), lqab075. 10.1093/nargab/lqab075.

Lespinet, O., Wolf, Y.I., Koonin, E.V., Aravind, L., 2002. The role of lineage-specific gene family expansion in the evolution of eukaryotes. Genome Res. 12 (7), 1048–1059. https://doi.org/10.1101/gr.174302.

Ma, Z., Geng, J., Yi, L.i., Bin, X.u., Jia, R., Li, Y., Meng, Q., Fan, H., Songnian, H.u., 2013. Insight into the specific virulence related genes and toxin-antitoxin virulent pathogenicity islands in swine streptococcosis pathogen *Streptococcus equi* Ssp. Zooepidemicus Strain ATCC35246. BMC Genomics 14 (1), 377. https://doi.org/10.1186/1471-2164-14-377.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., et al., 2020. Pfam: the protein families database in 2021. Nucleic Acids Res. 49 (D1), D412–D419. https://doi.org/10.1093/nar/gkaa913.

Moehlman, Andrew T, Casey, Amanda K., Servage, Kelly, Orth, Kim, Krämer, Helmut, 2018. Adaptation to Constant Light Requires Fic-Mediated AMPylation of BiP to Protect against Reversible Photoreceptor Degeneration. In: Burd, Christopher G. VijayRaghavan, K. (Eds.), eLife 7(July), e38752. 10.7554/eLife.38752.

Mushegian, A., Karin, E.L., Pupko, T., 2018. Sequence analysis of malacoherpesvirus proteins: pan-herpesvirus capsid module and replication enzymes with an ancient connection to 'Megavirales'. Virology 513, 114–128. https://doi.org/10.1016/j.virol.2017.10.009.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32 (1), 268–274. https://doi.org/10.1093/molbev/msu300.

Oliveira, P.H., Touchon, M., Cury, J., Rocha, E.P.C., 2017. The chromosomal organization of horizontal gene transfer in bacteria. Nat. Commun. 8 (1), 841. https://doi.org/10.1038/s41467-017-00808-w.

On, S.L.W., Althaus, D., Miller, W.G., Lizamore, D., Wong, S.G.L., Mathai, A.J., Chelikani, V., Carter, G.P., 2019. *Arcobacter cryaerophilus* isolated from New Zealand mussels harbor a putative virulence plasmid. Front. Microbiol. 10 (August), 1802. https://doi.org/10.3389/fmicb.2019.01802.

Perera, L.A., Rato, C., Yan, Y., Neidhardt, L., McLaughlin, S.H., Read, R.J., Preissler, S., Ron, D., 2019. An oligomeric state-dependent switch in the ER enzyme FICD regulates AMPylation and deAMPylation of BiP. EMBO J. 38 (21), e102177. https://doi.org/10.15252/embj.2019102177.

Preissler, S., Rato, C., Perera, L., Saudek, V., Ron, D., 2017. FICD acts bi-functionally to AMPylate and de-AMPylate the endoplasmic reticulum chaperone BiP. Nat. Struct. Mol. Biol. 24 (1), 23–29. https://doi.org/10.1038/nsmb.3337.

Rosani, Umberto, Gaia, Morgan, Delmont, Tom O., Krupovic, Mart, 2023. Tracing the invertebrate herpesviruses in the global sequence datasets. Front. Mar. Sci. 10. https://www.frontiersin.org/articles/10.3389/fmars.2023.1159754.

Rosani, U., Venier, P., 2017. Oyster RNA-Seq data support the development of malacoherpesviridae genomics. Front. Microbiol. 8, 1515. https://doi.org/10.3389/fmicb.2017.01515.

Roy, Craig R., Mukherjee, Shaeri, 2009. Bacterial FIC proteins AMP up infection. Sci. Signal. 2(62), pe14. 10.1126/scisignal.262pe14.

Roy, C.R., Cherfils, J., 2015. Structure and function of fic proteins. Nat. Rev. Microbiol. 13 (10), 631–640. https://doi.org/10.1038/nrmicro3520.

Salazar, G., Paoli, L., Alberti, A., Huerta-Cepas, J., Ruscheweyh, H.-J., Cuenca, M., Field, C.M., et al., 2019. Gene expression changes and community turnover differentially shape the global ocean metatranscriptome. Cell 179 (5), 1068–1083. e21. https://doi.org/10.1016/j.cell.2019.10.014.

Sanyal, A., Chen, A.J., Nakayasu, E.S., Lazar, C.S., Zbornik, E.A., Worby, C.A., Koller, A., Mattoo, S., 2015. A novel link between fic (filamentation induced by cAMP)-mediated adenylylation/ampylation and the unfolded protein response. J. Biol. Chem. 290 (13), 8482–8499. https://doi.org/10.1074/jbc.M114.618348.

Schulz, F., Roux, S., Paez-Espino, D., Jungbluth, S., Walsh, D.A., Denef, V.J., McMahon, K.D., et al., 2020. Giant virus diversity and host interactions through global metagenomics. Nature 578 (7795), 432–446. https://doi.org/10.1038/s41586-020-1957-x.

Sunagawa, S., Acinas, S.G., Bork, P., Bowler, C., Eveillard, D., Gorsky, G., Guidi, L., et al., 2020. Tara oceans: towards global ocean ecosystems biology. Nat. Rev. Microbiol. 18 (8), 428–445. https://doi.org/10.1038/s41579-020-0364-5.

Truttmann, M.C., Cruz, V.E., Guo, X., Engert, C., Schwartz, T.U., Ploegh, H.L., 2016. The *Caenorhabditis elegans* protein FIC-1 is AN AMPylase that covalently modifies heat-shock 70 family proteins, translation elongation factors and histones. PLoS Genet. 12 (5), e1006023.

Truttmann, M.C., Pincus, D., Ploegh, H.L., 2018. Chaperone AMPylation modulates aggregation and toxicity of neurodegenerative disease-associated polypeptides. PNAS 115 (22), E5008–E5017. https://doi.org/10.1073/pnas.1801989115.

Utsumi, R., Nakamoto, Y., Kawamukai, M., Himeno, M., Komano, T., 1982. Involvement of cyclic AMP and its receptor protein in filamentation of an *Escherichia coli* Fic mutant. J. Bacteriol. 151 (2), 807–812. https://doi.org/10.1128/jb.151.2.807-812.1982.

Dijk, Bram van, Hogeweg, Paulien, Doekes, Hilje M., Takeuchi, Nobuto, 2020. Slightly beneficial genes are retained by bacteria evolving DNA uptake despite selfish elements. In: Mitri, Sara, Wittkopp, Patricia J., Higgs, Paul G. eLife 9(May), e56801. 10.7554/eLife.56801.

Veyron, Simon, Peyroche, Gérald, Cherfils, Jacqueline, 2018. FIC proteins: from bacteria to humans and back again. Pathog. Dis. 76(2). 10.1093/femspd/fty012.

Vogan, A.A., Higgs, P.G., 2011. The advantages and disadvantages of horizontal gene transfer and the emergence of the first species. Biol. Direct 6 (January), 1. https://doi.org/10.1186/1745-6150-6-1.

Woolery, A.R., Luong, P., Broberg, C.A., Orth, K., 2010. AMPylation: something old is new again. Front. Microbiol. 1, 113. https://doi.org/10.3389/fmicb.2010.00113.

Yarbrough, Melanie L., Li, Yan, Kinch, Lisa N., Grishin, Nick V., Ball, Haydn L., Orth, Kim, 2009. AMPylation of Rho GTPases by vibrio VopS disrupts effector binding and downstream signaling. Science (New York, N.Y.) 323(5911), 269–272. https://doi.org/10.1126/science.1166382.

Zajac, Natalia, Zoller, Stefan, Seppälä, Katri, Moi, David, Dessimoz, Christophe, Jokela, Jukka, Hartikainen, Hanna, Glover, Natasha, 2021. Gene duplication and gain in the trematode *Atriophallophorus winterbourni* contributes to adaptation to parasitism. Genome Biol. Evol. 13(3), evab010. 10.1093/gbe/evab010.

Zhang, H., Ning, K., 2015. The Tara oceans project: new opportunities and greater challenges ahead. Genom. Proteom. Bioinform. 13 (5), 275–327. https://doi.org/10.1016/j.gpb.2015.08.003.