



## OPEN ACCESS

EDITED BY  
Javier Pascual,  
Darwin Bioprospecting Excellence,  
Spain

REVIEWED BY  
Bing Tang,  
Wuhan University, China  
Dominik Łagowski,  
University of Life Sciences of Lublin,  
Poland

\*CORRESPONDENCE  
Petra Siegert  
siegert@fh-aachen.de

SPECIALTY SECTION  
This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

RECEIVED 12 August 2022  
ACCEPTED 30 August 2022  
PUBLISHED 26 September 2022

CITATION  
Falkenberg F, Bott M, Bongaerts J and  
Siegert P (2022) Phylogenetic survey  
of the subtilase family  
and a data-mining-based search  
for new subtilisins from *Bacillaceae*.  
*Front. Microbiol.* 13:1017978.  
doi: 10.3389/fmicb.2022.1017978

COPYRIGHT  
© 2022 Falkenberg, Bott, Bongaerts  
and Siegert. This is an open-access  
article distributed under the terms of  
the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution  
or reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s)  
are credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Phylogenetic survey of the subtilase family and a data-mining-based search for new subtilisins from *Bacillaceae*

Fabian Falkenberg<sup>1</sup>, Michael Bott<sup>2</sup>, Johannes Bongaerts<sup>1</sup> and Petra Siegert<sup>1\*</sup>

<sup>1</sup>Institute of Nano- and Biotechnologies, Aachen University of Applied Sciences, Jülich, Germany, <sup>2</sup>Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich, Jülich, Germany

The subtilase family (S8), a member of the clan SB of serine proteases are ubiquitous in all kingdoms of life and fulfil different physiological functions. Subtilases are divided in several groups and especially subtilisins are of interest as they are used in various industrial sectors. Therefore, we searched for new subtilisin sequences of the family *Bacillaceae* using a data mining approach. The obtained 1,400 sequences were phylogenetically classified in the context of the subtilase family. This required an updated comprehensive overview of the different groups within this family. To fill this gap, we conducted a phylogenetic survey of the S8 family with characterised holotypes derived from the MEROPS database. The analysis revealed the presence of eight previously uncharacterised groups and 13 subgroups within the S8 family. The sequences that emerged from the data mining with the set filter parameters were mainly assigned to the subtilisin subgroups of true subtilisins, high-alkaline subtilisins, and phylogenetically intermediate subtilisins and represent an excellent source for new subtilisin candidates.

## KEYWORDS

*Bacillaceae*, S8 protease family, subtilisin, data mining, subtilase, phylogenetic analysis

## Introduction

The subtilase family or subtilisin-like proteases defined by the MEROPS database as S8 family is the third largest family of serine proteases, both in terms of number of sequences and characterised peptidases, which are represented in microorganisms (archaea, bacteria, fungi, yeast) as well as in higher eukaryotes (Rawlings et al., 2014). The MEROPS database is a comprehensive source of information for proteases. It uses a hierarchical, structure-based classification, and based on statistically significant amino acid sequence similarities, proteases are grouped into families and clans (Rawlings et al., 2014). Here, the S8 family belongs to the clan SB, which is one of 13 clans of serine

proteases and, in addition to the S8 family, also contains the S53 family (sedolisin family) (Page and Di Cera, 2008). The two families differ in their catalytic mechanism in that within S8 they form a catalytic triad with their active residues in the order Asp, His, Ser, referred to as the “classical” D-H-S family (Siezen et al., 2007). While protein folding of members of the family S53 is very similar to that of subtilisins, the catalytic triad has been altered to Glu, Asp, and Ser, which is referred to as the ED-S family (Siezen et al., 2007). Furthermore, the S8 family is subdivided into the two subfamilies S8A (subtilisin as a type example) and S8B (kexin as a type example) (Rawlings, 2020). In addition, subtilases were classified by Siezen and Leunissen (1997) into six groups based on sequence alignments of the catalytic domain, namely subtilisins, thermitases, proteinase K, lantibiotic peptidases, pyrolysins and kexins, while mentioning that further subdivision may become useful with more available sequences. The kexins form the subfamily S8B, while the other five groups belong to the subfamily S8A. In the MEROPS database, well-characterised specimens are selected and designated as “holotypes” at the subfamily level (Rawlings et al., 2014). There are currently 186 holotypes listed for S8A and 21 for S8B (Rawlings and Bateman, 2021). Uncharacterised homologs of a holotype were assigned to the same MEROPS identifier (Rawlings and Bateman, 2021). Here, only the catalytically active part of the protease (peptidase unit) is considered, and a new holotype is created when a protein is characterised that either has a different specificity than another protein in the subfamily or the same specificity but a different cellular location, has a different architecture, or the sequence in a phylogenetic tree does not cluster with that of an existing holotype with similar specificity (Rawlings and Bateman, 2021).

The known members of the S8 family are endopeptidases, with the exception of TPPII (tripeptidyl peptidase II), which releases tripeptides from the N-terminus of peptides (Siezen and Leunissen, 1997; Renn et al., 1998). In most bacteria, archaea, and lower eukaryotes they are mostly unspecific proteases and are involved in nutrition (Siezen and Leunissen, 1997). They fulfil other functions as well, as they are involved in developmental processes and immune responses in plants (Schaller, 2013), play a role in the metabolism of neuropeptides in *Drosophila melanogaster* (Renn et al., 1998), or are involved in pathogenesis (García-Sánchez et al., 2004). Several subtilases contain a C-terminal extension, relative to the subtilisins, with additional properties such as sequence repeats, Cys-rich domains as cell surface anchors, or transmembrane segments (Siezen and Leunissen, 1997). Except for the subtilase ASP (*Aeromonas sobria* protease), an N-terminal propeptide acts as an intramolecular chaperone during maturation, supporting the folding of the catalytic domain (Zhu et al., 1989; Eder et al., 1993; Kobayashi et al., 2015). Members of the subtilase family find a wide range of applications in industry, such as lactocepins playing an important economic role in the industrial production of cheese and fermented milk (Broadbent and Steele, 2013). Of

particular interest within this study are subtilisins, which find applications in several industrial sectors such as in detergents, leather processing, food, wastewater treatment, cosmetics, and pharmaceuticals (Kalisz, 1988; Solanki et al., 2021; Azrin et al., 2022). They are typically isolated from various species of the genus *Bacillus* such as *B. subtilis*, *B. licheniformis*, *Shouhella clausii* (formerly *Bacillus clausii*), etc. (Kalisz, 1988; Christiansen et al., 2003; Maurer, 2004; Joshi et al., 2021). They consist of about 270 amino acids and are secreted *via* the Sec-secretion pathway in a precursor form containing a signal peptide of about 28 amino acids and a propeptide of about 75 amino acids (Markland and Smith, 1971; Power et al., 1986; Siezen et al., 1991; Tjalsma et al., 2000).

Due to the increasing number of genome sequencing projects, the amount of data on uncharacterised proteins is growing exponentially (Rawlings, 2013). This rapidly increasing online database provides an excellent resource for broadening the sequence space of subtilisins. Our research on uncharacterised subtilisin sequences began with the analysis of the MEROPS S8 dataset in a phylogenetic tree containing only characterised proteases. The analysis quickly revealed the presence of previously uncharacterised groups and subgroups within the S8 family and motivated us to update the phylogeny of this family, which was necessary to place the uncharacterised sequences in this context. Sequences from a data mining approach for new subtilisin proteases from the *Bacillaceae* family were then evaluated and placed in the context of the S8 subfamilies and groups.

## Materials and methods

### Sequence-based phylogenetic analysis

The amino acid sequences of the mature part, referred to as the “peptidase unit” in the MEROPS database, comprising the structural domain of the protein directly responsible for peptidase activity and substrate binding, including the larger insertions compared to other subtilases (Rawlings et al., 2018). Other structural domains, if present, were excluded, such as the signal peptide, the propeptide and C-terminal domains. The sequences were aligned using MAFFT v 7.490 with L-INS-I parameter<sup>1</sup> (Katoh and Standley, 2013; Katoh et al., 2019). The alignment was trimmed using trimAi v1.2 with the “gappyout” parameter<sup>2</sup> (Sánchez et al., 2011). The phylogeny was made using iqtree v1.6.12<sup>3</sup> (Trifinopoulos et al., 2016) with automated ModelFinder (Kalyaanamoorthy et al., 2017) and ultrafast bootstrap (Hoang et al., 2018) options. Phylogenetic trees were

1 <https://mafft.cbrc.jp/alignment/server/>

2 <http://phylemon.bioinfo.cipf.es/>

3 <http://iqtree.cibiv.univie.ac.at/>

displayed and annotated with the iTOL software<sup>4</sup> (Letunic and Bork, 2021).

## Data mining

Holotype protein sequences from the S8 family for the analysis were obtained from the MEROPS database<sup>5</sup> (Rawlings et al., 2014). In addition, a selection of sequences of characterised subtilases was chosen from the Protein Data Bank (PDB). Only the mature part of the proteases was used as described above.

To search for uncharacterised subtilisin sequences from *Bacillaceae*, new amino acid sequences for the analysis were obtained from the NCBI Identical protein groups database<sup>6</sup> by searching for “S8 peptidase *Bacillaceae*.” To selectively search for subtilisins, a filter was set for peptide sequences with a length of 350–410 amino acids. The resulting dataset was clustered with a identity threshold of 85% by using CD-HIT<sup>7</sup> (Huang et al., 2010). Intracellular proteases were excluded by analysing the sequences with the SignalP 6.0 prediction tool<sup>8</sup> and including only protein with a predicted Sec signal peptide (Teufel et al., 2022). The propeptide was removed manually after alignment with Clustal Omega<sup>9</sup> (Sievers et al., 2011), using the JalView alignment annotation software<sup>10</sup> (Waterhouse et al., 2009). MSA was drawn with ESPrnt 3.0 using %strict option (percentage of strictly conserved residues per column) for the colouring scheme<sup>11</sup> (Robert and Gouet, 2014).

## Bioinformatic analysis

The isoelectric point of a protein was calculated with the sequence manipulation suite v2<sup>12</sup> using pK<sub>a</sub> values from DTAselct (Stothard, 2000).

## Results and discussion

### Phylogenetic tree analysis of the MEROPS S8 holotype dataset

The first part of this study aimed at gaining a comprehensive overview of the S8A family in order to be able to categorise

the sequences obtained from the data mining approaches in the second part. Therefore, a phylogenetic tree was first constructed containing only the MEROPS holotype dataset of the S8 peptidase unit and a selection of sequences of biochemically characterised proteases from the PDB with a total of 168 sequences. The additional PDB sequences were added to support some of the subfamilies described in literature. The amino acid sequences of the mature part, referred to as the “peptidase unit” in the MEROPS database, were used as described in methods. Other structural domains, if present, were excluded, such as the signal peptide, the propeptide and C-terminal domains. In uncurated datasets, it becomes difficult to identify N- and C-terminal extensions in a wide range of different sequences. During the alignment curation, trimAI reduced the alignment length to 248 positions as opposed to 2,330 positions without curation. Curation with trimAI, which uses a less stringent algorithm, was preferred to more stringent filtering methods, as these often result in the deletion of positions in the alignment that contain a gap (Tan et al., 2015). However, gaps can contain significant phylogenetic information (Dessimoz and Gil, 2010), not to mention that a sequence dataset may contain an incomplete or incorrect sequence, which can result in a large loss of information if left undetected. According to Tan et al. a less stringent filtering algorithm has little impact on tree accuracy and is a trade-off in terms of computation time saved for phylogenetic tree computation (Tan et al., 2015). An overview of the workflow of the analysis and the methods used is shown in **Figure 1**.

The curated alignment was used to create a maximum likelihood tree, the standard option for constructing phylogenetic trees, which, together with the Bayes method, is widely recognised as the most accurate approach in molecular phylogenetics (Kuhner and Felsenstein, 1994; Dereeper et al., 2008). In addition, it is important to statistically evaluate the reliability of the tree, which is usually done using a bootstrap-based bias correction method that calculates the branch support of the tree by repeating the tree construction (Hoang et al., 2018). However, the different parameters for the alignment, the curation methods, and the different tree generation methods result in different phylogenetic trees, making a detailed comparison of a generated tree with literature data difficult. Since the constructed tree is not rooted, an outgroup must be selected for restructuring the tree, which contains a set of sequences that are outside the ingroup but closely related to it (Sanderson and Shaffer, 2002). **Figure 2** shows the phylogenetic tree of S8A subfamily, which is closely related to the S8B subfamily, which was selected as the outgroup (Siezen and Leunissen, 1997). In this tree, we identified the groups proteinase K, pyrolysins, thermitases, subtilisins, lantibiotic peptidases and kexins as described by Siezen and Leunissen (1997), and several subgroups within these groups. However, our analysis revealed that the S8A proteases form more groups and subgroups than previously

4 <https://itol.embl.de/>

5 <https://www.ebi.ac.uk/merops/>; published before January 19, 2022.

6 <https://www.ncbi.nlm.nih.gov/ipg/>; published before January 31, 2022.

7 <http://weizhong-lab.ucsd.edu/cdhit-web-server>

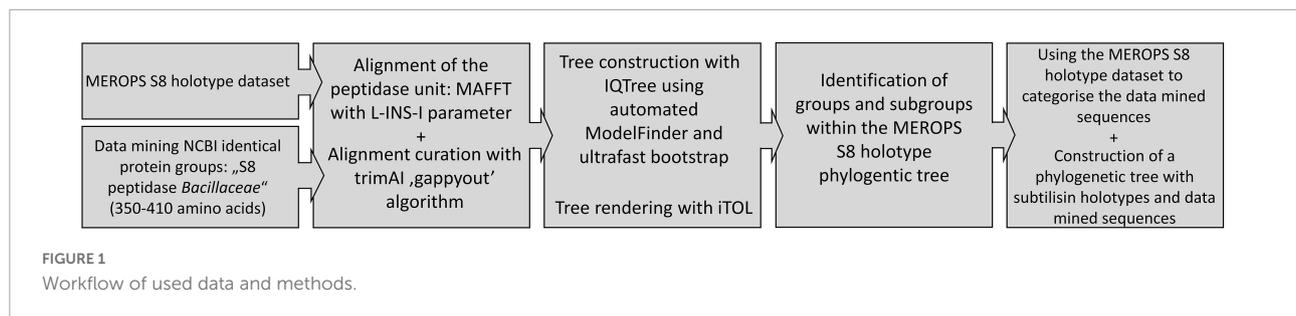
8 <https://services.healthtech.dtu.dk/service.php?SignalP-6.0>

9 <https://www.ebi.ac.uk/Tools/msa/clustalo/>

10 <https://www.jalview.org/>

11 <https://esprnt.ibcp.fr/ESPrnt/ESPrnt/>

12 <https://www.bioinformatics.org/sms2>



described. To account for the diversity resulting from their different positions in the phylogenetic tree, their biochemical properties, biological functions, their structural similarity, and the taxa- and species-specific clusters formed, we propose a revision of the subtilase groups and smaller, better defined subgroups. The naming is based on the criteria mentioned if a connection is recognisable, otherwise they are named according to the protease first described in this group. These groups and subgroups are discussed below in order to place the subtilisins in the context of the subtilases and **Table 1** provides an overview of them. Additionally, each group is shown as a pruned tree (**Supplementary Figures 1–17**).

## Pyrolysin group

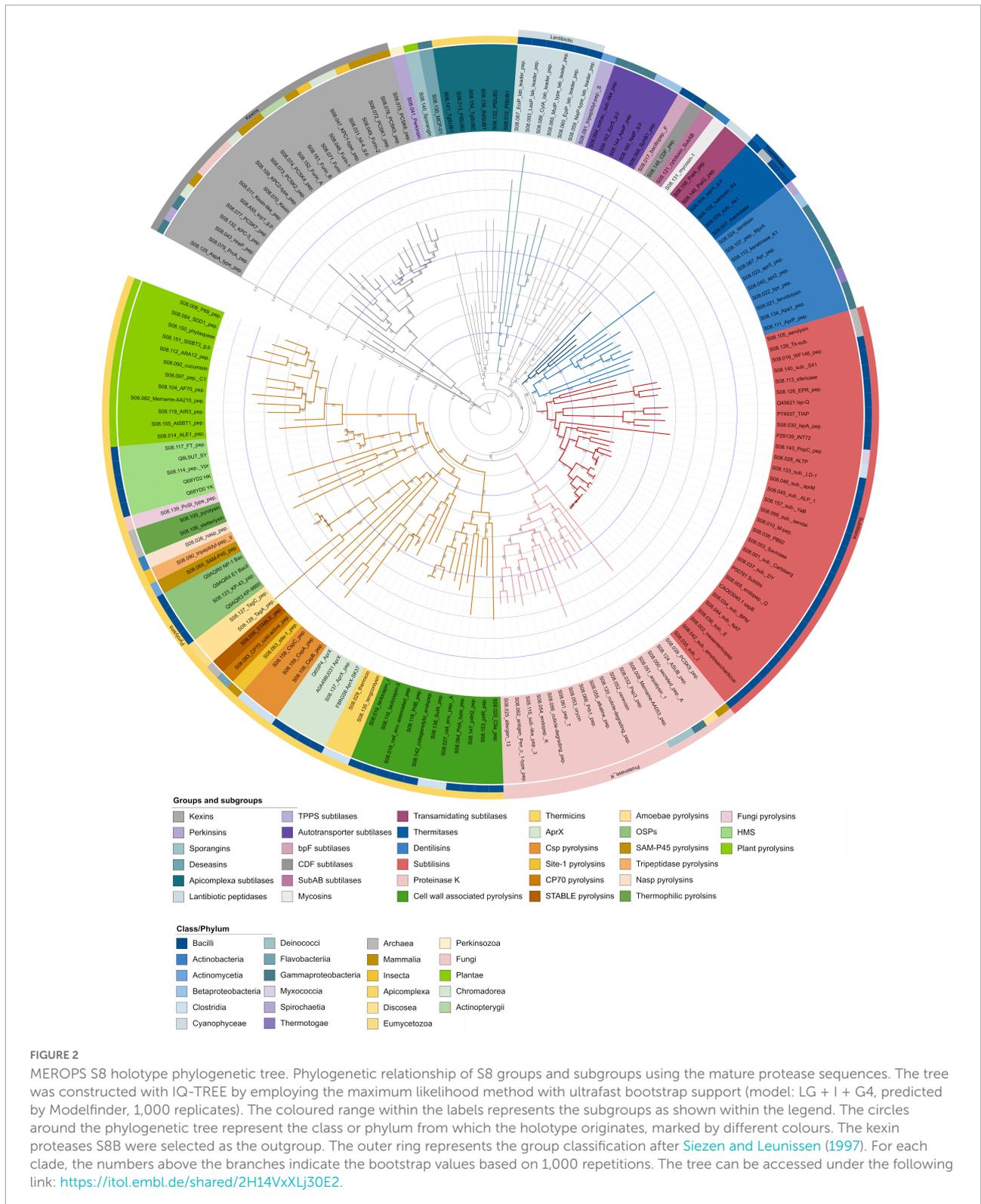
The pyrolysin group clustered within the phylogenetic tree in 90% of the replicates. Pyrolysins are a heterogeneous group of enzymes of diverse origin and low sequence conservation (Siezen and Leunissen, 1997). The average sequence identity of the pyrolysin group present in the phylogenetic tree is 38% (**Figure 2**). Within this branch, lactocepin 1 (S08.116) and lactocepin 3 (S08.019) are described by Siezen and Leunissen as pyrolysins in the gram-positive subgroup (Siezen and Leunissen, 1997; Broadbent and Steele, 2013). However, in the MEROPS phylogenetic tree, together with other sequences of gram-positive bacteria, they form a subgroup of their own, named here cell wall-associated pyrolysins with a bootstrap support of 100% and 51% sequence identity (**Supplementary Figure 1**). Lactocepins are cell envelope-associated endopeptidases of *Lactococci*, which play an important economic role in the industrial production of cheese and fermented milk due to their use as starter bacteria. Rapid growth is ensured by lactocepin, which provides amino acids from milk proteins and ensures autolysis, which is important for cheese ripening (Broadbent and Steele, 2013). They cluster together with other cell wall-associated proteases: S08.020 (Kagawa and Cooney); S08.153 (Pastar et al., 2003); S08.147 (Genay et al., 2009); S08.064 (Bethe et al., 2001); S08.027 (Lawrenson and Sriskandan); S08.138 (Karlsson et al., 2007); S08.118 (Gilbert et al., 1996). An exception is the thermophilic collagenolytic protease from *Geobacillus collagenovorans* MO-1 (S08.142),

which contains a collagen-binding segment and is secreted into the culture supernatant without being displayed on the cell surface (Okamoto et al., 2001; Itoi et al., 2006).

Thermicin (S08.029), tengconlysin (S08.135), and the AprX proteases (S08.137) cluster together in the tree in 99% of the replicates (**Supplementary Figure 2**). Thermicin from the extremely thermophilic bacterium *Thermoanaerobacter yonseiensis* KB-1 (Jang et al., 2002) and tengconlysin from *Thermoanaerobacter tengcongensis* (Koma et al., 2007) show both high-temperature optima above 90°C. As mentioned by Jang et al. (2002) thermicin is a novel enzyme that differs from other thermostable proteases and here forms the new subgroup of thermicins. The Bacilli-derived AprX subtilases form another subgroup as previously described, because they lack a signal peptide and exhibit a mesophilic temperature optimum (Valbuzzi et al., 1999; Phrommao et al., 2011). The intracellular subtilase AprX-SK37 from *Virgibacillus* sp. SK37 is a halotolerant, oxidation-stable, and moderately thermophilic alkaline serine protease with properties that could be attractive for various biotechnological applications (Phrommao et al., 2011).

The three proteases CspB (S08.108), CspA (S08.159), and CspC (S08.158) (Clostridial serine proteases) from *Clostridium perfringens* form a new subgroup (Csp pyrolysins) with 45% sequence identity, related to germination and synthesised in the mother cell compartment of spore-forming cells (**Supplementary Figure 3**; Masayama et al., 2006).

Site-1 peptidase (S08.063) from *Cricetulus griseus* is an important processing enzyme of the endoplasmic reticulum/Golgi lumen that acts on sterol regulatory element binding proteins (SREBPs) to regulate cholesterol and fatty acid biosynthesis in addition to other cellular functions (Seidah, 2013b). CP70 (S08.083) from *Flavobacterium balustinum* is a cold-active extracellular protease (Morita et al., 1998); STABLE (S08.096) is a hyperthermostable protease bound to the surface layer of the archaeon *Staphylothermus marinus* and is responsible for the generation of the peptides required in the energy metabolism of the cell (Mayr et al., 1996). Because of their different physiological functions and origins, they will most likely all form new individual subgroups as more homologs are added (Site-1 pyrolysins, CP70 pyrolysins, STABLE pyrolysins). TagA (S08.128) and TagC (S08.127) are



produced by the amoeba *Dictyostelium discoideum*. While TagA is involved in the differentiation of cell types (Good et al., 2003), TagC is part of a transmembrane protein of the ABC

family, which is expressed during the aggregation stage of development (Anjard and Loomis, 2005). TagA and TagC are forming the new subgroup of amoebae pyrolysins with a 100%

TABLE 1 Group and subgroups of the subtilase family.

Group	Subgroup	References
Pyrolysins	Plant pyrolysins	Siezen and Leunissen, 1997
	High-mass subtilases (HMS)	Okuda et al., 2004
	Fungi pyrolysins	Faraco et al., 2005
	Thermophilic pyrolysins	Siezen and Leunissen, 1997
	Nasp pyrolysins	+
	Tripeptidase pyrolysins	Siezen and Leunissen, 1997
	SAM-P45 pyrolysins	+
	Oxidatively stable proteases (OSP)	Saeki et al., 2000
	Amoebae pyrolysins	+
	STABLE pyrolysins	+
	CP70 pyrolysins	+
	Site-1 pyrolysins	+
	Csp pyrolysins	+
	AprX	Phrommao et al., 2011
	Thermicins	+ / Jang et al., 2002
	Cell-wall associated pyrolysins	+
	Proteinase K	Gram -
Fungal		Siezen and Leunissen, 1997
Yeast		Siezen and Leunissen, 1997
Subtilisins	EPR subtilisins	+
	PopC subtilisins	+
	Extremophilic subtilisins	+
	Archaea subtilisins	+
	True subtilisins	Siezen and Leunissen, 1997
	High-alkaline subtilisins	Siezen and Leunissen, 1997
	Phylogenetically intermediate subtilisins (PIS)	Saeki et al., 2003
Dentilisins	+	
Thermitases	Extremophilic thermitases	Siezen and Leunissen, 1997
Transamidating subtilases		Agarwal et al., 2012
Mycosins		Dave et al., 2002
SubAB subtilases		+
CDF subtilases		+
bpF subtilases		+
Autotransporter subtilases		Henderson and Nataro, 2001
TPPS subtilases		+
Lantibiotic peptidases		Siezen and Leunissen, 1997
Apicomplexa subtilases		+
Deseasins		Chen et al., 2007
Sporangins		+
Perkinsins		+
Kexins (S8B)	PC1	Siezen and Leunissen, 1997
	PC2	Siezen and Leunissen, 1997
	Furins	Siezen and Leunissen, 1997
	Yeast Kexins	Siezen and Leunissen, 1997

+Indicates that the group or subgroup was defined within this work.

bootstrap support and 42% sequence identity (**Supplementary Figure 3**).

KP-43 (S08.123) from *Bacillus* strain KSM-KP43 and the sequences within this clade form the subgroup of oxidatively stable serine proteases (OSPs) as described by Saeki et al., with a sequence identity of 92% (Saeki et al., 2000, 2002). However, it should be mentioned that subtilisins not belonging to this subgroup were reported to have higher stability against H<sub>2</sub>O<sub>2</sub> (Joshi and Satyanarayana, 2013; Falkenberg et al., 2022). The C-terminal half of KSM-KP43 downstream of the putative catalytic residue, Ser-255, is homologous to the internal segments of TagC (Saeki et al., 2002). While the MEROPS dataset suggests that OSPs are only of bacterial origin, Li et al. (2017) described sequences originating from some species of *Pezizomycotina* fungi.

SAM-P45 (S08.069), a membrane-anchored protease from *Streptomyces albogriseolus* which is considered to be an evolutionary link between primitive bacterial subtilisins and highly diversified eukaryotic proteases, forms its own new subgroup (SAM-P45 pyrolysins) (Suzuki et al., 1997). TPPII, isolated from *Drosophila melanogaster*, has an elongated C-terminus compared to other subtilases and is involved in the metabolism of neuropeptides, which are important signalling molecules in insects and belongs to the tripeptidase pyrolysins subgroup (Siezen and Leunissen, 1997; Renn et al., 1998). The extracellular serine protease Nasp (S08.026) from *Dermatophilus congolensis* is involved in pathogenesis and forms the new subgroup Nasp pyrolysins (Garcia-Sanchez et al., 2004). The two thermophilic proteases from archaea *Thermococcus stetteri* (stetterlysin; S08.106) (Klingeberg et al., 1995) and *Pyrococcus furiosus* (pyrolysin; S08.100), the eponym of the whole group (Blumentals et al., 1990) are both resistant against SDS (1% w/v) and have a high-temperature optimum (85°C, 115°C) forming the thermophilic pyrolysins subgroup (**Supplementary Figure 4**; Siezen and Leunissen, 1997).

PoSI (*P. ostreatus* extracellular protease) (S08.139), a protease from the fungus *Pleurotus ostreatus*, is involved in the activation of other secreted proteases and the post-translational regulation of laccase (Faraco et al., 2005). It shows a high sequence identity with the minor extracellular serine protease from *Bacillus subtilis*, Vpr (S08.114) (31%), and together with other proteases from Ascomycetes and Basidiomycetes forms a separate pyrolysin subgroup (fungi pyrolysins) (**Supplementary Figure 5**; Faraco et al., 2005). A further subdivision of fungal subtilases has been made by others and is beyond the scope of this study (Hu and Leger, 2004; Muszewska et al., 2011; Li et al., 2017).

According to Okuda et al. Vpr (S08.114) belongs to the subgroup of high-molecular-mass subtilisins, which can be divided into at least two classes (**Supplementary Figure 5**; Okuda et al., 2004). One class is less alkaline, its stability depends on Ca<sup>2+</sup> ions and it is resistant to proteolysis. The other class is strongly alkaline, its stability also depends on Ca<sup>2+</sup> ions and is

sensitive to proteolysis (Okuda et al., 2004). The subgroup name is slightly misleading, as they do not cluster together with the subtilisins. Therefore, they were named here high-molecular-mass subtilases (HMS). Here, the average sequence identity is 82%.

Plant subtilases are a widely distributed subgroup involved in plant developmental processes and immune responses (Schaller, 2013). The average sequence identity between the investigated sequences is 56% (Figure 2 and Supplementary Figure 6). The first subtilase cloned from plants was the extracellular alkaline protease cucumisin (S08.092) from melon fruit (Kaneda and Tominaga, 1975). The plant subtilases have been divided into seven classes (Xu et al., 2019). A detailed discussion of each of these classes is beyond the scope of this study. Good overviews of the classes and the plant subtilases were provided by Schaller (2013), Taylor and Qiu (2017), and Xu et al. (2019).

## Proteinase K group

The alkaline proteinase secreted into the culture medium by the mould *Tritirachium album* Limber is commonly known as proteinase K (S08.054) and is the type example for this group (Ebeling et al., 1974). It can be used to synthesise peptides (Ageitos et al., 2013), and besides peptide bonds, it can also hydrolyse esters (Borhan et al., 1996). In contrast to subtilisins, which contain no cysteine residues, proteinase K contains five Cys residues, four which form two disulfide bridges (Betzel et al., 1990). Because of its remaining activity at higher temperatures (> 60°C) in the presence of urea, 0.5% (w/v) SDS, or 1% (w/v) Triton X100, proteinase K is used for the degradation of proteins and in the preparation of nucleic acids (Sweeney and Walker, 1993; Goldenberger et al., 1995). Most of the proteinase K holotypes found in the phylogenetic tree derive from fungi (Figure 1). Workflow of used data and methods.

There, the fungal proteinase K subgroup is separated from the other proteinase K members and may play an important role in the evolution of pathogenicity, as several entomopathogenic and nematophagous fungi have been characterised as having the ability to destroy the structural integrity of insect or nematode cuticle during invasion and colonisation. Therefore, they are also referred to as cuticle-degrading proteases (S08.120, S08.056) (Leger et al., 1987; Tunlid and Jansson, 1991; Li et al., 2010). For many saprophytes, the subtilases as broad-spectrum proteases play a role in nutrition acquisition, such as digesting proteins to release peptides and amino acids (Gunkel and Gassen, 1989; Hu and Leger, 2004). Further phylogenetic analysis by Li et al. of 138 fungal proteinase K genes revealed a subdivision into five distinct classes (Li et al., 2017). The fungal proteinase K-like proteases are separated from the bacterial ones including aqualysin (S08.051) from the thermophilic bacterium *Thermus aquaticus* (Sakaguchi), the *Amoebzoa* protease ASUB

(S08.124) from *Acanthamoeba healyi* (Kong et al., 2000), and the proprotein convertase PCSK9 (S08.039) from *Mus musculus* (Seidah, 2013a; Supplementary Figure 7). The average sequence identity within the proteinase K family in the phylogenetic tree is 48%.

## Thermitase/subtilisin group

Thermitase, the type enzyme for this group, is an extracellular, thermostable protease of the thermophilic microorganism *Thermoactinomyces vulgaris* (Frömmel et al., 1978). For an in-depth review of this protease see Betzel (2004). The other three enzymes of the thermitase-type WprA (S08.004) (Margot and Karamata, 1996), halolysin (S08.102) (Kamekura et al., 1996), and Subtilisin AK1 (S08.009) (Toogood et al., 2000) were also identified by Siezen and Leunissen (1997) and forming the subgroup of extremophilic thermitases. While halolysin derives from the halophilic archaeon *Haloferax mediterranei* (Kamekura et al., 1996), all other members come from Bacilli (see Supplementary Figure 8 and Figure 2). Figure 2 (light blue) shows a more distinctive group with two clades. Here, Siezen and Leunissen identified bpr (S08.022) of *Dichelobacter nodosus* as a thermitase and AprP of *Pseudomonas* sp. KFCC 10818 as a subtilisin (Lilley et al., 1992; Jang et al., 1996; Siezen and Leunissen, 1997). However, the bootstrap value for these two clades is only 18%, which is why they are treated as an intermediate subgroup between thermitases and subtilisins. This intermediate new subgroup is named here as dentilisins, since this protease was already described in 1990 (Que and Kuramitsu, 1990). In general, thermitases are co-located in the clade with subtilisins, highlighting their similarity.

The subtilisin Carlsberg (S08.001) is the type example of the subtilisins and the entire S8 family and belongs to the subgroup of true subtilisins, along with the high-alkaline subtilisins, the intracellular subtilisins, and the phylogenetically intermediate subtilisins (PIS) (Smith et al., 1966; Siezen and Leunissen, 1997; Saeki et al., 2003). Extracellular subtilisins play an important role in nutrition, whereas intracellular subtilisins (Isp), such as IspA (S08.030), play a role in protein turnover and processing during sporulation or are involved in the heat shock response (Reysset and Millet, 1972; Koide et al., 1986). As shown in Figure 2 all holotypes are derived from microorganisms, mainly from Bacilli, while aerolysin (S08.105) (Völkl et al., 1994), Tk-subtilisin (*Thermococcus kodakaraensis* subtilisin) (S08.129) (Kannan et al., 2001), PopC, (S08.143) (Rolbetzki et al., 2008), and ALTP (*Alkaliphilus transvaalensis* protease) (S08.028) (Kobayashi et al., 2007) derive from Archaea, Myxococci, and Clostridia, respectively. However, *Bacillus* as the most prominent source of subtilisins spawned alkaline proteases such as subtilisin Carlsberg, BPN', and Savinase, which have their major application as detergent enzymes

with excellent properties, including high stability toward extreme temperatures, pH, organic solvents, detergents, and oxidising compounds (Kalisz, 1988; Contesini et al., 2018). Besides the application in detergents, subtilisins are applied for example in leather processing, food, wastewater treatment, and cosmetics (Kalisz, 1988; Solanki et al., 2021). These subtilisin holotype sequences will be used for the classification of the new sequences provided by a data mining approach.

## Various diverse groups

Several holotypes form a clade together within the phylogenetic tree (Figure 2 and Supplementary Figure 9). Due to their different origins, their different biological functions, and their low bootstrap value (40%), they most likely form individual groups. Bacillopeptidase F (bpF) (S08.017) from *B. subtilis* is a cell envelope protein and contributes to nutrition in the soil environment (bpF subtilases) (Hageman). While Siezen and Leunissen grouped bpF within the pyrolysins, it is separated in this study, which could be due to the fact that Siezen and Leunissen only analysed the amino acids around the catalytically active ones (Siezen and Leunissen, 1997). CDF (S08.149) from *Thermoactinomyces* sp. CDF is a protease located on the surface of the spore coat (CDF subtilases) (Cheng et al., 2009). Cytotoxin SubAB (S08.121) is a toxin from *Escherichia coli* with two subunits, where subunit B binds to the surface receptor of target cells and subunit A, the enzymatically active moiety, is responsible for cytotoxicity and has a very narrow substrate specificity (SubAB subtilases) (Yahiro et al.).

Mycosin-1 (S08.131) from *Mycobacterium tuberculosis* is an extracellular protein that is membrane- and cell wall-associated and is expressed after infection of macrophages, forming the group of mycosins (Dave et al., 2002). PatA (S08.156) and PatG (S08.146) from *Prochloron didemni* are involved in the maturation of cyanobactins in *Cyanobacteria* and form the group of transamidating subtilases (Lee et al., 2009; Agarwal et al., 2012).

A separated clade can be also observed for the new group of tripeptidyl peptidase subtilases (TPPS) and the group of autotransporter subtilases (AT) (Henderson and Nataro, 2001; Supplementary Figure 10). Like tripeptidyl peptidase II (TPPII, S08.090) from the pyrolysin group, tripeptidyl peptidase S (TPPS) (S08.091) from *Streptomyces lividans* is an exopeptidase that cleaves tripeptide units from oligopeptides or polypeptides and probably forms its own group (Butler, 2013). The extracellular *Serratia* serine protease (SSP) (S08.094) from *Serratia marcescens* (Yanagida et al., 1986) was grouped by Siezen and Leunissen (1997) to gram-negative pyrolysins. However, SSP together with EprS (S08.162) (Kida et al., 2013), AasP (autotransported serine protease A) (S08.1449) (Ali et al., 2008), NalP (Neisserial autotransporter lipoprotein) (S08.160) (Turner et al., 2002), and SphB1 (S08.068) (Coutte et al., 2001),

are forming the group of autotransporter subtilases in gram-negative bacteria with a low sequence identity of 35%.

An additional clade with 94% bootstrap support is formed by perkinsin (S08.041) from the protist *Perkinsus marinus*, which is an enzyme of unknown function but may be involved in cell invasion of the eastern oyster *Crassostrea virginica* (Brown and Reece, 2003; Supplementary Figure 11). Sporangin (S08.145) from the alga *Chlamydomonas reinhardtii* is localised to the flagella of daughter cells within the sporangial cell wall and is released into the culture medium where it is involved in the digestion of the sporangial cell wall (Kubo et al., 2009). Perkinsin and sporangin are forming two new groups (perkinsins, sporangins). MCP-01 (S08.130), the extracellular cold-adapted protease from the deep-sea bacterium *Pseudoalteromonas* sp. SM9913 forms the group of deseasins secreted mainly by bacteria in deep-sea or lake sediments (Chen et al., 2007; Zhao et al., 2008). It is a multidomain protein with a collagen-binding domain at its C-terminus that exhibits collagenolytic activity and therefore plays an important role in the degradation of particulate organic nitrogen from deep-sea sediments (Zhao et al., 2008). The proteases produced by the parasites in the phylum Apicomplexa are forming a new group (Apicomplexa subtilases) with TgSub1 (*Toxoplasma gondii*) (S08.141) (Miller et al., 2001), PfSUB2 (*Plasmodium falciparum*) (S08.013) (Hackett et al., 1999), TgSUB2 (S08.154) (Miller et al., 2003), BdSUB1 (*Babesia divergens*) (S08.136) (Montero et al., 2006), PfSUB3 (S08.122) (Withers-Martinez et al., 2004), and PfSUB1 (S08.012) (Withers-Martinez et al., 2004), with a sequence identity of 34%. These proteases are involved in the host-cell invasion (Silmon de Monerri et al.).

## Lantibiotic peptidase group

The lantibiotic peptidase group within S8A subfamily comprises highly specialised enzymes for cleavage of leader peptides from precursors of the antimicrobial peptides (lantibiotics) (Sahl et al., 1995; Supplementary Figure 12). ElkP (S08.095) and PepP (S08.85) from *Staphylococcus epidermidis* are not included in the dataset because only sequence fragments were available. Lantibiotic peptidases are found intracellularly, extracellularly and membrane-anchored (Bierbaum et al.). The sequences included in the phylogenetic tree show 31% sequence identity.

## Kexin subfamily (S8B)

Kexin the type example for the subfamily S8B and was first identified in *Saccharomyces cerevisiae*. It can process the yeast precursors of alpha-mating factor and killer toxin and plays a significant role in post-translational modification in eukaryotes (Rogers et al., 1979). For a review see Fuller (2013).

AspA (S08.125) is one of the two subtilases which is lacking a propeptide and stands apart within the S8B family (Figure 2; Mellergaard, 1983; Kobayashi et al., 2015). Siezen and Leunissen (1997) also mentioned that AspA is a more distant member. As mentioned above, the S8B subfamily was used as an outgroup for the phylogenetic tree as it is closely related to the S8A subfamily (Siezen and Leunissen, 1997). Within the phylogenetic tree, it forms a clearly defined clade comprising all 21 holotype sequences of the MEROPS S8B subfamily, with an average sequence identity of 53% (Supplementary Figure 13). The clustering with a bootstrap support of 89% supports that kexins are a distinct subfamily (S8B) within the subtilase subset (Siezen and Leunissen, 1997). Kexins are also divided into at least four subgroups: PC1, PC2, furins and yeast kexins, but the subdivision will not be discussed further here as the focus is on the S8A subfamily (Siezen and Leunissen, 1997).

## Data mining and phylogenetic tree analysis of subtilisins from *Bacillaceae*

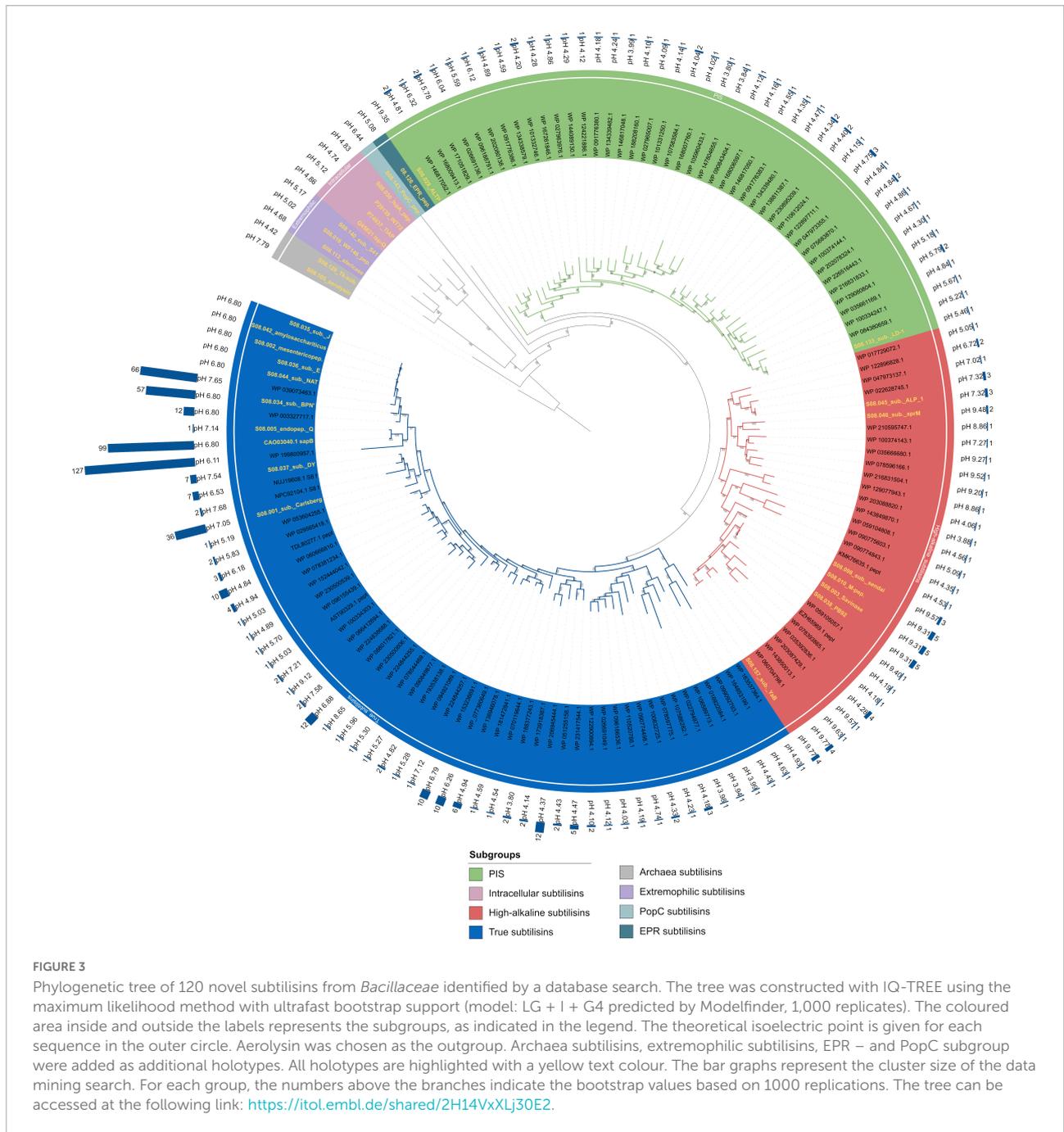
Due to the increasing number of genome sequencing projects, the amount of data on uncharacterised proteins is growing exponentially (Rawlings, 2013). Many genomes encode multiple secreted proteases and many proteases can be found in different species (Takimura et al., 2007). The great potential of the data mining approach becomes clear when looking at the huge number of 247.897 hits (January 31, 2022) that were found when searching for S8 peptidases within the NCBI identical protein groups database. Within this second part of our study, we focused on subtilisins derived from *Bacillaceae* because they are right now the most relevant industrial proteases (Maurer, 2004; Azrin et al., 2022).

To search for new subtilisin sequences from *Bacillaceae*, the database search was performed as described above and in Figure 1. The search yielded 1,424 sequences with the set values. With the length specification of 350–410 amino acids, sequences typical of AprX, lantibiotic peptidases, kexins, OSP, and HMS are excluded, while typical thermitases, intracellular subtilisins, proteinase K, and high/true/PIS subtilisins can still be found. The size exclusion was set to reduce the number of sequences (18.881 without size exclusion) and was chosen because typical subtilisin sequences derived from *Bacillaceae* are around 380 amino acids long, including the signal peptide and the propeptide (Markland and Smith, 1971; Power et al., 1986; Siezen et al., 1991; Tjalsma et al., 2000). Without the size exclusion, many additional new subtilases from *Bacillaceae* could probably be found. CD-HIT clustering with an identity threshold of 85% yielded 375 clusters. For each cluster, one representative was used for further analysis. The number of sequences within one cluster is displayed as a bar chart around the phylogenetic tree in Figure 3. Signal peptide analysis identified 135 sequences without a signal peptide,

reducing the dataset to 240 sequences, as we are only interested in extracellular proteases for further analysis and potential biochemical characterisation. The remaining sequences were aligned and the propeptide was manually removed as described above. Sequences that could be directly visually classified as thermitases after alignment were discarded, leaving a sequence set of 120 sequences within the sequence space of subtilisins as shown in Figure 3 (Supplementary Table 1). Sequences from the first phylogenetic tree comprising all 168 MEROPS holotypes, which build the subfamilies of subtilisins, were used again and aligned with the 120 sequences from the data mining approach. The sequence alignment was refined with TrimAl, which reduced the alignment length to 260 positions in contrast to 448 positions without refinement. Here, the two archaea subtilisins were used as an outgroup to reroot the tree. Figure 3 shows that all sequences derived from *Bacillaceae* in the data mining set represent the three main subgroups within the subtilisins, the true subtilisins, the high-alkaline subtilisins, and the phylogenetically intermediate subtilisins.

In addition to the above mentioned three subgroups, the following proteases form additional subgroups (Supplementary Figure 14): Aerolysin (S08.105) and Tk-subtilisin (S08.129) from the hyperthermophilic archaea *Pyrobaculum aerophilum* (Völkl et al., 1994) and *Thermococcus kodakaraensis* (Kannan et al., 2001) are forming an own new subgroup (Archaea subtilisins) with a sequence identity of 47%. Subtilisin S41 (S08.140) a psychrophilic protease from antarctic *Bacillus* TA41 (Almog et al., 2009), WF146 (S08.016), a thermophilic protease from *Bacillus* sp. WF146 (Wu et al., 2004), and Sfericase (S08.113), a psychrophilic protease from *Lysinibacillus sphaericus*, are forming a new subgroup named here extremophilic subtilisins with a sequence identity of 70%. The fact that no similar sequences from the group of extremophilic subtilisins were found in the data mining search could be due to the fact that all three representatives are larger than 410 amino acids, which also applies to EPR (S08.126), an extracellular protease from *B. subtilis* involved in cell motility (Dixit et al., 2002). PopC (S08.143) is involved in the cell signalling cascade for forming *Myxococcus xanthus* cells into fruiting bodies and sporulation (Rolbetzki et al., 2008). EPR and PopC are forming two new individual subgroups (EPR subtilisins, PopC subtilisins). The intracellular subtilisins form their own known subgroup with a sequence identity of 72% (Siezen and Leunissen, 1997). Since all sequences without signal peptides were excluded from the data mining set, no sequences are clustered with the holotypes. In general, subtilisins are mainly found in Bacilli and none in fungi (Siezen et al., 2007; Muszewska et al., 2011). In the following, all amino acid positions refer to the BPN<sup>2</sup> numbering.

The subgroup of **true subtilisins** includes subtilisin Carlsberg (S08.001) from *Bacillus licheniformis* (Linderström-Lang and Ottesen, 1947; Güntelberg and Ottesen, 1952; Smith et al., 1966), which together with BPN<sup>2</sup> (S08.032) from *Bacillus*



*amyloliquefaciens* were the first two subtilisins to be studied in detail (Matsubara et al., 1965; Smith et al., 1966). Their group is supported by a 91% bootstrap value within the phylogenetic tree (Figure 3 and Supplementary Figure 17). Interestingly the data mining search revealed the most similar sequences within the subgroup around these holotypes, as indicated by the bar chart with cluster sizes up to 127 sequences. Several newly found sequences are phylogenetically more distinct from any known holotype, which suggests that these sequences

could have other biochemical characteristics and may form new classes. The calculated isoelectric point of the sequences within the true subtilisins is on average rather acidic to neutral. The representatives of this subgroup characterised so far are more sensitive and less active under high-alkaline conditions compared to the high-alkaline subtilisins (Nakamura et al., 1973; Maeda et al., 2001).

The subgroup of **phylogenetically intermediate subtilisins** (PIS) was introduced by Saeki et al. with the biochemical

characterisation of the subtilisin LD1 (S08.133) from the alkaliphilic *Bacillus* sp. KSM-LD1. Due to its properties and the phylogenetic position, LD1 forms a subgroup at an intermediate position between true subtilisins and high-alkaline subtilisins (Saeki et al., 2003). LD1 has a C-terminal extension of 29 amino acids, suggesting an association with the cell surface of *Bacillus* sp. KSM-LD1 (Saeki et al., 2003). Interestingly, the sequences WP\_100334247.1 from *Bacillus alkalisoli* and WP\_084380659.1 from *Sutcliffiella cohnii* have a C-terminal extension like LD1 (Spanka and Fritze, 1993; Liu et al., 2019). LD1 and other PIS have multiple amino acid insertions compared to BPN<sup>+</sup>, but this does not affect substrate specificity toward synthetic substrates (Saeki et al., 2003). The protease ALTP (S08.028) from the anaerobic and extremely alkaliphilic *Alkaliphilus transvaalensis* is the first high-alkaline protease reported from a strict anaerobe (Kobayashi et al., 2007). ALTP is 66% identical to LD1 and, according to Kobayashi et al., it is in an intermediate position between the true and the highly alkaline subtilisins (Kobayashi et al., 2007). This assignment is supported by the phylogenetic tree (Figure 2) constructed in our study. In the phylogenetic tree with the newly mined database sequences (Figure 3 and Supplementary Figure 15), ALTP is separated from the *Bacillaceae*-derived phylogenetically intermediate subtilisins, as it is derived from the bacterial class *Clostridia*. ALTP has solely an alkaline isoelectric point, while the other sequences within this subgroup all have an acidic pI (Figure 3).

The subgroup of **high-alkaline subtilisins** was discovered in the 1980s and originates from alkaliphilic *Bacilli* (Ito et al., 1998; Maurer, 2004). Since the first discovery of protease no. 221, an increasing number of high-alkaline subtilisins have been characterised (Nakamura et al., 1973). Alkaline subtilisins, such as Savinase, are much more stable in an alkaline environment than true subtilisins such as BPN<sup>+</sup> or subtilisin Carlsberg and can be used to adapt to harsh industrial conditions, especially in modern detergents (Maurer, 2004). Within the phylogenetic tree they are forming a distinct subgroup with a branch support of 93% (Figure 3 and Supplementary Figure 16). ALP-1 (S08.045) from *Bacillus* sp. NKS-21 (Yamagata et al., 1995), WP\_017729072.1 from *Halalkalibacterium ligniniphilum* (Zhu et al., 2014; Joshi et al., 2021), WP\_122896828.1 from *Alteribacter keqinensis* (Liu et al., 2022), WP\_047973137.1 from *Bacillus* sp. LL01 (Vilo et al., 2015), and WP\_022628745.1 from *Alkalihalophilus marmarensis* (Denizci et al., 2010; Joshi et al., 2021) form a more separated clade, as they lack the four amino acid deletion around position 160, in contrast to the other high-alkaline proteases. This position corresponds to a loop near the P1 binding site (Wells et al., 1987; Betzel et al., 1992). They form another class of the ALP-1-type subtilisins, as mentioned by Yamagata et al. (2002). Additionally, the theoretical isoelectric point of these proteins is neutral to acidic in contrast to the majority of the other high-alkaline proteases. High-alkaline proteases adapt to higher alkaline conditions by an altered surface charge at higher pHs, as indicated by an

increased pI value caused by a higher number of Arg and a decreased number of Lys residues (Masui et al., 1998). The substrate specificity of ALP-1 toward the B-chain of insulin differs from that of other alkaline subtilisins, but is similar to that of neutrophilic subtilisins, which may be related to the deletion of four amino acids around position 160 (Tsuchida et al., 1986; Yamagata et al., 1995). For ALP-1, an enzyme engineering study identified amino acids in the C-terminal region that increased stability 120-fold under alkaline conditions after replacement (Yamagata et al., 2002). Some of the high-alkaline proteases including Savinase have an extra proline at position 131, which provides extra active-site rigidity compared with other subtilisins (Betzel et al., 1996). Recently, we reported about SPAO from *Alkalihalobacillus okhensis* Kh10-101<sup>T</sup>, which showed high stability against hydrogen peroxide and NaCl concentrations up to 5.0 M (Falkenberg et al., 2022). SPAO can be assigned here to the holotype subtilisin sendai (S08.098) (Figure 3).

The average sequence identity between the sequences within the three subgroups of high-alkaline, PIS, and true subtilisins was calculated to be 67, 72, and 66%, respectively (sequences from Figure 3). The identity between true and high-alkaline subtilisins is 58%, between true and PIS 57%, and between high-alkaline subtilisins and PIS 55%.

A detailed investigation of all insertions and deletions within the three subgroups PIS, high-alkaline, and true subtilisins showed that the four amino acid deletion in the clade of aprM (S08.046) (Takami et al., 1990; Masui et al., 1994) is between Ser<sup>161</sup> and Thr<sup>174</sup>, while for the other high-alkaline proteases, except for the ALP-1 clade mentioned above, the deletion is between Gly<sup>160</sup> and Thr<sup>174</sup> (Supplementary Figure 18). Interestingly, all high-alkaline subtilisins have a deletion of one amino acid at positions 37 and 57 (Supplementary Figure 18). The loop of amino acids 50–59 is known to be one of the most variable parts of subtilisin structures (van der Laan et al., 1992). Therefore, deletions within this loop could be detected in several sequences within the three subfamilies. Several PIS sequences have a double insertion between positions 42 and 43 in common. All sequences within the PIS subgroup share the insertion between positions 159 and 160, while high-alkaline subtilisins have a deletion of four amino acids around this position. Position 160 is localised in a loop that, as mentioned above, takes part in the conformation of the P1 pocket and might be involved in the P1 preference, and the recognition of steric conformation (Yamagata et al., 1995). Additionally, shorter loops can increase the stability of an enzyme (Gavrilov et al., 2015). In general, all insertions or deletions are located at the surface of the protease, which could be due to the fact that the overall structure within the subtilisins is highly conserved (Goddette et al., 1992).

For a distinct further subdivision of true subtilisins, high-alkaline subtilisins and PIS into classes, supporting biochemical data might be necessary. However, based on the phylogenetic

tree, the deletion and insertion analysis, and the isoelectric point, a further subdivision into classes is most likely.

## Conclusion

Phylogenetic studies of the S8 family within the MEROPS holotype dataset revealed a large number of different subtilases forming new groups and subgroups. In addition to the known groups of proteinase K, pyrolysins, kexins, subtilisins, thermitases and lantibiotic peptidases, the analysis revealed new groups or subgroups within the S8A subfamily depending on their position in the phylogenetic tree, their biochemical properties or their origin. This analysis was used in the second part of this study to categorise 120 newly identified predicted S8 protease sequences derived from *Bacillaceae*. They were found to represent the three main subgroups within the subtilisins, the true subtilisins, the high-alkaline subtilisins, and the phylogenetically intermediate subtilisins. However, without the specified filter parameters for data mining, more new subtilases outside the group of subtilisins from *Bacillaceae* could probably be found. In the absence of experimental characterisation for most of the found subtilisin sequences, a subdivision needs further experimental studies, because with bioinformatic analysis alone, a prediction of their biological and biochemical properties is possible only to a limited extent. For the newly found enzymes it is thus possible that they possess unique specificities and are of high interest for biotechnological applications.

## Data availability statement

The original contributions presented in this study are included in the article/**Supplementary material**, further inquiries can be directed to the corresponding author.

## References

- Agarwal, V., Pierce, E., McIntosh, J., Schmidt, E. W., and Nair, S. K. (2012). Structures of cyanobactin maturation enzymes define a family of transamidating proteases. *Chem. Biol.* 19, 1411–1422. doi: 10.1016/j.chembiol.2012.09.012
- Ageitos, J. M., Baker, P. J., Sugahara, M., and Numata, K. (2013). Proteinase K-catalyzed synthesis of linear and star oligo(L-phenylalanine) conjugates. *Biomacromolecules* 14, 3635–3642. doi: 10.1021/bm4009974
- Ali, T., Oldfield, N. J., Wooldridge, K. G., Turner, D. P., and Ala'Aldeen, D. A. A. (2008). Functional characterization of AasP, a maturation protease autotransporter protein of *Actinobacillus pleuropneumoniae*. *Infect. Immun.* 76, 5608–5614. doi: 10.1128/IAI.00085-08
- Almog, O., González, A., Godin, N., Leeuw, M., de Mekele, M. J., Klein, D., et al. (2009). The crystal structures of the psychrophilic subtilisin S41 and the mesophilic subtilisin Sph reveal the same calcium-loaded state. *Proteins* 74, 489–496. doi: 10.1002/prot.22175
- Anjard, C., and Loomis, W. F. (2005). Peptide signaling during terminal differentiation of Dictyostelium. *Proc. Natl. Acad. Sci. U.S.A.* 102, 7607–7611. doi: 10.1073/pnas.0501820102
- Azrin, N. A. M., Ali, M. S. M., Rahman, R. N. Z. R. A., Oslan, S. N., and Noor, N. D. M. (2022). Versatility of subtilisin: A review on structure, characteristics, and applications. *Biotechnol. Appl. Biochem.* [Epub ahead of print]. doi: 10.1002/bab.2309
- Bethe, G., Nau, R., Wellmer, A., Hakenbeck, R., Reinert, R. R., Heinz, H. P., et al. (2001). The cell wall-associated serine protease PrtA: A highly conserved virulence factor of *Streptococcus pneumoniae*. *FEMS Microbiol. Lett.* 205, 99–104. doi: 10.1111/j.1574-6968.2001.tb10931.x
- Betzel, C. (2004). "Thermitase," in *Handbook of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3167–3169. doi: 10.1016/B978-0-12-382219-2.00696-7

## Author contributions

FF collected and analyzed the data and wrote the original draft. JB, MB, and PS supervised the study and revised the manuscript. All authors contributed to the final manuscript.

## Funding

This work was supported by the FH Aachen University of Applied Sciences within the framework of the program for the promotion of young scientists.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.1017978/full#supplementary-material>

- Betzel, C., Klupsch, S., Branner, S., and Wilson, K. S. (1996). Crystal structures of the alkaline proteases savinase and esperase from *Bacillus lentus*. *Adv. Exp. Med. Biol.* 379, 49–61. doi: 10.1007/978-1-4613-0319-0\_7
- Betzel, C., Klupsch, S., Papendorf, G., Hastrup, S., Branner, S., and Wilson, K. S. (1992). Crystal structure of the alkaline proteinase Savinase from *Bacillus lentus* at 1.4 Å resolution. *J. Mol. Biol.* 223, 427–445. doi: 10.1016/0022-2836(92)90662-4
- Betzel, C., Teplyakov, A. V., Harutyunyan, E. H., Saenger, W., and Wilson, K. S. (1990). Thermitase and proteinase K: A comparison of the refined three-dimensional structures of the native enzymes. *Protein Eng.* 3, 161–172. doi: 10.1093/protein/3.3.161
- Bierbaum, G., Jack, R. W., and Sahl, H.-G. (2013). “Lantibiotic leader peptidases,” in *Handbook of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3220–3222. doi: 10.1016/B978-0-12-382219-2.00709-2
- Blumentals, I. I., Robinson, A. S., and Kelly, R. M. (1990). Characterization of sodium dodecyl sulfate-resistant proteolytic activity in the hyperthermophilic archaeobacterium *Pyrococcus furiosus*. *Appl. Environ. Microbiol.* 56, 1992–1998. doi: 10.1128/aem.56.7.1992-1998.1990
- Borhan, B., Hammock, B., Seifert, J., and Wilson, B. W. (1996). Methyl and phenyl esters and thioesters of carboxylic acids as surrogate substrates for microassay of proteinase K esterase activity. *Anal. Bioanal. Chem.* 354, 490–492. doi: 10.1007/s0021663540490
- Broadbent, J. R., and Steele, J. L. (2013). “Lactocepine: The cell envelope-associated endopeptidase of lactococci,” in *Handbook of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press.), 3188–3195. doi: 10.1016/B978-0-12-382219-2.00703-1
- Brown, G. D., and Reece, K. S. (2003). Isolation and characterization of serine protease gene(s) from *Perkinsus marinus*. *Dis. Aquatic Organisms* 57, 117–126. doi: 10.3354/dao057117
- Butler, M. J. (2013). “Tripeptidyl-peptidase S,” in *Handbook Of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3222–3223. doi: 10.1016/B978-0-12-382219-2.00710-9
- Chen, X.-L., Xie, B.-B., Lu, J.-T., He, H.-L., and Zhang, Y. (2007). A novel type of subtilase from the psychrotolerant bacterium *Pseudoalteromonas* sp. SM9913: Catalytic and structural properties of deseasin MCP-01. *Microbiology* 153, 2116–2125. doi: 10.1099/mic.0.2007/006056-0
- Cheng, G., Zhao, P., Tang, X.-F., and Tang, B. (2009). Identification and characterization of a novel spore-associated subtilase from *Thermoactinomyces* sp. CDF. *Microbiology* 155, 3661–3672. doi: 10.1099/mic.0.031336-0
- Christiansen, T., Michaelsen, S., Wümpelmann, M., and Nielsen, J. (2003). Production of savinase and population viability of *Bacillus clausii* during high-cell-density fed-batch cultivations. *Biotechnol. Bioeng.* 83, 344–352. doi: 10.1002/bit.10675
- Contesini, F. J., Melo, R.R., and Sato, H. H. (2018). An overview of *Bacillus* proteases: From production to application. *Crit. Rev. Biotechnol.* 38, 321–334. doi: 10.1080/07388551.2017.1354354
- Coutte, L., Antoine, R., Drobecq, H., Loch, C., and Jacob-Dubuisson, F. (2001). Subtilisin-like autotransporter serves as maturation protease in a bacterial secretion pathway. *EMBO J.* 20, 5040–5048. doi: 10.1093/emboj/20.18.5040
- Dave, J. A., van Gey Pittius, N. C., Beyers, A. D., Ehlers, M. R. W., and Brown, G. D. (2002). Mycosin-1, a subtilisin-like serine protease of *Mycobacterium tuberculosis*, is cell wall-associated and expressed during infection of macrophages. *BMC Microbiol.* 2:30. doi: 10.1186/1471-2180-2-30
- Denizci, A. A., Kazan, D., and Erarslan, A. (2010). *Bacillus marmarensis* sp. nov., an alkaliphilic, protease-producing bacterium isolated from mushroom compost. *Int. J. Syst. Evol. Microbiol.* 60, 1590–1594. doi: 10.1099/ijs.0.012369-0
- Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., et al. (2008). Phylogeny.fr: Robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 36:W465–W469. doi: 10.1093/nar/gkn180
- Dessimoz, C., and Gil, M. (2010). Phylogenetic assessment of alignments reveals neglected tree signal in gaps. *Genome Biol.* 11:R37. doi: 10.1186/gb-2010-11-4-r37
- Dixit, M., Murudkar, C. S., and Rao, K. K. (2002). Epr is transcribed from a final sigma(D) promoter and is involved in swarming of *Bacillus subtilis*. *J. Bacteriol.* 184, 596–599. doi: 10.1128/JB.184.2.596-599.2002
- Ebeling, W., Hennrich, N., Klockow, M., Metz, H., Orth, H. D., and Lang, H. (1974). Proteinase K from *Tritirachium album* Limber. *Eur. J. Biochem.* 47, 91–97. doi: 10.1111/j.1432-1033.1974.tb03671.x
- Eder, J., Rheinhecker, M., and Fersht, A. R. (1993). Folding of subtilisin BPN<sup>+</sup>: Role of the pro-sequence. *J. Mol. Biol.* 233, 293–304. doi: 10.1006/jmbi.1993.1507
- Falkenberg, F., Rahba, J., Fischer, D., Bott, M., Bongaerts, J., and Siegert, P. (2022). Biochemical characterization of a novel oxidatively stable, halotolerant and high-alkaline subtilisin from *Alkalihalobacillus okhensis* Kh10-101<sup>T</sup>. *FEBS Open Bio.* [Epub ahead of print]. doi: 10.1002/2211-5463.13457
- Faraco, V., Palmieri, G., Festa, G., Monti, M., Sannia, G., and Giardina, P. (2005). A new subfamily of fungal subtilases: Structural and functional analysis of a *Pleurotus ostreatus* member. *Microbiology* 151, 457–466. doi: 10.1099/mic.0.27441-0
- Frömmel, C., Hausdorf, G., Höhne, W. E., Behnke, U., and Ruttloff, H. (1978). Characterization of a protease from *Thermoactinomyces vulgaris* (thermitase). *Acta Biol. Medica Germanica* 37, 1193–1204.
- Fuller, R. S. (2013). “Kexin,” in *Handbook of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3270–3277. doi: 10.1016/B978-0-12-382219-2.00722-5
- Garcia-Sanchez, A., Cerrato, R., Larrasa, J., Ambrose, N. C., Parra, A., Alonso, J. M., et al. (2004). Characterisation of an extracellular serine protease gene (nasp gene) from *Dermatophilus congolensis*. *FEMS Microbiol. Lett.* 231, 53–57. doi: 10.1016/S0378-1097(03)00958-3
- Gavrilov, Y., Dagan, S., and Levy, Y. (2015). Shortening a loop can increase protein native state entropy. *Proteins* 83, 2137–2146. doi: 10.1002/prot.24926
- Genay, M., Sadat, L., Gagnaire, V., and Lortal, S. (2009). prtH2, not prtH, is the ubiquitous cell wall proteinase gene in *Lactobacillus helveticus*. *Appl. Environ. Microbiol.* 75, 3238–3249. doi: 10.1128/AEM.02395-08
- Gilbert, C., Atlan, D., Blanc, B., Portailier, R., Germond, J. E., Lapiere, L., et al. (1996). A new cell surface proteinase: Sequencing and analysis of the prtB gene from *Lactobacillus delbruekii* subsp. bulgaricus. *J. Bacteriol.* 178, 3059–3065. doi: 10.1128/jb.178.11.3059-3065.1996
- Goddette, D. W., Paech, C., Yang, S. S., Mielenz, J. R., Bystroff, C., Wilke, M. E., et al. (1992). The crystal structure of the *Bacillus lentus* alkaline protease, subtilisin BL, at 1.4 Å resolution. *J. Mol. Biol.* 228, 580–595. doi: 10.1016/0022-2836(92)90843-9
- Goldenberger, D., Perschil, I., Ritzler, M., and Altwegg, M. (1995). A simple “universal” DNA extraction procedure using SDS and proteinase K is compatible with direct PCR amplification. *PCR Methods Appl.* 4, 368–370. doi: 10.1101/gr.4.6.368
- Good, J. R., Cabral, M., Sharma, S., Yang, J., van Driessche, N., Shaw, C. A., et al. (2003). TagA, a putative serine protease/ABC transporter of *Dictyostelium* that is required for cell fate determination at the onset of development. *Development* 130, 2953–2965. doi: 10.1242/dev.00523
- Gunkel, F. A., and Gassen, H. G. (1989). Proteinase K from *Tritirachium album* Limber. Characterization of the chromosomal gene and expression of the cDNA in *Escherichia coli*. *Eur. J. Biochem.* 179, 185–194. doi: 10.1111/j.1432-1033.1989.tb14539.x
- Güntelberg, A. V., and Ottesen, M. (1952). Preparation of crystals containing the plalalbumin-forming enzyme from *Bacillus subtilis*. *Nature* 170:802. doi: 10.1038/170802a0
- Hackett, F., Sajid, M., Withers-Martinez, C., Grainger, M., and Blackman, M. J. (1999). PfSUB-2: A second subtilisin-like protein in *Plasmodium falciparum* merozoites. *Mol. Biochem. Parasitol.* 103, 183–195. doi: 10.1016/S0166-6851(99)00122-X
- Hageman, J. H. (2013). “Bacillopeptidase F” in *Handbook Of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3170–3171. doi: 10.1016/B978-0-12-382219-2.00697-9
- Henderson, I. R., and Nataro, J. P. (2001). Virulence functions of autotransporter proteins. *Infect. Immun.* 69, 1231–1243. doi: 10.1128/IAI.69.3.1231-1243.2001
- Hoang, D. T., Chernomor, O., Haeseler, A., von Minh, B. Q., and Le Vinh, S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281
- Hu, G., and Leger, R. J. S. (2004). A phylogenomic approach to reconstructing the diversification of serine proteases in fungi. *J. Evol. Biol.* 17, 1204–1214. doi: 10.1111/j.1420-9101.2004.00786.x
- Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010). CD-HIT Suite: A web server for clustering and comparing biological sequences. *Bioinformatics* 26, 680–682. doi: 10.1093/bioinformatics/btq003
- Ito, S., Kobayashi, T., Ara, K., Ozaki, K., Kawai, S., and Hatada, Y. (1998). Alkaline detergent enzymes from alkaliphiles: Enzymatic properties, genetics, and structures. *Extremophiles* 2, 185–190. doi: 10.1007/s007920050059
- Itoi, Y., Horinaka, M., Tsujimoto, Y., Matsui, H., and Watanabe, K. (2006). Characteristic features in the structure and collagen-binding ability of a thermophilic collagenolytic protease from the thermophile *Geobacillus collagenovorans* MO-1. *J. Bacteriol.* 188, 6572–6579. doi: 10.1128/JB.00767-06
- Jang, H. J., Kim, B. C., Pyun, Y. R., and Kim, Y. S. (2002). A novel subtilisin-like serine protease from *Thermoanaerobacter yonseiensis* KB-1: Its cloning,

- expression, and biochemical properties. *Extremophiles* 6, 233–243. doi: 10.1007/s00792-001-0248-1
- Jang, W. H., Kim, E. K., Lee, H. B., Chung, J. H., and Yoo, O. J. (1996). Characterization of an alkaline serine protease from an alkaline-resistant *Pseudomonas* sp.: Cloning and expression of the protease gene in *Escherichia coli*. *Biotechnol. Lett.* 18, 57–62. doi: 10.1007/BF00137811
- Joshi, A., Thite, S., Karodi, P., Joseph, N., and Lodha, T. (2021). *Alkalihalobacterium elongatum* gen. nov. sp. nov.: An antibiotic-producing bacterium isolated from lonar lake and reclassification of the genus *Alkalihalobacillus* into seven novel genera. *Front. Microbiol.* 12:722369. doi: 10.3389/fmicb.2021.722369
- Joshi, S., and Satyanarayana, T. (2013). Characteristics and applications of a recombinant alkaline serine protease from a novel bacterium *Bacillus lehensis*. *Bioresour. Technol.* 131, 76–85. doi: 10.1016/j.biortech.2012.12.124
- Kagawa, T. F., and Cooney, J. C. (2013). “C5a Peptidase,” in *Handbook Of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3202–3208. doi: 10.1016/B978-0-12-382219-2.00705-5
- Kalisz, H. M. (1988). Microbial proteinases. *Adv. Biochem. Eng. Biotechnol.* 36, 1–65. doi: 10.1007/BFb0047944
- Kalyanamoorthy, S., Minh, B. Q., Wong, T. K. F., Haeseler, A., and Jermini, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285
- Kamekura, M., Seno, Y., and Dyll-Smith, M. (1996). Halolysin R4, a serine proteinase from the halophilic archaeon *Haloflex mediterranei*; gene cloning, expression and structural studies. *Biochim. Biophysica Acta* 1294, 159–167. doi: 10.1016/0167-4838(96)00016-7
- Kaneda, M., and Tominaga, N. (1975). Isolation and characterization of a proteinase from the sarcocarp of melon fruit. *J. Biochem.* 78, 1287–1296. doi: 10.1093/oxfordjournals.jbchem.a131026
- Kannan, Y., Koga, Y., Inoue, Y., Haruki, M., Takagi, M., Imanaka, T., et al. (2001). Active subtilisin-like protease from a hyperthermophilic archaeon in a form with a putative prosequence. *Appl. Environ. Microbiol.* 67, 2445–2452. doi: 10.1128/AEM.67.6.2445-2452.2001
- Karlsson, C., Andersson, M.-L., Collin, M., Schmidtchen, A., Björck, L., and Frick, I.-M. (2007). SufA—a novel subtilisin-like serine proteinase of *Finegoldia magna*. *Microbiology* 153, 4208–4218. doi: 10.1099/mic.0.2007/010322-0
- Kato, K., Rozewicki, J., and Yamada, K. D. (2019). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinf.* 20, 1160–1166. doi: 10.1093/bib/bbx108
- Kato, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kida, Y., Taira, J., Yamamoto, T., Higashimoto, Y., and Kuwano, K. (2013). EprS, an autotransporter protein of *Pseudomonas aeruginosa*, possessing serine protease activity induces inflammatory responses through protease-activated receptors. *Cell. Microbiol.* 15, 1168–1181. doi: 10.1111/cmi.12106
- Klingeberg, M., Galunsky, B., Sjöholm, C., Kasche, V., and Antranikian, G. (1995). Purification and properties of a highly thermostable, sodium dodecyl sulfate-resistant and stereospecific proteinase from the extremely *Thermophilic Archaeon Thermococcus stetteri*. *Appl. Environ. Microbiol.* 61, 3098–3104. doi: 10.1128/aem.61.8.3098-3104.1995
- Kobayashi, H., Yoshida, T., Miyakawa, T., Tashiro, M., Okamoto, K., Yamanaka, H., et al. (2015). Structural basis for action of the external chaperone for a propeptide-deficient serine protease from *Aeromonas sobria*. *J. Biol. Chem.* 290, 11130–11143. doi: 10.1074/jbc.M114.622852
- Kobayashi, T., Lu, J., Li, Z., Hung, V. S., Kurata, A., Hatada, Y., et al. (2007). Extremely high alkaline protease from a deep-subsurface bacterium, *Alkaliphilus transvaalensis*. *Appl. Microbiol. Biotechnol.* 75, 71–80. doi: 10.1007/s00253-006-0800-0
- Koide, Y., Nakamura, A., Uozumi, T., and Beppu, T. (1986). Cloning and sequencing of the major intracellular serine protease gene of *Bacillus subtilis*. *J. Bacteriol.* 167, 110–116. doi: 10.1128/jb.167.1.110-116.1986
- Koma, D., Yamanaka, H., Moriyoshi, K., Ohmoto, T., and Sakai, K. (2007). Overexpression and characterization of thermostable serine protease in *Escherichia coli* encoded by the ORF TTE0824 from *Thermoanaerobacter tengcongensis*. *Extremophiles* 11, 769–779. doi: 10.1007/s00792-007-0103-0
- Kong, H.-H., Kim, T.-H., and Chung, D.-I. (2000). Purification and characterization of a secretory serine proteinase of *Acanthamoeba healyi* isolated from Gae. *J. Parasitol.* 86, 12–17. doi: 10.1645/0022-3395(2000)086[0012:PACOAS]2.0.CO;2
- Kubo, T., Kaida, S., Abe, J., Saito, T., Fukuzawa, H., and Matsuda, Y. (2009). The *Chlamydomonas* hatching enzyme, sporangin, is expressed in specific phases of the cell cycle and is localized to the flagella of daughter cells within the sporangial cell wall. *Plant Cell Physiol.* 50, 572–583. doi: 10.1093/pcp/pcp016
- Kuhner, M. K., and Felsenstein, J. (1994). A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Mol. Biol. Evol.* 11, 459–468.
- Lawrenson, R. A., and Sriskandan, S. (2013). “Cell envelope proteinase A (*Streptococcus*),” in *Handbook Of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3195–3202. doi: 10.1016/B978-0-12-382219-2.00704-3
- Lee, J., McIntosh, J., Hathaway, B. J., and Schmidt, E. W. (2009). Using marine natural products to discover a protease that catalyzes peptide macrocyclization of diverse substrates. *J. Am. Chem. Soc.* 131, 2122–2124. doi: 10.1021/ja8092168
- Leger, R., Charnley, A. K., and Cooper, R. M. (1987). Characterization of cuticle-degrading proteases produced by the entomopathogen *Metarhizium anisopliae*. *Arch. Biochem. Biophys.* 253, 221–232. doi: 10.1016/0003-9861(87)90655-2
- Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49:W293–W296. doi: 10.1093/nar/gkab301
- Li, J., Gu, F., Wu, R., Yang, J., and Zhang, K.-Q. (2017). Phylogenomic evolutionary surveys of subtilase superfamily genes in fungi. *Sci. Rep.* 7:45456. doi: 10.1038/srep45456
- Li, J., Yu, L., Yang, J., Dong, L., Tian, B., Yu, Z., et al. (2010). New insights into the evolution of subtilisin-like serine protease genes in Pezizomycotina. *BMC Evol. Biol.* 10:68. doi: 10.1186/1471-2148-10-68
- Lilley, G. G., Stewart, D. J., and Kortt, A. A. (1992). Amino acid and DNA sequences of an extracellular basic protease of *Dichelobacter nodosus* show that it is a member of the subtilisin family of proteases. *Eur. J. Biochem.* 210, 13–21. doi: 10.1111/j.1432-1033.1992.tb17385.x
- Linderström-Lang, K., and Ottesen, M. (1947). A new protein from ovalbumin. *Nature* 159:807. doi: 10.1038/159807a0
- Liu, G.-H., Narsing Rao, M. P., Dong, Z.-Y., Wang, J.-P., Chen, Z., Liu, B., et al. (2019). Two novel alkaliphiles, *Bacillus alkalisolis* sp. nov., and *Bacillus solitudinis* sp. nov., isolated from saline-alkali soil. *Extremophiles* 23, 759–764. doi: 10.1007/s00792-019-01127-2
- Liu, J., Zhang, X., Cao, H., Guo, L., Zhao, B., Zhang, X., et al. (2022). *Alteribacter keqinensis* sp. nov., a moderately halophilic bacterium isolated from a soda lake. *Int. J. Syst. Evol. Microbiol.* 72. doi: 10.1099/ijsem.0.005351
- Maeda, H., Mizutani, O., Yamagata, Y., Ichishima, E., and Nakajima, T. (2001). Alkaline-resistance model of subtilisin ALP I, a novel alkaline subtilisin. *J. Biochem.* 129, 675–682. doi: 10.1093/oxfordjournals.jbchem.a002906
- Margot, P., and Karamata, D. (1996). The wprA gene of *Bacillus subtilis* 168, expressed during exponential growth, encodes a cell-wall-associated protease. *Microbiology* 142, 3437–3444. doi: 10.1099/13500872-142-12-3437
- Markland, F. S., and Smith, E. L. (1971). “16 Subtilisins: Primary structure, chemical and physical properties,” in *The Enzymes*, ed. P. D. Boyer (New York, NY: Academic Press), 3, 561–608. doi: 10.1016/S1874-6047(08)60407-2
- Masayama, A., Hamasaki, K., Urakami, K., Shimamoto, S., Kato, S., Makino, S., et al. (2006). Expression of germination-related enzymes, CspA, CspB, CspC, SleC, and SleM, of *Clostridium perfringens* S40 in the mother cell compartment of sporulating cells. *Genes Genetic Syst.* 81, 227–234. doi: 10.1266/ggs.81.227
- Masui, A., Fujiwara, N., and Imanaka, T. (1994). Stabilization and rational design of serine protease AprM under highly alkaline and high-temperature conditions. *Appl. Environ. Microbiol.* 60, 3579–3584. doi: 10.1128/aem.60.10.3579-3584.1994
- Masui, A., Fujiwara, N., Yamamoto, K., Takagi, M., and Imanaka, T. (1998). Rational design for stabilization and optimum pH shift of serine protease AprN. *J. Ferment. Bioeng.* 85, 30–36. doi: 10.1016/S0922-338X(97)80349-2
- Matsubara, H., Kasper, C. B., Brown, D. M., and Smith, E. L. (1965). Subtilisin BPN<sup>+</sup>. *J. Biol. Chem.* 240, 1125–1130. doi: 10.1016/S0021-9258(18)97548-4
- Maurer, K.-H. (2004). Detergent proteases. *Curr. Opin. Biotechnol.* 15, 330–334. doi: 10.1016/j.copbio.2004.06.005
- Mayr, J., Lupas, A., Kellermann, J., Eckerskorn, C., Baumeister, W., and Peters, J. (1996). A hyperthermostable protease of the subtilisin family bound to the surface layer of the archaeon *Staphylothermus marinus*. *Curr. Biol.* 6, 739–749. doi: 10.1016/S0960-9822(09)00455-2
- Møllergaard, S. (1983). Purification and characterization of a new proteolytic enzyme produced by *Aeromonas salmonicida*. *J. Appl. Bacteriol.* 54, 289–294. doi: 10.1111/j.1365-2672.1983.tb02619.x
- Miller, S. A., Binder, E. M., Blackman, M. J., Carruthers, V. B., and Kim, K. (2001). A conserved subtilisin-like protein TgSUB1 in microneme organelles

- of *Toxoplasma gondii*. *J. Biol. Chem.* 276, 45341–45348. doi: 10.1074/jbc.M106665200
- Miller, S. A., Thathy, V., Ajioka, J. W., Blackman, M. J., and Kim, K. (2003). TgSUB2 is a *Toxoplasma gondii* rhoptry organelle processing proteinase. *Mol. Microbiol.* 49, 883–894. doi: 10.1046/j.1365-2958.2003.03604.x
- Montero, E., Gonzalez, L. M., Rodriguez, M., Oksov, Y., Blackman, M. J., and Lobo, C. A. (2006). A conserved subtilisin protease identified in *Babesia divergens* merozoites. *J. Biol. Chem.* 281, 35717–35726. doi: 10.1074/jbc.M604344200
- Morita, Y., Hasan, Q., Sakaguchi, T., Murakami, Y., Yokoyama, K., and Tamiya, E. (1998). Properties of a cold-active protease from psychrotrophic *Flavobacterium balustinum* P104. *Appl. Microbiol. Biotechnol.* 50, 669–675. doi: 10.1007/s002530051349
- Muszewska, A., Taylor, J. W., Szczesny, P., and Grynberg, M. (2011). Independent subtilases expansions in fungi associated with animals. *Mol. Biol. Evol.* 28, 3395–3404. doi: 10.1093/molbev/msr176
- Nakamura, K., Matsushima, A., and Horikoshi, K. (1973). The state of amino acid residues in alkaline protease produced by *Bacillus* No. 221. *Agricul. Biol. Chem.* 37, 1261–1267. doi: 10.1080/00021369.1973.10860834
- Okamoto, M., Yonejima, Y., Tsujimoto, Y., Suzuki, Y., and Watanabe, K. (2001). A thermostable collagenolytic protease with a very large molecular mass produced by thermophilic *Bacillus* sp. strain MO-1. *Appl. Microbiol. Biotechnol.* 57, 103–108. doi: 10.1007/s002530100731
- Okuda, M., Sumitomo, N., Takimura, Y., Ogawa, A., Saeki, K., Kawai, S., et al. (2004). A new subtilisin family: Nucleotide and deduced amino acid sequences of new high-molecular-mass alkaline proteases from *Bacillus* spp. *Extremophiles* 8, 229–235. doi: 10.1007/s00792-004-0381-8
- Page, M. J., and Di Cera, E. (2008). Serine peptidases: Classification, structure and function. *Cell. Mol. Life Sci.* 65, 1220–1236. doi: 10.1007/s00018-008-7565-9
- Pastar, I., Tonic, I., Golic, N., Kojic, M., van Kranenburg, R., Kleerebezem, M., et al. (2003). Identification and genetic characterization of a novel proteinase, PrTR, from the human isolate *Lactobacillus rhamnosus* BGT10. *Appl. Environ. Microbiol.* 69, 5802–5811. doi: 10.1128/AEM.69.10.5802-5811.2003
- Phrommao, E., Yongsawatdigul, J., Rodtong, S., and Yamabhai, M. (2011). A novel subtilase with NaCl-activated and oxidant-stable activity from *Virgibacillus* sp. SK37. *BMC Biotechnol.* 11:65. doi: 10.1186/1472-6750-11-65
- Power, S. D., Adams, R. M., and Wells, J. A. (1986). Secretion and autoproteolytic maturation of subtilisin. *Proc. Natl. Acad. Sci. U.S.A.* 83, 3096–3100. doi: 10.1073/pnas.83.10.3096
- Que, X. C., and Kuramitsu, H. K. (1990). Isolation and characterization of the *Treponema denticola* prtA gene coding for chymotrypsin like protease activity and detection of a closely linked gene encoding PZ-PLGPA-hydrolyzing activity. *Infect. Immun.* 58, 4099–4105. doi: 10.1128/iai.58.12.4099-4105.1990
- Rawlings, N. D. (2013). Identification and prioritization of novel uncharacterized peptidases for biochemical characterization. *Database* 2013:bat022. doi: 10.1093/database/bat022
- Rawlings, N. D. (2020). Twenty-five years of nomenclature and classification of proteolytic enzymes. *Biochim. Biophys. Acta Proteins Proteom* 1868:140345. doi: 10.1016/j.bbapap.2019.140345
- Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., and Finn, R. D. (2018). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* 46:D624–D632. doi: 10.1093/nar/gkx1134
- Rawlings, N. D., and Bateman, A. (2021). How to use the MEROPS database and website to help understand peptidase specificity. *Protein Sci.* 30, 83–92. doi: 10.1002/pro.3948
- Rawlings, N. D., Waller, M., Barrett, A. J., and Bateman, A. (2014). MEROPS: The database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* 42:D503–D509. doi: 10.1093/nar/gkt953
- Renn, S. C., Tomkinson, B., and Taghert, P. H. (1998). Characterization and cloning of tripeptidyl peptidase II from the fruit fly, *Drosophila melanogaster*. *J. Biol. Chem.* 273, 19173–19182. doi: 10.1074/jbc.273.30.19173
- Reysset, G., and Millet, J. (1972). Characterization of an intracellular protease in *B. subtilis* during sporulation. *Biochem. Biophys. Res. Commun.* 49, 328–334. doi: 10.1016/0006-291X(72)90414-7
- Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42:W320–W324. doi: 10.1093/nar/gku316
- Rogers, D. T., Saville, D., and Bussey, H. (1979). *Saccharomyces cerevisiae* killer expression mutant *kek2* has altered secretory proteins and glycoproteins. *Biochem. Biophys. Res. Commun.* 90, 187–193. doi: 10.1016/0006-291X(79)91607-3
- Rolbetzki, A., Ammon, M., Jakovljevic, V., Konovalova, A., and Søgaard-Andersen, L. (2008). Regulated secretion of a protease activates intercellular signaling during fruiting body formation in *M. xanthus*. *Dev. Cell* 15, 627–634. doi: 10.1016/j.devcel.2008.08.002
- Saeki, K., Hitomi, J., Okuda, M., Hatada, Y., Kageyama, Y., Takaiwa, M., et al. (2002). A novel species of alkaliphilic *Bacillus* that produces an oxidatively stable alkaline serine protease. *Extremophiles* 6, 65–72. doi: 10.1007/s007920100224
- Saeki, K., Magallones, M. V., Takimura, Y., Hatada, Y., Kobayashi, T., Kawai, S., et al. (2003). Nucleotide and deduced amino acid sequences of a new subtilisin from an alkaliphilic *Bacillus* isolate. *Curr. Microbiol.* 47, 337–340. doi: 10.1007/s00284-002-4018-9
- Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H., et al. (2000). Novel oxidatively stable subtilisin-like serine proteases from alkaliphilic *Bacillus* spp.: Enzymatic properties, sequences, and evolutionary relationships. *Biochem. Biophys. Res. Commun.* 279, 313–319. doi: 10.1006/bbr.2000.3931
- Sahl, H. G., Jack, R. W., and Bierbaum, G. (1995). Biosynthesis and biological activities of lantibiotics with unique post-translational modifications. *Eur. J. Biochem.* 230, 827–853. doi: 10.1111/j.1432-1033.1995.tb20627.x
- Sakaguchi, M. (2013). “Aqualysin I,” in *Handbook Of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3174–3176. doi: 10.1016/B978-0-12-382219-2.00699-2
- Sánchez, R., Serra, F., Tárraga, J., Medina, I., Carbonell, J., and Pulido, L. (2011). Phylemon 2.0: A suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. *Nucleic Acids Res.* 39:W470–W474. doi: 10.1093/nar/gkr408
- Sanderson, M. J., and Shaffer, H. B. (2002). Troubleshooting Molecular Phylogenetic Analyses. *Annu. Rev. Ecol. Syst.* 33, 49–72. doi: 10.1146/annurev.ecolsys.33.010802.150509
- Schaller, A. (2013). “Plant Subtilisins,” in *Handbook Of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3247–3254. doi: 10.1016/B978-0-12-382219-2.00717-1
- Seidah, N. G. (2013a). “Proprotein Convertase PCSK9,” in *Handbook Of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3315–3322. doi: 10.1016/B978-0-12-382219-2.00732-8
- Seidah, N. G. (2013b). “Site-1 Protease,” in *Handbook Of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3265–3270. doi: 10.1016/B978-0-12-382219-2.00721-3
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75
- Siezen, R. J., and Leunissen, J. A. (1997). Subtilases: The superfamily of subtilisin-like serine proteases. *Protein Sci.* 6, 501–523. doi: 10.1002/pro.5560060301
- Siezen, R. J., Renckens, B., and Boekhorst, J. (2007). Evolution of prokaryotic subtilases: Genome-wide analysis reveals novel subfamilies with different catalytic residues. *Proteins* 67, 681–694. doi: 10.1002/prot.21290
- Siezen, R. J., Vos, W. M., de Leunissen, J. A., and Dijkstra, B. W. (1991). Homology modelling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteinases. *Protein Eng.* 4, 719–737. doi: 10.1093/protein/4.7.719
- Silmon de Monerri, N. C., Ruecker, A., and Blackman, M. J. (2013). “Plasmodium Subtilisins,” in *Handbook Of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3260–3265. doi: 10.1016/B978-0-12-382219-2.00720-1
- Smith, E. L., Markland, F. S., Kasper, C. B., DeLange, R. J., Landon, M., and Evans, W. H. (1966). The complete amino acid sequence of two types of subtilisin, BPN’ and carlsberg. *J. Biol. Chem.* 241, 5974–5976. doi: 10.1016/S0021-9258(18)96365-9
- Solanki, P., Putatunda, C., Kumar, A., Bhatia, R., and Walia, A. (2021). Microbial proteases: Ubiquitous enzymes with innumerable uses. *3 Biotech* 11:428. doi: 10.1007/s13205-021-02928-z
- Spanka, R., and Fritze, D. (1993). *Bacillus cohnii* sp. nov., a new, obligately alkaliphilic, oval-spore-forming *Bacillus* species with ornithine and aspartic acid instead of diaminopimelic acid in the cell wall. *Int. J. Syst. Bacteriol.* 43, 150–156. doi: 10.1099/00207173-43-1-150
- Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *BioTechniques* 28, 1102–1104. doi: 10.2144/002861r01
- Suzuki, M., Taguchi, S., Yamada, S., Kojima, S., Miura, K. I., and Momose, H. (1997). A novel member of the subtilisin-like protease family from *Streptomyces albobriseolus*. *J. Bacteriol.* 179, 430–438. doi: 10.1128/jb.179.2.430-438.1997

- Sweeney, P. J., and Walker, J. M. (1993). Proteinase K (EC 3.4.21.14). *Methods Mol. Biol.* 16, 305–311. doi: 10.1385/0-89603-234-5:305
- Takami, H., Akiba, T., and Horikoshi, K. (1990). Characterization of an alkaline protease from *Bacillus* sp. no. AH-101. *Appl. Microbiol. Biotechnol.* 33, 519–523. doi: 10.1007/BF00172544
- Takimura, Y., Saito, K., Okuda, M., Kageyama, Y., Saeki, K., Ozaki, K., et al. (2007). Alkaliphilic *Bacillus* sp. strain KSM-LD1 contains a record number of subtilisin-like serine proteases genes. *Appl. Microbiol. Biotechnol.* 76, 395–405. doi: 10.1007/s00253-007-1022-9
- Tan, G., Muffato, M., Ledergerber, C., Herrero, J., Goldman, N., Gil, M., et al. (2015). Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. *Syst. Biol.* 64, 778–791. doi: 10.1093/sysbio/syv033
- Taylor, A., and Qiu, Y.-L. (2017). Evolutionary history of subtilases in land plants and their involvement in symbiotic interactions. *Mol. Plant Microbe Interact.* 30, 489–501. doi: 10.1094/MPMI-10-16-0218-R
- Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gislason, M. H., Pihl, S. I., Tsirigos, K. D., et al. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* 40, 1023–1025. doi: 10.1038/s41587-021-01156-3
- Tjalma, H., Bolhuis, A., Jongbloed, J. D., Bron, S., and van Dijk, J. M. (2000). Signal peptide-dependent protein transport in *Bacillus subtilis*: A genome-based survey of the secretome. *Mol. Biol. Rev.* 64, 515–547. doi: 10.1128/MMBR.64.3.515-547.2000
- Toogood, H. S., Smith, C. A., Baker, E. N., and Daniel, R. M. (2000). Purification and characterization of Ak.1 protease, a thermostable subtilisin with a disulphide bond in the substrate-binding cleft. *Biochem. J.* 350, 321–328. doi: 10.1042/bj3500321
- Trifinopoulos, J., Nguyen, L.-T., Haeseler, A., and von Minh, B. Q. (2016). W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* 44:W232–W235. doi: 10.1093/nar/gkw256
- Tsuchida, O., Yamagata, Y., Ishizuka, T., Arai, T., Yamada, J.-I., Takeuchi, M., et al. (1986). An alkaline proteinase of an alkalophilic *Bacillus* sp. *Curr. Microbiol.* 14, 7–12. doi: 10.1007/BF01568094
- Tunlid, A., and Jansson, S. (1991). Proteases and their involvement in the infection and immobilization of nematodes by the nematophagous fungus *Arthrobotrys oligospora*. *Appl. Environ. Microbiol.* 57, 2868–2872. doi: 10.1128/aem.57.10.2868-2872.1991
- Turner, D. P. J., Wooldridge, K. G., and Ala'Aldeen, D. A. A. (2002). Autotransported serine protease A of *Neisseria meningitidis*: An immunogenic, surface-exposed outer membrane, and secreted protein. *Infect. Immun.* 70, 4447–4461. doi: 10.1128/IAI.70.8.4447-4461.2002
- Valbuzzi, A., Ferrari, E., and Albertini, A. M. (1999). A novel member of the subtilisin-like protease family from *Bacillus subtilis*. *Microbiology* 145, 3121–3127. doi: 10.1099/00221287-145-11-3121
- van der Laan, J. M., Teplyakov, A. V., Kelders, H., Kalk, K. H., Misset, O., Mulleners, L. J., et al. (1992). Crystal structure of the high-alkaline serine protease PB92 from *Bacillus alcalophilus*. *Protein Eng.* 5, 405–411. doi: 10.1093/protein/5.5.405
- Vilo, C., Galetovic, A., Araya, J. E., Gómez-Silva, B., and Dong, Q. (2015). Draft genome sequence of a *Bacillus* bacterium from the atacama desert wetlands metagenome. *Genome Announcements* 3:e955–e915. doi: 10.1128/genomeA.00955-15
- Völkl, P., Markiewicz, P., Stetter, K. O., and Miller, J. H. (1994). The sequence of a subtilisin-type protease (aerolysin) from the hyperthermophilic archaeum *Pyrobaculum aerophilum* reveals sites important to thermostability. *Protein Sci.* 3, 1329–1340. doi: 10.1002/pro.5560030819
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinf.* 25, 1189–1191. doi: 10.1093/bioinformatics/btp033
- Wells, J. A., Powers, D. B., Bott, R. R., Graycar, T. P., and Estell, D. A. (1987). Designing substrate specificity by protein engineering of electrostatic interactions. *Proc. Natl. Acad. Sci. U.S.A.* 84, 1219–1223. doi: 10.1073/pnas.84.5.1219
- Withers-Martinez, C., Jean, L., and Blackman, M. J. (2004). Subtilisin-like proteases of the malaria parasite. *Mol. Microbiol.* 53, 55–63. doi: 10.1111/j.1365-2958.2004.04144.x
- Wu, J., Bian, Y., Tang, B., Chen, X., Shen, P., and Peng, Z. (2004). Cloning and analysis of WF146 protease, a novel thermophilic subtilisin-like protease with four inserted surface loops. *FEMS Microbiol. Lett.* 230, 251–258. doi: 10.1016/S0378-1097(03)00914-5
- Xu, Y., Wang, S., Li, L., Sahu, S. K., Petersen, M., Liu, X., et al. (2019). Molecular evidence for origin, diversification and ancient gene duplication of plant subtilases (SBTs). *Sci. Rep.* 9:12485. doi: 10.1038/s41598-019-48664-6
- Yahiro, K., Moss, J., and Noda, M. (2013). “Subtilase Cytotoxin (SubAB),” in *Handbook Of proteolytic enzymes*, eds A. J. Barrett, J. F. Woessner, and N. D. Rawlings (London: Elsevier Academic Press), 3155–3161. doi: 10.1016/B978-0-12-382219-2.00694-3
- Yamagata, Y., Maeda, H., Nakajima, T., and Ichishima, E. (2002). The molecular surface of proteolytic enzymes has an important role in stability of the enzymatic activity in extraordinary environments. *Eur. J. Biochem.* 269, 4577–4585. doi: 10.1046/j.1432-1033.2002.03153.x
- Yamagata, Y., Sato, T., Hanzawa, S., and Ichishima, E. (1995). The structure of subtilisin ALP I from alkalophilic *Bacillus* sp. NKS-21. *Curr. Microbiol.* 30, 201–209. doi: 10.1007/BF00293634
- Yanagida, N., Uozumi, T., and Beppu, T. (1986). Specific excretion of *Serratia marcescens* protease through the outer membrane of *Escherichia coli*. *J. Bacteriol.* 166, 937–944. doi: 10.1128/jb.166.3.937-944.1986
- Zhao, G.-Y., Chen, X.-L., Zhao, H.-L., Xie, B.-B., Zhou, B.-C., and Zhang, Y.-Z. (2008). Hydrolysis of insoluble collagen by desasin MCP-01 from deep-sea *Pseudoalteromonas* sp. SM9913: Collagenolytic characters, collagen-binding ability of C-terminal polycystic kidney disease domain, and implication for its novel role in deep-sea sedimentary particulate organic nitrogen degradation. *J. Biol. Chem.* 283, 36100–36107. doi: 10.1074/jbc.M804438200
- Zhu, D., Tanabe, S.-H., Xie, C., Honda, D., Sun, J., and Ai, L. (2014). *Bacillus ligniniphilus* sp. nov., an alkaliphilic and halotolerant bacterium isolated from sediments of the South China Sea. *Int. J. Syst. Evol. Microbiol.* 64, 1712–1717. doi: 10.1099/ijs.0.058610-0
- Zhu, X. L., Ohta, Y., Jordan, F., and Inouye, M. (1989). Pro-sequence of subtilisin can guide the refolding of denatured subtilisin in an intermolecular process. *Nature* 339, 483–484. doi: 10.1038/339483a0