Computational and Structural Biotechnology Journal 19 (2021) 2905–2920

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

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

The repertoire of ABC proteins in *Clostridioides difficile*

Methinee Pipatthana [a], Phurt Harnvoravongchai [b], Pisut Pongchaikul [c], Somsak Likhitrattanapisal [d], Matthew Phanchana [e], Surang Chankhamhaengdecha [b], Tavan Janvilisri [a,*]

[a] *Department of Biochemistry, Faculty of Science, Mahidol University, Bangkok, Thailand*
[b] *Department of Biology, Faculty of Science, Mahidol University, Bangkok, Thailand*
[c] *Chakri Naruebodindra Medical Institute, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Samut Prakarn, Thailand*
[d] *Thailand Bioresource Research Center (TBRC), National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency (NSTDA), Pathumthani, Thailand*
[e] *Department of Molecular Tropical Medicine and Genetics, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand*

ARTICLE INFO

Article history:
Received 14 February 2021
Received in revised form 2 May 2021
Accepted 5 May 2021
Available online 8 May 2021

Keywords:
ATP-binding cassette transporter
Active transporter
In silico analysis
ABC classification
Antimicrobial resistance

ABSTRACT

ATP-binding cassette (ABC) transporters belong to one of the largest membrane protein superfamilies, which function in translocating substrates across biological membranes using energy from ATP hydrolysis. Currently, the classification of ABC transporters in *Clostridioides difficile* is not complete. Therefore, the sequence-function relationship of all ABC proteins encoded within the *C. difficile* genome was analyzed. Identification of protein domains associated with the ABC system in the *C. difficile* 630 reference genome revealed 226 domains: 97 nucleotide-binding domains (NBDs), 98 transmembrane domains (TMDs), 30 substrate-binding domains (SBDs), and one domain with features of an adaptor protein. Gene organization and transcriptional unit analyses indicated the presence of 78 ABC systems comprising 28 importers and 50 exporters. Based on NBD sequence similarity, ABC transporters were classified into 12 subfamilies according to their substrates. Interestingly, all ABC exporters, accounting for 64% of the total ABC systems, are involved in antibiotic resistance. Based on analysis of ABC proteins from 49 *C. difficile* strains, the majority of core NBDs are predicted to be involved in multidrug resistance systems, consistent with the ability of this organism to survive exposure to an array of antibiotics. Our findings herein provide another step toward a better understanding of the function and evolutionary relationships of ABC proteins in this pathogen.

1. Introduction

ATP-binding cassette (ABC) transporters contribute to many biological processes in all three domains of life [1]. The vast majority of ABC proteins are involved in the transport of substrates across biological membranes. However, exceptions to this are ABC proteins that participate in DNA excision repair proteins [2,3], as well as the ABC-F proteins, ART protein family, REG protein subfamily, and human ABCE proteins [4,5]. The archetypal ABC transporter consists of four functional domains comprising two nucleotide-binding domains (NBDs) and two hydrophobic transmembrane domains (TMDs). The NBDs, localized in the cytoplasmic face of the cellular membrane, are responsible for binding and hydrolyzing ATP to generate the energy for substrate translocation. Several highly conserved elements are present within different regions of the protein chain such as the Walker A, ABC signature, Walker B, and H-loop motifs [6]. The TMDs are responsible for substrate translocation across the cellular membrane. Unlike NBDs, TMDs do not exhibit significant levels of sequence similarity, providing the potential to transport various substrates. Translocation of substrates by an ABC transporter is facilitated by conformational changes to protein upon ATP binding and hydrolysis at the NBDs, promoting an alteration to the TMD structure and finally release of the substrate from its binding site [7]. In eukaryotes, ABC transporters predominantly contain a full domain structure, harboring all four functional domains, or a half structure, with

Abbreviations: ABC transporter/protein, ATP-binding cassette transporter/protein; NBD, Nucleotide-binding domain; TMD, Transmembrane domain; SBD, Substrate-binding domain; NP, Nucleotide-binding domain (NBD)-containing protein; TP, Transmembrane protein; SP, Substrate-binding domain (SBD)-containing protein; ECF-type transporter, Energy coupling factor-type transporter.

* Corresponding author at: 272 Rama VI Road, Ratchathewi, Bangkok 10400, Thailand.
E-mail address: tavan.jan@mahidol.ac.th (T. Janvilisri).

https://doi.org/10.1016/j.csbj.2021.05.012

an NBD is fused to a TMD [8]. Unlike the arrangement in eukaryotes, most bacterial ABC domains are encoded independently into individual polypeptides, which can be fused with other domains to obtain the full structure of the ABC transporter. The TMD or NBD can also form a homodimer and interact with the corresponding partner to achieve the formation of a functional transporter [9,10]. Apart from the NBD and TMD, auxiliary proteins can be recruited in certain ABC transporters. For example, the substrate-binding domain (SBD) provides the primary substrate binding site for solute uptake systems, delivering the substrate to the membrane-associated domains of the transporter. In addition, the ABC exporters in gram-negative bacteria require additional outer-membrane proteins, such as TolC and HlyD, to facilitate translocation across the periplasm and through the outer membrane [11].

ABC transporters can function either as importers or exporters. Exporters have been found in all living organisms; in contrast, importers appear to be largely restricted to prokaryotes [12]. The import systems can transport a wide variety of substrates, including sugars, amino acids, oligopeptides, osmoprotectants, organic and inorganic phosphates, iron (III), molybdenum, and vitamins [13]. To capture substrates present in the extracellular environment, the SBD-containing protein (SP) facilitates substrate-binding before translocation to the central cavity of TMD. Although an SP is not required for energy-coupling factor (ECF)-type importers [14–16]. ABC importers are important for several biological processes and they appear to be involved in bacterial pathogenesis and virulence. Some pathogenic bacteria contain special import systems for the specific uptake of micronutrients, promoting survival and allowing propagation in the host [17–19]. Bacterial efflux systems also promote the ability to survive in various environments and are major factors conferring antibiotic resistance. Efflux of complex carbohydrates, polysaccharides, proteases, and xenobiotics have been implicated in virulence and survival following antibiotic exposure [20]. The best-characterized bacterial MDR associated ABC transporters are LmrA [23], LmrCD in *Lactococcus lactis* [24], MacAB in *Escherichia coli* [25], EfrCD from *Enterococcus faecalis* [26], and AbcA in *Staphylococcus aureus*. These bacterial transporters confer resistance in a given organism through the export of multiple antibiotics. Identification of ABC systems associated with drug resistance in other organisms has provided a better understanding of the function and mechanism of these bacterial proteins. For example, human multidrug resistance (MDR) ABC transporters, such as P-glycoproteins, are capable of transporting numerous, structurally unrelated compounds and are involved in promoting drug resistance in cancer cells [21,22].

The increasing availability of genome sequence data has enabled the identification and classification of ABC proteins from many organisms. Although several publications have reported on bacterial ABC transporters [9,27–29], classifications were typically established based on the type or direction of transport, protein fold, and domain architecture, such as LanFEG [30]. Few classifications have considered the substrates of the transport system, e.g. *E. coli* [31] and *Bacillus subtilis* [32]. Importantly, while many ABC systems from *C. difficile* have been identified and functionally validated, no effort has been made to classify the complete ABC systems within *C. difficile*. Therefore, this work presents a classification of ABC systems using NBD sequence similarity and construction of ABC systems based on their substrates. We also analyzed and identified core and accessory ABC proteins in the gram-positive bacterium *C. difficile*, using the reference strain 630. Our data present an overview of the *C. difficile* ABC proteins and should aid in the design of experimental investigations toward a better understanding of the functional and evolutionary relationships among these transporters.

2. Results and discussion

2.1. Identification of ABC proteins in C. difficile 630

ABC proteins are ubiquitous in all living organisms, many with known functions. Identification of ABC proteins in *C. difficile* 630 was therefore filtered by searching for protein names annotated as "ABC" proteins. Search results identified 246 ABC proteins annotated within the *C. difficile* 630 genome. Among the candidate proteins, two exhibited a discrepancy with the ABC protein name, pabC (cd630_14470) and coaBC (cd630_25870), and were excluded from this study. Seven proteins identified as subunits of DNA excision repair UvrABC system, UvrA (cd630_02030, cd630_04560, cd630_31950, cd630_32170, and cd630_34110), UvrC (cd630_34100), and UvrB (cd630_34120), were also excluded from our analyses as they lack the ABC conserved motif and are not associated with transport activity [2]. Eleven proteins (cd630_03880, cd630_04690, cd630_08160, cd630_22690, cd630_30750, cd630_30860, cd630_30970, cd630_31160, cd630_31250, cd630_31340, and cd630_31370) annotated as phosphoryl transfer protein subunit IIABC were also omitted from further analyses. Although the phosphotransferase system (PTS) is involved in the sugar uptake in bacteria, its transport activity is activated by phosphorylation that is different from ABC transport system. After exclusions, 226 putative proteins encoded from *C. difficile* 630 genome were involved in the ABC system with either transport or non-transport activities. Analyses of all putative ABC genes revealed that the coding sequence of ABC proteins accounted for ~6% of the *C. difficile* 630 genome, and these proteins occupied ~6% of the total chromosomal proteins. ABC proteins ranged from 94 (cd630_03250) to 1142 (cd630_02030) amino acid residues. The hallmark(s) of the ABC domain: Walker boxes, transmembrane domain, or signal peptide, were identified to distinguish between NBD, TMD, and SBD prior to the construction of a complete ABC system.

2.2. Functional domains of ABC transporters in C. difficile 630

An archetypal ABC transporter commonly comprises two NBDs and two TMDs (Fig. 1). Most bacterial ABC domains are encoded as independent polypeptides from genes within an operon or from neighboring genes, although a few exceptions to this have been reported [33,34]. Moreover, both domains can form the full structure of an ABC complex through several types of organization (Fig. 1a–h, Table 1). In cases where one of the four domains is absent, the remaining NBD or TMD can function as a homodimer to complete the transport activity [35]. Fig. 1 illustrates the 8 different structures of ABC system and domain organizations found in *C. difficile* 630. Domain organization with NBD-TMD (Fig. 1a) was the major structure, accounted for ~53% of total ABC structures, followed by ~24% with NBD-TMD1-TMD2 structure (Fig. 1b). A tandem NBD that was not genetically associated with TMD partner (Fig. 1h), which is unlikely to be involved in transport activity, was found at ~8% of total structures. Interestingly, the fused TMD/NBD structures (Fig. 1f, 1e) were predicted to function solely as exporters in *C. difficile* 630, although this configuration was present in only ~2%–6% of total ABC structures. Only one ABC system consisting of four individual domains (Fig. 1d) was observed in this bacterium. The import systems found in *C. difficile* 630, were categorized into three domain organizations, as illustrated in Fig. 1i–iii. For the bacterial import systems, an SBD is generally coupled with a TMD to complete the import activity. Separate SBD (Fig. 1i) and fused SBD/TMD (Fig. 1ii) configurations were the major types of importer identified. However, an SP is not required for ECF-type ABC importers as they contain an S compo-

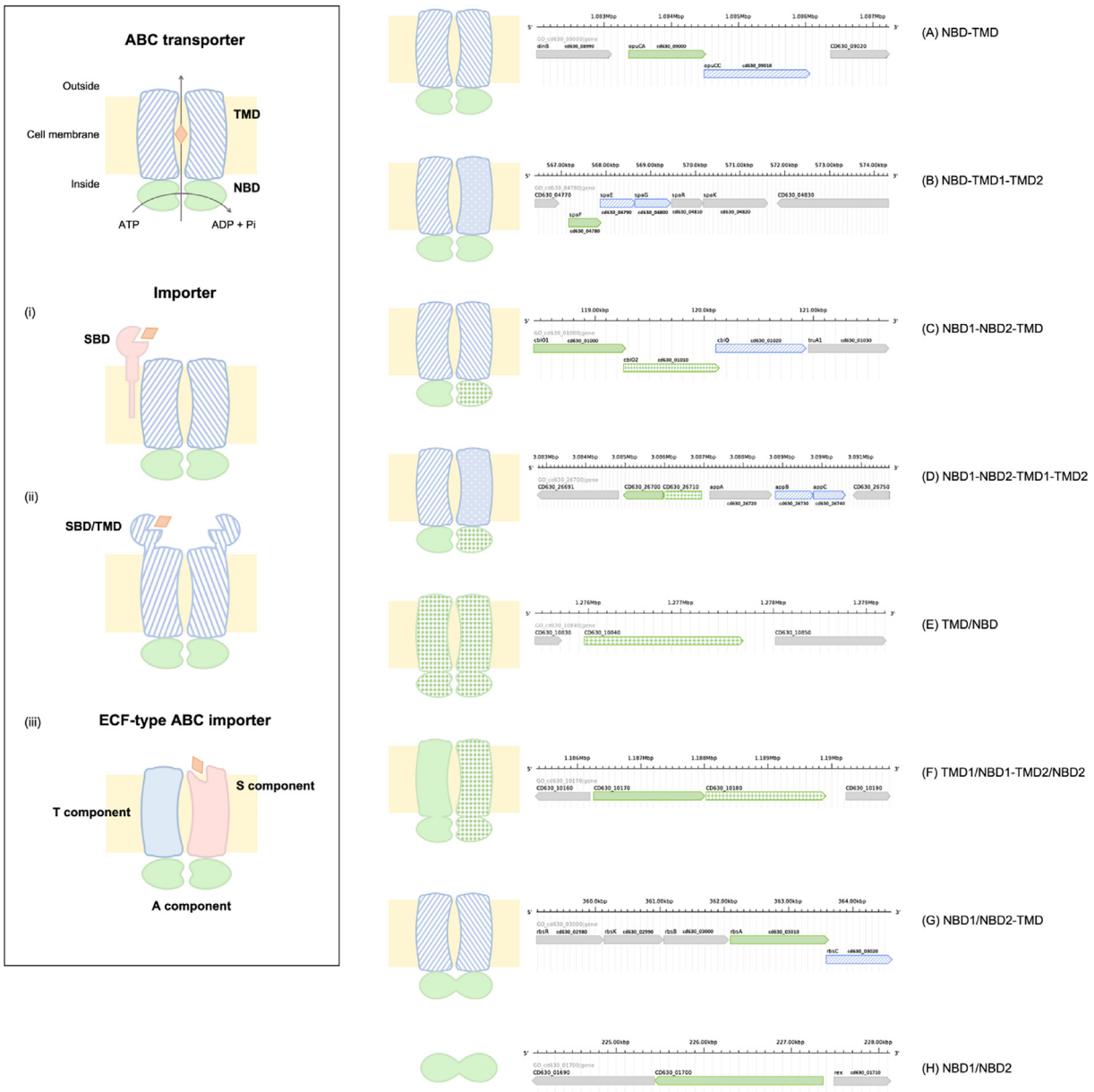


Fig. 1. Representatives of ATP-binding cassette structure and genetic organization in *C. difficile*. Genetic organization of representative ABC genes is predicted using Annotation sketch (http://genometools.org/annotationsketch.html). The genome of *C. difficile* reference strain 630 was obtained from NCBI database (NC_009089.1). The chromosomal location is indicated on the upper line. The gene name or locus name is presented above the color filled box. Other genes in the chromosomal area are depicted in grey filled box. The domain architecture is illustrated, which the nucleotide-binding domain (NBD) is depicted as an oval shape, the transmembrane domain (TMD) is represented by a barrel shape and substrate-binding domain (SBD) is illustrated by ball-and-stick shape. Different colored patterns indicate the homo- or heterologous assembled complex. Fused domain either NBDs or NBD and TMD is represented by slash (/) symbol. The tentative domain assembly is indicated using dash (-) symbol. Eight domain architectures of ABC systems are depicted including (a) a homodimeric complex both in NBD and TMD; (b) homodimeric NBDs with two separated TMDs; (c) homodimeric TMDs with two separated NBDs; (d) all four domains in individual polypeptides; (e) a homodimeric protein containing fused NBD and TMD within each polypeptide; (f) a heterodimeric protein containing fused NBD and TMD within each polypeptide; (g) tandem NBD with homodimeric TMDs; (h) tandem NBD without identifiable TMD. Three domain structures of ABC import systems are illustrated comprising; (i) separated SBD domain anchored on the cell membrane; (ii) fused SBD/TMD protein; (iii) energy-coupling factor (ECF)-type ABC importer consisting of A, T, and S components.

nent for capturing solute substrate [14–16]. Few import systems were ECF-type transporters, an example is the cobalt/vitamin uptake system. Each predicted domain was further investigated.

2.2.1. Nucleotide-binding domain

The NBD contains several conserved motifs responsible for ATP binding and hydrolysis. All 97 NBD-containing proteins (NPs) within *C. difficile* 630 possessed protein domain PF00005, a protein sequence of approximately 200 amino acids. Conserved sequences for Walker A (GxxGxGKS/T) region, followed by ABC signature (LSGGQ) and Walker B (hhhhLDE) (Fig. 2a) were observed in NBD regions. Nine of these proteins harbored tandem NBD within their structures, making a total of 106 NBDs analyzed. Domain organization analysis of the 97 NPs revealed 3 structures: (i) nine proteins exhibited tandem NBD in a single polypeptide (NBD/NBD); (ii) 12 proteins possessed an NBD fused with TMD (TMD/NBD); and (iii) 76 proteins contained a single NBD within their structure, as shown in Table 1. A diversity of NP structures in *C. difficile* 630 was not distinct from other organisms, such as *E. coli* and *B. subtilis*, although the proportions of each NP type are different. The propor-

Table 1
A complete catalog of ABC proteins encoded on the *C. difficile* 630 reference genome.

No.	Locus name	Gene name	Domain	Sub-family	Length	Predicted domain partners				Gene organization	Annotated substrate(s)	Function
						SBD1(2)	NBD2	TMD1	TMD2			
1	cd630_01000	*cbiO1*	NBD	12	277		cd630_01010	cd630_01020		NBD-NBD-TMD	Vitamin	Importer
2	cd630_01010	*cbiO2*	NBD	12	288		cd630_01000	cd630_01020		NBD-NBD-TMD		
3	cd630_01610		TMD/NBD	Unknown	590					TMD/NBD	Multidrug	Exporter
4	cd630_01700		NBD/NBD	3	637					NBD/NBD	Unknown	N/A
5	cd630_02930		NBD	7	305			cd630_02940		NBD-TMD	Bacitracin	Exporter
6	cd630_03010	*rbsA*	NBD/NBD	1	503	cd630_03000		cd630_03020		SBD-NBD/NBD-TMD	Ribose	Importer
7	cd630_03180		NBD	9	307			cd630_03170	cd630_03160	TMD-TMD-NBD	Bacitracin	Exporter
8	cd630_03270	*cbiO*	NBD	12	273	cd630_03250		cd630_03240	cd630_03260	TMD-SBD-TMD-NBD	Cobalt	Importer
9	cd630_03360		NBD	Unknown	222			cd630_03370		NBD-TMD	Unknown	Exporter
10	cd630_03640		NBD	Unknown	256			cd630_03630		TMD-NBD	Unknown	Exporter
11	cd630_03660		NBD	9	308			cd630_03650		TMD-NBD	Multidrug	Exporter
12	cd630_04310		NBD/NBD	12	490			cd630_04300		TMD-NBD	Cobalt	Importer
13	cd630_04320		TMD/NBD	6	571		cd630_04330			TMD/NBD-TMD/NBD	Multidrug	Exporter
14	cd630_04330		TMD/NBD	6	582		cd630_04320					
15	cd630_04590		NBD	Unknown	227			cd630_04600		NBD-TMD	Unknown	Exporter
16	cd630_04780	*spaF*	NBD	9	234			cd630_04790	cd630_04800	NBD-TMD-TMD	Lantibiotic	Exporter
17	cd630_04840		NBD	Unknown	222			cd630_04830		TMD-NBD	Unknown	Exporter
18	cd630_06130		NBD	Unknown	250			cd630_06120		TMD-NBD	Unknown	Exporter
19	cd630_06450		NBD	9	308			cd630_06460		NBD-TMD	Lantibiotic	Exporter
20	cd630_06530		NBD	7	285			cd630_06540		NBD-TMD	Antibiotic	Exporter
21	cd630_06670	*cdd4*	NBD	9	304			cd630_06660	cd630_06650	TMD-TMD-NBD	Lantibiotic	Exporter
22	cd630_07520		NBD	4	240	cd630_07500		cd630_07510		SBD-TMD-NBD	Polar amino acid	Importer
23	cd630_07850		NBD/NBD	3	516					NBD/NBD	Unknown	N/A
24	cd630_08220		NBD	9	307			cd630_08230	cd630_08240	NBD-TMD-TMD	Bacitracin	Exporter
25	cd630_08560	*oppD*	NBD	2	341	cd630_08550		cd630_08530	cd630_08540	TMD-TMD-SBD-NBD	Oligopeptide	Importer
26	cd630_08710	*modC*	NBD	11	358	cd630_08690		cd630_08700		SBD-TMD-NBD	Molybdate	Importer
27	cd630_08740		NBD	Unknown	251	cd630_08730		cd630_08750		SBD-NBD-TMD	Sugar	Importer
28	cd630_08770		NBD	Unknown	251	cd630_08760		cd630_08780		SBD-NBD-TMD	Sugar	Importer
29	cd630_09000	*opuCA*	NBD	5b	378			cd630_09010		NBD-TMD	Glycine betaine, carnitine, choline	Importer
30	cd630_10010		NBD	Unknown	232	cd630_09990		cd630_10000		SBD-TMD-NBD	Nitrate/sulfonate/taurine	Importer
31	cd630_10170		TMD/NBD	6	575		cd630_10180			TMD/NBD-TMD/NBD	Multidrug	Exporter
32	cd630_10180		TMD/NBD	6	620		cd630_10170					
33	cd630_10240	*potA*	NBD	5c	347	cd630_10270		cd630_10250	cd630_10260	NBD-TMD-TMD-SBD	Spermidine/putrescine	Importer
34	cd630_10500		NBD	7	286			cd630_10510		NBD-TMD	Antibiotic	Exporter
35	cd630_10840		TMD/NBD	6	566					TMD/NBD	Multidrug	Exporter
36	cd630_10970		NBD	9	308			cd630_10960	cd630_10950	TMD-TMD-NBD	Bacitracin	Exporter
37	cd630_12680		NBD	7	311			cd630_12670	cd630_12660	TMD-TMD-NBD	Antibiotic	Exporter
38	cd630_13490		NBD	9	235			cd630_13500	cd630_13510	NBD-TMD-TMD	Lantibiotic	Exporter
39	cd630_14050		NBD/NBD	3	639					NBD/NBD	Unknown	N/A
40	cd630_14660		NBD	Unknown	225			cd630_14670		NBD-TMD	Unknown	Exporter
41	cd630_14720		TMD/NBD	6	590		cd630_14730			TMD/NBD-TMD/NBD	Multidrug	Exporter
42	cd630_14730		TMD/NBD	6	607		cd630_14720					
43	cd630_14830	*ssuB*	NBD	11	243	cd630_14840		cd630_14820		TMD-NBD-SBD	Sulfonate	Importer
44	cd630_14890	*metN*	NBD	4	321	cd630_14910		cd630_14900		NBD-TMD-SBD	D-methionine	Importer
45	cd630_15040		NBD	7	291			cd630_15050		NBD-TMD	Antibiotic	Exporter
46	cd630_15280		NBD	Unknown	224			cd630_15270	cd630_15290	TMD-NBD-TMD	Unknown	Exporter
47	cd630_15320		NBD	Unknown	255			cd630_15330		NBD-TMD	Unknown	Exporter
48	cd630_15390		NBD	Unknown	248			cd630_15400		NBD-TMD	Unknown	Exporter
49	cd630_15870		NBD/NBD	1	498	cd630_15890		cd630_15880		NBD/NBD-TMD-SBD	Ribose/sugar	Importer
50	cd630_16040		NBD	7	320			cd630_16030		TMD-NBD	Antibiotic	Exporter
51	cd630_16070		NBD	7	238			cd630_16080		NBD-TMD	Antibiotic	Exporter
52	cd630_16140		NBD	7	264			cd630_16150		NBD-TMD	Sodium	Exporter
53	cd630_16180		NBD	7	241			cd630_16190		NBD-TMD	Antibiotic	Exporter

Table 1 (*continued*)

No.	Locus name	Gene name	Domain	Sub-family	Length	Predicted domain partners				Gene organization	Annotated substrate(s)	Function
						SBD1(2)	NBD2	TMD1	TMD2			
54	cd630_16490		NBD	8	251	cd630_16500		cd630_16470	cd630_16480	TMD-TMD-NBD-SBD	Iron (III)	Importer
55	cd630_16690		NBD	7	295			cd630_16700		NBD-TMD	Antibiotic	Exporter
56	cd630_17530		NBD	7	281			cd630_17540	cd630_17550	NBD-TMD-TMD	Fluoroquinolone	Exporter
57	cd630_17760		NBD	4	223	cd630_17740		cd630_17750		SBD-TMD-NBD	Polar amino acid	Importer
58	cd630_18740	*vexP2*	NBD	Unknown	218			cd630_18730	cd630_18750	TMD-NBD-TMD	Unknown	Exporter
59	cd630_18890		NBD	7	288			cd630_18900		NBD-TMD	Antibiotic	Exporter
60	cd630_19470		NBD	Unknown	228			cd630_19480		NBD-TMD	Unknown	Exporter
61	cd630_19550		NBD	Unknown	228			cd630_19540		TMD-NBD	Unknown	Exporter
62	cd630_19610		NBD	Unknown	221			cd630_19620		NBD-TMD	Unknown	Exporter
63	cd630_19780		NBD	Unknown	245	cd630_19790		cd630_19800		NBD-SBD-TMD	Unknown	Importer
64	cd630_20240		NBD	7	287					NBD	Antibiotic	N/A
65	cd630_20680		NBD/NBD	3	541					NBD/NBD	Unknown	N/A
66	cd630_21050		NBD	7	235			cd630_21030	cd630_21040	TMD-TMD-NBD	Antibiotic	Exporter
67	cd630_21720		NBD	4	249	cd630_21740 cd630_21770		cd630_21750	cd630_21760	NBD-SBD-TMD-TMD-SBD	Cystine/amino acid	Importer
68	cd630_22100		TMD/NBD	6	610		cd630_22110			TMD/NBD-TMD/NBD	Multidrug	Exporter
69	cd630_22110		TMD/NBD	6	748		cd630_22100			TMD/NBD-TMD/NBD	Multidrug	
70	cd630_23130		NBD	11	231	cd630_23110		cd630_23120		SBD-TMD-NBD	Molybdenum	Importer
71	cd630_23160		NBD	11	250					NBD	Nitrate/sulfonate/taurine	N/A
72	cd630_24560		NBD	5a	369					NBD	Multiple sugar/sugar	N/A
73	cd630_25340		NBD	Unknown	227			cd630_25330		TMD-NBD	Unknown	Exporter
74	cd630_25930		NBD/NBD	3	550					NBD/NBD	Pristinamycin/macrolide/multidrug	N/A
75	cd630_26700		NBD	2	323	cd630_26720	cd630_26710	cd630_26730	cd630_26740	NBD-NBD-SBD-TMD-TMD	Peptide/nickel	Importer
76	cd630_26710		NBD	2	325	cd630_26720	cd630_26700	cd630_26730	cd630_26740			
77	cd630_28170		TMD/NBD	6	596		cd630_28180			TMD/NBD-TMD/NBD	Multidrug	Exporter
78	cd630_28180		TMD/NBD	6	575		cd630_28170					
79	cd630_28750	*fhuC*	NBD	8	260	cd630_28780		cd630_28770	cd630_28760	NBD-TMD-TMD-SBD	Iron/ferrichrome	Importer
80	cd630_29850		NBD	Unknown	248			cd630_29840		TMD-NBD	Unknown	Exporter
81	cd630_29900	*ssuB2*	NBD	11	292	cd630_29890		cd630_29910		SBD-NBD-TMD	Sulfonate	Importer
82	cd630_29970		NBD	Unknown	398	cd630_29990		cd630_29980		NBD-TMD-SBD	Iron (III)	Importer
83	cd630_30560		NBD	Unknown	250			cd630_30550		TMD-NBD	Unknown	Exporter
84	cd630_31620		NBD	7	290			cd630_31610	cd630_31600	TMD-TMD-NBD	Antibiotic	Exporter
85	cd630_32010		NBD	7	281			cd630_31990	cd630_32000	TMD-TMD-NBD	Antibiotic	Exporter
86	cd630_32150		NBD	5b	253			cd630_32160		NBD-TMD	Glycine betaine, carnitine, choline	Importer
87	cd630_32610	*pstB*	NBD	Unknown	254	cd630_32680		cd630_32620	cd630_32630	NBD-TMD-TMD-SBD	Phosphate	Importer
88	cd630_33590		NBD/NBD	3	512					NBD/NBD	Unknown	N/A
89	cd630_33640		NBD	7	331			cd630_33620	cd630_33630	TMD-TMD-NBD	Antibiotic	Exporter
90	cd630_34170		NBD	5a	356	cd630_34140		cd630_34160	cd630_34150	SBD-TMD-TMD-NBD	*sn*-glycerol 3-phosphate/sugar	Importer
91	cd630_35270		NBD	13	331	cd630_35250		cd630_35260		SBD-TMD-NBD	Iron (III)	Importer
92	cd630_35300		NBD	13	331	cd630_35280		cd630_35290		SBD-TMD-NBD	Iron (III)	Importer
93	cd630_35330	*phnL*	NBD	Unknown	237		cd630_35340			NBD-NBD	Alpha-D-ribose 1-methylphosphonate/ 5-triphosphate synthase/phosphonate	N/A
94	cd630_35340	*phnK*	NBD	Unknown	288		cd630_35330			NBD-NBD		N/A
95	cd630_35850		NBD	Unknown	228			cd630_35840		TMD-NBD	Unknown	Exporter
96	cd630_36080		NBD	Unknown	242					NBD	Fe-S cluster assembly	N/A
97	cd630_36230		NBD	Unknown	205			cd630_36240		NBD-TMD	Unknown	Exporter

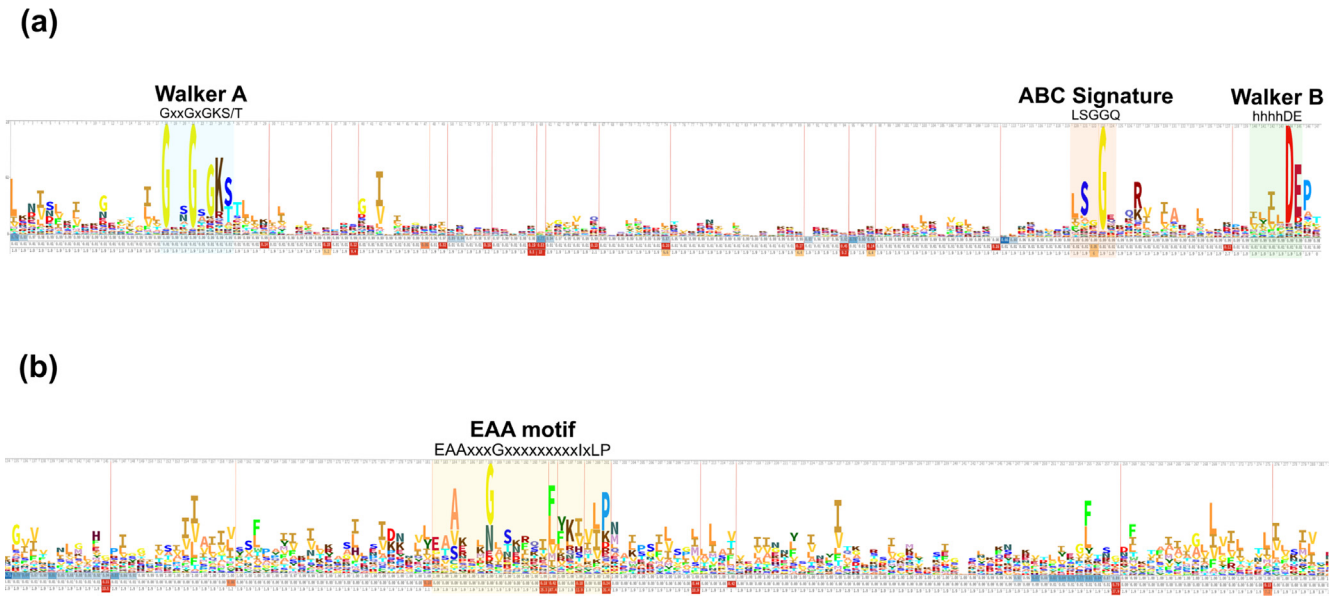


Fig. 2. Conserved amino acid sequences within the protein domain. The HMM logo representing both sequence alignments and profile hidden Markov model was generated from Skylign (http://skylign.org/). Protein sequences within conserved motifs were extracted and aligned prior subjected to HMM logo construction. (a) Protein sequences from 97 NPs within domain PF00005 with approximately 200 amino acids. (b) Partial amino acid sequences from 42 importer TPs with approximately of 150 amino acids from the C-terminus. A high degree of similarity is depicted as a bold letter of that amino acid residue. Well characterized conserved motifs including Walker A (Gx_2GxGKS/T), ABC signature (LSGGQ), Walker B (hhhhDE), and EAA ($EAAx_3Gx_9IxLP$) are indicated in the colored-box above aligned sequences. G; glycine, K; lysine, S; serine, T; threonine, L; leucine, Q; glutamine, D; aspartic acid, E; glutamic acid, E; glutamic acid, A; alanine, I; isoleucine, P; proline, h; hydrophobic amino acid, and x; any amino acid.

tion of proteins associated with drug resistance was over-represented in *C. difficile* 630 and *B. subtilis*, compared to *E. coli*. Our findings imply a role for these NPs to limit antibiotic susceptibility, although experimental evidence regarding drug resistance provided by these transporters is needed. A homology search of *C. difficile* 630 NPs was also performed using the NCBI database, as shown in Supplement 1.

2.2.2. Transmembrane domain

The hydrophobic transmembrane proteins (TPs) have been recognized as the substrate-transporting region of ABC transporter systems. Unlike NPs, these protein regions share few significant conserved structural domains, this diversity in structure appears to enable the transport of a wide variety of structurally diverse compounds. Our results showed that 98 ABC proteins encoded from the *C. difficile* 630 genome possessed a transmembrane region, as predicted by TMHMM and Phobius embedded in InterProscan 5.29 [36]. In addition to 98 individual TPs, 12 proteins were TMD/NBD fusion proteins. The TPs possessed three to 13 hydrophobic helices. Most TPs (43%) possessed six helices, followed by eight helices (14%) and five helices (11%). Although the numbers of transmembrane helices cannot be exclusively used to categorize the ABC system sub-family, their clustering can be useful to support the NBD classification. Apart from the hydrophobic regions, another conserved motif present in the TMD of some bacterial ABC importers is a cytoplasmic hydrophilic conserved loop known as EAA motif ($EAAx_3Gx_9IxLP$) (PROSITE: PDOC00364). This loop was found to physically interact with the NBD [9,37,38] (Fig. 2b). The EAA motif was identified within the sequences of 29 TMDs from *C. difficile* 630, located approximately 100 amino acids from the C-terminus. These TMDs were therefore predicted to be involved in solute uptake systems.

2.2.3. Substrate-binding domain

Although exporters can be found in all kingdoms, importers appear to be essentially restricted to bacteria and plants [9]. The directionality of transport was primarily predicted using the presence or absence of SBD(s). SP is commonly required as part of the solute uptake system and functions as the primary substrate-binding site prior to transferring the substrate to the transmembrane region, although the SP domain is absent in ECF-type ABC importers (Fig. 1i-1iii). The presence of an SBD motif was evaluated in all predicted ABC protein sequences using SignalP, SignalP II, and Phobius programs embedded in InterProscan 5.29 [36]. The motif search showed that 30 *C. difficile* ABC proteins harbored the signal peptide motif (PROSITE: PDOC00013), suggesting SBD in these proteins. Results from the SignalP prediction indicated the signal peptide was located within the first 23 amino acids from the N-terminus of the protein chain. In gram-positive bacteria, the SBD can be expressed as either an individual protein capable of anchoring on the cell membrane or a fused SBD/TMD protein [39]. In *C. difficile* 630, a separate SBD structure was observed as a major domain architecture (Fig. 1i, Table 1). An SBD/TMD fusion structure was observed in the glycine betaine uptake system (Fig. 1ii, Table 1). The TPs of this sub-family contained conserved domains (PF4069 and IPR007210), which represent the fusion of SBD with TP, a unique characteristic of glycine betaine transport systems. This fusion of SBD to the C-terminus of the TMD (SBD/TMD) is also observed in the glycine betaine transporter, OpuABC system in *L. lactis* [40]. The ECF-type system is another type of ABC importer, although structurally different from canonical ABC transporters (Fig. 1iii). In the ECF-type ABC importer, the S component replaces the function of the SP region. From our prediction, an ECF-type ABC importer was only found in cobalt/vitamin import system.

One *C. difficile* NP, cd630_15290, was annotated as a periplasmic adaptor subunit (HlyD family secretion protein). This NP contained only a coiled-coil structure, which is not consistent with an SBD, NBD, or TMD. Across species, this protein shared up to 50% sequence similarity with periplasmic adaptor proteins. However, only one protein from *B. subtilis*, YknX, with high similarity was documented. YknX belongs to the membrane fusion protein (MFP) family, which interacts with the ABC transporter YknYZ complex. Although the MFP can be found in gram-positive bacteria, its transport function remains unclear. However, the YknX was

found to be involved in the YknWXYZ antimicrobial stress response [41]. We, therefore, suggested NP cd630_15290 be a part of a transport system with an unknown function.

2.3. Construction of complete ABC systems in C. difficile 630

The complete catalog of ABC system established in *C. difficile* 630 is shown in Table 1. Among the 226 ABC proteins identified in *C. difficile* 630, there were 97 NPs, 98 TPs, 30 SPs, and one adaptor protein. A functional ABC system requires at least one NBD and one TMD to complete the transport activity. To construct a complete ABC transport system in *C. difficile*, we first identified the NBDs due to their high degree of conservation. Then, the ABC partners including TMD(s) and SBD(s) encoded from the genes in the vicinity of the corresponding NBD were searched and included as part of the complete ABC transport system. We initially hypothesized the presence of 97 ABC systems operating in *C. difficile* 630 based on the number of NPs. However, after the complete construction of ABC systems, only 85 NPs were found to have the dedicated TP partner. From these 85 NPs, 7 systems were predicted to possess 2 distinct NPs, allowing the prediction of 78 complete ABC systems in *C. difficile* 630 (Table 1).

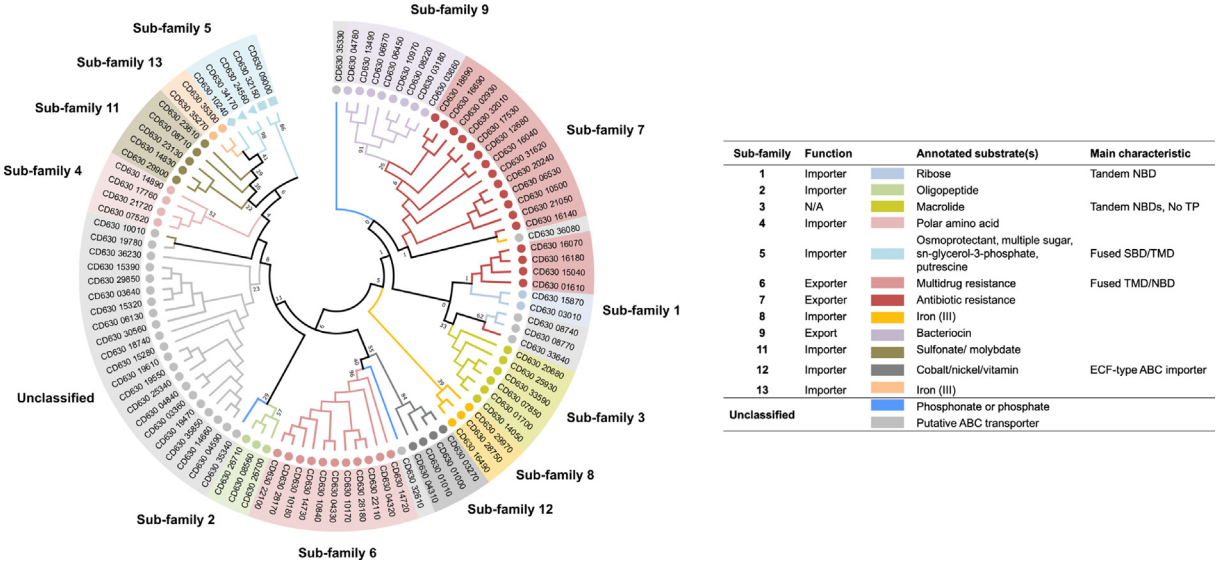
ABC domain partners were not identified for the other 12 NPs. The absence of a TP partner for these NPs was reinvestigated searching for other non-annotated ABC proteins in the vicinity and analyses of protein-protein interaction networks. Based on protein associations, it was revealed that three NPs (cd630_23610, cd630_24560, and cd630_25930) showed interactions with several ABC domains. The association of an NP with several TPs has been previously reported and occurs with NPs involved in the sugar uptake ABC systems, e.g. MsmK in *Streptococcus pneumoniae* [42] and MsmX in *B. subtilis* [43]. This finding supports the possibility of an annotated sugar NP (cd630_24560) to function with other sugar ABC importers in *C. difficile* 630. Based on gene localization analysis, it was revealed that three NPs (cd630_1700 and cd630_07850 and cd630_20240) were separately located within the same transcriptional unit with an un-annotated ABC putative permease, suggesting that they may form a complete transport system. Membrane partners were not identified for six NPs (cd630_14050, cd630_20680, cd630_33590, cd630_25330, cd630_25340, and cd630_36080). From these predictions, more than 78 ABC systems may be operating in *C. difficile* 630.

Generally, ABC transporters are unidirectional, either import or export substrates. The direction of transport is usually predicted based on the presence of SBD in the system, except for the ECF-type ABC transporter. Our prediction revealed that of 78 *C. difficile* 630 ABC systems, 28 were predicted to be import systems, while the remaining 50 systems were presumed as exporters (Table 1).

2.4. Classification and characteristics of ABC transporters in C. difficile 630

This is the first report on the classification of the entire complement of ABC proteins in *C. difficile* 630. The ABC classification was constructed using sequence similarity of the NBD, which contains several conserved motifs, as proposed by CF Higgins [44]. The numbering of sub-families was conducted in accordance with those proposed in *B. subtilis* and *E. coli* [31,32]. Based on the sequence similarity analysis, 97 NPs were classified into 12 sub-families following putative substrates and protein topology, as shown in Fig. 3. Unlike in *E. coli*, a sub-family 10 was absent in our classification. Interestingly, a new ABC sub-family, sub-family 13, was present in *C. difficile* 630 compared with the classification in *E. coli* and *B. subtilis*. Therefore, 12 sub-families, denoted as sub-families 1–13 with the absence of sub-family 10, were established in *C. difficile* 630. Note that most ABC proteins were categorized in sub-family 7 (17%), followed by sub-family 6 (11%) and sub-family 9 (7%) (Fig. 3). Interestingly, these three sub-families, accounting for 35% of the ABC systems, were exporters involved in antimicrobial extrusion.

Even though each domain of bacterial ABC systems is usually encoded as a single, separate polypeptide and then assembles as a complex. However, some systems exhibit domain fusions of TMD and NBD (TMD/NBD). In *C. difficile* 630, 12 TMD/NBD fusion



Sub-family	Function	Annotated substrate(s)	Main characteristic
1	Importer	Ribose	Tandem NBD
2	Importer	Oligopeptide	
3	N/A	Macrolide	Tandem NBDs, No TP
4	Importer	Polar amino acid	
5	Importer	Osmoprotectant, multiple sugar, sn-glycerol-3-phosphate, putrescine	Fused SBD/TMD
6	Exporter	Multidrug resistance	Fused TMD/NBD
7	Exporter	Antibiotic resistance	
8	Importer	Iron (III)	
9	Export	Bacteriocin	
11	Importer	Sulfonate/ molybdate	
12	Importer	Cobalt/nickel/vitamin	ECF-type ABC importer
13	Importer	Iron (III)	
Unclassified		Phosphonate or phosphate	
		Putative ABC transporter	

Fig. 3. Unrooted tree of nucleotide-binding domain containing proteins (NP) encoded in *C. difficile*. An unrooted tree was generated using the protein sequence within protein domain PF00005 from 97 NPs from *C. difficile* 630 reference strain. Multiple sequence alignment was performed by CLUSTALW embedded in MEGA version X [105]. Phylogenetic and molecular evolutionary analyses were conducted using MEGA X with the neighbor-joining (NJ) method with 1000 bootstrap test. Each sub-family is indicated by a colored-circle symbol and colored-text box. The NPs were classified into 12 sub-families according to protein similarity. The numbering of sub-family followed classification as proposed by Linton, et al. [31] and Quentin, et al. [32]. Sub-family 10 is absent in *C. difficile* while sub-family 13 was not reported in corresponding classification references. For sub-family 5, colored-triangle symbol represents group 5a; colored-rectangle symbol represents group 5b; colored-diamond symbol represents group 5c. Bootstrap score is indicated above a given clade. The putative substrate of each protein is represented by colored branch. A unique characteristic of a given sub-family is noted. N/A referred as not applicable.

proteins formed a distinct clade of sub-family 6, which was predicted to function in multidrug extrusion. Another combination of fused proteins was also found in nine proteins containing tandem NBD (NBD/NBD). About half of these proteins were classified in the macrolide resistance system (sub-family 3), while the others were distributed to other systems such as the sugar system (sub-family 1) and the cobalt system (sub-family 12). The ABC importers were classified into eight sub-families (sub-family 1, 2, 4, 5, 8, 11, 12, and 13) responsible for ribose, oligopeptides, polar amino acids, osmoprotectants, iron (III), sulfonate/molybdate, cobalt uptake, and iron uptake, respectively. Among 97 NPs, 27 NPs, accounting for ~28% of total NPs, were not assigned to any ABC sub-families.

As SP and TP contain few conserved regions, their protein sequence cannot be clustered using the multiple sequence alignment approach. Therefore, protein sequences of SP and TP were partitioned and grouped separately using the correlation matrix. The predicted SPs in *C. difficile* 630 were separated into five groups, named as group i to v, while TPs were classified into six groups designated as a to f (Supplement 2). The inter-relationship between all domains consisting of NBD, SBD, and TMD is represented in Table 2. It was observed that the NPs of sub-family 13, which were involved in iron transport, showed associations with SP group i and TP group d. Moreover, NPs of sub-family 2, which were involved in the peptide/nickel import system, specifically associated with the SP group iii. Other NP sub-families formed connections with the largest group of each SP (group v) and TP (group f). The remaining SP and TP groups were randomly associated with unclassified NPs.

In this section, the main features of each sub-family are explained. To provide details for each sub-family, *C. difficile* ABC systems were compared with those in *B. subtilis* [32] and *E. coli* [31]. As experimental evidence on most ABC proteins in *C. difficile* is limited, the known ABC proteins from *B. subtilis* and *E. coli* were used as a reference for a functional assignment for *C. difficile* ABC proteins.

Sub-family 1: ribose importer

The cd630_03010 (RbsA) and cd630_15870 were only two NPs classified into sub-family 1. These NPs contain tandem NBD within their protein sequences, which are also present in ABC sub-family 1 (ribose uptake) of *E. coli* and *B. subtilis* [31,32]. Amino acid sequence analysis of cd630_03010 and cd630_15780 revealed 42.3% identity with each other; and at least 40% to *E. coli* (RbsA) and *B. subtilis* (RbsA and YufO). Their SPs (cd630_03000 and cd630_15890) and TPs (cd630_03020 and cd630_15880) also exhibited sequence similarity (>30%) to each other. It is previously reported that the *rbs* operon of *C. difficile* composes of five functional genes, namely *rbsRKBAC* system. The *rbsB*, *rbsA,* and *rbsC* (*cd630_03000*, *cd630_03010,* and *cd630_03020*) involved in substrate transport, while *rbsR* and *rbsK* (*cd630_02980* and *cd630_02990*) play a role in gene regulation. The functional role of this RbsRKBAC system was experimentally validated as an importer of catabolizable carbohydrates in *C. difficile* [45].

Sub-family 2: peptide/nickel importer

Three NPs (cd630_08560, cd630_26700, and cd630_26710) were grouped into sub-family 2, which is functionally verified as a peptide/nickel importer [46,47]. It was noted that NPs of sub-family 2 specifically associated with SP in group iii (Table 2), indicating a unique substrate specificity of this importer. Although the domain partners of the *oppD* gene (*cd630_08560*) were not found under the same promoter, the SBD gene (*oppA*; *cd630_08550*) and the TMD operon (*oppB*; *cd630_ 08530* and *oppC*; *cd630_08530*) were located within close proximity. These genes were recruited to fulfill the function of OppABCD system for nickel uptake [47]. The NP cd630_26700 shared ~37% of protein sequence identity with NP cd630_26710, which was located within the same operon. Although the SBD and the TMD partners were not identified within the NBD operon, the nearby SP (AppA; *cd630_26720*) and TPs (AppBC; *cd630_26730* and *cd630_26740*) were predicted as their protein partners. Across species, these two NPs showed a sequence similarity of at least 40% with oligopeptide NPs, AppF and AppD, of *B. subtilis* [48]. In *C. difficile*, both NPs are found to be energized within a single system as the AppABCDF complex (cd630_26700-cd630_26740) responsible for oligopeptide import [46]. In addition, the crystal structure of this complex revealed the preferred substrate was limited to a restricted set of peptides [47].

Sub-family 3: non-transport ABC protein

Six *C. difficile* NPs (cd630_01700, cd630_07850, cd630_14050, cd630_20680, cd630_25930, and cd630_33590) were clustered in this sub-family. They have a unique tandem NBD protein topology. None of these NPs harbored a predicted TMD partner(s) in the same transcriptional unit or in the vicinity. It is likely that the absence of TMD partners is a nature of NPs of this sub-family, which has been reported in many organisms [49]. Across species, these NPs exhibited amino acid sequence similarity up to 30% with energy-dependent translational throttle protein, EttA from *E. coli* [50]; elongation factor 3 (EF-3); the GCN20 protein from *Saccharomyces cerevisiae*; tylosin resistance ATP-binding protein TlrC from *Streptomyces fradiae*; the SrmB protein from *Streptomyces ambofaciens*; and the carbomycin resistance protein CarA protein from *Streptomyces thermotolerans* [51]. Many functional studies have suggested that these proteins are not involved in substrate transport [49,52–54]. Interestingly, these NPs exhibited diverse functions; from protection against drug targets to the regulation of protein translation [4,52,55–57]. Although most studies have been focused on the role of these proteins in non-transport activity, reports suggest that they may hijack another membrane protein(s) to form an efflux transporter [33,34,57,58]. As the putative transmembrane partner(s) of these NPs has not yet been experimentally elucidated, we therefore cannot reject a transport role for these proteins. Hence, this class of proteins was considered to function as either an active transport system, defender of drug targets, or a regulator of protein translation. Interestingly, there is a functional study on one of the *C. difficile* proteins in sub-family 3, namely CD2068. The protein is encoded from *cd630_20680* and showed a considerable degree of sequence homology, up to 67% identity to YkpA, an uncharacterized ATPase protein in sub-family 3 of *B. subtilis*. Although the YkpA was assumed to not fully function in *B. subtilis* [32], the CD2068 was previously characterized to confer multidrug resistance using heterologous expression in *E. coli* and gene disruption and complementation in *C. difficile* [59]. It appears possible that this protein may couple with other membrane protein(s) for transport activity [59]. Therefore, a comprehensive study with direct assays of the NPs in sub-family 3

Table 2
Inter-relationship of NBD sub-family and its SBD and TMD partners.

NBD sub-family	Associated domain partners		Annotated substrate
	SBD group	TMD group	
1	v	f	Ribose sugar
2	iii	f	Peptide/nickel
3	N/A	N/A	Macrolide
4	v	f	Amino acids
5	v	f	Multiple sugars/ osmoprotectants/polyamine
6	N/A	N/A	Multidrugs
7	N/A	f	Antibiotic
8	v	f	Iron (III)
9	N/A	f	Bacteriocin
11	v	f	Sulfonate/molybdate
12	v	f	Cobalt/vitamins
13	i	d	Iron (III)

would be helpful to fully understand the function of this class of transporters.

Sub-family 4: amino acid(s)/D-methionine/cystine importer

Four NPs (cd630_07520, cd630_14890, cd630_17760, and cd630_21720) belonging to sub-family 4, were annotated as amino acid transporters. Two NPs (cd630_07520 and cd630_17760) were annotated to function in the uptake of polar amino acids, while two other NPs (cd630_14890 and cd630_21720) are responsible for the import of sulfur-containing amino acids. Both NPs cd630_07520 and cd630_17760 possessed their own TP and SP regions forming two complete ABC importers (cd630_07500-cd630_07520 and cd630_17740-cd630_17760). These two NPs (cd630_07520 and cd630_17760) exhibited a high degree of sequence similarity, up to 48% identity, suggesting a corresponding role in the transport of similar substrate(s). These NPs showed approximately 50% identity to the arginine transporter ArtM of *B. subtilis* and >45% identity to the glutamine transporter GlnQ from *B. subtilis* and *E. coli*. Consistent with their SPs and TPs, these NPs also showed sequence similarity to either glutamine or arginine transport proteins in other organisms. From the prediction, it was, therefore, noted that these *C. difficile* transporters may be involved in the uptake of glutamine, arginine, or both of these amino acids. However, experimental validations are needed to confirm the type of amino acid substrates for these transporters. The NPs cd630_14890 (*metN*) formed an ABC importer with the potential partner TP (*metI*; cd630_14900) and SP (*metO*; cd630_14910). This system exhibited a substantial degree of sequence homology of more than 44% identity to the methionine importer MetNPQ of *B. subtilis* [60]. Our findings suggested that the MetNIQ system in *C. difficile* operates in methionine transport. NP cd630_21720 appeared as an individual protein under the transcriptional unit. Two nearby operons, each consisting of a SBD and TMD (*cd630_21740-cd630_21750* and *cd630_21770-cd630_21760*) could be the potential domain partners. From the protein–protein interaction network, cd630_21720 could function as one multicomplex or operated with protein partners encoded from each operon. All predicted domains from cd630_21720 displayed more than 27% identity to two L-cystine uptake systems from *B. subtilis*, TcyABC and TcyJKLMN [61]. These findings suggest that the functional role of this system(s) is the uptake of L-cysteine within *C. difficile*.

Sub-family 5

Five NPs were classified within this sub-family, which were further subdivided into three groups according to their predicted substrates. Sub-family 5a is composed of two NPs encoded from *cd630_24560* and *cd630_34170*. Genome analysis revealed that *cd630_24560* resided individually in its transcriptional unit, with the SBD and TMD absent in proximity. In contrast, NP cd630_34170 was regulated by the same promoter as its TP and SP partners, forming a complete ABC import system (cd630_34140-cd630_cd630_34170). These two NPs share a high degree of similarity at both DNA and amino acid levels (greater than 45%) to the maltodextrin transporter MsmX of *B. subtilis* [62], *sn*-glycerol-3-phosphate transporter UpgC of *Salmonella* Typhimurium [63], and multiple sugar transporter MsmK of *Streptococcus mutans* [64]. Analysis of protein–protein interaction networks uncovered interactions of NP cd630_24560 with several ABC proteins involved in sugar uptake including SPs (cd630_26450 and cd630_25500) and TPs (cd630_25490 and cd630_25480), which were encoded from distant genes. Moreover, this protein was also predicted to interact with TPs (cd630_34150 and cd630_34160) responsible for *sn*-glycerol-3-phosphate transport. Thus, it appears this protein may be required by several systems similar to the functional exchange of maltose NP MalK and *sn*-glycerol-3-phosphate NP UgpC in *E. coli* [65]. Hence, the NPs in sub-family 5a are predicted to participate in carbohydrate transport in *C. difficile*.

Sub-family 5b contained two annotated glycine betaine, carnitine and choline NPs cd630_09000 and cd630_32150. Interestingly, it was observed that TP partners (cd630_9010 and cd630_32160) for these systems contained a high-affinity SBD fused within their proteins (PF4069, IPR007210) and no individual SP was identified in the operon. This appearance of fusion proteins is frequently observed in Gram-positive rather than gram-negative bacteria [66], especially with transport systems involved in bacterial osmoregulation [67]. The OpuCA gene (*cd630_09000*) exhibited more than 46% sequence similarity to the osmoprotectant glycine betaine transporter OpuCA of *B. subtilis* [68] and L-carnitine transporter OpuCA of *Listeria monocytogenes* [69]. These proteins are regulated in response to osmolarity change in bacterial cells and are expressed to increase the accumulation of osmoprotectants from exogenous sources, preventing high-osmolarity stress [67,70]. This finding may enable the assignment of the functional role of these NPs in sub-family 5b to be involved transport of osmoprotectants.

The third group, sub-family 5c, contained only one NP (cd630_10240) involved in polyamine compound transport. The PotA protein encoded by *cd630_10240* gene formed a distinct phylogenetic clade apart from other proteins in sub-family 5. It was found that the *C. difficile* PotABCD system (cd630_10240-cd630_10270) exhibited a sequence similarity of more than 35% identity to the *E. coli* PotABCD complex. This system has been reported to be associated with the import of polyamines (putrescine, spermidine, and spermine), which play an important role in cell proliferation and differentiation in *E. coli* [71]. These results suggest that the functional role of the PotABCD system is in the uptake of polyamines in *C. difficile*.

Sub-family 6: multidrug exporter

Sub-family 6 contained 11 NPs (Table 1), making this group the second most abundant among all sub-families. All systems belonging to this sub-family possessed characteristics of TMD-NBD fusions. This is a common feature among MDR proteins from eukaryotes and drug resistance-associated proteins in prokaryotes, including LmrA and LmrCD from *L. lactis* [20,72], BmrCD from *B. subtilis* [73], EfrCD from *E. faecalis* [26], and AbcA from *S. aureus* [74]. These 11 *C. difficile* NPs showed more than 25% homology to the listed MDR ABC proteins. The *C. difficile* complexes consisted of two half-transporter (TMD-NBD) proteins, which could form either homo- or heterodimeric ABC transporters (Fig. 1e, f). These NPs contributed to six complete ABC systems. The majority of these systems were formed from two half-transporter domains, constituting heterodimer ABC transporters, except for cd630_10840. Experimental studies of all MDR transporters in other bacteria have revealed that these transporters are a major determinant for both acquired and intrinsic multidrug resistance in many bacterial hosts. The MDR transporters have played an important role in reducing the efficacy of drugs including hoechst 33342, mitoxantrone, doxorubicin, ethidium, daunomycin, kanamycin, macrolides, bacitracin, colistin, and phenol-soluble modulins through enhanced efflux activity [20,26,72–74]. Hence, numerous proteins contained in sub-family 6 may be key contributors to the high level of antibiotic resistance seen in *C. difficile*.

Sub-family 7: antibiotic exporter

This sub-family is the largest cluster with 17 NPs (Table 1) and accounts for ~18% of total NPs from *C. difficile* 630. The NPs in this sub-family were also annotated to be involved in antibiotic resistance, similar to sub-family 6. However, the protein topology of sub-family 7 is a homodimeric NP coupled with either homo- or heterodimeric TP. These transport systems are therefore distinctly clustered in a separate group from the sub-family 6. Most NPs were predicted to associate with a putative TP partner, except for NP cd630_20240. Protein-protein interaction analysis for NP cd630_20240 revealed an interaction with a neighboring putative

non-ABC annotated TP (cd630_20250), suggesting that this TP is the functional partner required for full transport activity. To our knowledge, none of the systems in this sub-family have been experimentally verified for a role in antibiotic efflux activity. However, the protein sequence of these proteins displays similarity with other drug efflux pumps, such as tetracycline-resistant YbhF of *E. coli* and *B. subtilis*, linearmycins-resistant YfiL of *B. subtilis* [75], and daunorubicin/doxorubicin-resistant YadG of *E. coli* [76]. Therefore, these ABC transporters could contribute to antibiotic resistance in *C. difficile*.

Sub-family 8: iron (III) importer

According to the sub-family tree (Fig. 3), *C. difficile* may possess two groups of ABC systems, sub-family 8 and 13, responsible for iron (III) uptake present in a distantly associated phylogenetic clade. The dissimilarity of SPs between the two sub-families (<10% similarity) indicates that these two transport systems play a role in the specific uptake of different substrates. Based on the findings of Delepelaire, P. [77], many forms of iron-chelating siderophores are produced by microorganisms to scavenge the soluble iron in the environment. Several SPs are discovered to be associated with the different forms of iron-containing substrate. This finding strongly supports the possibility that distantly related SPs in sub-family 8 and 13 may contribute to iron uptake through the transport of distinct iron siderophore complexes. We, therefore, assigned a new sub-family (sub-family 13), which will be discussed in a subsequent section.

Three putative iron NPs (cd630_16490, cd630_28750 and cd630_29970) were classified in sub-family 8. All three NPs were present in their own complete ABC importer (cd630_16470-cd630_16500, cd630_28750-cd630_28780 and cd630_29970-cd630_29990). Across species, these two *C. difficile* iron ABC systems exhibited up to 65% identity to iron (III)-petrobactin uptake YclNOPQ system and iron (III)-hydroxamate import FhuCGBD system of *B. subtilis*. It has been experimentally validated that the two *C. difficile* iron (III) regulons, *yclNOPQ* (*cd630_16470-cd630_16500*) and *fhuCGBD* (*cd630_28750-cd630_28780*) are involved in the uptake of extracellular iron (III) into the cell. These genes are upregulated under iron (III) limited conditions, allowing the cells to compete for iron in the environment [78–80].

Sub-family 9: bacteriocin resistance exporter

Eight NPs (cd630_03180, cd630_03660, cd630_04780, cd630_06450, cd630_06670, cd630_08220, cd630_10970, and cd630_13490) were classified within this sub-family. All NPs harbored a putative TP partner, contributing to 9 individual systems. These NPs showed more than 40% similarity to the lantibiotic transport ATP-binding protein SrtF of *Streptococcus pyogenes* [81] and the bacitracin export ATP-binding protein BceA of *B. subtilis* [82]. These systems in other bacteria are involved in resistance to bacteriocins, including lantibiotic and bacitracin. The bacteriocins are polycyclic antimicrobial peptides produced by specific bacteria, allowing them to compete with closely related bacteria under limited resources [83]. Interestingly, the *cprABC* operon (*cd630_13490-cd630_13510*) has been reported to be responsible for antimicrobial peptide resistance in *C. difficile* [84]. It has been observed that expressing this ABC transporter confers the ability for bacteria to survive exposure to an array of antimicrobial peptides, including nisin, gallidermin, and polymyxin B [85]. These findings suggest possible targets for further experimental validation of other members of this protein sub-family.

Sub-family 11: molybdate/sulfonate importer

Five NPs (cd630_08710, cd630_14830, cd630_23130, cd630_23610 and cd630_29900) were grouped in sub-family 11. Among these five NPs, only one NP cd630_23610 did not have an ABC protein partner that could be identified. Each of the other four NPs was associated with their corresponding SP and TP contributing to four complete ABC importers. Although *C. difficile* molybdate transport *ModABC* operon (*cd630_08690-cd630_08710*) has yet to be experimentally verified for its role in molybdate uptake, it exhibits protein sequence similarity to the molybdate transport ModABC system of *E. coli* K12. Regarding the protein partners of NP cd630_08710 and cd630_23130, the SPs (cd630_08690 and cd630_23110) showed similarity of up to 36% with the molybdate-binding protein ModA of *E. coli* K12 [86]. Likewise, the sequences of TPs (cd630_08700 and cd630_23120) were also highly similar (reaching 43%) with the molybdenum permease YvgM protein of *B. subtilis* [87] and ModB protein of *E. coli* [88]. From these results, we suggest that these two NPs are involved in molybdate import systems. Some bacteria utilize aliphatic sulfonate as a source of sulfur for growth. Starvation for sulfur leads to the expression of *SsuEADCB* genes, which express an ABC transporter required for aliphatic sulfonate uptake [89,90]. In *C. difficile* 630, two NPs (cd630_14830 and cd630_29900) were predicted to contribute to two sulfonate transport SsuABC systems (cd630_14820-cd630_14840 and cd630_29890-cd630_29910), which showed more than 30% protein sequence similarity to the SsuABC systems from *B. subtilis* and *E. coli* K12. Although the NP cd630_23610 is annotated as a nitrate/sulfonate/taurine ABC protein, ABC domain partners have not been genetically identified. Protein-protein interaction network analysis revealed associations with other ABC proteins involved in nitrate/sulfonate/taurine SP (cd630_23650) and TPs (cd630_19040 and cd630_23670). Therefore, we suggest that the NP cd630_23610 forms a complete ABC system with the corresponding partners responsible for nitrate/-sulfonate/taurine uptake.

Sub-family 12: cobalt/vitamin(s) importer

Four putative NPs (cd630_01000, cd630_01010, cd630_03270, and cd630_04310) were identified within this sub-family. Among them, two NPs (cd630_01000 and cd630_01010) were predicted to work cooperatively as heterodimeric domains in the CbiO1O2Q system, making a total of three transport systems identified in this sub-family. Cobalt is an essential trace element for assembly of cobalt-dependent enzymes, which play key roles in several biological processes. To gain sufficient cobalt, bacteria usually express a high affinity import system to acquire available cobalt from environment [91]. The amino acid sequence of *C. difficile* CbiMNOQ system (cd630_03240, cd630_03250, cd630_03270, and cd630_03260) was found to be similar to the cobalt transport CbiMNOQ system of *Salmonella* Typhimurium [92], with up to 60% identity. The CbiMNOQ system was predicted to be an ECF-type ABC transporter composed of an ATP-binding protein (A component; cd630_03270), a transmembrane protein (T component; cd630_03260), and possible substrate capture proteins (S component; cd630_03240 and cd630_03250). The absence of an SP in the import system is a unique feature for EFC-type ABC importers. It has been reported in *S.* Typhimurium that the uptake of cobalt from the ABC transporter is not dependent on an extracellular SP [92,93]. Also, the NP encoded from *cd630_04310* and its TP encoded from *cd630_04300* shared protein sequence homology to uncharacterized proteins involved in cobalt transport and energy ECF-type systems of many organisms. Experimental investigation is still required to confirm the functional role(s) for these transporters in *C. difficile*. Two heterodimeric NPs (cd630_01000 and cd630_01010) are present in the CbiO1O2Q (or EcfA1A2T) system, which was annotated to be involved in cobalt uptake in *C. difficile*. However, the protein sequences of these NP were more similar (>44%) to EcfA1A2T system in *L. lactis* than to cobalt transport systems. *L. lactis* EcfA1A2T has a functional role in the uptake of vitamins and their precursors, and the TP can interact with several SPs [94]. The use of an ABC transporter by various genetically unrelated SP is often observed with vitamin transport in Gram-positive bacteria, especially pathogens [95]. This situation was also predicted for the CbiO1O2Q system in *C. difficile* using the protein

network from STRING. The CbiO1O2Q system showed an interaction with components of the ECF-type transporters involved in riboflavin (vitamin B2) import and proteins in the biotin (vitamin D) uptake system. These findings suggest that the *C. difficile* CbiO1O2Q system shares SP(s) or S component, or both, with other system(s). We, therefore, assigned a functional role for the CbiO1O2Q in vitamin import in *C. difficile*.

Sub-family 13: iron (III) importer

Only two NPs (cd630_35270 and cd630_35300), annotated as iron (III) transporters, were clustered within sub-family 13. Both of these NPs were predicted to form independent ABC systems, with their SBD and TMD located within the same transcriptional unit as the corresponding NBD. Domain association results revealed a unique interaction of the NPs of this sub-family with the SP group i and TP group d, which differs from the NPs of sub-family 8 (Table 2). These findings support the classification of sub-family 8 and 13 into separate clades. These two NPs shared substantial sequence similarity, of up to 85% identity, suggesting a possible gene duplication event. Although the two NPs were annotated as iron (III) transporters, their SPs and TPs also possessed aminoethylphosphonate-binding domains (IPR017663 and IPR017664). Aminoethylphosphonate is a naturally occurring phosphonate, and members of this class of compounds are typically conjugated with glycans, lipids, and proteins. The degradation of these compounds yields usable forms of carbon, nitrogen, and phosphorus that are beneficial for specialized pathogens enabling host infection and persistence. As the specificity of SP in this group can be described as either iron (III) or 2-aminoethylphosphonate, functional verification of this system for iron (III) or 2-aminoethylphosphonate transport is required. In addition, analysis of the protein–protein interaction network revealed associations between all ABC proteins in sub-family 13 and the iron (III)-associating SP (cd630_29530) and TP (cd630_28880). The association of an ABC system with more than one set of domain partners has been proposed by Quentin et al for the iron (III) *fhu* system of *B. subtilis,* which possesses two binding proteins encoded by the *fhuD* and *foxD* genes [32]. We, therefore, suggested that SP (cd630_29530) and TP (cd630_28880) are part of an iron (III) ABC importer in *C. difficile* 630.

Unclassified

The phylogenetic relationship of 97 NPs was classified into 12 different sub-families related to their predicted substrate(s) and protein topology; however, 27 NPs were not fit to any of the sub-families. Most of these NPs were grouped in the same phylogenetic clade with unrelated or unpredicted substrates (Table 1). Twenty four of the 27 NPs exhibited associations with their putative domain partner for five ABC importers and 19 exporters. ABC protein partners were not identified for the remaining NPs (cd630_35330, cd630_35340, and cd630_36080). For NP cd630_36080, the sequence analysis showed a 30% identity to ATP-dependent iron-sulfur cluster assembly protein, SufC, of *E. coli* K12. The protein–protein interaction network also indicated the interaction with another iron-sulfur cluster assembly protein encoded from *cd630_36070*. The *E. coli* SufBCD complex has been found to serve as a scaffold for iron-sulfur cluster biogenesis [96]. We, therefore, suggest that the cd630_36080 protein is not involved in transport activity but was rather a part of a complex involved in iron-sulfur transfer for redox, catalytic, or regulatory functions and/or as an oxygen or iron sensor. Among the 24 complete ABC systems, five systems containing NPs cd630_08740, cd630_08770, cd630_10010, PstB (cd630_32610), and cd630_33640 had putative substrate(s) predicted by both protein annotation and BLASTP analysis. These NPs were predicted to be involved in the transport of sugars, nitrate/sulfonate/taurine, phosphonate, drugs, and phosphate. However, the potential substrates for the majority of unclassified NPs, which made up to ~20% of total NPs in *C. difficile* 630, have yet to be elucidated. Moreover, some NPs containing an annotated substrate (cd630_35300, cd630_36080, cd630_08740, cd630_08770, cd630_33640, and cd630_32610) were not classified to their group but distributed in other sub-families. This finding might reflect the limitation of classification solely based on NP sequence similarity, as the protein structures might differ. Moreover, it was observed that some of the unclassified NPs shared sequence similarity with other known ABC proteins, suggesting the possibility that they may be involved in these substrate transport systems. For example, NP cd630_19780 exhibited interactions with sulfonate/nitrate transport components. Its TP (cd630_19790) showed sequence similarity to known sulfonate TPs, SsuC and SsuC2. Based on these observations, this ABC system may play a role in sulfonate import. However, classifications of other ABC domains, to include not only sequence similarity but also protein folding information, could provide for the better functional assignment of ABC systems with unknown substrates.

In conclusion, the construction of the entire ABC systems in *C. difficile* 630, based on computational analyses, allowed us to classify complete ABC systems into 12 different sub-families. Each sub-family was primarily characterized based on their annotated substrate(s) and experimental validation of ABC systems in *C. difficile* greatly supported our construction [45,47,78–80,84,85]. Although the function of most *C. difficile* ABC systems has yet to be elucidated, protein comparison with known ABC proteins and other classification approaches has the potential to facilitate functional assignment and can be a guideline for further validation analyses.

3. Stress responses mediated by the ABC systems in *C. difficile* 630

To evaluate the response of ABC systems to various conditions such as antibiotic exposure, stress, and infection, we retrieved transcriptomic data from previous reports (see Method 7.) and robustly normalized and compared the data sets, as shown in Fig. 4. Under antibiotic treatment, the most upregulated ABC genes were from sub-families 2, 3, 4, 12, and 13. These sub-families are responsible for the transport of peptide/nickel, macrolide resistance, amino acids, cobalt/vitamins, and iron (III). It has been reported that the peptide importer operon, *OppABCDF,* and amino acid production related-genes are significantly upregulated following exposure to polymyxin B derivatives in *S. aureus*, although their role in resistance to this stress is not well understood [97]. The ABC genes in sub-families 6 and 7, which were functionally assigned to have roles in drug resistance, were not over-pronounced due to exposure to certain antibiotics. This may be due to *C. difficile* possessing an array of resistance mechanisms to protect against different antibiotics [98]. It has also been hypothesized that the tested antibiotics were not the dedicated substrates for those ABC transporters.

Under various stress conditions, *C. difficile* 630 specifically responded to each pressure through altered expression of sets of ABC systems. For example, transcriptomic data revealed a significant upregulation of ABC systems associated with oligopeptides, multiple sugars, and spermine/putrescine uptake (sub-families 2 and 5) under conditions that promote spore germination. Spermine/putrescine is found to be an essential source of polyamine used for cell proliferation, cell differentiation, and spore germination in both prokaryotic and eukaryotic cells [71,99]. Genes associated with amino acid and peptide import system (sub-families 2 and 4) were highly expressed upon alkali exposure, which has also been observed in the response to alkali stress by *L. monocytogenes* [100]. The role of peptide import has

been suggested to help generate a proton source for pH neutralization, through digestion to the free amino acids [101]. There was no apparent correlation between ABC systems and the remaining stresses examined.

For *in vivo* infection, it was observed that the ABC transporters in sub-family 4, responsible for the uptake of amino acids, were upregulated to the highest level at all time points. The important role of amino acid uptake has previously been discussed and catabolism of amino acids is the main resource for growth during infection [102].

4. Identification of core and accessory ABC proteins in *C. difficile*

The available bacterial genomes sequences are expanding, providing more resources to study ABC protein repertoires. As ABC systems can be found in all living organisms, identification of core and accessory ABC proteins can facilitate the prediction of the biological niche and help to reveal relationships among organisms.

To examine the relationship among *C. difficile* strains, genomes from 49 *C. difficile* strains were retrieved from the NCBI database. Most *C. difficile* strains exhibited an antimicrobial resistance (AMR) background and the identification of genes associated with AMR was predicted using the CARD database [103]. It was observed that all *C. difficile* strains harbor between 75 and 95 AMR genes, accounting for ~2% of their genomes (Supplement 3). Among the identified AMR genes, ABC genes were present ranging from 12 to 18% of total AMR genes. In *C. difficile* 630, 93 AMR genes were predicted within the genome. ABC genes were identified as 16% of total AMR genes, and these AMR-ABC genes accounted for 6.5% of total ABC genes in the genome. These findings on the number of ABC associated AMR genes might reflect the adaptability of *C. difficile* to survive under antibiotic treatment.

Available, complete *C. difficile* genomes provide more details on the intra-genome comparison of ABC protein repertoires. Putative ABC genes were extracted from all *C. difficile* genomes. The reciprocal BLASTP was performed to obtain ABC orthologues prior to being subjected to core and accessory ABC analysis. Within *C. difficile* 630, it was discovered that 18 NPs (approximately 18% of total NPs) were identified as core ABC proteins, which could be found in all selected strains (Fig. 4). These NPs were predominantly categorized in sub-families associated with drug resistance, including sub-family 3, 6, 7, and 9. Moreover, NPs from *C. difficile* 630 which shared more than 80% with other strains were significantly detected in antibiotic-resistance related sub-families. It was not surprising that approximately 50% of NPs were annotated as involved in drug resistance were found in core ABC proteins, as these ABC transport systems may impact the drug resistance background of *C. difficile*.

Fig. 4. NBD gene expression level under physical and biological stresses and conservation of NBD among *C. difficile* strains. Heat map was generated with quantile normalized of differentially expressed NBD genes, which are listed along the Y-axis. Each column represents expression value of genes under an individual stress condition. The color is represented according to the differential expression level as red indicates the induced genes after treatment, while blue denotes the repressed genes. For the conservation panel, the highest percentage identification is colored red, whereas the lowest percent identification is shaded as green. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5. Conclusions

The ABC proteins establish one of the largest groups of membrane proteins found in all living organisms. Genome analysis of *C. difficile* 630 allowed us to identify 226 putative ABC proteins, which accounted for ~6% of total proteins. Determination of functional domains revealed 97 NPs, 30 SPs, 98 TPs, and one undefined domain protein. The NPs containing highly conserved motifs were classified into 12 different sub-families consistent with their annotated substrate. Bioinformatic analysis facilitated the construction of at least 78 ABC systems, functioning in import, export, and non-transport activities. The majority of ABC systems (64% of total ABC systems) used by *C. difficile* 630 were predicted to be associated with antimicrobial extrusion, consistent with the ability of this organism to resist an array of antibiotic treatments. Therefore, ABC proteins associated with drug resistance in the sub-families 3, 6, 7, and 9 warrant further investigation for their potential impact on understanding *C. difficile* pathogenesis. Moreover, core identified NPs from *C. difficile* 630 were predominantly found within sub-families responsible for antibiotic extrusion, emphasizing the role of this system to confer drug resistance in *C. difficile*. Our construction was based on bioinformatic analyses; however, substrates for most ABC proteins in *C. difficile* have yet to be experimentally validated. Our work herein provides an overview of ABC systems employed by *C. difficile* and should facilitate the functional analyses for a better understanding of the role of ABC systems in this organism.

6. Materials and methods

6.1. Identification of putative ABC proteins in C. difficile 630

We used the NCBI databases as a primary source of information for *C. difficile* 630 (NC_009089.1) proteins and DNA sequences. All ABC protein sequences were extracted from genome using the protein annotation keyword "ABC". The conserved motifs were searched on the primary amino acid sequence of all ABC proteins using InterProscan 5.29 [36]. This program utilizes several databases such as PFAM, SignalP, Phobius, and TMHMM to identify motifs in a set of unaligned protein sequences. ABC transporter domain PF00005 (IPR027417), which was present in all NBDs, was used to identify the NPs. Within protein domain PF00005, there were several well-characterized signature motifs of NBD comprising Walker A (Gx_2GxGKS/T), ABC signature (LSGGQ), and Walker B (hhhhDE), which are responsible for ATP-binding and hydrolysis. In addition, identification of NPs was also performed by searching protein family PF00005 from all encoded proteins in *C. difficile* 630 combining with ABC proteins from the TransportDB database [104]. Although transmembrane proteins display non-significant conserved motifs, they harbor hydrophobic regions along their protein structure. Therefore, TMHMM (Transmembrane Helices; Hidden Markov Model) and Phobius (a combined transmembrane topology and signal peptide) predictor in InterProscan was used as a tool for identification of membrane integral domain among the pool ABC proteins. Another common conserved motif for TP importers, EAA loop (PROSITE: PDOC00364) ($EAAx_3Gx_9IxLP$), was identified in all predicted TPs. The TP containing EAA motif was then proposed to associate with import systems. Substrate-binding proteins also do not exhibit substantial sequence conservation among different substrates; however, all gram-positive SPs contain hydrophobic domain as an anchor for the external surface of the cell membrane. These proteins are expressed with a precursor signal peptide, which is then cleaved by signal peptidase during translocation across the membrane. Therefore, signalP (signal peptide motif predictor) and Phobius in InterProscan were used to scan all predicted ABC protein sequences encoded from *C. difficile* 630 genome. With signalP, the signalP II recognized as a short-conserved cleavage site before exporting out of the cell (PROSITE: PDOC00013) was also used to identify the SBD.

6.2. Classification of ABC transporter

To categorize and establish a complete set of ABC systems, a computer-assisted method was performed to better understand the diversity and complexity of the ABC superfamilies in *C. difficile*. To classify the ABC proteins encoded from *C. difficile* 630, a phylogenetic tree of NPs, which exhibited a high degree of conserved structures, was generated based on sequence similarity. Amino acid sequences within the protein domain PF00005 (approximately 200 amino acids in length) of 97 NPs were extracted prior to subjecting to alignment. Multiple alignments were obtained from CLUSTRALW included in the MEGA version X [105]. Phylogenetic and molecular evolutionary analyses were conducted using MEGA X with the neighbor-joining (NJ) method with a 1000 bootstrap test. The phylogenetic tree was color-modified using Inkscape 0.92.2 program.

As both TMDs and SBDs possessed very poor conserved protein sequences, neither TMDs nor SBDs were solely used to build a classification of ABC proteins based on sequence similarity. Clustering matrix was used as an algorithm for TMD and SBD classification. Clustering matrix was performed using k-means prior construction of dendrogram with Euclidean distance and ward D2 clustering method (R 3.6.1, gplots 3.0.1). The clusters of both TMD and SBD were further used to support the classification of NBD (Supplement 2).

6.3. Construction of ABC transport system

Each bacterial ABC domain is typically encoded by genes residing under the same promoter or gene/operon in adjacent regions. The gene organization was analyzed to identify all compartments of the ABC system including NBD, TMD and SBD. Construction of ABC systems in *C. difficile* 630 was performed among 226 annotated ABC proteins. After identification of NBD, the domain partners were further manually searched based on operon data of *C. difficile* 630 obtained from the Prokaryotic Operon DataBase (ProOpDB, http://operons.ibt.unam.mx/OperonPredictor) [106]. Annotated ABC domain partner genes, operons, or both residing nearby NBD were then collected. A complete ABC system was required to incorporate at least one NBD and one TMD.

To facilitate the identification of ABC systems, the functional assignment of the *C. difficile* ABC system was facilitated by BLASTP analysis and protein annotation of all functional domains. Protein similarity alignment search using the BLASTP platform was performed against other reference proteomes with a *P*-value threshold lower than e-10. The BLAST hit protein was accepted when the protein coverage was more than 80% and percent sequence similarity was greater than 30% identity.

6.4. Prediction of protein–protein network interactions

Protein interaction was used to predict the association of an isolated NP, SP, and TP with their putative partner(s). The protein–protein interaction network, STRING, was used as a source and prediction tool for both direct (physical) phylogenetic events and indirect (functional) associations. The interactions in STRING are derived from five main sources including genomic context prediction, automated text mining, previous knowledge in databases, (conserved) co-expression, and high-throughput lab experiments [107]. The protein name was imported to the database for the pre-

diction. The protein(s) with the highest e-value score was further predicted as an interacting partner.

6.5. Identification of antimicrobial resistance proteins

Antibiotic resistance homolog prediction was performed using the comprehensive antibiotic resistance database (CARD) as a reference [103]. Complete 49 *C. difficile* strains genomes were retrieved from NCBI databases (see genome reference number in Supplement 3). All proteins were extracted from 49 *C. difficile* genomes prior to being subjected to CARD. The Diamond program (version 0.9.26.127) was used for the alignment search tool [108].]. The positive hits had homolog with E-values less than 1e-5, identity more than 35%, query cover more than 75%, and subject cover more than 75% were then selected. Among hit AMR genes, ABC genes were sorted with keyword "ABC" and then calculated as a percentage against total AMR genes and total ABC genes in each *C. difficile* strain.

6.6. Identification of core and accessory ABC

Core and accessory ABC genes in *C. difficile* were identified in 49 complete genomes of *C. difficile* strains. Gene annotation was obtained from either publicly available database or in-house annotation using PROKKA [109]. OrthoMCL was used to perform orthologous protein clustering [110]. The output of OrthoMCL was used to create a table showing the presence or absence of each OrthoMCL cluster within every genome using an in-house python script. Core genes are defined as genes, which are present in all genomes of tested *C. difficile* strains, accessory genes are defined as not present in all strains. ABC transporter genes were obtained from the core genes set and accessory genes set by searching for the gene name with the term "ABC", then each gene was submitted to NCBI blast to ascertain the annotation. The percentage of identified NPs in *C. difficile* 630 among the tested strains was colored ranging from the highest percentage as red to the lowest percentage as green.

6.7. Transcriptomic profiles of ABC genes in C. difficile 630

Transcriptomic profiles of *C. difficile* 630 under various conditions were retrieved from many reports. The exposure conditions were categorized into three groups comprising antibiotic treatment (amoxicillin, clindamycin, and metronidazole) [111], abiotic treatment (heat, acid, alkali, oxygen, osmotic shock, nutrient shift, and spore germination) [111–113], and *in vivo* infection [114]. All differentially expressed ABC genes were collected and manipulated using median normalization within the dataset and quantile normalization for between datasets to facilitate the comparison among all treatments. A heatmap was generated with the highest expression level as red and the down-regulated expression as blue.

CRediT authorship contribution statement

Methinee Pipatthana: Investigation, Visualization, Formal analysis, Writing - original draft. **Phurt Harnvoravongchai:** Supervision, Methodology, Writing - review & editing. **Pisut Pongchaikul:** Software, Formal analysis, Writing - review & editing. **Somsak Likhitrattanapisal:** Software, Formal analysis. **Matthew Phanchana:** Writing - review & editing. **Surang Chankhamhaengdecha:** Writing - review & editing. **Tavan Janvilisri:** Conceptualization, Supervision, Project administration, Methodology, 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.

Acknowledgements

This research project is supported by Mahidol University. MP is a recipient of the Science Achievement Scholarship of Thailand (SAST). We thank Assoc. Prof. Laran T. Jensen, Department of Biochemistry, Faculty of Science, Mahidol University for critically reading the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.05.012.

References

[1] Higgins CF. ABC transporters: from microorganisms to man. Annu Rev Cell Biol 1992;8:67–113.
[2] Truglio JJ, Croteau DL, Van Houten B, Kisker C. Prokaryotic nucleotide excision repair: the UvrABC system. Chem Rev 2006;106:233–52.
[3] Wilkens S. Structure and mechanism of ABC transporters. F1000Prime Rep 2015;7.
[4] Fostier CR, Monlezun L, Ousalem F, Singh S, Hunt JF, Boël G. ABC-F translation factors: from antibiotic resistance to immune response. FEBS Lett 2020.
[5] Sharom FJ, Dassa E. Natural history of ABC systems: not only transporters. Essays Biochem 2011;50:19–42.
[6] Walker JE, Saraste M, Runswick MJ, Gay NJ. Distantly related sequences in the alpha-and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J 1982;1:945–51.
[7] ter Beek J, Guskov A, Slotboom DJ. Structural diversity of ABC transporters. J Gen Physiol 2014;143:419–35.
[8] Hyde SC, Emsley P, Hartshorn MJ, Mimmack MM, Gileadi U, Pearce SR, et al. Structural model of ATP-binding proteing associated with cystic fibrosis, multidrug resistance and bacterial transport. Nature 1990;346:362–5.
[9] Davidson AL, Dassa E, Orelle C, Chen J. Structure, function, and evolution of bacterial ATP-binding cassette systems. Microbiol Mol Biol Rev 2008;72:317–64.
[10] Xiong J, Feng J, Yuan D, Zhou J, Miao W. Tracing the structural evolution of eukaryotic ATP binding cassette transporter superfamily. Sci Rep 2015;5:1–15.
[11] Boos W. Periplasmic binding protein-dependent ABC transporters. Eshrichia coli and Salmonera typhimurium 1999:1175–209.
[12] Lewinson O, Livnat-Levanon N. Mechanism of action of ABC importers: conservation, divergence, and physiological adaptations. J Mol Biol 2017;429:606–19.
[13] Davidson AL, Chen J. ATP-binding cassette transporters in bacteria. Annu Rev Biochem 2004;73:241–68.
[14] Jochim A, Adolf L, Belikova D, Schilling NA, Setyawati I, Chin D, et al. An ECF-type transporter scavenges heme to overcome iron-limitation in *Staphylococcus lugdunensis*. Elife 2020;9:e57322.
[15] Finkenwirth F, Eitinger T. ECF-type ABC transporters for uptake of vitamins and transition metal ions into prokaryotic cells. Res Microbiol 2019;170:358–65.
[16] Rempel S, Stanek W, Slotboom D. ECF-type ATP-binding cassette transporters. Annu Rev Biochem 2019;88:551–76.
[17] Karlyshev A, Oyston P, Williams K, Clark G, Titball R, Winzeler E, et al. Application of high-density array-based signature-tagged mutagenesis to discover novel yersinia virulence-associated genes. Infect Immun 2001;69:7810–9.
[18] Tanaka KJ, Song S, Mason K, Pinkett HW. Selective substrate uptake: the role of ATP-binding cassette (ABC) importers in pathogenesis. Biochim Biophys Acta Biomembr 2018;1860:868–77.
[19] Wang Z, Bie P, Cheng J, Lu L, Cui B, Wu Q. The ABC transporter YejABEF is required for resistance to antimicrobial peptides and the virulence of *Brucella melitensis*. Sci Rep 2016;6:31876.
[20] Chang G. Multidrug resistance ABC transporters. FEBS Lett 2003;555:102–5.
[21] Ambudkar SV, Kimchi-Sarfaty C, Sauna ZE, Gottesman MM. P-glycoprotein: from genomics to mechanism. Oncogene 2003;22:7468.
[22] Silverton L, Dean M, Moitra K. Variation and evolution of the ABC transporter genes ABCB1, ABCC1, ABCG2, ABCG5 and ABCG8: implication for pharmacogenetics and disease. Drug Metabol Drug Interact 2011;26:169–79.

[23] Poelarends GJ, Mazurkiewicz P, Konings WN. Multidrug transporters and antibiotic resistance in *Lactococcus lactis*. Biochim Biophys Acta 2002;1555:1–7.
[24] Lubelski J, De Jong A, Van Merkerk R, Agustiandari H, Kuipers OP, Kok J, et al. LmrCD is a major multidrug resistance transporter in *Lactococcus lactis*. Mol Microbiol 2006;61:771–81.
[25] Kobayashi N, Nishino K, Yamaguchi A. Novel macrolide-specific ABC-type efflux transporter in *Escherichia coli*. J Bacteriol 2001;183:5639–44.
[26] Hürlimann LM, Corradi V, Hohl M, Bloemberg GV, Tieleman DP, Seeger MA. The heterodimeric ABC transporter EfrCD mediates multidrug efflux in *Enterococcus faecalis*. Antimicrob Agents Chemother 2016;60:5400–11.
[27] Thomas C, Aller SG, Beis K, Carpenter EP, Martinoia E. Structural and functional diversity calls for a new classification of ABC transporters. FEBS Lett 2020;594:3767–75.
[28] Paulsen IT, Sliwinski MK, Saier Jr MH. Microbial genome analyses: global comparisons of transport capabilities based on phylogenies, bioenergetics and substrate specificities. J Mol Biol 1998;277:573–92.
[29] Saurin W, Hofnung M, Dassa E. Getting in or out: early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters. J Mol Evol 1999;48:22–41.
[30] Gebhard S. ABC transporters of antimicrobial peptides in F irmicutes bacteria–phylogeny, function and regulation. Mol Microbiol 2012;86:1295–317.
[31] Linton KJ, Higgins CF. The *Escherichia coli* ATP-binding cassette (ABC) proteins. Mol Microbiol 1998;28:5–13.
[32] Quentin Y, Fichant G, Denizot F. Inventory, assembly and analysis of *Bacillus subtilis* ABC transport systems. J Mol Biol 1999;287:467–84.
[33] Reynolds E, Ross JI, Cove JH. Msr (A) and related macrolide/streptogramin resistance determinants: incomplete transporters?. Int J Antimicrob Agents 2003;22:228–36.
[34] Chesneau O, Ligeret H, Hosan-Aghaie N, Morvan A, Dassa E. Molecular analysis of resistance to streptogramin A compounds conferred by the Vga proteins of staphylococci. Antimicrob Agents Chemother 2005;49:973–80.
[35] Davidson AL, Nikaido H. Purification and characterization of the membrane-associated components of the maltose transport system from *Escherichia coli*. J Biol Chem 1991;266:8946–51.
[36] Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics 2014;30:1236–40.
[37] Mourez M, Hofnung M, Dassa E. Subunit interactions in ABC transporters: a conserved sequence in hydrophobic membrane proteins of periplasmic permeases defines an important site of interaction with the ATPase subunits. EMBO J 1997;16:3066–77.
[38] Kerr ID. Structure and association of ATP-binding cassette transporter nucleotide-binding domains. Biochim Biophys Acta Biomembr 2002;1561:47–64.
[39] Berntsson R-P-A, Smits SH, Schmitt L, Slotboom D-J, Poolman B. A structural classification of substrate-binding proteins. FEBS Lett 2010;584:2606–17.
[40] van der Heide T, Poolman B. ABC transporters: one, two or four extracytoplasmic substrate-binding sites?. EMBO Rep 2002;3:938–43.
[41] Yamada Y, Tikhonova EB, Zgurskaya HI. YknWXYZ is an unusual four-component transporter with a role in protection against sporulation-delaying-protein-induced killing of *Bacillus subtilis*. J Bacteriol 2012;194:4386–94.
[42] Marion C, Aten AE, Woodiga SA, King SJ. Identification of an ATPase, MsmK, which energizes multiple carbohydrate ABC transporters in *Streptococcus pneumoniae*. Infect Immun 2011;79:4193–200.
[43] Ferreira MJ, de Sá-Nogueira I. A multitask ATPase serving different ABC-type sugar importers in *Bacillus subtilis*. J Bacteriol 2010;192:5312–8.
[44] Higgins CF. ABC transporters: physiology, structure and mechanism – an overview. Res Microbiol 2001;152:205–10.
[45] Antunes A, Camiade E, Monot M, Courtois E, Barbut F, Sernova NV, et al. Global transcriptional control by glucose and carbon regulator CcpA in *Clostridium difficile*. Nucleic Acids Res 2012;40:10701–18.
[46] Edwards AN, Nawrocki KL, McBride SM. Conserved oligopeptide permeases modulate sporulation initiation in *Clostridium difficile*. Infect Immun 2014;82:4276–91.
[47] Hughes A, Wilson S, Dodson EJ, Turkenburg JP, Wilkinson AJ. Crystal structure of the putative peptide-binding protein AppA from *Clostridium difficile*. Acta Crystallogr F Struct Biol Commun 2019;75.
[48] Koide A, Hoch JA. Identification of a second oligopeptide transport system in *Bacillus subtilis* and determination of its role in sporulation. Mol Microbiol 1994;13:417–26.
[49] Murina V, Kasari M, Takada H, Hinnu M, Saha CK, Grimshaw JW, et al. ABCF ATPases involved in protein synthesis, ribosome assembly and antibiotic resistance: structural and functional diversification across the tree of life. J Mol Biol 2019;431:3568–90.
[50] Boël G, Smith PC, Ning W, Englander MT, Chen B, Hashem Y, et al. The ABC-F protein EttA gates ribosome entry into the translation elongation cycle. Nat Struct Mol Biol 2014;21:143.
[51] Schoner B, Geistlich M, Rosteck Jr P, Rao RN, Seno E, Reynolds P, et al. Sequence similarity between macrolide-resistance determinants and ATP-binding transport proteins. Gene 1992;115:93–6.
[52] Murina V, Kasari M, Hauryliuk V, Atkinson GC. Antibiotic resistance ABCF proteins reset the peptidyl transferase centre of the ribosome to counter translational arrest. Nucleic Acids Res 2018;46:3753–63.
[53] Hirose T, Horvitz HR. The translational regulators GCN-1 and ABCF-3 act together to promote apoptosis in *C. elegans*. PLoS Genet 2014;10.
[54] Crowe-McAuliffe C, Graf M, Huter P, Takada H, Abdelshahid M, Nováček J, et al. Structural basis for antibiotic resistance mediated by the *Bacillus subtilis* ABCF ATPase VmlR. Proc Natl Acad Sci U S A 2018;115:8978–83.
[55] Chen B, Boël G, Hashem Y, Ning W, Fei J, Wang C, et al. EttA regulates translation by binding the ribosomal E site and restricting ribosome-tRNA dynamics. Nat Struct Mol Biol 2014;21:152.
[56] Sharkey LK, Edwards TA, O'Neill AJ. ABC-F proteins mediate antibiotic resistance through ribosomal protection. mBio 2016;7:e01975–e2015.
[57] Kerr ID, Reynolds ED, Cove JH. ABC proteins and antibiotic drug resistance: is it all about transport?. Biochem Soc Trans 2005;33:1000–2.
[58] Nunez-Samudio V, Chesneau O. Functional interplay between the ATP binding cassette Msr (D) protein and the membrane facilitator superfamily Mef (E) transporter for macrolide resistance in Escherichia coli. Res Microbiol 2013;164:226–35.
[59] Ngernsombat C, Sreesai S, Harnvoravongchai P, Chankhamhaengdecha S, Janvilisri T. CD2068 potentially mediates multidrug efflux in Clostridium difficile. Sci Rep 2017;7:9982.
[60] Hullo M-F, Auger S, Dassa E, Danchin A, Martin-Verstraete I. The metNPQ operon of *Bacillus subtilis* encodes an ABC permease transporting methionine sulfoxide, D-and L-methionine. Res Microbiol 2004;155:80–6.
[61] Burguiere P, Auger S, Hullo M-F, Danchin A, Martin-Verstraete I. Three different systems participate in L-cystine uptake in *Bacillus subtilis*. J Bacteriol 2004;186:4875–84.
[62] Schönert S, Seitz S, Krafft H, Feuerbaum E-A, Andernach I, Witz G, et al. Maltose and maltodextrin utilization by *Bacillus subtilis*. J Bacteriol 2006;188:3911–22.
[63] Hengge R, Larson T, Boos W. sn-Glycerol-3-phosphate transport in *Salmonella typhimurium*. J Bacteriol 1983;155:186–95.
[64] Russell R, Aduse-Opoku J, Sutcliffe I, Tao L, Ferretti J. A binding protein-dependent transport system in *Streptococcus mutans* responsible for multiple sugar metabolism. J Biol Chem 1992;267:4631–7.
[65] Hekstra D, Tommassen J. Functional exchangeability of the ABC proteins of the periplasmic binding protein-dependent transport systems Ugp and Mal of *Escherichia coli*. J Bacteriol 1993;175:6546–52.
[66] Fulyani F, Schuurman-Wolters GK, Žagar AV, Guskov A, Slotboom D-J, Poolman B. Functional diversity of tandem substrate-binding domains in ABC transporters from pathogenic bacteria. Structure 2013;21:1879–88.
[67] Kappes RM, Kempf B, Bremer E. Three transport systems for the osmoprotectant glycine betaine operate in *Bacillus subtilis*: characterization of OpuD. J Bacteriol 1996;178:5071–9.
[68] Kappes RM, Kempf B, Kneip S, Boch J, Gade J, Meier-Wagner J, et al. Two evolutionarily closely related ABC transporters mediate the uptake of choline for synthesis of the osmoprotectant glycine betaine in *Bacillus subtilis*. Mol Microbiol 1999;32:203–16.
[69] Fraser KR, Harvie D, Coote PJ, O'Byrne CP. Identification and characterization of an ATP binding cassette l-carnitine transporter in *Listeria monocytogenes*. Appl Environ Microbiol 2000;66:4696–704.
[70] Biemans-Oldehinkel E, Poolman B. On the role of the two extracytoplasmic substrate-binding domains in the ABC transporter OpuA. EMBO J 2003;22:5983–93.
[71] Kashiwagi K, Innami A, Zenda R, Tomitori H, Igarashi K. The ATPase activity and the functional domain of PotA, a component of the sermidine-preferential uptake system in *Escherichia coli*. J Biol Chem 2002;277:24212–9.
[72] van Veen HW, Venema K, Bolhuis H, Oussenko I, Kok J, Poolman B, et al. Multidrug resistance mediated by a bacterial homolog of the human multidrug transporter MDR1. Proc Natl Acad Sci U S A 1996;93:10668–72.
[73] Torres C, Galián C, Freiberg C, Fantino J-R, Jault J-M. The YheI/YheH heterodimer from *Bacillus subtilis* is a multidrug ABC transporter. Biochim Biophys Acta Biomembr 2009;1788:615–22.
[74] Yoshikai H, Kizaki H, Saito Y, Omae Y, Sekimizu K, Kaito C. Multidrug-resistance transporter AbcA secretes *Staphylococcus aureus* cytolytic toxins. J Infect Dis 2016;213:295–304.
[75] Stubbendieck RM, Straight PD. Escape from lethal bacterial competition through coupled activation of antibiotic resistance and a mobilized subpopulation. PLoS Genet 2015;11:e1005722.
[76] Jin M, Lu J, Chen Z, Nguyen SH, Mao L, Li J, et al. Antidepressant fluoxetine induces multiple antibiotics resistance in *Escherichia coli* via ROS-mediated mutagenesis. Environ Int 2018;120:421–30.
[77] Delepelaire P. Bacterial ABC transporters of iron containing compounds. Res Microbiol 2019;170:345–57.
[78] Jahn D, Berges M, Michel A-M, Lassek C, Nuss AM, Beckstette M, et al. Iron regulation in *Clostridioides difficile*. Front Microbiol 2018;9:3183.
[79] Ho TD, Ellermeier CD. Ferric uptake regulator fur control of putative iron acquisition systems in *Clostridium difficile*. J Bacteriol 2015;197:2930–40.
[80] Hastie JL, Hanna PC, Carlson PE. Transcriptional response of *Clostridium difficile* to low iron conditions. Pathog Dis 2018;76. fty009.
[81] Karaya K, Shimizu T, Taketo A. New gene cluster for lantibiotic streptin possibly involved in streptolysin S formation. J Biochem 2001;129:769–75.
[82] Johnson BA, Anker H, Bacitracin MFL. A new antibiotic produced by a member of the *B. subtilis* group. Science 1945;102:376–7.

[83] Riley MA. Molecular mechanisms of bacteriocin evolution. Annu Rev Genet 1998;32:255–78.
[84] Suárez JM, Edwards AN, McBride SM. The *Clostridium difficile* cpr locus is regulated by a noncontiguous two-component system in response to type A and B lantibiotics. J Bacteriol 2013;195:2621–31.
[85] McBride SM, Sonenshein AL. Identification of a genetic locus responsible for antimicrobial peptide resistance in *Clostridium difficile*. Infect Immun 2011;79:167–76.
[86] Rech S, Deppenmeier U, Gunsalus RP. Regulation of the molybdate transport operon, modABCD, of *Escherichia coli* in response to molybdate availability. J Bacteriol 1995;177:1023–9.
[87] Wipat A, Brignell SC, Guy BJ, Rose M, Emmerson PT, Harwood CR. The yvsA-yvqA (293–289) region of the *Bacillus subtilis* chromosome containing genes involved in metal ion uptake and a putative sigma factor. Microbiology (Reading) 1998;144:1593–600.
[88] Maupin-Furlow JA, Rosentel JK, Lee JH, Deppenmeier U, Gunsalus R, Shanmugam K. Genetic analysis of the modABCD (molybdate transport) operon of *Escherichia coli*. J Bacteriol 1995;177:4851–6.
[89] van der Ploeg JR, Eichhorn E, Leisinger T. Sulfonate-sulfur metabolism and its regulation in *Escherichia coli*. Arch Microbiol 2001;176:1–8.
[90] Endoh T, Kasuga K, Horinouchi M, Yoshida T, Habe H, Nojiri H, et al. Characterization and identification of genes essential for dimethyl sulfide utilization in *Pseudomonas putida* strain DS1. Appl Microbiol Biotechnol 2003;62:83–91.
[91] Cheng J, Poduska B, Morton RA, Finan TM. An ABC-type cobalt transport system is essential for growth of *Sinorhizobium melilotiat* trace metal concentrations. J Bacteriol 2011;193:4405–16.
[92] Rodionov DA, Hebbeln P, Gelfand MS, Eitinger T. Comparative and functional genomic analysis of prokaryotic nickel and cobalt uptake transporters: evidence for a novel group of ATP-binding cassette transporters. J Bacteriol 2006;188:317–27.
[93] Chai C, Yu Y, Zhuo W, Zhao H, Li X, Wang N, et al. Structural basis for a homodimeric ATPase subunit of an ECF transporter. Protein Cell 2013;4:793–801.
[94] Ter Beek J, Duurkens RH, Erkens GB, Slotboom DJ. Quaternary structure and functional unit of energy coupling factor (ECF)-type transporters. J Biol Chem 2011;286:5471–5.
[95] Rodionov DA, Hebbeln P, Eudes A, Ter Beek J, Rodionova IA, Erkens GB, et al. A novel class of modular transporters for vitamins in prokaryotes. J Bacteriol 2009;191:42–51.
[96] Hirabayashi K, Yuda E, Tanaka N, Katayama S, Iwasaki K, Matsumoto T, et al. Functional dynamics revealed by the structure of the SufBCD complex, a novel ATP-binding cassette (ABC) protein that serves as a scaffold for iron-sulfur cluster biogenesis. J Biol Chem 2015;290:29717–31.
[97] Zhao J, Cheah S-E, Roberts KD, Nation RL, Thompson PE, Velkov T, et al. Transcriptomic analysis of the activity of a novel polymyxin against *Staphylococcus aureus*. mSphere 2016;1:e00119–e216.
[98] Harnvoravongchai P, Pipatthana M, Chankhamhaengdecha S, Janvilisri T. Insights into drug resistance mechanisms in *Clostridium difficile*. Essays Biochem 2017;61:81–8.
[99] Ruiz-Herrera J. Polyamines, DNA methylation, and fungal differentiation. Crit Rev Microbiol 1994;20:143–50.
[100] Giotis ES, Muthaiyan A, Natesan S, Wilkinson BJ, Blair IS, McDowell DA. Transcriptome analysis of alkali shock and alkali adaptation in *Listeria monocytogenes* 10403S. Foodborne Pathog Dis 2010;7:1147–57.
[101] Padan E, Bibi E, Ito M, Krulwich TA. Alkaline pH homeostasis in bacteria: new insights. Biochim Biophys Acta Biomembr 2005;1717:67–88.
[102] Jenior ML, Leslie JL, Young VB, Schloss PD. *Clostridium difficile* colonizes alternative nutrient niches during infection across distinct murine gut microbiomes. mSystems 2017;2.
[103] Alcock BP, Raphenya AR, Lau TT, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res 2020;48:D517–25.
[104] Elbourne LD, Tetu SG, Hassan KA, Paulsen IT. TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. Nucleic Acids Res 2017;45:D320–4.
[105] Stecher G, Tamura K, Kumar S. Molecular evolutionary genetics analysis (MEGA) for macOS. Mol Biol Evol 2020;37:1237–9.
[106] Mao F, Dam P, Chou J, Olman V, Xu Y. DOOR: a database for prokaryotic operons. Nucleic Acids Res 2008;37:D459–63.
[107] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2014;43:D447–52.
[108] Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods 2015;12:59.
[109] Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2014;30:2068–9.
[110] Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 2003;13:2178–89.
[111] Emerson JE, Stabler RA, Wren BW, Fairweather NF. Microarray analysis of the transcriptional responses of *Clostridium difficile* to environmental and antibiotic stress. J Med Microbiol 2008;57:757–64.
[112] Scaria J, Mao C, Chen J-W, McDonough SP, Sobral B, Chang Y-F. Differential stress transcriptome landscape of historic and recently emerged hypervirulent strains of *Clostridium difficile* strains determined using RNA-seq. PLoS ONE 2013;8.
[113] Dembek M, Stabler RA, Witney AA, Wren BW, Fairweather NF. Transcriptional analysis of temporal gene expression in germinating *Clostridium difficile* 630 endospores. PLoS ONE 2013;8.
[114] Janvilisri T, Scaria J, Chang Y-F. Transcriptional profiling of *Clostridium difficile* and Caco-2 cells during infection. J Infect Dis 2010;202:282–90.