Computational Biology and Bioinformatics
2017; 5(3): 27-35
http://www.sciencepublishinggroup.com/j/cbb
doi: 10.11648/j.cbb.20170503.11
ISSN: 2330-8265 (Print); ISSN: 2330-8281 (Online)

In Silico Analysis of Occurrence of Tricorn Protease and Its Homologs

Florence Ng'ong'a*, Steven Nyanjom, Fred Wamunyokoli

Department of Biochemistry, School of Biomedical Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

Email address:
fngonga@jkuat.ac.ke (F. Ng'ong'a)
*Corresponding author

To cite this article:
Florence Ng'ong'a, Steven Nyanjom, Fred Wamunyokoli. *In Silico* Analysis of Occurrence of Tricorn Protease and Its Homologs. *Computational Biology and Bioinformatics*. Vol. 5, No. 3, 2017, pp. 27-35. doi: 10.11648/j.cbb.20170503.11

Received: May 30, 2017; **Accepted**: July 24, 2017; **Published**: August 15, 2017

Abstract: Tricorn protease is an archaeal protease acting downstream of the proteasome and together with its interacting aminopeptidases, degrades oligopeptides to free amino acids thus playing an important role in protein turnover. This study reports a wide distribution of tricorn protease and its homologs in archaea and bacteria. The homologs were identified through a combination of PSI-BLAST, orthology clustering and domain predictions. Functionally important sites were identified through multiple sequence alignment conducted by MAFFT v. 7. The aligned sequences were used to predict the phylogenetic relationship of tricorn protease and its homologs using MEGA v. 7. The functional associations of tricorn protease were predicted through STRING network v.10.0. This study identified several tricorn protease homologs in archaea and in all the bacterial phyla complete with β-propeller, PDZ and catalytic domains. However, in eukaryotes, tricorn protease-like homologs seemed limited to viridiplantae, stramenopile and in a basal metazoa and were classified as non-peptidase homologs with unknown functions. Conserved domain architecture retrieval revealed detectable homology of tricorn protease C-terminal half with the carboxyl-terminal proteases with similar PDZ domains. Therefore, this study predicts functional conservation of tricorn core catalytic domain in prokaryotes and given its role in cellular functions, targeting this protein or its functional homologs in prokaryotic pathogens could lead to development of alternative therapeutic agents.

Keywords: Proteolysis, Tricorn Protease, Homologs, *In Silico*

1. Introduction

The degradation of cytosolic proteins is mainly carried out by ATP-dependent proteases employing molecular sieving techniques. These proteases include the proteasomes which are responsible for removing misfolded or unneeded proteins, however, the lengths of these peptides range from 7 to 9 amino acid long thus require further processing for them to be of any use to the cell [1]. Studies have shown that these peptides get rapidly cleared since their accumulation of in the cytosol could interfere with important protein-protein interactions and also their degradation provides amino acids for use in the synthesis of new proteins thus is essential for cell viability [2]. In *Thermoplasma acidophilum*, tricorn protease and its interacting factors, F1, F2, and F3 act *in vivo* downstream of the proteasome degrading proteasomal products to free amino acids for other metabolic processes [3]. Tricorn protease is a hexameric protease of 720 kDa and can assemble into a giant icosahedral capsid, which might serve as the organizing center of a multi-proteolytic complex [4 - 5]. The monomeric 120 kDa polypeptide has a mosaic structure with two open Velcro beta-propeller structures of six and seven blades, a helical bundle, a PDZ domain and an alpha-beta sandwich structure (PDB ID: 1 K32 A). These five domains combine to form one of six sub-units, which further assemble to form 3-2 symmetric core protein. [6 - 7]. Tricorn protease N-terminal domain folds as a six-bladed β-propeller (β6) followed by a seven-bladed β-propeller (β7). The β6 represents a gated exit from while β7 represents a passage route into the catalytic chamber and the PDZ domain is inter-sparsed between two carboxyl-terminal mixed α-β domains C1 and C2 [7]. The nucleophilic serine is positioned at a helix entrance within subdomain C2 and the arrangement of the tetrad active site molecules suggests peptide bond hydrolysis following the classical trypsin-like serine

proteases [7].

Tricorn protease acts as a carboxypeptidase with di- and tripeptidase activity [8] and it seems limited to some archaea and eubacteria thus functional analogues exist. Indeed, tetrahedral aminopeptidase (TET) has been shown to be a complementary protein degradation machinery to tricorn protease in *Pyrococcus horikoshii* and it degrades peptides to free amino acids [9]. In eukaryotes, tripeptidyl peptidase II (TPP II) is a tricorn protease functional analog, and has been shown to act downstream of the proteasome [10] [11] [12], where it is implicated in peptide processing of MHC class I antigens [13] [14]. Since peptides for antigen presentation are generated through the degradation of proteasomal products, the parasite secreted oligopeptidases targeting such peptides would limit host immune response thus aid in immune evasion [15].

Tricorn protease C-terminal region has been shown to have detectable structural homology with several C-terminal processing proteases (CTP) of bacterial and eukaryotic origin [7] for example the eukaryotic D1 protease, which cleaves 8-16 residues peptide from C-terminal of D1 protein allowing light driven assembly of the tetranuclear Mn cluster responsible for photosynthetic water oxidation in PSII [6]. C-terminal processing proteases have been shown to contribute to virulence in some pathogenic bacteria for instance *Staphylococcus aureus* [16]. In *Escherichia coli,* a tail specific protease, also known as periplasmic protease is involved in C-terminal processing of penicillin binding protein 3 and mutations of this gene have resulted in altered cell morphology and increased susceptibility to thermal and osmotic stress due to reduced cell-wall integrity [16]. In *Brucella suis,* CTP has been associated with protecting cells against osmotic pressure, determining cell morphology and survival during acute and chronic phases of infection [17]. This study therefore investigated the occurrence of tricorn protease and its homologs in archaea, bacteria and eukaryotes.

2. Methods

2.1. Blast Identification of Tricorn Protease Homologs

Position specific iterated basic local alignment search tool (PSI-BLAST) (https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE=Proteins&PROGRAM=blastp&RUN_PSIBLAST=on) [18] search was initiated using the *T. acidophillum* tricorn protease (AAC 44621.1) protein sequence against the NCBI's non-redundant (NR) protein database. The search was performed to determine patterns of conservation which aid in the recognition of distant similarities with a cut-off e-value of 1 e-5 for significant matches.

Identification of orthologous groups was done through Ortho MCL database (http://orthomcl.org/orthomcl/) [19] which employs Markov Cluster algorithm to group putative orthologs. The algorithms involved BLASTp comparisons of the query sequences (AAC44621.1) and sequences from other eukaryotic genomes (fungi, plants, animals, protists and protozoans). The between species and within species relationships of the putative orthologs were identified by reciprocal best similarity pairs with a cut-off e-value of 1 e-5.

Protein homology through domain architecture was determined through Conserved Domain Architecture Retrieval Tool (CDART) (https://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi) [20]. The hits obtained represented proteins with significant structural similarities to the structure of tricorn protease.

2.2. Multiple Sequence Alignment

Multiple sequence alignment was performed by MAFFT version 7.0 (Multiple alignment program for amino acid sequences) at http://mafft.cbrc.jp/alignment/server/index.html [21] integrating the BLOSUM62 matrix which scores alignments between evolutionarily divergent protein sequences. MAFFT version 7.0 default gap penalty (1.53) and a cut-off e-value (1 e-5) were used to obtain only significant matches.

2.3. Protein-Protein Interactions

Protein-protein interactions of the tricorn protease with other proteins were analyzed through STRING database version 10.0 (Search Tool for Retrieval of Interacting Proteins) (http://string-db.org) [22] to determine their associations in-terms of co-expression, text-mining, protein homology and gene neighborhood. The searches were performed at high confidence levels ($\geq$ 0.7) and included only first shell interactions. Protein-protein interaction enrichment p-value of $\leq$ 0.05 was considered significant.

2.4. Constructing the Phylogenetic Tree of Tricorn Protease and Homologs

The amino acid sequences of tricorn protease and its homologs were aligned using MAFFT version 7.0 at http://mafft.cbrc.jp/alignment/server/index.html [21]. Phylogenetic analysis of aligned amino acid sequences was performed using the MEGA version 7.0 package [23]. Phylogenetic tree was constructed using Maximum Likelihood method [24] and bootstrap resampling (1000 replicates) was used to assess the robustness of the groupings obtained.

3. Results

3.1. Blast Identification of Tricorn Protease Homologs

Protein homology through sequence similarity conducted through BLAST homology searches predicted 159 sequences in archaea, 62,710 sequences in bacteria, 609 sequences in plants, 302 sequences in protists, 128 sequences in animals, 85 sequences in fungi and 9 sequences in viruses. The archaeon sequences were mainly annotated as tricorn proteases and tricorn-like proteases while the bacterial hits

included tricorn protease homologs and C-terminal proteases. The hits in plants included D 1 processing proteases, C-terminal proteases, peptidase S 41 family, proteases, tricorn protease homologs and predicted proteins. The hits in animals included protease m1 Zn metalloproteases, Ctp A-like serine proteases, membrane alanyl aminopeptidase, glutamyl aminopeptidase, endoplasmic reticulum aminopeptidase, aminopeptidase N, aminopeptidase Q, PDZ containing proteins, WD40-like containing proteins and C-terminal proteases. The fungal hits included tricorn protease N- terminal domain (6-bladed beta propeller domain) and tricorn protease domain II (7- bladed beta propeller domain) while all the hits in viruses were hypothetical proteins with WD40 domains. Further analysis of tricorn protease-like sequences through PSI-BLAST displayed low sequence similarity among the homologs with the highest similarity being 54%, tricorn protease, *Thermoplasma volcanium,* a close relative of *Thermoplasma aciodophillum.* Other archaeon hits, example, tricorn protease from *Ferroplasma acidarmanus* had 39% sequence identity to the query sequence (Table 1).

Table 1. *BLAST homologies between tricorn protease, Thermoplasma acidophilum and its homologs.*

Description	Max score	E value	Identity	Accession
Tricorn protease, *Thermoplasma volcanium*	1138	0	54%	WP_010917153.1
Hypothetical protein, *Thermoplasmatales archaeon*	776	0	42%	EQB68734.1
Tricorn core peptidase, *Cuniculiplasma divulgatum*	774	0	42%	SIM28710.1
Tricorn protease, *Ferroplasma acidarmanus*	756	0	39%	WP_019841748.1
Peptidase S41, *Picrophilus torridus*	753	0	41%	WP_011177615.1
Peptidase S41, *Thermoproteus uzoniensis*	656	0	37%	WP_013679574.1
Tricorn protease, *Pyrobaculum aerophilum*	643	0	36%	AAL63388.1
peptidase S41, *Sulfolobus solfataricus*	604	0	37%	WP_009989776.1
Hypothetical protein, *Scytonema hofmannii*	602	0	34%	WP_017743915.1
Peptidase S41, *Armatimonadetes bacterium*	562	1.00 E-179	34%	WP_072260369.1
Peptidase S41, *Streptosporangium roseum*	561	2.00 E-179	34%	WP_012890710.1
Tricorn protease, *Nonomuraea solani*	559	7.00 E-179	33%	SEH01969.1
Tricorn protease homolog, *Candidatus C. acidaminovorans*	557	5.00 E-178	33%	WP_015423959.1
Peptidase S41, *Herbidospora cretacea*	556	1.00 E-177	33%	WP_061299287.1
Hypothetical protein, *Longispora albida*	554	1.00 E-176	34%	WP_018350392.1
Periplasmic protease, *Chlorobi bacterium*	550	9.00 E-175	34%	KXK57595.1
Peptidase S41, *Petrotoga mobilis*	541	8.00 E-172	34%	KUK16346.1
Hypothetical protein, *Kutzneria sp.*	443	3.00 E-135	31%	WP_052396111.1
Tricorn protease, *Lentzea waywayandensis*	429	1.00 E-129	30%	SFR27171.1
Tricorn protease, *Actinokineospora terrae*	429	2.00 E-129	30%	SES10996.1
Tricorn protease, *Alloactinosynnema album*	429	3.00 E-129	29%	SDI48145.1
Tricorn protease, *Streptomyces davawensis*	428	5.00 E-129	29%	WP_015660463.1
Predicted protein, *Micromonas pusilla*	257	1.00 E-67	25%	XP_003064769.1
Tricorn protease homolog, *Ostreococcus tauri*	200	3.00 E-49	29%	XP_003080833.1
Protease, *Bathycoccus prasinos*	164	6.00 E-38	29%	XP_007515703.1
Predicted protein, *Phaeodactylum tricornutum*	157	1.00 E-35	27%	XP_002181061.1
Tricorn protease-like 1, *Symbiodinium microadriaticum*	153	3.00 E-34	34%	OLP82703.1
Predicted protein, *Thalassiosira pseudonana*	128	7.00 E-27	36%	XP_002295534.1
Hypothetical protein, *Nematostella vectensis*	117	1.00 E-26	33%	XP_001617415.1
Protease-related protein, *Vibrio cholera*	110	7.00 E-24	30%	vcho|NP_232446
Carboxyl-terminal protease, *Clostridium perfringens*	110	9.00 E-07	31%	cper|YP_694758
Carboxyl-terminal protease, *Thermotoga maritima*	110	1.00 E-04	30%	tmar|NP_228556
Carboxyl-terminal protease, *Aquifex aeolicus*	110	1.00 E-04	28%	aaeo|NP_213546
Carboxyl-terminal protease, *Dehalococcoides ethenogenes*	108	1.00 E-03	27%	deth|YP_181110
Carboxyl-terminal protease, *Chlorobium tepidum*	107	2.00 E-03	26%	ctep|NP_661945
Carboxyl-terminal protease, *Staphylococcus aureus*	107	1.10 E-02	29%	saur|NP_371944
Tail specific periplasmic protease, *Rickettsia typhi*	105	1.40 E-01	27%	rtyp|YP_067184
Japonica cultivar-group, *Oryza sativa*	102	2.90 E-01	26%	osat|NP_001051846

The alignments were obtained from PSI-BLAST with tricorn protease (AAC44621.1) running against NCBI's non-redundant protein database. The results indicate low sequence similarity among the tricorn protease homologs.

The BLAST results also revealed a wide distribution of tricorn protease homologs across the bacterial genome with significant hits found in all bacterial phyla. These hits included hypothetical proteins in *Aminicenantes bacterium* (Aminicenantes) and in *Scytonema hofmanni* (Cyanobacteria), Tricorn protease homologs in Bacteriodes fragilis (Bacteriodetes), *Chloroherpeton thalassium*, Chlorobi, *Roseiflexus sp.,* Chloroflexi, *Sulfobacillus acidophilus,* Firmicutes, *Streptomyces coelicolar*, Actinobacteria, *Vibrio cholerae*, Proteobacteria and MdsD protein in *Gemmata obscuriglobus*, Planctomycetes (Table 1). The sequence similarity in the bacterial homologs was remarkably low and was within the range of 29-34% but with significant E-values (Table 1). The similarities of these bacterial homologs with the archaeon tricorn protease were also reflected in domain composition and organization

(Figure 1).

Significant hits were also present in some eukaryotes; example viridiplantae, tricorn protease, *Ostreococcus tauri,* protease, *Bathycoccus prasinos* and predicted protein, *Micromonas pusilla*. Predicted proteins in *Thalasiossira pseudonana* and in *Phaeododactylum tricornutum* had significant matches with the archaeon tricorn protease as well as hypothetical protein in *Nematostella vectensis,* a basal metazoa (Table 1). In-terms of domain organization, the viridiplantae hits had similar domains composition and domain organization to the archaeon tricorn protease while the stramenopiles (*Thalasiossira pseudonana* and *Phaeododactylum tricornutum*) and the basal metazoan hits lacked the N-terminal domain (Figure 1).

Orthology predictions through OrthoMCL database categorized the hits into three ortholog groups where OG 5_164838 included the archaeon tricorn protease, bacterial and eukaryotic peptidase homologs. Ortholog group OG5_204488 included non-peptidase homologs in *Thalassiosira pseudonana*, *Ostreococcus tauri.* Ortholog group OG5_130275 included the carboxyl-terminal proteases, carboxyl-terminal processing proteases and tail specific proteases widely distributed in other bacteria and higher eukaryotes (Table 2 and Appendix 1).

Tricorn protease structural similarities through Conserved Domain Architecture Retrieval Tool (CDART) showed structural similarities with 4 C 2 EAc (Carboxyl-terminal processing protease), 1 dfc 6 ac 4 (Photosystem II D1 protease), dj 7 xa (Retinal binding protein), 4 QL 6 Ab (Carboxyl-terminal protease) and 3 K 50 Aa (Putative S41 protease). Detectable homology with these proteins occurred within tricorn protease catalytic domain (C-terminal region) (Figure 2).

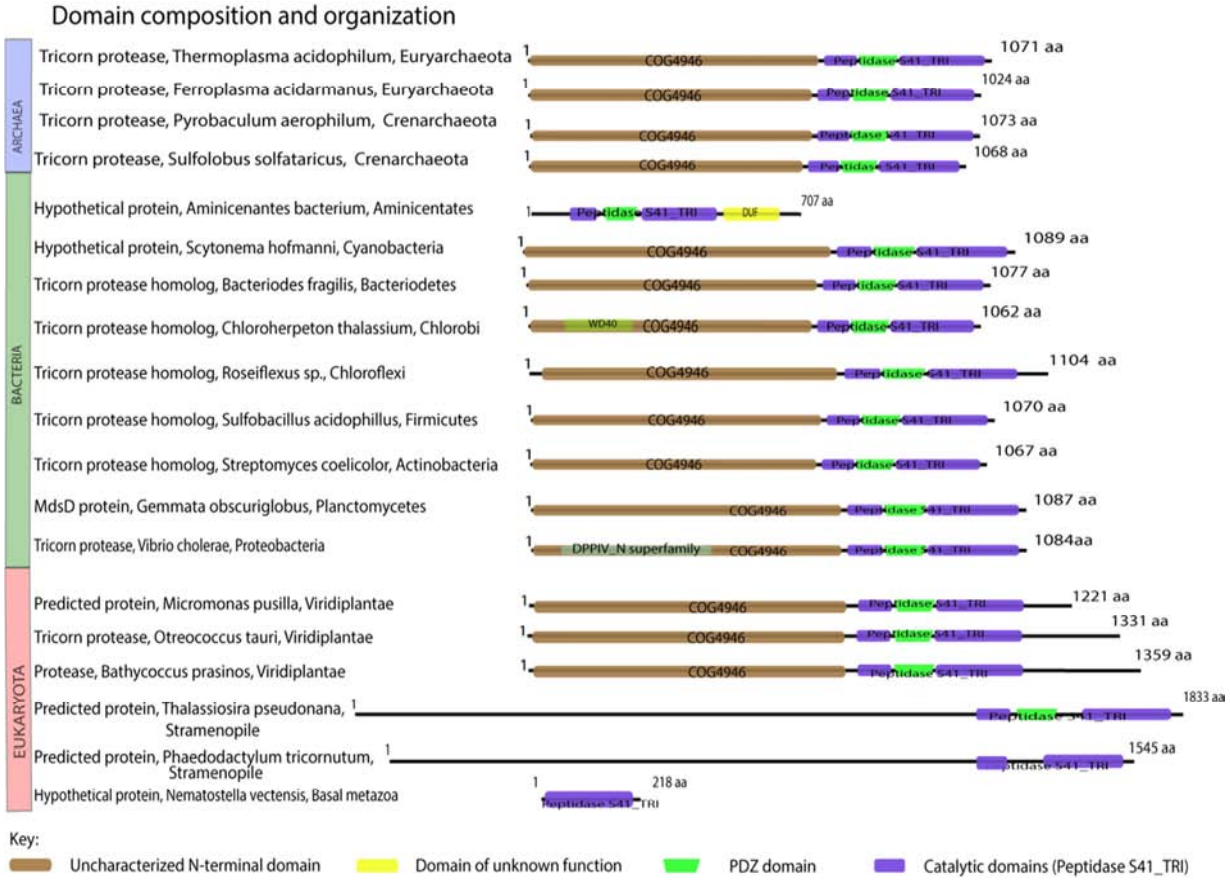

Figure 1. *The domain composition and organization of tricorn protease and its homologs. The similarity in C-terminal half of the proteins is remarkably high with the PDZ domain intersparsing the catalytic domain 1 and 2 with PDZ domain seemingly lacking in the predicted protein in P. tricornutum and the hypothetical protein in N. vectensis.*

Table 2. *Identified ortholog groups from OrthoMCL database.*

Group ID	Accession	Taxon name	P. L (aa)	Product	E-value
OG5_164838	tvol|NP_111414	*Thermoplasma volcanium*	1030	Tricorn protease	0.00 E+00
	ssol|NP_343486	*Sulfolobus solfataricus*	1068	Tricorn protease	0.00 E+00
	otau|0800010263	*Ostreococcus tauri*	1333	N/A	2.00 E-50
	micr|ACO65681	*Micromonas sp. RCC299*	1249	predicted protein	2.00 E-50
	nvec|fgenesh1	*Nematostella vectensis*	219	N/A	4.00 E-27
	vcho|AOF84260.1	*Vibrio cholerae*	1084	Tricorn protease	7.00 E-24
OG5_204488	micr|ACO70435	*Micromonas sp. RCC299*	1132	predicted protein	1.00 E-45

Group ID	Accession	Taxon name	P. L (aa)	Product	E-value
	tpse|fgenesh1	*Thalassiosira pseudonana*	1834	N/A	4.00 E-28
	otau|1100010345	*Ostreococcus tauri*	1468	N/A	9.00 E-24
OG5_130275	cper|YP_694758	*Clostridium perfringens*	428	carboxyl-terminal protease	9.00 E-07
	tmar|NP_228556	*Thermotoga maritima*	402	carboxyl-terminal protease	1.00 E-04
	aaeo|NP_213546	*Aquifex aeolicus*	408	carboxyl-terminal protease	1.00 E-04
	deth|YP_181110	*Dehalococcoides ethenogenes*	377	carboxyl-terminal protease	1.00 E-03
	ctep|NP_661945	*Chlorobium tepidum*	675	carboxyl-terminal protease	2.00 E-03
	rtyp|YP_067184	*Rickettsia typhi*	442	tail specific periplasmic protease precursor	1.40 E-01
	vcho|NP_231137	*Vibrio cholerae*	665	carboxyl-terminal protease	1.50 E-01
	cjej|YP_002343943	*Campylobacter jejuni*	444	putative secreted protease	2.60 E-01
	saur|NP_371944	*Staphylococcus aureus*	496	carboxyl-terminal processing proteinase	1.10 E-02
	bmal|YP_104685	*Burkholderia mallei*	524	carboxyl-terminal protease	4.00 E-01
	bpse|YP_332075.1	*Burkholderia pseudomallei*	530	C-terminal processing protease-3	3.60 E-01
	rsol|NP_518473	*Ralstonia solanacearum*	549	putative carboxyl-terminal processing protease	6.30 E-01
	syne|NP_898059	*Synechococcus*	425	putative carboxyl-terminal processing protease	9.00 E-01
	atha|NP_191327	*Arabidopsis thaliana*	516	peptidase S41 family protein	1.60 E+00
	bant|YP_022076	*Bacillus anthracis*	469	carboxyl-terminal protease	1.90 E+00
	cbot|YP_001788718.1	*Clostridium botulinum*	401	carboxyl-terminal protease	1.90 E+00
	rpro|NP_220614.1	*Rickettsia prowazekii*	441	tail-specific protease precursor	1.60 E+00
	lmon|NP_465376	*Listeria monocytogenes*	496	hypothetical protein	2.30 E+00
	gsul|NP_952023	*Geobacter sulfurreducens*	450	carboxyl-terminal processing protease	3.00 E+00
	atum|NP_355704	*Agrobacterium tumefaciens*	442	carboxyl-terminal protease	3.90 E+00

OG5_164838 includes archaeon tricorn protease, its bacterial and eukaryotic homologs having the Peptidase_S41_TRI domain. The OG5_204488 group includes the non-peptidase homologs with peptidase_S41_TRI domain. The OG_130275 group includes the tail specific proteases, carboxyl-terminal processing proteases with the peptidase_S41_CPP domain. Key: P. L (aa): Protein length (Amino acid).

3.2. Multiple Sequence Alignment

Multiple sequence alignment revealed conservation of active site tetrad in close homologs with major variation occurring in the hypothetical protein in *Nematostella vectensis* which seemed to lack both the N-terminal domain and the catalytic domain 1 (C1) (Figure 2). The active site nucleophile S965 (position corresponding to tricorn protease in *Thermoplasma acidophilum*) was conserved in all the homologs (Figure 3).

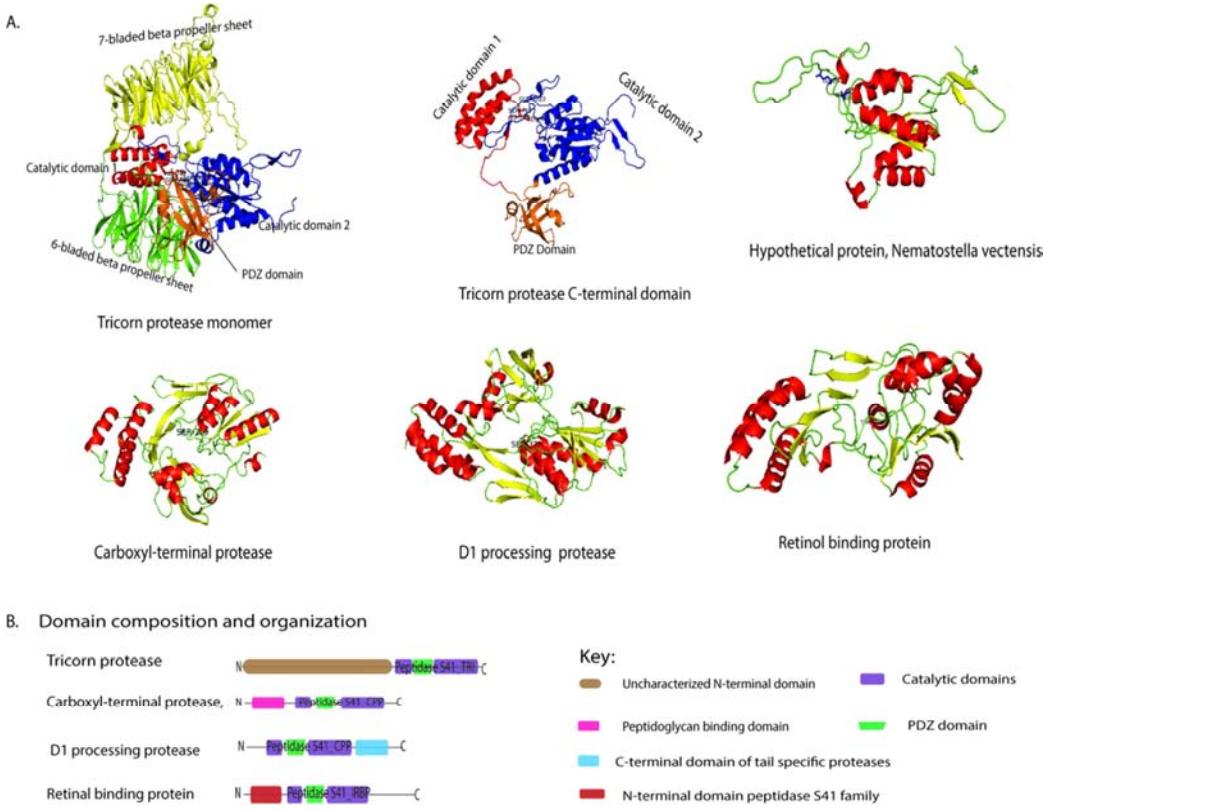

Figure 2. *A. The 3 D structure of tricorn protease and its structural analogues from CDART. Tricorn protease monomer was colored by domain while the structural analogues were coloured by secondary sructures. B: The domain composition and organization of tricorn protease and its structural analogues where similarities occurred in the conserved catalytic domain.*

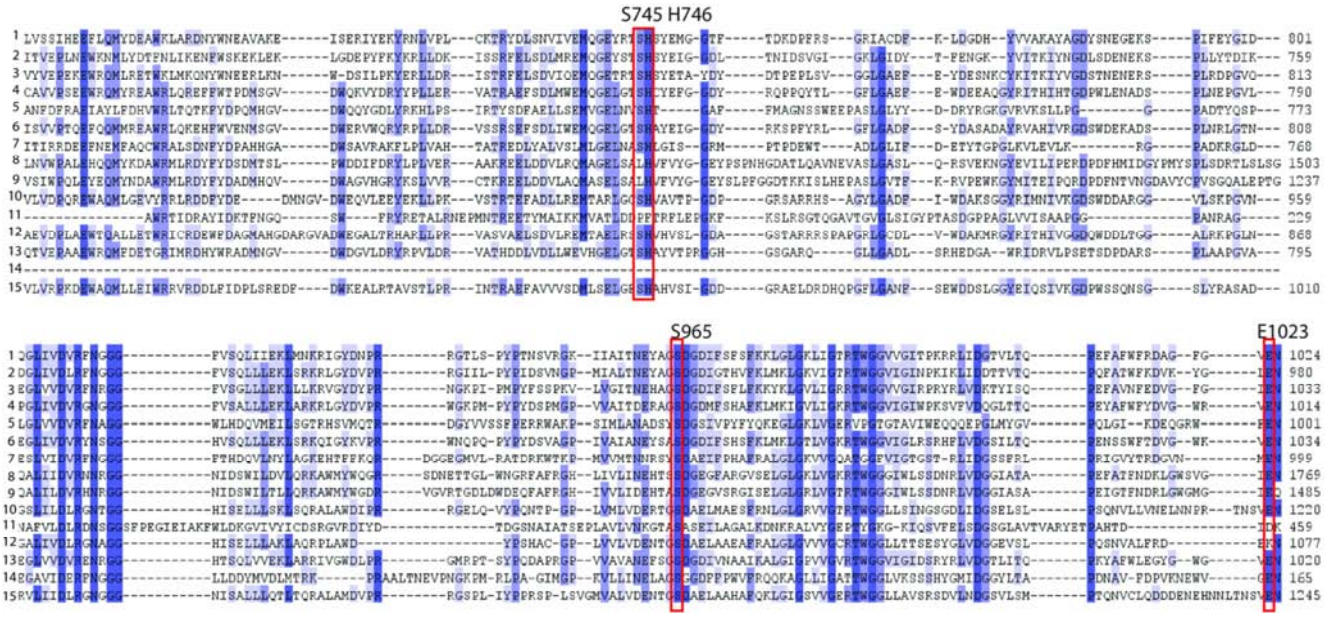

Figure 3. *Multiple sequence alignment of catalytic domain 1 (C 1) and catalytic domain 2 (C 2) of tricorn protease and its homologs. The active site tetrad S 745, H 746 are located in C 1 while S 965 and E 1023 are located in C 2 domain (as shown in the red boxes) with the positions corresponding to tricorn protease in Thermoplasma acidophilum. The sequences were retrieved from GenBank: 1; Tricorn protease, Thermoplasma acidophilum, 2; tricorn protease, Ferroplasma acidarmanus, 3; Tricorn protease, Sulfolobus salfataricus, 4; Tricorn protease, Pyrobaculum aerophilum, 1; hypothetical protein, Aminicenantes bacterium 5; tricorn protease, Vibrio cholerae, 6; tricorn protease, Streptomyces coelicor, 7; Hypothetical protein, Scytonema hofmanni, 8; MdsD protein, Gemmata obscuriglobus, 9; Predicted protein, Thalasiossira pseudonana, 10; Predicted protein, Phaedodactylum tricornutum, 11; Tricorn protease, Ostreococcus tauri, 12; D 1 processing protease, Arabidopsis thaliana, 13; Predicted protein, Micromonas pusilla, 14; Hypothetical protein, Nematostella vectensis, 1; Protease, Bathycoccus prasinos.*

3.3. Protein-Protein Interactions

The network of protein-protein interactions of tricorn protease, *Thermoplasma acidophilum* through STRING database revealed 11 nodes with 12 edges and average node degree of 2.18. The clustering coefficient was 0.273 with a protein-protein interaction enrichment p-value of 1.52 e-06. However, only 2 edges were expected thus significantly more interactions than expected. Based on gene ontology annotations, the network's major functional enrichments was proteolysis (GO: 0051603) as biological process with an e-value of 1.47 e-06 with cellular component GO: 0005737) being cytoplasm. The major KEGG pathway for the network was proteasome (03050) (Figure 4).

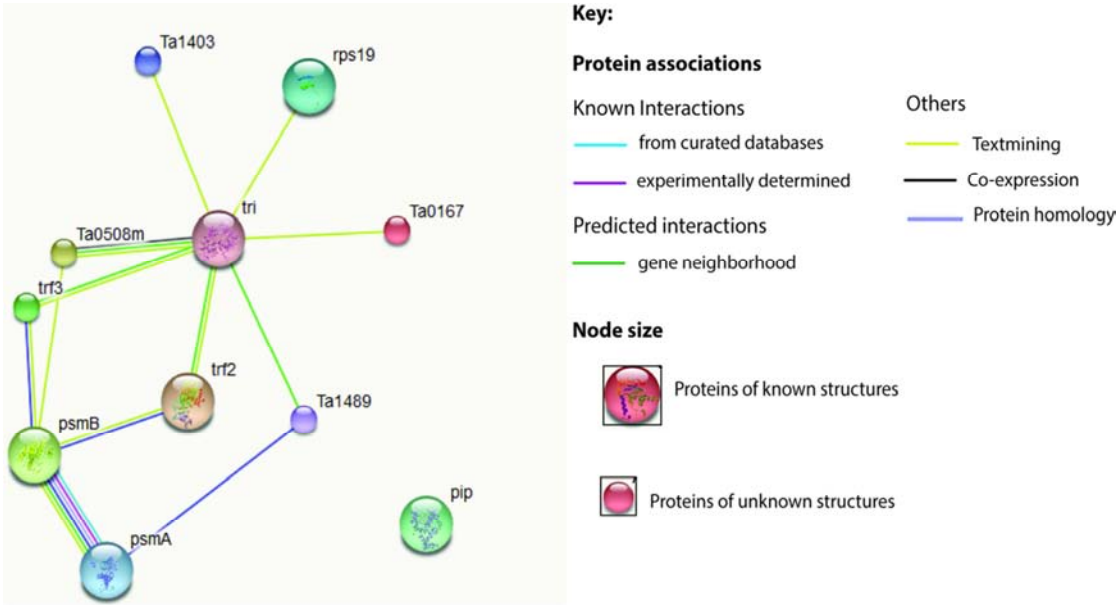

Figure 4. *Protein-protein interaction network of tricorn protease, Thermoplasma acidophillum, in STRING database. Key: tri: Tricorn protease, pip: proline iminopeptidase, trf 2: Tricorn protease interacting factor F 2, trf 3: tricorn protease interacting factor F 3, Tal 1403: thermopsin, precursor, Ta 0167: thermopsin precursor, rps 19: 30 S ribosomal protein S 19, Ta 0508: cell division protein FtsZ; Ta 1489: GTP-binding protein, psmA: proteasome subunit alpha, psmB: proteasome subunit beta.*

3.4. Constructing the Phylogenetic Tree of Tricorn Protease and Homologs

The phylogenetic tree displayed inferred evolutionary relationship with nodes indicating separate paths. The node value given as percentages represent the measure of support for the node where 100 represents maximal support, that is, sequences to the right node cluster together to the exclusion of any other. The phylogenetic tree showed two main evolutionary nodes where tricorn protease clustered with the homologs having the Peptidase_S 41_TRI domain while the carboxyl-terminal proteases having Peptidase_S 41_CPP domain also clustered together (Figure 5).

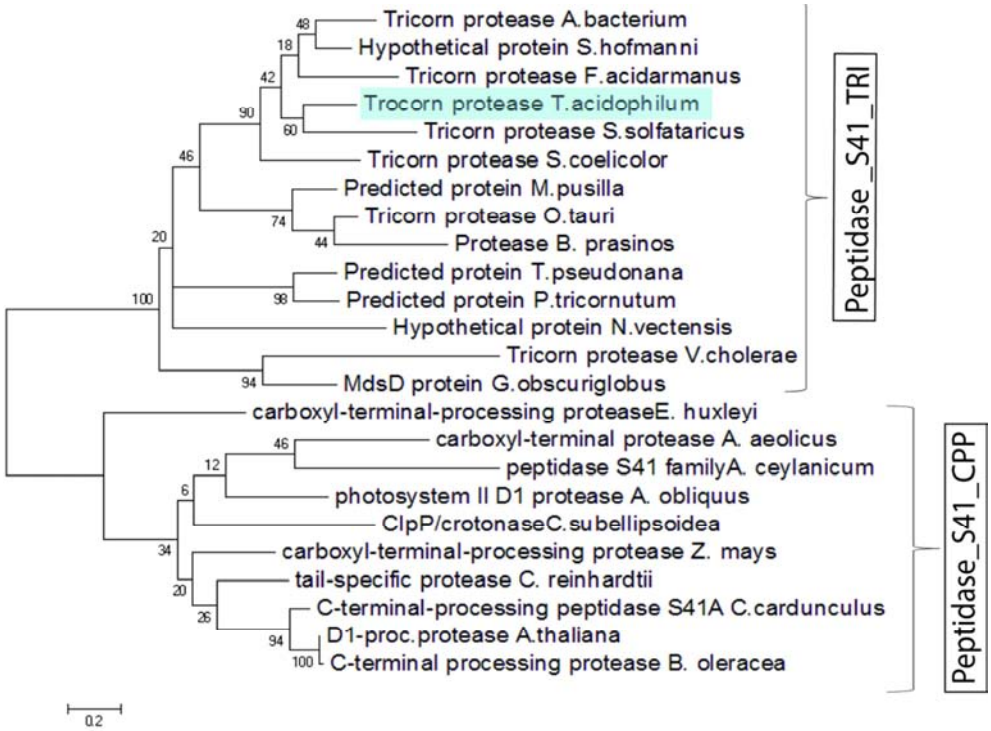

Figure 5. *Phylogenetic analysis of archaeon tricorn protease, its bacterial and eukaryotic homologs and the carboxyl-terminal protease functional analogues. The evolutionary history was inferred by using the Maximum Likelihood method. The bootstrap scores are provided at each branch of the tree. The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 82 positions in the final dataset. In general, the the archaeon tricorn protease closely evolved with its bacterial homologs as indicated with a bootstrap score of 100%.*

4. Discussion

This study predicts a wide distribution of tricorn protease within archaeon with presence of bacterial homologs and therefore could indicate the importance of this protein in cellular functions. Proteasomes are ubiquitous in archaea which also seem to have AAA ATPases closely related to 19 S regulatory proteins in eukaryotes [3]. However, studies have shown that the proteasome is dispensable under normal conditions in *Thermoplasma acidophilum.* [25]. This is could only show the existence of variant pathways that contribute to the pool of oligopeptides in the cytosol thus the well-developed tricorn protease machinery in these organisms [26]. Studies have reported that in bacteria, the occurrence of a genuine proteasome is limited to actinomycetes with majority of the bacteria having the ATP- dependent proteasome-related HsIUV [8].

The identification criteria used in this study did not find tricorn protease or its homologs in methanogens, desulforococales, nanoarchaeota and korarchaeota which have been shown to have complementary pathways [9]. However, this study predicts the existence of tricorn protease/ tricorn protease homologs in all the bacterial phyla with complete similar functional domains (beta-propeller, PDZ and catalytic domains). Indeed, some of these homologs have been characterized [27]. In bacterial species where tricorn gene seemed missing for example *E. coli*, the carboxyl-terminal protease functional analog was present. The beta-propeller domain was however not well conserved in the carboxyl-terminal protease functional analogs, a fact that has also been reported in other studies [27]. This shows functional conservation of tricorn core catalytic domain since the beta-propeller domains mainly serve as gated exit for substrates into and products out of the catalytic chamber [7].

Other studies have shown that in eukaryotes, detecting tricorn analogues might be difficult since tricorn is built from five folding domains thus it is possible that tricorn analogues might assemble non covalently from different gene products and could have additional functionalities [6]. Therefore, based on sequence and structural homology, this study predicted tricorn protease homologs in few eukaryotic groups, but they seemed to lack the beta-propeller domain while others also lacked PDZ domain. An

isolated tricorn protease C 2 subdomain, a subdomain associated with the catalytic residue also existed in a basal metazoa, *Nematostella vectensis.* Other eukaryotes like fungi were shown to have tricorn protease beta propeller domains which also seemed to occur in kinetoplastids. These organisms also seemed to have genes encoding homologs of tricorn protease interacting factors which cooperate with tricorn in degrading proteasomal products [28] [29]. This further raises the possibility of tricorn analogues assembling from different gene products in eukaryotes.

5. Conclusion

Given the predicted wide distribution of tricorn protease and its homologs archaea and bacteria, this protease machinery could have been conserved during evolution thus is essential in cellular protein degradation. Therefore, targeting it or its functional homologs in pathogens could lead to development of alternative therapeutic agents.

Acknowledgements

This work was supported by the Jomo Kenyatta University of Agriculture and Technology, Research Production and Extension division (JKU/2/4/RP/139) and National Commission for Science, Technology and Innovation (NACOSTI/RCD/ST & I 5th CALL PhD/145).

Appendix

Table 3. *Occurrence of tricorn protease homologs in Plants.*

PLANTS	Accession number	Type of hit	CD ID	Domain
D 1 protease precursor, *Triticum aestivum*	AF 487525_1	specific	Cd 07560	Peptidase_S 41_CPP
C-terminal protease, *Spinacia oleracea*	BAA 09134.1	specific	Cd 07560	Peptidase_S 41_CPP
Carboxyl-terminal-processing protease, *Gossypium arboreum*	KHG 11574.1	specific	Cd 07560	Peptidase_S 41_CPP
Photosystem II D 1 protease, *Acutodesmus obliquus*	AAC 49799.1	specific	Cd 07560	Peptidase_S 41_CPP
carboxyl-terminal processing protease, *Monoraphidium neglectum*	XP_013900651.1	specific	Cd 07560	Peptidase_S 41_CPP
Photosystem II D 1 protease, *Micromonas pusilla*	EEH 59872.1	specific	Cd 07560	Peptidase_S 41_CPP
Carboxyl-terminal-processing protease, *Glycine soja*	KHN 01016.1	specific	Cd 07560	Peptidase_S 41_CPP
Carboxyl-terminal-processing protease, *Morus notabilis*	XP_010103469.1	specific	Cd 07560	Peptidase_S 41_CPP
Putative protease, *Oryza sativa Japonica*	BAD 61578.1	specific	Cd 07560	Peptidase_S 41_CPP
Hypothetical protein, *Chlorella variabilis*	EFN 57183.1	specific	Cd 07560	Peptidase_S 41_CPP
Carboxyl-terminal-processing peptidase, *Citrus sinensis*	XP_006493999.1	specific	Cd 07560	Peptidase_S 41_CPP
Peptidase S 41 family protein, *Klebsormidium flaccidum*	GAQ 77612.1	specific	Cd 07560	Peptidase_S 41_CPP
Hypothetical protein, *Phaseolus vulgaris*	ESW 16143.1	specific	Cd 07560	Peptidase_S 41_CPP
Hypothetical protein, *Populus trichocarpa*	EEE 89071.2	specific	Cd 07560	Peptidase_S 41_CPP
C-terminal-processing peptidase S 41 A, *Cynara cardunculus*	KVI 00398.1	specific	Cd 07560	Peptidase_S 41_CPP
Unknown, *Zea mays*	CAN 30696.1	specific	Cd 00988	PDZ_CTP_protease
		superfamily	C l02526	Peptidase_S 41 superfamily
Unknown, *Picea sitchensis*	ABR 16649.1	specific	Cd 07560	Peptidase_S 41_CPP
ClpP/crotonase, *Coccomyxa subellipsoidea*	XP_005648181.1	specific	Cd 07560	Peptidase_S 41_CPP
Hypothetical protein, *Sorghum bicolor*	EES 01215.1	superfamily	C l02526	Peptidase_S 41 superfamily
		specific	Cd 00988	PDZ_CTP_protease
Peptidase S 41 family protein isoform, *Theobroma cacao*	EOY 06248.1	specific	Cd 07560	Peptidase_S 41_CPP
Hypothetical protein, *Prunus persica*	EMJ 23544.1	specific	Cd 07560	Peptidase_S 41_CPP
Hypothetical protein, *Selaginella moellendorffii*	EFJ 13574.1	specific	Cd 07560	Peptidase_S 41_CPP
Predicted protein, *Physcomitrella patens*	EDQ 59224.1	specific	Cd 07560	Peptidase_S 41_CPP
Hypothetical protein, Capsella rubella	XP_006285651.1	specific	Cd 07560	Peptidase_S 41_CPP
D 1 protease precursor, *Nicotiana plumbaginifolia*	AF 487528_1	specific	Cd 07560	Peptidase_S 41_CPP
D 1-processing protease, partial, *Arabidopsis thaliana*	CAA 10694.1	specific	Cd 07560	Peptidase_S 41_CPP
Carboxyl-terminal-processing protease, *Zea mays*	NP_001148747.1	specific	Cd 07560	Peptidase_S 41_CPP
Serine protease, *Cicer arietinum*	AIA 26575.1	specific	Cd 07560	Peptidase_S 41_CPP
D 1 protease, *Genlisea aurea*	EPS 73244.1	specific	Cd 07560	Peptidase_S 41_CPP
Carboxyl-terminal-processing protease, *Zostera marina*	KMZ 68881.1	specific	Cd 07560	Peptidase_S 41_CPP
Carboxyl-terminal-processing protease, *Auxenochlorella protothecoides*	KFM 23854.1	specific	Cd 07560	Peptidase_S 41_CPP
Carboxyl-terminal-processing protease, *Aegilops tauschii*	EMT 06506.1	specific	Cd 07560	Peptidase_S 41_CPP
Protease, putative, *Ricinus communis*	EEF 35650.1	specific	Cd 07560	Peptidase_S 41_CPP
Tail-specific protease, *Chlamydomonas reinhardtii*	EDP 07844.1	specific	Cd 07560	Peptidase_S 41_CPP
C-terminal processing protease, *Brassica oleracea*	ABD 65160.1	specific	Cd 07560	Peptidase_S 41_CPP

*CD ID: Conserved Domain Identity

References

[1] M. Groll and T. Clausen, "Molecular shredders: How proteasomes fulfill their role," *Curr. Opin. Struct. Biol.*, vol. 13, no. 6, pp. 665–673, 2003.

[2] A. Paugam, A.-L. Bulteau, J. Dupouy-Camet, C. Creuzet, and B. Friguet, "Characterization and role of protozoan parasite proteasomes," *Trends Parasitol.*, vol. 19, no. 2, pp. 55–59, Feb. 2003.

[3] N. Tamura, F. Lottspeich, W. Baumeister, and T. Tamura, "The role of tricorn protease and its aminopeptidase-interacting factors in cellular protein degradation.," *Cell*, vol. 95, no. 5, pp. 637–48, 1998.

[4] T. Tamura, N. Tamura, F. Lottspeich, and W. Baumeister, "Tricorn protease (TRI) interacting factor 1 from Thermoplasma acidophilum is a proline iminopeptidase," *FEBS Lett.*, vol. 398, no. 1, pp. 101–105, 1996.

[5] J. Walz, T. Tamura, N. Tamura, R. Grimm, W. Baumeister, and A. J. Koster, "Tricorn protease exists as an icosahedral supermolecule in vivo," *Mol. Cell*, vol. 1, no. 1, pp. 59–65, 1997.

[6] H. Brandstetter, J. S. Kim, M. Groll, and R. Huber, "Crystal structure of the tricorn protease reveals a protein disassembly line.," *Nature*, vol. 414, no. 6862, pp. 466–470, 2001.

[7] J. S. Kim *et al.*, "Navigation inside a protease: substrate selection and product exit in the tricorn protease from Thermoplasma acidophilum," *J. Mol. Biol.*, vol. 324, no. 5, pp. 1041–1050, 2002.

[8] M. Groll, M. Bochtler, H. Brandstetter, T. Clausen, and R. Huber, "Molecular machines for protein degradation," *ChemBioChem*, vol. 6, no. 2, pp. 222–256, 2005.

[9] L. Borissenko and M. Groll, "Crystal structure of TET protease reveals complementary protein degradation pathways in prokaryotes.," *J. Mol. Biol.*, vol. 346, no. 5, pp. 1207–19, Mar. 2005.

[10] B. Rockel, È. Peters, B. Ku, R. M. Glaeser, and W. Baumeister, "A giant protease with a twist : the TPP II complex from Drosophila studied by electron microscopy," *EMBO J.*, vol. 21, no. 22, pp. 5979–5984, 2002.

[11] B. Rockel and W. Baumeister, "A tale of two giant proteases," *Ernst Schering Found. Symp. Proc.*, vol. 1, no. 1, pp. 17–40, 2008.

[12] B. Tomkinson, "Tripeptidyl peptidases: enzymes that count," *Trends Biochem. Sci.*, vol. 24, no. 9, pp. 355–359, Dec. 1999.

[13] E. Reits *et al.*, "A major role for TPPII in trimming proteasomal degradation products for MHC class I antigen presentation," *Immunity*, vol. 20, no. 4, pp. 495–506, 2004.

[14] R. Glas, G. Preta, R. De Klark, and R. Gavioli, "The enigma of tripeptidyl-peptidase II: Dual roles in housekeeping and stress," *J. Oncol.*, vol. 2010, 2010.

[15] C. Muñoz *et al.*, "Role of the Ubiquitin-Proteasome Systems in the Biology and Virulence of Protozoan Parasites," *Biomed Res. Int.*, vol. 2015, pp. 1–13, 2015.

[16] L. N. Shaw, "The lone S41 family C-terminal processing protease in Staphylococcus aureus is localized to the cell wall and contributes to virulence," *Microbiology*, vol. 160, pp. 1737–1748, 2014.

[17] A. B. Bandara, N. Sriranganathan, G. G. Schurig, and S. M. Boyle, "Carboxyl-Terminal Protease Regulates," *Microbiology*, vol. 187, no. 16, pp. 5767–5775, 2005.

[18] S. F. Altschul *et al.*, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Res*, vol. 25, no. 17, pp. 3389–3402, 1997.

[19] F. Chen, "OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups," *Nucleic Acids Res.*, vol. 34, no. 90001, pp. D363–D368, 2006.

[20] L. Y. Geer, M. Domrachev, D. J. Lipman, and S. H. Bryant, "CDART : Protein Homology by Domain Architecture CDART : Protein Homology by Domain Architecture," pp. 1619–1623, 2002.

[21] K. Katoh and D. M. Standley, "MAFFT multiple sequence alignment software version 7: Improvements in performance and usability," *Mol. Biol. Evol.*, vol. 30, no. 4, pp. 772–780, 2013.

[22] D. Szklarczyk *et al.*, "The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored," *Nucleic Acids Res*, Nov. 2010.

[23] S. Kumar, G. Stecher, and K. Tamura, "MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets.," *Mol. Biol. Evol.*, vol. 33, no. 7, p. msw054, 2016.

[24] S. Whelan and N. Goldman, "A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach," pp. 691–699, 1995.

[25] W. Ruepp, A., Eckerskorn, C., Bogyo, M., and Baumeister, "Proteasome function is dispensable under normal bit not under heat shock conditions in Thermoplasma acidophilum.," *FEBS Lett.*, vol. 425, pp. 87–90, 1998.

[26] T. Tamura, N. Tamura, Z. Cejka, R. Hegerl, F. Lottspeich, and W. Baumeister, "Tricorn protease--the core of a modular proteolytic system.," *Science*, vol. 274, no. 5291, pp. 1385–1389, 1996.

[27] N. Tamura, G. Pfeifer, W. Baumeister, and T. Tamura, "Tricorn protease in bacteria: Characterization of the enzyme from Streptomyces coelicolor," *Biol. Chem.*, vol. 382, no. 3, pp. 449–458, 2001.

[28] D. R. Caprioglio, C. Padilla, and M. Werner-washburnes, "Isolation and Characterization of," *J. Biol. Chem.*, vol. 268, no. 19, pp. 14310–14315, 1993.

[29] O. J. P. Kyrieleis, P. Goettig, R. Kiefersauer, R. Huber, and H. Brandstetter, "Crystal structures of the tricorn interacting factor F3 from Thermoplasma acidophilum, a zinc aminopeptidase in three different conformations," *J. Mol. Biol.*, vol. 349, no. 4, pp. 787–800, 2005.

Biography

Florence Ng'ong'a, is a postgraduate student in Biochemistry with research interests in Computational Biology and Molecular Parasitology focusing on biological processes that drive pathogenesis in parasitic infections.