Title

Analysis of 58 Families of Holins Using a Novel Program, PhyST

Journal

Microbial Physiology, 26(6)

ISSN

2673-1665

Authors

Kuppusamykrishnan, Harikrishnan
Chau, Larry M
Moreno-Hagelsieb, Gabriel
et al.

Publication Date

2016

DOI

10.1159/000448040

Peer reviewed

Research Article

Journal of Molecular Microbiology and Biotechnology

J Mol Microbiol Biotechnol 2016;26:381–388
DOI: 10.1159/000448040

Received: January 29, 2016
Accepted: June 28, 2016
Published online: August 24, 2016

Analysis of 58 Families of Holins Using a Novel Program, PhyST

Harikrishnan Kuppusamykrishnan Larry M. Chau Gabriel Moreno-Hagelsieb
Milton H. Saier Jr.

Department of Molecular Biology, Division of Biological Sciences, University of California at San Diego, La Jolla, Calif., USA

Key Words
Holin · Membrane protein · Topology · Phylogeny · Phyla

Abstract
We have designed a freely accessible program, PhyST, which allows the automated characterization of any family of homologous proteins within the Transporter Classification Database. The program performs an NCBI-PSI-BLAST search and reports (1) the average protein sequence length with standard deviations, (2) the average predicted number of transmembrane segments, (3) the total number of homologues retrieved, (4) a quantitative list of all source phyla, and (5) potential fusion proteins of sizes considerably exceeding the average size of the proteins retrieved. We have applied this program to 58 families of holins, and the results are presented. The results show that holins are very rarely fused to other protein domains, suggesting that holins form transmembrane pores as homooligomers without the participation of other proteins or protein domains.

Introduction

Proteins fall into families, members of which, by definition, share a common ancestry. As the protein databases expand, there is a need to provide quantitative information about these families, i.e. average characteristics of the proteins that comprise them. Programs that provide such information are not easy to find because they are highly customized and adapted to the interests of individual researchers. To satisfy the needs of the scientific community with specific interests in transport proteins, the focus of our laboratory, we have designed a program called PhyST (Phylum-Size/Topology). Although designed for our specific needs in the characterization of transport protein families, this program should prove applicable to the study of all types of protein families.

Our laboratory maintains the IUBMB-approved Transporter Classification Database (TCDB; http://www.tcdb.org), which classifies and presents information about transport proteins and their family and superfamily associations [Saier et al., 2009, 2014, 2016]. The database also provides a collection of software that facilitates analysis of the proteins and protein families included in TCDB [Reddy and Saier, 2012; Zhai and Saier, 2001a, b]. One subclass of particular interest includes the holins,

KARGER

E-Mail karger@karger.com
www.karger.com/mmb

1464–1801/16/0266–0381$39.50/0

Milton H. Saier Jr.
Department of Molecular Biology, Division of Biological Sciences
University of California at San Diego, 9500 Gilman Drive
La Jolla, CA 92093-0116 (USA)
E-Mail msaier@ucsd.edu

which are small proteins originally identified in bacteriophages. Holins facilitate the release of phage particles from infected bacteria during viral-induced cell lysis [Young, 2014]. They have been used for a variety of medical and industrial purposes [Gao et al., 2013; Yan et al., 2013] and have numerous physiological functions in noninfected bacterial and eukaryotic cells [Saier and Reddy, 2015]. They are exceptionally diverse in sequence, topology, and mechanism of action [Catalao et al., 2013; Young, 2013]. They have been classified within seven superfamilies, but many holin families, all belonging to subclass 1.E in TCDB, are not members of these superfamilies [Reddy and Saier, 2013]. We apply PhyST to the characterization of the 58 holin families currently included in TCDB.

Results

Properties of 58 Holin Families

Table 1 presents the properties of 58 holin families obtained using the PhyST program. These small integral membrane proteins form 'holes' in bacterial membranes, allowing the release of endolysins [Wang et al., 2000; Young and Blasi, 1995]. They have sizes ranging from less than 50 amino acyl residues (aas) to over 200 aas with 1–4 transmembrane segments (TMSs) [Reddy and Saier, 2013]. Family sizes range from just a few proteins retrieved from the NCBI NR protein database, to over 2,000 proteins, depending on the family (table 1). While the majority of these proteins are derived from Proteobacteria, Firmicutes, and double-stranded DNA (dsDNA) viruses, several have been isolated from other bacterial phyla, and a few are from eukaryotes and Archaea; they can serve a variety of cellular functions [Saier and Reddy, 2015]. The bacterial phyla from which these holins have been identified are tabulated in tables 1 and 2. In decreasing order of prevalence, these phyla are: Proteobacteria (21 families), Firmicutes (20 families), Actinobacteria (9 families), Chloroflexi, Fusobacteria, and Spirochaetes (4 families each), *Deinococcus-Thermus* (3 families), Bacteroidetes, Cyanobacteria, and Tenericutes (2 families each), and Acidobacteria and Thermotogae (1 family each). Fourteen bacterial phyla are listed in table 2. It should be recalled that phylum representation is probably related to the intensity with which that phylum has been studied as well as the frequency with which genomes of representative members have been sequenced. Consequently, the results should not be interpreted strictly in terms of the phyla most frequently subject to holin action.

Correlation of Protein Size with Topology

Figure 1 presents the 58 holin families according to average protein size, and correlates size with numbers of TMSs. It can be seen that there is not an absolute correlation as some of the smallest proteins have 2 TMSs and some of the largest proteins have only 1 TMS or 2 TMSs. Nevertheless, there is a general correlation. For the most part, the transmembrane domains of 1–4 TMSs represent the entirety of these proteins. Extra soluble domains probably resulted from fusions between dissimilar domains, but the occurrence of such fusions is rare. These fusions will be considered in the next section.

Fusions of Holins with Other Protein Domains

The program allows the identification of homologues whose sequences are, for example, two, three, or four times longer than the average size observed for members of a holin family. Many of these were examined manually. Twenty-three fusion proteins, judged to probably be genuine rather than artifactual, were identified from a pool of about 105,000 total proteins, representing about 0.2% (1 in 500). In virtually all such 'oversized' proteins, the holin domain was either N-terminal or C-terminal, but never in the middle of the proteins. Many other large proteins appeared to be artifacts arising due to the erroneous loss of translational termination (stop) codons, probably due to sequencing errors (these are not included in table 1). These putative artifactual proteins were usually found in single copy in the entire NCBI NR protein database. The most common of the apparently genuine fusions were holins fused to cell wall lysins, particularly in phage, where these two proteins are usually encoded by adjacent genes within an operon [Shi and Sun, 2012; Young, 1992]. However, other fusions involved holins linked to other protein domains including putative antiterminators, proteases, amidases, aminotransferases, transglycosylase-associated domains, and domains of unknown function [Ortega et al., 2012; Reddy and Saier, 2013] (e.g. 1.E.40.5.1 and 1.E.2.1.11). In a few cases, the longer proteins resulted from intragenic duplication of a holin domain, giving rise, for example, to a 4-TMS holin from a 2-TMS holin (see proteins with TC No. 1.E.18.1.9 and 1.E.18.1.10, as well as 1.E.36.6.2 and 1.E.36.6.3). It was concluded that holins are rarely fused to other protein domains, suggesting that these proteins form transmembrane pores as homooligomers without the participation of other proteins or protein domains [Wang et al., 2003].

Table 1. Characteristics of the 58 holin families

TCDB No.	Family name and abbreviation	Phyla represented	Average sequence length, aa	Standard deviations, aa	Number of potential fusion proteins	Number of TMSs predicted	Total number of proteins analyzed per family
1.E.1	P21 holin S (P21 holin)	Proteobacteria (99.67%) dsDNA viruses (0.33%)	65	36	7	1;2	900
1.E.2	Lambda holin S (λ holin)	Proteobacteria (96.50%) dsDNA viruses (1.75%) Unclassified phages (1.40%) Euryarchaeota (0.35%)	107	14	1	1	286
1.E.3	P2 holin (P2 holin)	Proteobacteria (99.29%) dsDNA viruses (0.71%	92	14	0	3	283
1.E.4	LydA holin (LydA holin)	Proteobacteria (97.44%) dsDNA viruses (2.56%)	95	16	0	2	39
1.E.5	PRD1 phage P35 holin (P35 holin)	dsDNA viruses (60%) Proteobacteria (40%)	114	7	0	2	5
1.E.6	T7 holin (T7 holin)	dsDNA viruses (100%)	64	8	0	2	34
1.E.7	HP1 holin (HP1 holin)	Proteobacteria (87.50%) dsDNA viruses (12.50%)	74	3	0	1	8
1.E.8	T4 holin (T4 holin)	dsDNA viruses (95.12%) Unclassified phages (3.66%) Artificial sequences (1.22%) (Proteobacteria)	207	39	0	1	82
1.E.9	T4 immunity (T4 imm)	Proteobacteria (43.13%) dsDNA viruses (37.25) Actinobacteria (3.92%) Unclassified phages (1.96%) Acidobacteria (1.96%) Cyanobacteria (9.80%) Planctomycetes (1.96%)	89	45	2	2	51
1.E.10	*Bacillus subtilis* φ29 holin (φ29 holin)	Firmicutes (87.70%) dsDNA viruses (10.50%) Unclassified phages (1.75%)	133	8	0	2	57
1.E.11	φ11 holin (φ11 holin)	Firmicutes (89.94%) dsDNA viruses (7.90%) Actinobacteria (1.44%) Unclassified phages (0.72%)	106	36	2	2	557
1.E.12	φAdh holin (φAdh holin)	Firmicutes (91.67%) dsDNA viruses (8.33%)	128	21	0	1	12
1.E.13	φU53 holin (φU53 holin)	dsDNA viruses (64.40%) Firmicutes (33.30%) Bacteroidetes (2.22%)	114	10	0	3	46
1.E.14	CidA/LrgA holin (CidA/LrgA holin)	Firmicutes (72.82%) Proteobacteria (14.36%) Bacteroidetes (9.10%) Fusobacteria (1.03%) Euryarchaeota (0.96%) Spirochaetes (0.77%) Synergistes (0.19%) Chloroflexi (0.19%) *Deinococcus-Thermus* (0.13%) Tenericutes (0.13%) Deferribacteres (0.06%) Environmental samples (0.06%)	130	15	0	4	1,561
1.E.15	ArpQ holin (ArpQ holin)	Firmicutes (100%)	58	1	0	2	6
1.E.16	Cph1 holin (Cph1 holin)	Firmicutes (88.13%) dsDNA viruses (10.17%) Fusobacteria (1.69%)	138	9	0	3	59
1.E.17	BlyA holin (BlyA holin)	Spirochaetes (100%)	61	11	0	1	47
1.E.18	*Lactococcus lactis* phage r1t holin (r1t holin)	Actinobacteria (60.98%) Firmicutes (21.95%) dsDNA viruses (17.07%)	83	25	3	2	82

Table 1 (continued)

TCDB No.	Family name and abbreviation	Phyla represented	Average sequence length, aa	Standard deviations, aa	Number of potential fusion proteins	Number of TMSs predicted	Total number of proteins analyzed per family
1.E.19	*Clostridium difficile* TcdE holin (TcdE holin)	Firmicutes (89.53%) Fusobacteria (3.73%) dsDNA viruses (1.53%) Proteobacteria (0.93%) Tenericutes (0.33%) Chloroflexi (0.33%) Unclassified phages (0.27%) *Deinococcus-Thermus* (0.07%) Thermotogae (0.07%)/ Environmental samples (0.07%)	138	29	2	3;4	1,509
1.E.20	*Pseudomonas aeruginosa* Hol holin (Hol holin)	Proteobacteria (99.56%) Artificial sequences (0.44%)	113	14	0	2	229
1.E.21	*Listeria* phage A118 holin (Hol118)	Chloroflexi (100%)	108	0.0	0	3	5
1.E.22	Neisserial phage-associated holin (NP-holin)	Proteobacteria (100%)	48	8	0	1	25
1.E.23	*Bacillus* Spore morphogenesis and germination holin (BSH)	Firmicutes (100%)	89	6	0	3	38
1.E.24	Bacterophase Dp-1 holin (Dp-1 holin)	Firmicutes (94.59%) dsDNA viruses (5.40%)	72	6	0	2	37
1.E.25	Pseudomonas phage F116 holin (F116 holin)	Proteobacteria (98.70%) Unclassified phages (1.30%)	87	11	0	1	77
1.E.26	holin LLH (holin LLH)	Firmicutes (100%)	109	9	0	1	45
1.E.27	BlhA holin (BlhA holin)	Firmicutes (96.39%) dsDNA viruses (3.61%)	69	11	0	1	83
1.E.28	*Streptomyces aureofaciens* Phage Mu1/6 holin (Mu1/6 holin)	Actinobacteria (95.65%) dsDNA viruses (4.35%)	75	7	0	2	23
1.E.29	Holin Hol44 (Hol44)	Firmicutes (97.28%) Unclassified phages (2.17%) dsDNA viruses (0.54%)	79	16	0	3	184
1.E.30	Vibrio holin (Vibrio holin)	dsDNA viruses (100%)	53	0.0	0	1	1
1.E.31	SPP1 holin (SPP1 holin)	Firmicutes (94.72%) dsDNA viruses (3.30%) Actinobacteria (1.98%)	89	25	2	3	303
1.E.32	Actinobacterial 1 TMS holin (A-1 holin)	Actinobacteria (88.89%) dsDNA viruses (11.11%)	107	5	0	1	9
1.E.33	2 or 3 TMS putative holin (2/3 holin)	Proteobacteria (98.71%) dsDNA viruses (1.29%)	111	23	2	3	233
1.E.34	Putative actinobacterial holin-X (Hol-X)	Actinobacteria (100%)	100	31	0	2	2
1.E.35	Mycobacterial 1 TMS phage holin (M1 Hol)	dsDNA viruses (100%)	81	6	0	1	17
1.E.36	Mycobacterial 2 TMS phage holin (M2 Hol)	dsDNA viruses (85.71%) Actinobacteria (14.29%)	137	9	0	2	14
1.E.37	Phage T1 holin (T1 holin)	dsDNA viruses (100%)	72	4	0	1	13
1.E.38	*Staphylococcus* phage P68 putative holin (P68 Hol)	dsDNA viruses (100%)	92	0	0	2	1
1.E.39	Mycobacterial phage PBI1 Gp36 holin (Gp36 Hol)	dsDNA viruses (100%)	112	5	0	2	3

Table 1 (continued)

TCDB No.	Family name and abbreviation	Phyla represented	Average sequence length, aa	Standard deviations, aa	Number of potential fusion proteins	Number of TMSs predicted	Total number of proteins analyzed per family
1.E.40	Mycobacterial 4 TMS phage holin (MP4 holin)	dsDNA viruses (100%)	194	24	0	3	7
1.E.41	Deinococcus-thermus holin (D/T-Hol)	*Deinococcus-Thermus* (100%)	106	2	0	2	2
1.E.42	Putative holin-like toxin (Hol-Tox)	Firmicutes (98.57%) Environmental samples (1.43%)	38	7	0	1	71
1.E.43	Putative 3–4 TMS transglycosylase-associated holin (T-A Hol)	Proteobacteria (52.11%) Actinobacteria (34.77%) Cyanobacteria (5.88%) Bacterioidetes (2.94%) Firmicutes (1.03%) Fusobacteria (0.62%) Planctomycetes (0.58%) Acidobacteria (0.49%) Verrucomicrobia (0.36%) Chloroflexi (0.27%)	86	9	2	3	2,249
1.E.44	Putative *Lactococcus lactis* holin (LLHol)	dsDNA viruses (80%) Firmicutes (20%)	67	10	0	2;1	5
1.E.45	*Xanthomonas* phage holin (XanPHol)	dsDNA viruses (100%)	64	0.0	0	2	1
1.E.46	Prophage Hp1 Hol (Hp1Hol)	Firmicutes (100%)	69	0.0	0	1	1
1.E.47	Caulobacter phage holin (CauHol)	dsDNA viruses (100%)	158	0.0	0	2	5
1.E.48	Enterobacterial holin (EBHol)	Proteobacteria (100%)	107	1	0	1	26
1.E.49	Putative *Treponema* 4 TMS holin (Tre4Hol)	Spirochaetes (93.75%) dsDNA viruses (6.25%)	100	16	0	3	16
1.E.50	β-Proteobacterial holin (BP-Hol)	dsDNA viruses (80%) Proteobacteria (20%)	80	16	0	1	5
1.E.51	Putative *Listeria* phage holin (LP-Hol)	Firmicutes (83.30%) dsDNA viruses (16.67%)	41	0.0	0	1	6
1.E.52	The Flp/Fap pilin putative holin (FFPP-Hol)	Firmicutes (100%)	63	8	0	1	111
1.E.53	Toxic Hok/Gef protein (Hok/Gef)	Proteobacteria (99.86%) Artificial sequences (0.14%)	59	12	0	1	735
1.E.54	Gene transfer agent-release holin (GTA-Hol)	Proteobacteria (99.40%) Environmental samples (0.60%)	177	22	0	2	168
1.E.55	*Brachyspira* holin (B-Hol)	Spirochaetes (100%)	87	3	0	2	15
1.E.56	Putative 3 TMS holin (3-Hol)	Proteobacteria (98.40%) dsDNA viruses (1.60%)	96	8	0	3	313
1.E.57	Actinobacterial phage holin (APH)	dsDNA viruses (77.91%) Actinobacteria (20.93%) Unclassified phages (1.16%)	112	13	0	1	86
1.E.58	*Erwinia* Phage Phi-Ea1h holin (EPPE-Hol)	dsDNA viruses (100%)	112	6	0	1	2

Family numbers represent the TC number under subclass 1.E.

Table 2. Bacterial phyla from which the 58 holin families were derived

Bacterial phylum	TC families listed under subclass 1.E
Acidobacteria	43
Actinobacteria	9, 11, 18*, 28*, 31, 32*, 34*, 43, 57
Bacteroidetes	13, 43
Chloroflexi	14, 19, 21*, 43
Cyanobacteria	9, 43
Deinococcus-Thermus	14, 19, 41*
Firmicutes	10*, 12*, 13*, 14*, 15*, 16*, 18, 19*, 23*, 24*, 26*, 27*, 29*, 31*, 42*, 43, 44, 46, 51*, 52*
Fusobacteria	14, 16, 19, 43
Planctomycetes	9, 43
Proteobacteria	1*, 2*, 3*, 4*, 5*, 6*, 7*, 8*, 9*, 14, 19, 20*, 22*, 25*, 33*, 43*, 48*, 50, 53, 54, 56
Spirochaetes	14, 17*, 49*, 55
Tenericutes	14, 19
Thermatogae	19
Verrucomicrobia	43

Bacterial phyla are arranged in alphabetical order with all 58 holin families indicated if represented in that family, regardless of frequency. A TC number with an asterisk indicates that this phylum is the dominant phylum from which members of that family derive.

Conclusions

We have described how PhyST tabulates a variety of familial characteristics including (1) average protein sequence length with standard deviation, (2) average number of TMSs, (3) total number of homologues retrieved, (4) a list of all source phyla, and (5) candidate fusion proteins. PhyST is able to give researchers a statistical picture of homologous proteins that comprise the families. Researchers can use PhyST as a preliminary tool to study the diversity of protein properties within a family. We have applied this program to 58 families of holins currently in TC subclass 1.E. Our results show that holins very rarely fuse to other protein domains, implying that holins form transmembrane pores as homooligomers without the participation of other proteins or protein domains. Although this program was designed for use in the characterization of (putative) transport protein families, it should be noted that it can be applied to any family of proteins found in nature. We hope PhyST will prove to be value for many purposes.

Methods

Compatibility

PhyST has been tested on Ubuntu Linux 14.04+, OS X 10.11 Capitan, and Windows 8. PhyST requires the same modules as the BioV suite of programs [Reddy and Saier, 2012]. The full list of modules and external dependencies required to run PhyST, as well as details regarding their installation, can be found online (http://biotools.tcdb.org/barphyst.html). In addition, the website provides a tutorial illustrating how to run PhyST and suggests types of analysis that can be performed on the output produced by this program.

The PhyST Program: Description and Functionality

PhyST relies on PSI-BLAST used with an established member of the family of interest as the input sequence to retrieve homologues. In order to capture a large fraction of the sequence diversity among family members, it may be necessary to run PSI-BLAST using two or more iterations. The program (filename: physt.py) takes one or more TC family number(s) (i.e. 1.E.1, 1.E.2, 1.E.3, etc.) as input, and retrieves the sequence of one member of each family in order of appearance in TCDB. Thus, the program selects the first protein of that family (e.g. 1.E.1.1.1) to conduct PSI-BLAST searches, unless an alternative TC protein number is given (i.e. 1.E.1.2.3). If it is not possible to access the sequence of the first family member, due to network-related errors, PhyST skips that protein and moves on to the next one (e.g. 1.E.1.1.2). Once PhyST finds a protein that returns PSI-BLAST hits, it exits PSI-BLAST and starts calculating the statistics for the homologues retrieved. PhyST discards all proteins smaller than a minimal cutoff length relative to the length of the query (by default the minimal cutoff length is 50%, but can be adjusted to any value) with the purpose of eliminating fragmentary sequences and false positives. PhyST then computes the number of TMSs inferred for each protein using HMMTOP [Tusnady and Simon, 2001] and outputs the total number of protein homologues from every phylum found as well as the average sequence length with standard deviation and the average number of TMSs. It also predicts putative fusion proteins, which are defined as proteins that are at least twice the average length of the BLAST hits retrieved.

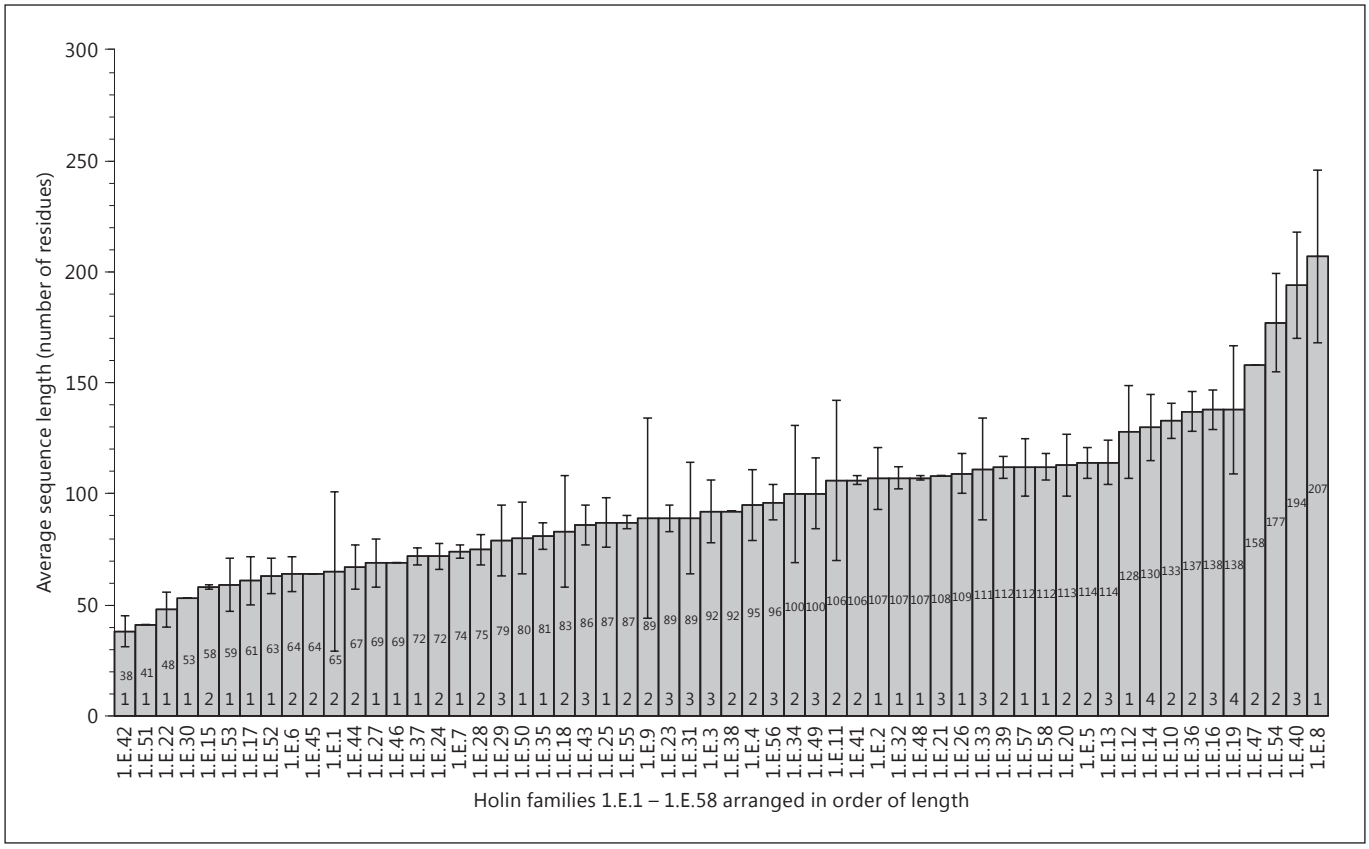

Fig. 1. Average holin size arranged from smallest to largest, correlated with topology. Error bars indicate standard deviations of measurements, with the average sequence length located in the middle of each bar. The TC family number is given on the x-axis, and the number of TMSs for the proteins of each family is indicated below the average sequence length within each bar.

The sequences of homologous proteins retrieved by PSI-BLAST can be useful to perform additional analyses (e.g. construction of multiple alignments, plotting average hydropathy, amphipathicity and similarity, construction of phylogenetic trees, etc.). For this reason, PhyST provides a file with all the sequences of the homologues retrieved in FASTA format. If specified by the user, PhyST can run any multiple alignment program, but the default program is ClustalΩ. The details regarding these functions can be found on the program's website (http://biotools.tcdb.org/barphyst.html).

Basic Usage

From the Unix terminal, the program runs by typing ./physt.py, followed by the TC number of the family (i.e. 1.E.1). The user can customize several parameters including E-value, number of iterations, the minimal sequence length of PSI-BLAST hits relative to the query protein, etc. The program is available for download in the biotools section of TCDB as an addition to the BioV suite of programs [Reddy and Saier, 2012; Zhai and Saier, 2001a, b]. Furthermore, the website offers a tutorial with several examples illustrating usage of the command line options with links to sample output files.

PhyST is also able to process other types of protein families that are not necessarily listed in TCDB. For this purpose, the user needs to provide one or more files with sequences in FASTA format, and each file will be used by PhyST to extract data it normally processes for any TCDB family. However, if more than one protein is chosen to represent a family, it must first be established that all such proteins are homologous [Park and Saier, 1996].

A sample results page is provided in the supplementary materials for a holin family (1.E.5 which is equivalent to inputting 1.E.5.1.1) (online suppl. appendix 1; for all online suppl. material, see www.karger.com/doi/10.1159/000448040). This presents a visual representation of the data returned by physt.py.

Acknowledgements

We thank Professor Tsai-Tien Tseng and Arturo Medrano-Soto for assistance with the description of the methods and manuscript preparation. The work reported in this paper was supported by NIH grants GM077402 and GM109895 to M.H.S.

References

Catalao MJ, Gil F, Moniz-Pereira J, Sao-Jose C, Pimentel M: Diversity in bacterial lysis systems: bacteriophages show the way. FEMS Microbiol Rev 2013;37:554–571.

Gao Y, Feng X, Xian M, Wang Q, Zhao G: Inducible cell lysis systems in microbial production of bio-based chemicals. Appl Microbiol Biotechnol 2013;97:7121–7129.

Ortega AD, Gonzalo-Asensio J, Garcia-del Portillo F: Dynamics of *Salmonella* small RNA expression in non-growing bacteria located inside eukaryotic cells. RNA Biol 2012;9:469–488.

Park JH, Saier MH Jr: Phylogenetic characterization of the mip family of transmembrane channel proteins. J Membr Biol 1996;153:171–180.

Reddy BL, Saier MH Jr: Topological and phylogenetic analyses of bacterial holin families and superfamilies. Biochim Biophys Acta 2013;1828:2654–2671.

Reddy VS, Saier MH Jr: BioV suite – a collection of programs for the study of transport protein evolution. FEBS J 2012;279:2036–2046.

Saier MH Jr, Reddy BL: Holins in bacteria, eukaryotes, and archaea: multifunctional xenologues with potential biotechnological and biomedical applications. J Bacteriol 2015;197:7–17.

Saier MH Jr, Reddy VS, Tamang DG, Vastermark A: The transporter classification database. Nucleic Acids Res 2014;42:D251–D258.

Saier MH Jr, Reddy VS, Tsu BV, Ahmed MS, Li C, Moreno-Hagelsieb G: The Transporter Classification Database (TCDB): recent advances. Nucleic Acids Res 2016;44:D372–D379.

Saier MH Jr, Yen MR, Noto K, Tamang DG, Elkan C: The Transporter Classification Database: recent advances. Nucleic Acids Res 2009;37:D274–D278.

Shi Y, Sun J: Current advance in the topological structure and function of holin encoded by bacteriophage lambda – a review (in Chinese). Wei Sheng Wu Xue Bao 2012;52:141–145.

Tusnady GE, Simon I: The hmmtop transmembrane topology prediction server. Bioinformatics 2001;17:849–850.

Wang IN, Deaton J, Young R: Sizing the holin lesion with an endolysin-beta-galactosidase fusion. J Bacteriol 2003;185:779–787.

Wang IN, Smith DL, Young R: Holins: The protein clocks of bacteriophage infections. Annu Rev Microbiol 2000;54:799–825.

Yan J, Fan X, Xie J: Emerging biomedicines based on bacteriophages. Crit Rev Eukaryot Gene Expr 2013;23:299–308.

Young R: Bacteriophage lysis: mechanism and regulation. Microbiol Rev 1992;56:430–481.

Young R: Phage lysis: do we have the hole story yet? Curr Opin Microbiol 2013;16:790–797.

Young R: Phage lysis: three steps, three choices, one outcome. J Microbiol 2014;52:243–258.

Young R, Blasi U: Holins: form and function in bacteriophage lysis. FEMS Microbiol Rev 1995;17:191–205.

Zhai Y, Saier MH Jr: A web-based program for the prediction of average hydropathy, average amphipathicity and average similarity of multiply aligned homologous proteins. J Mol Microbiol Biotechnol 2001a;3:285–286.

Zhai Y, Saier MH Jr: A web-based program (WHAT) for the simultaneous prediction of hydropathy, amphipathicity, secondary structure and transmembrane topology for a single protein sequence. J Mol Microbiol Biotechnol 2001b;3:501–502.