Isolation and characterization of a new keratinolytic Fervidobacterium pennivorans strain from a hot spring in Tajikistan

UNIVERSITY OF BERGEN
Faculty of Mathematics and Natural Sciences

RUBÉN JAVIER LÓPEZ

Supervisor: Prof. Nils-Kåre Birkeland

Master's Thesis in Microbiology
Department of Biological Sciences
Faculty of Mathematics and Natural Sciences

University of Bergen
June 2018

Acknowledgments

This project work fulfils the requirement for the Master's degree at the Department of Biology, University of Bergen (UiB), Bergen, Norway. The laboratory part of this project was carried out at the laboratory of General Microbiology, Department of Biology, UiB, from August 2017 to May 2018.

First, I would like to give my deepest gratitude to my supervisor, Professor Nils-Kåre Birkeland. I am truly thankful to Nils-Kåre for his invaluable guidance and help during this project. His vast knowledge is always inspiring, and his motivation and ideas have made this work successful. I am grateful to Nils-Kåre for his help and patience during the writing process.

I would also like to thank Dr. Antonio García Moyano for his supervision and guidance, especially during the cloning and expression sections of this work. Antonio has also been a source of good ideas throughout this study, and I am really thankful for his interest in my work. Also, my most sincere acknowledgment to Adam Boulaich, for providing me with "cheap manpower" when needed.

My gratitude too to the members of General Microbiology group: Dr. Birte Tøpper, for helping me in the lab on a daily basis, especially during my first weeks; Dr. Thomas Kruse Hansen for his help in the Bioinformatics part of this project, and his patience; MSc. Chandini Murarilal Ratnadevi, for his help, support and good advices; and others, for all the help given me in lab.

I would thank to my fellow Microbiology Master students and friends, for their help and the good moments shared: Petra Hribovšek, Eirik Sæbø, William Erazo García, Sara Carolin Rundqvist, Anine Veronica Grønlund, Johanne Øyro and Eli Cholakova.

Finally, I am immensely grateful to my family and friends from Valencia and Bergen for their support and encouragement in the distance. I am thankful to my girlfriend (Dr. Jessica Furrior Palmer) for her help and care during my study at UiB, for her ideas, guidance and motivation.

Rubén Javier López
University of Bergen
Bergen, Norway

Table of Contents

Abstract

Thermophiles are organisms found among the Archaea and Eubacteria domains which are adapted to live at high temperature, the optimal growth temperature being between 50 °C and 79 °C. Phylum Thermotogae includes anaerobic and thermophilic members, and represents one of the deepest branching groups among the eubacterial line of descent, suggesting that the first eubacteria on Earth was thermophilic. Fervidobacteriaceae is a family within the Thermotogae phylum, some of which members can disintegrate chicken feathers, degrading the keratin, which is one of the most abundant proteins on Earth and often accumulates, causing a serious waste problem. Keratinases and keratinolytic microorganisms have biotechnological potential as they can convert this substrate into peptides and uncommon amino acids.

This work includes the isolation, biochemical characterization and genomic analysis of a new strain of *Fervidobacterium* pennivorans, named strain T, isolated from a terrestrial hot-spring in Tajikistan. This strain is an anaerobic and thermophilic bacterium, with an optimal growth temperature of 65 °C. It can degrade feathers and tolerates up to 40 g/L NaCl concentration. It is a neutrophilic microbe, optimally growing at pH 6.5. Its draft genome consisted of 28 contigs with a total size near 2 Mega bases.

The genomic analyses carried out include a Fervidobacteriaceae phylogenetic tree reconstruction and genomic comparisons of the studied strain with other *Fervidobacterium* pennivorans strains. These comparisons showed an 81 % overall sequence identity with the species type strain, isolated from the Azores, meaning that this new isolate, termed strain T, represents a new *Fervidobacterium pennivorans* strain.

Peptidases of this bacterium were investigated using bioinformatic approaches. A putative keratinase was identified, and its 3D structure predicted by homology modelling. Furthermore, the peptidase catalytic triad was found by aligning its sequence with other homologue proteins. Finally, this enzyme was cloned and expressed in *Escherichia coli* and its activity investigated, showing promising results and new insight into this thermophilic group.

Abbreviations

ANI	Average Nucleotide Identity
BLAST	Basic Local Alignment Search Tool
BODIPY	Boron-dipyrromethene
BRIG	BLAST Ring Image Generator
CSI	Conserved Signature Indels
DDH	DNA-DNA Hybridization
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
EMBOSS	European Molecular Biology Open Software Suite
Gepard	Genome Pair Rapid Dotter
GGDC	Genome to Genome Distance Calculator
GMQE	Global Model Quality Evaluation
HHBlits	HMM-HMM–Based Lightning-fast Iterative Sequence search
HMM	Hidden Markov Model
HSP	High Scoring Segment Pair
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Luria Bertani medium
LOBSTR	Low Backgrond Strain
MMF	Minimal medium, «fresh water»
MSA	Multiple Sequence Alignment
MQ	Milli-Q water
NCBI	National Center for Biotechnology Information
OD	Optical Density
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
PSI-BLAST	Position-Specific Iterated BLAST
PSSM	Position-Specific Scoring Matrix
QSQE	Quaternary Structure Quality Stimate
RAST	Rapid Annotation using Subsystem Technology
SDS	Sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEM	Scanning Electron Microscope
SMTL	SWISS-MODEL Template Library
SOC	Super Optimal broth with Catabolite repression
SVM	Support Vector Machines
UiB	Universitetet i Bergen

Introduction

Extremophiles are organisms adapted to live in harsh conditions, such as high or low temperatures (thermophiles and psychrophiles), high or low pH levels (acidophiles and alkaliphiles), high pressure (barophiles or piezophiles), high sugar or salt concentration (osmophiles), etc. These organisms can only be found among the Archaea and Eubacteria domains. A remarkable feature of extremophiles is that they do not only tolerate these harsh geochemical or physical conditions, but they need them to live and grow.

So, thermophiles, for instance, have optimal temperatures between 50 °C and 79 °C, and will not divide if the environment temperature is 40 °C or lower. Likewise, hyper-thermophiles need temperatures even higher to grow, with optimal growth temperatures above 80 °C (Godde et al., 2005, Madigan et al., 2006, Madigan et al., 2014).

Phylum Thermotogae includes both hyper-and thermophilic eubacteria (Frock et al., 2010). The members of this group can be easily identified morphologically because all of them are covered by a sheath-like envelope or "toga", present outside the cell wall (Bhandari and Gupta, 2014) (Figure 1). These gram-negative bacteria are strictly anaerobic, rod-shaped, and heterotrophic, and grow at neutral or slightly acidic pH. They can grow on a diverse kind of carbon substrates, including: pentoses, hexoses, disaccharides, glucans, xylans, etc. (Frock et al., 2010). These bacteria can be found in a variety of geothermal or volcanically heated environments, such as oil reservoirs, hydrothermal vents, or terrestrial hot springs (Gupta and Bhandari, 2011).

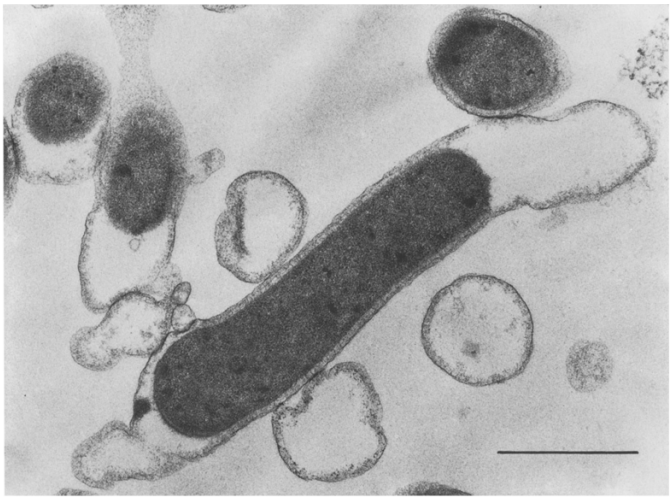


Figure 1. Thin section of *Thermotoga maritima.* Bar, 1 µm. (Huber et al., 1986)

Thermotogae represents one of the deepest branching groups among the Eubacterial line of descent, suggesting that the first eubacteria on earth was thermophilic (Achenbach-Richter et al., 1987). Furthermore, it has been found that some members of the Thermotogae group possess mosaic genomes, with genes from different origins, sharing up to 11 % of their genes with Archaea, suggesting that a significant horizontal gene transfer has occurred between these two groups (Frock et al., 2010, Cuecas et al., 2017), arising a debate over the phylogenetic placement of the Thermotogales in the tree of life (Figure 2) (Conners et al.). The metabolic versatility of the group, their thermophilic and hyper-thermophilic features, the possession of thermostable enzymes, involved in diverse carbohydrate utilization, and the increasement of the genomic and metagenomic studies have made the Thermotogales one of the most interesting groups for industrial processes (Conners et al., Miranda-Tello et al., 2004).

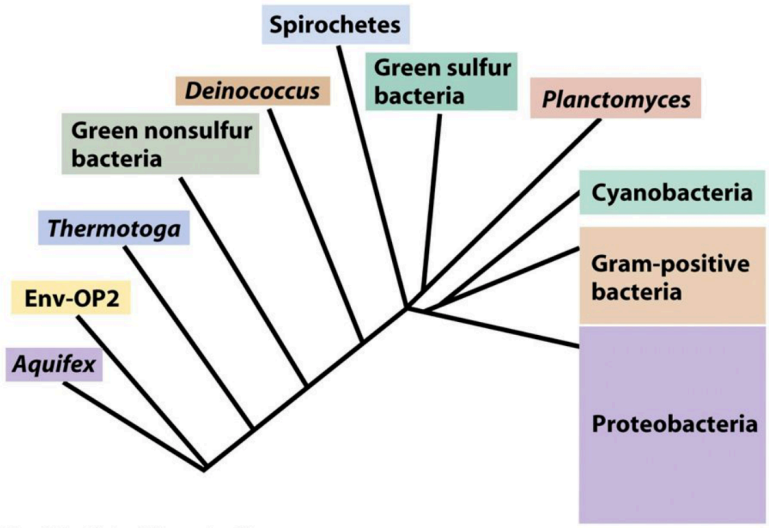


Figure 2. Phylogenetic tree of bacteria (Madigan et al., 2006)

Fervidobacteriaceae is a family within the Thermotogae phylum, conformed only by one genus: *Fervidobacterium*. Its name comes from the Latin word *fervidus*, *-i*, meaning "very hot" or "boiling" (Patel et al., 1985). Seven species comprise this genus, all of them isolated from terrestrial hot springs around the world: *F. islandicum*, isolated from a hot spring in Iceland (Huber et al.); *F. nodosum*, found in a New Zealand hot spring (Patel et al., 1985); *F. pennivorans*, isolated from a hot spring in Azores Islands (Friedrich and Antranikian, 1996); *F. changbaicum*, found in a hot spring in the Changbai Mountains (China) (Cai et al., 2007); *F. gondwanense*, isolated from geothermal waters of the Great Artesian Basin (Australia)

(Andrews and Patel, 1996); *F. riparium*, found in a hot spring of Kunashir Islands (Russia) (Podosokorskaya et al., 2011); and *F. thailandense*, isolated from a hot spring in Thailand (Kanoksilapatham et al., 2016) . Figure 3 shows the global distribution of the *Fervidobacterium* species with arrows indicating their isolation source.

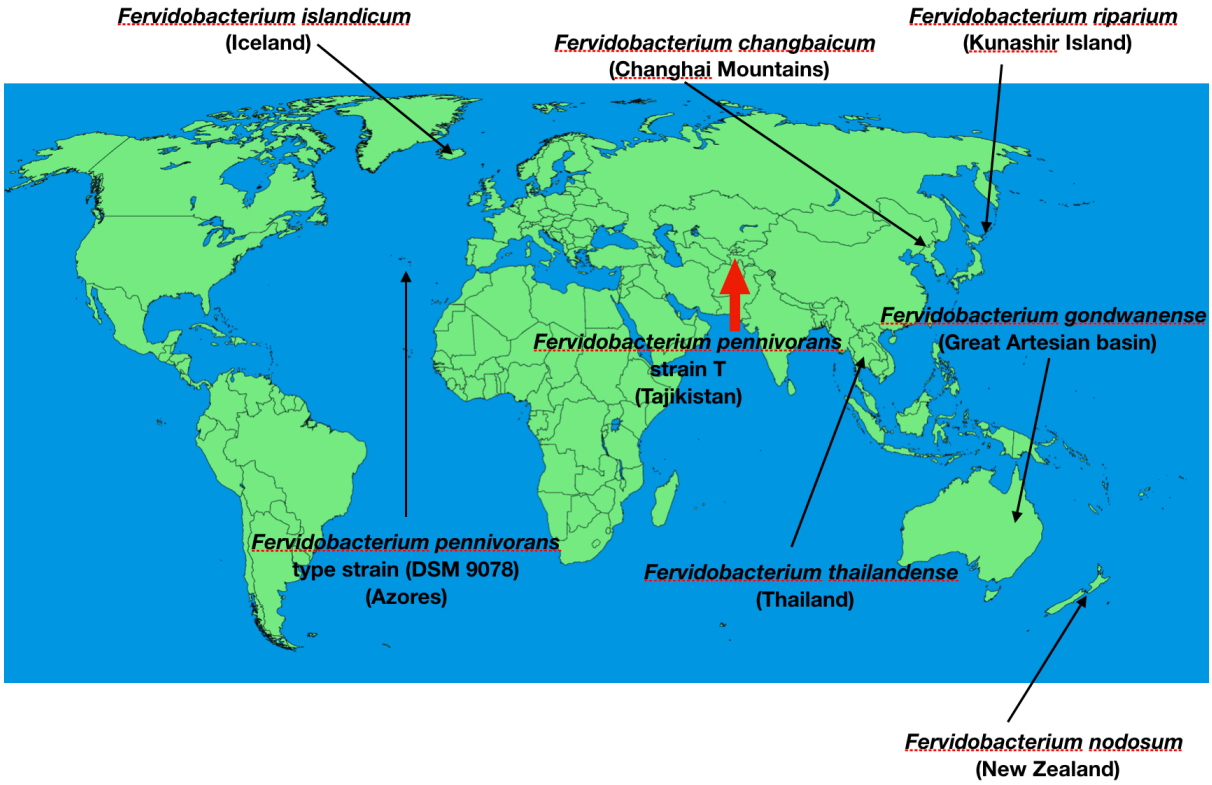


Figure 3. Locations of the described species of the genus *Fervidobacterium*. Black arrows point to the location where the species where first isolated. The red arrow points to Tajikistan, where the sample for this work was collected.

The members of this genus grow from 45 °C as the lowest temperature reported, until 90 °C, as the highest one, with optimal temperatures ranging from 65 °C to 80 °C. All of them can utilize different kind of carbon sources, such as pentoses, hexoses, disaccharides, glucans, xylans or sugar alcohols (Kanoksilapatham et al., 2016, Nam et al., 2002, Cai et al., 2007, Podosokorskaya et al., 2011, Andrews and Patel, 1996, Patel et al., 1985), the preferred substrates being: glucose, maltose, fructose and starch (Nam et al., 2002). In addition, they can grow on proteinaceous substrates, such as peptone, and at least *F. islandicum* and *F. pennivorans* have shown keratinolytic activity, being reported to have the ability to degrade feathers (Friedrich and Antranikian, 1996, Nam et al., 2002, Lee et al., 2015b). Figure 4 shows a phylogenetic tree of the *Fervidobacterium* genus.

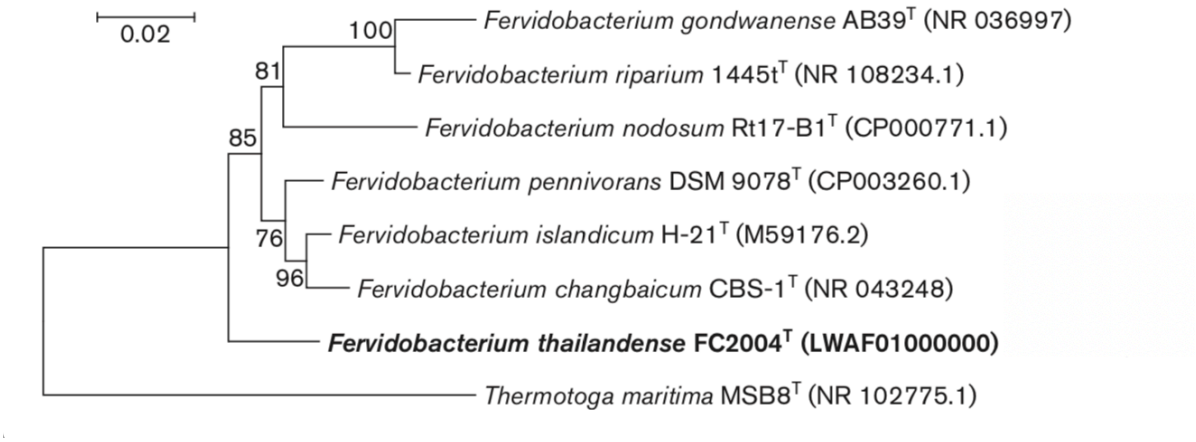


Figure 4. Neighbor-joining tree of 16S rRNA gene sequences of members of the family Fervidobacteriaceae showing the relationship the species belonging to genus *Fervidobacterium*. Bootstrap values as a percentage of 1000 replications are presented. Bar, 0.02 changes per nucleotide position (Kanoksilapatham et al., 2016).

Keratin is a protein present in animal structures such as feathers, horns, hair, skin and scales, and is one of the most abundant proteins on Earth: every year more than five million tons of chicken feathers are produced as waste by poultry farming, causing a serious waste problem (Lee et al., 2015a, Friedrich and Antranikian, 1996, Lee et al., 2015b). This protein has a secondary structure rich in α-helices with a high content of cysteine linked by disulfide bridges, which makes it chemically unreactive and mechanically durable (Parry et al., 1977, Lee et al., 2015b). Due to its rigid structure and disulfide bonds, it can accumulate (Friedrich and Antranikian, 1996, Lee et al., 2015a). Figure 5 showing inter- and intra-molecular bonding in keratin, including the mentioned disulfide bonds, which have an important role in determining the physicochemical properties of keratin (Shavandi et al., 2017).

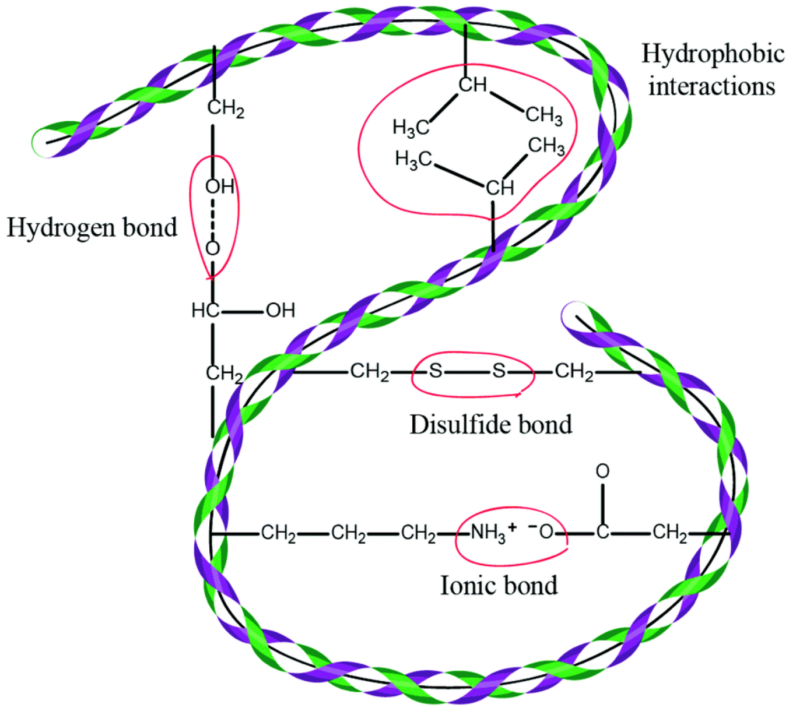


Figure 5. A diagram showing inter- and intra-molecular bonding in keratin. Various chemical bonds, e.g. hydrogen, ionic and disulfide bonds, result in increased strength and stability of the protein, and determine its structure (Shavandi et al., 2017).

Poultry feathers may be used as cheap animal feedstock, since they contain potentially useful proteins and amino acids (Williams et al., 1991), or even as fertilizers or soil conditioners (Lee et al., 2015b). Nevertheless, these mechanical and chemical processes allow using feather waste only on a limited basis as a dietary protein supplement (Papadopoulos, 1989) and, in addition, they increase greenhouse gas emissions and environmental pollution (Lee et al., 2015b). Due to these limitations and the industrial potential of poultry feathers and other waste stuff, many microbial thermostable proteases are being studied, isolated and characterized (Friedrich and Antranikian, 1996, Nam et al., 2002, Lee et al., 2015b). The use of thermostable proteases may lead to a more efficient and environmentally friendlier degradation of these feathers (Nam et al., 2002).

Proteases are one of the most important groups of industrial enzymes, accounting for about 60% of the total worldwide sale of enzymes, being used in detergents, beer, meat and leather industries for a long time (Niehaus et al., 1999, Kumar et al., 2005). But as high temperatures accelerate reaction rates and increase the solubility of solid reagents, thermoactive and thermostable proteases are very interesting from an industrial point of view, being an extra

advantage the consequent reduction of the contamination from mesophilic organisms, resulting in an increasement in the effectiveness of the process. (Chen et al., 2004). The mechanisms responsible of this thermostability are often a combination of different factors, such as an increased number of Van der Waals interactions, hydrogen bonds and ion-pairs (Godde et al., 2005).

Microbes produce both intracellular and extracellular proteases. The extracellular ones play an important role in the hydrolysis of environmental proteins and enabling the cell to absorb and use hydrolytic products (Kalisz, 1988). These extracellular proteases are the ones which have been exploited for industrial applications (Chen et al., 2004). Extracellular proteases commonly possess an N-terminal signal peptide targeting the protein to be secreted, where they hydrolyse proteins into peptides (Wandersman, 1989). Figure 6 shows the structure of a typical signal peptide.

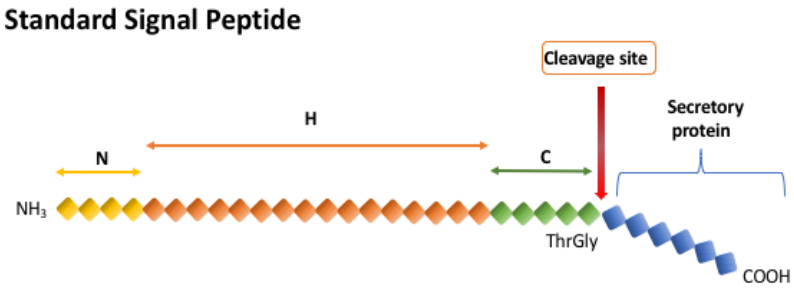


Figure 6. Structure of a standard signal peptide, showing its three-regions: a positively charged N-terminal region (N-region), a hydrophobic central region (H-region) and a neutral, polar C-terminal region (C-region).

Among these enzymes, the serine proteases comprise the majority of the thermostable proteases produced and characterized to date (Niehaus et al., 1999). More specifically, most of the serine proteases belong to the group of the subtilisin-like serine proteases, known as subtilases (Barrett and Rawlings, 1995). Subtilases are widely distributed in nature, are extracellularly secreted by different microbes and have been studied from an engineering

point of view (Rawlings and Barrett, 1994, Shaw and Pal, 2007). Subtilases are synthesized intracellularly as a precursor, called preprosubtilin, with a presequence and a prosequence attached to the N terminus of the mature protein (Jacobs M Fau - Eliasson et al.): the presequence acts as a signal peptide, mediating the secretion of the mature protein through the cellular membrane, and the prosequence guides the correct folding of the mature protein (Inouye, 1991). More than 200 subtilases have been described, but only a few belonging to thermophilic Archaea or Eubacteria have been characterized at the biochemical and genetic level (Kluskens et al., 2002). From an industrial and commercial point of view, these proteases have a remarkable utility; so, the search of new microbial sources of thermostable proteases is of continued value (Chen et al., 2004). With these enzymes, recalcitrant substrates such as feathers may be digested and converted to rare amino acids such as serine, cysteine or proline. In the past decades several thermophilic and hyper thermophilic organisms have been reported to be able to produce thermostable serine type proteases (Friedrich and Antranikian, 1996).

As the *Fervidobacterium* members can utilize different proteinaceous substrates, such as tryptone, peptone, casein, or casamino acids, they possess different proteases able to degrade these substrates under extreme conditions (Kluskens et al., 2002). Furthermore, as previously mentioned, *F. islandicum* and *F. pennivorans* have been reported to grow on chicken feathers and use keratin as a substrate at high temperatures. Keratinases have biotechnological potential as they can convert this substrate into peptides and rare amino acids. Different keratinases have been described, with sizes ranging from 16 kDa to 440 kDa (Kim et al., 2004). A keratinolytic enzyme belonging to *F. pennivorans* was purified and described in the past, although the recombinant protein was inactive (Friedrich and Antranikian, 1996, Kim et al., 2004, Kluskens et al., 2002). This enzyme, named *fervidolysin*, is encoded by a 2.1 kb gene called *fls*. The product is a medium size enzyme with 699 amino acids, a precursor of 73 kDa and a mature part of 58 kDa with a multi-domain structure, showing a high homology with the subtilisin family and a high thermo stability (Kim et al., 2004, Kluskens et al., 2002). Likewise, a keratinase from *F. islandicum* has also been characterized and expressed in *Escherichia coli*. This enzyme, called *islandisin* is encoded by a 2106 bp open reading frame, which leads to a mature protein of 668 amino acids with a signal sequence of 33 amino acids and three catalytic

residues (Asp177, His215 and Ser391). Islandisin has been successfully expressed in *E. coli*, showing the recombinant protein optimal activity at 80 °C and pH 8.0 (Godde et al., 2005).

Nevertheless, traditional cloning and expression methods require the purification of intermediate cleavage products, are of limited efficiency and do not avoid the occurrence of restriction sites in the sequences of targeted genes (Geertsma and Dutzler, 2011). Fragment exchange (FX) cloning was developed as an alternative to the traditional approaches. FX cloning is based on a class II restriction enzyme and negative selection markers. These restriction enzymes cleave the DNA at a fixed distance outside their asymmetric recognition site (Szybalski et al., 1991, Geertsma and Dutzler, 2011), making possible to leave only an extra amino acid to either side of the protein. The counterselection markers keep the background of untransformed vectors low, and allows unambiguous selection of positive clones in each step. Thus, this method has turned to be highly efficient and very economic, allowing the transfer of open reading frames into a high variety of expression vectors with a minimal handling (Geertsma and Dutzler, 2011).

Aims

This project focuses on the isolation and study of a new *F. pennivorans* strain, with following objectives:

1. Growth of this thermophilic Thermotogae bacterium, as well as its identification through 16S rRNA gene sequencing, its isolation and its biochemical characterization.
2. Genomic analysis of this strain, including genomic sequencing and comparison with other members of the genus.
3. Bioinformatic analyses, including a phylogenetic tree reconstruction to compare the *Fervidobacterium* genus, genomic alignment and search of possible conserved genetic markers.
4. Cloning and expression of a putative keratinase gene in *Escherichia coli.*
5. Activity assay to assess the activity of the expressed protein.

Materials and Methods

1. Biochemistry and morphology

The sample containing the studied strain was collected in a terrestrial hot-spring in Tajikistan. It was first enriched in an appropriate medium (MMF supplemented with peptone and yeast extract) and, when its growth was assessed, an isolation of the strain by dilution to extinction was carried out. Then, the rest of the analyses followed.

1.1 Medium preparation

The sample has been stored at room temperature for about a year in a sealed and corked 120 mL serum flask. 1 mL of this culture was inoculated into another 60 mL serum flask with MMF medium for enrichment. To prepare it, a basal medium with mineral salts, trace elements, carbonate buffer and sulphide as reducing agent was made. The composition of the mentioned medium is detailed in Table 1, and the composition of the trace elements added is detailed in Table 2. 0.2 % resazurin (5 mL/L) was also added to the medium as redox indicator.

Table 1. Composition of basal MMF medium.

Reagents	Amount
NaCl	3 g/L
$MgSO_4 \cdot 7H_2O$	0.7 g/L
KCl	0.34 g/L
NH_4Cl	0.25 g/L
$CaCl_2 \cdot 2H_2O$	0.14 g/L
KH_2PO_4	0.14 g/L
Yeast extract	0.2 g/L
Trace elements	1 mL/L

Table 2. Composition of the trace elements solution added to the basal medium.

Reagents	Amount
HCl (25%)	10 mL/L
$FeCl_2 \cdot 4H_2O$	1.5 g/L
$CoCl_2 \cdot 6H_2O$	190 mg/L
$MnCl_2 \cdot H_2O$	100 mg/L
$ZNCl_2$	70 mg/L
$Na_2MoO_4 \cdot 2H_2O$	36 mg/L
$NiCl_2 \cdot 6H_2O$	24 mg/L
H_3BO_3	6 mg/L
$CuCl_2 \cdot H_2O$	2 mg/L

After autoclaving at 121°C for 30 min, the medium was cooled to 40 – 50°C while gassing with sterile nitrogen. Then, 10 mL of vitamin solution and 4 mL of 0.5M Na_2S were added, as well as 20 mM $Na_2S_2O_3$ and 0.2 % resazurin. The composition of the vitamin solution is detailed in

Table 3. Finally, the pH was adjusted to 6.5 using 1M HCl; pH was checked with a pH 3110 pH-meter (WTW).

Table 3. Composition of the vitamin solution added to MMF medium

Reagents	**Amount**
4-Aminobenzoic acid	8 mg/L
D(+) Biotin	2 mg/L
Nicotinic acid	20 mg/L
Ca-D(+) pantothenate	10 mg/L
Pyridoxamine · 2HCl	30 mg/L
Thiamin dichloride	20 mg/L
Vitamin B12	10 mg/L

The medium was aliquoted into smaller flasks using Hungate technique (Macy et al., 1972) and the flasks sealed with rubber corks and aluminum crimps. The flasks were incubated at 65 °C and, after 48h of incubation, growth was checked with a phase-contrast microscope.

1.2 Enrichment and isolation

After enrichment and growth in MMF with peptone and yeast extract, the studied bacterium was isolated from the sample. To isolate the strain, the dilution to extinction method was conducted. A battery of eight flasks filled with 9 mL of MMF medium was made and labelled from 10^{-1} until 10^{-8}. This medium was supplemented with 0.5 % Peptone extract, and its concentration of yeast extract was increased up to 0.05 %. 1 mL of the original culture was transferred to the flask labelled as 10^{-1}, from this one 1 mL was transferred to the one labelled as 10^{-2}, and so on. After 48 hours growth was checked in all the flasks under a Nikon Eclipse E400 phase-contrast microscope. Growth was reported in flasks labelled as 10^{-1}, 10^{-2}, 10^{-3} and 10^{-4}, so the 10^{-4} flask was used to start a new dilution series: from this flask 1 mL was transferred to a new 10^{-1} flask, from this one to a new 10^{-2}, and so on, until 10^{-8}. Again, after 48 hours of incubation at 65 °C, growth was reported in flasks labelled as 10^{-1}, 10^{-2}, 10^{-3} and 10^{-4}. So, the 10^{-4} flask was assumed to contain a pure culture and used to inoculate a new flask with fresh MMF medium to obtain biomass and isolate DNA. The medium was supplemented

with 0.5 % peptone extract and 0.05 % yeast extract. The flask was incubated at 65 °C for 48 hours.

1.3 Carbon utilization test

The studied organism was tested against different carbon sources to assess its metabolic capabilities. The following substrates were used: cellulose, dextrin, galactose, lactose, arabinose, mannose, mannitol, starch, sorbitol, sucrose and raffinose. For each substrate, a different flask was made with 10 mL of MMF medium supplemented with the respective substrate at a final concentration of 0.5 %, with a yeast extract concentration of 0.05%. A negative control flask was also prepared, containing 10 mL of MMF medium and a final concentration of yeast extract of 0.05 %. After 48 hours of incubation at 65 °C, a new battery of flasks was set up, and 1 mL from each previous flask was transferred to its correspondent one with fresh MMF medium, 0.5% of the respective carbon source and a final yeast extract concentration of 0.05 %. A second negative control flask was prepared, again filled with 10 mL of MMF medium and a final yeast extract concentration of 0.05 %. After 48 h of incubation, growth was checked under the phase-contrast microscope.

1.4 Temperature tolerance

The organism was incubated at different temperatures to determine the minimum and maximum temperature growth. Three different temperatures were tested (additionally to 65 °C): 38 °C, 55 °C and 80 °C. Two flasks with 10 mL of MMF medium supplemented with peptone (0.5%) and Yeast Extract (0.1%) were prepared for each temperature. After 48 h of incubation, growth was checked under the phase-contrast microscope.

1.5 Osmotic stress test

A battery of flasks with different NaCl concentrations were made to test the tolerance of this bacterium to salinity. Two flasks were prepared for each NaCl concentration: 1 %, 2 %, 3 %, 4 % and 5 %. The medium used was MMF supplemented with 0.5 % peptone and 0.1% Yeast Extract. After 48 h of incubation at 65 °C growth was checked under the phase-contrast microscope.

1.6 pH test

To test the tolerance to different pH of this organism, four flasks were prepared with 10 mL of MMF medium supplemented with 0.5 % Peptone and 0.1 % Yeast Extract. Two of these flasks were adjusted to pH 5.5 with HCl (1N) and the other two were adjusted to pH 8.0 with NaOH (1N). After 48 h of incubation at 65 °C growth was checked under the phase-contrast microscope.

1.7 Keratinase activity test

A battery of flasks filled with approximately 50 mL of MMF medium supplemented with peptone (0.5 %) and yeast extract (0.05 %) was made to test this activity. One chicken feather of 15 ± 5 mg was introduced in every flask, as well as 0.5 mL of a *F. pennivorans* culture as inoculum. Previously, the feathers had been sterilized by immersion in a methanol:ethanol solution (1:1) and posterior tyndallisation for 60 minutes at 100 °C (Friedrich and Antranikian, 1996). Feathers with different rigidity, from breast and wings were tested. The flasks were incubated at 65 °C. A negative control flask was also set up: MMF medium with the same concentration of peptone and yeast extract, but without inoculum. Every 24 hours the flasks were monitored to check for changes in the integrity of the feathers.

1.8 Scanning electron microscopy

A sample from a dense culture was prepared for scanning electron microscopy (SEM) and for further morphological study of the isolated strain. First, 1 mL of a dense culture was fixed with 80 µl 25% glutaraldehyde (final glutaraldehyde concentration of 2%) and incubated for 1 hour at room temperature. Then, the sample was centrifuged at 12500 rpm for 7 minutes, the supernatant discarded, and the pellet resuspended in 500 µL of 0.1 M cacodylate buffer. The prepared samples were then delivered to the microscopy service of the Molecular Imaging Center Platform (https://www.uib.no/en/rg/mic).

2. Genomic and Bioinformatic analyses

2.1 DNA isolation and PCR

After growing the bacterium in MMF medium supplemented with peptone as described, its growth was checked under phase-contrast microscope, transferred to a 50 mL Falcon tube

and then centrifuged at 4 °C for 15 minutes at 6000 rpm. The supernatant was discarded. The genomic DNA was isolated from the pelleted cells with a *NA2100 SIGMA* GenElute™ Bacterial Genomic DNA extraction kit, by Sigma-Aldrich, following the standard protocol. The DNA quality and quantity were determined using a NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer, by Thermofisher Scientific. Following, a 0.8 % agarose gel electrophoresis was run to check the DNA size, comparing to GeneRuler DNA Ladder Mix, by ThermoFisher (catalog number: SM0333). The gel was run for 45 minutes at 5 V/cm.

The strain identification starts with a PCR amplification of the 16S rRNA gene (Woese et al., 1983, Peake, 1989); the sequence of the primers used in the amplification and the PCR program are detailed in Table 4 and Table 5, respectively (Rainey et al., 1992). Genomic DNA of another Thermotoga, *Thermosipho africanus*, was used a positive control. The amount of the DNA templates to amplify were of approximately 120 ng. The PCR master mix without any DNA template was used as a negative control. Two duplicates of each control were used, and the same for the sample. The DNA polymerase used in this reaction was Taq DNA Polymerase by BioLabs (catalog number M0273).

Table 4. Primers used in the PCR amplification of the 16S rRNA gene

Primer name	Primer sequence
16S Forward (27 F)	5'-GAGTTTGATCCTGGCTCAG
16S Reverse (1525 R)	5'-GAAAGGAGGTGATCCAGCC

Table 5. PCR program used to amplify the 16S rRNA gene.

Step	Temperature (Celsius)	Time (seconds)
Initial denaturation	94	30
30 Cycles	94	30
	68	45
	68	60
Final extension	68	300

The size and quality of the PCR product were checked running a 1.5% agarose gel electrophoresis for 45 minutes at 5 V/cm, comparing to GeneRuler DNA Ladder Mix, by ThermoFisher (catalog number: SM0333). Then, the PCR products were used to perform a cycle sequencing reaction according to the Big Dye v3.1 protocol (Sequencing facility), and then sequenced in the UiB Sequencing facility; the sequence of the primers used in the cycle sequencing reaction is detailed in Table 6. The obtained sequences were corrected using 4Peaks software (Griekspoor and Groothuis, 2015). Then, they were merged and a consensus sequence obtained using the EMBOSS software package (Rice et al., 2000). This consensus sequence was then compared in BLAST through the Blastn suite, using the default options for megablast query and excluding the "uncultured/environmental sample sequences" from the search.

Table 6. Primers used in the cycle sequencing reaction of 16S rRNA gene.

Primer name	Primer sequence
16S Forward	5'-GAGTTTGATCCTGGCTCAG
575 Forward	5'-CGGAATTACTGGGCKTAAAG
K517 Reverse	5'-ATTACCGCGGCTCCTGG
16S Reverse	5'-GAAAGGAGGTGATCCAGCC

2.2 Phylogenetic trees construction

A phylogenetic tree based on the 16S rRNA gene sequence was built to study the evolutionary relationship between the experimental strain characterized in this work and the rest of the described species of the *Fervidobacterium* genus. The 16S rRNA gene sequences of the different *Fervidobacterium* species were obtained from the Genbank database (Benson et al., 2005, Tatusova et al., 2016). The sequences were then aligned using the ClustalO tool through the CLC Genomic Workbench suite, version 11.01 (Sievers et al., 2011). SeaView software (Gouy et al., 2010) was used to build the phylogenetic tree itself with the aligned sequences, using the Bio Neighbor-Joining clustering algorithm (Gascuel, 1997). The nucleotide distance was measured by "Kimura 80" method (Kimura, 1983). A bootstrap analysis with 100 repetitions was conducted to check the quality of the tree, and the 16S rRNA gene sequence of *Thermosipho africanus* was employed as an outgroup lineage. In the same way, another

phylogenetic tree was reconstructed, based on the sequence of the peptidase S8 that has been expressed in *E. coli* in this work, comparing the homologue sequences of this protein found in all the available genomes of the genus *Fervidobacterium*.

2.3 Genomic analyses

2.3.1 Genome assembly and annotation

A genomic analysis of the strain was performed by shot-gun sequencing of its whole genome using a commercial Next Generation DNA sequence provider (Bertoni et al., 2017) based on Illumina technology. 1µg of genomic DNA was shipped. The sequences obtained were assembled with CLC GenomicWorkbench 11 software (QIAGEN Bioinformatics) , using "de novo assembly" tool. The optimal k-mer length for the genome de novo assembly was determined using KmerGenie software (Chikhi and Medvedev, 2014). Following, the genome was uploaded to RAST (Rapid Annotation using Subsystem Technology) for annotation (Aziz et al., 2008) and analysed using the SEED viewer (Overbeek et al., 2014). The genome was then downloaded and compared with the *F. pennivorans* type strain genome (DSM 9078, accession number CP003260.1) and with *F. pennivorans* strain DYC (accession number CP011393.1). These genomes were downloaded from the NCBI database (Benson et al., 2005, Tatusova et al., 2016). The comparisons included a genomic alignment, an Average Nucleotide Identity calculation (ANI), a Genome-to-Genome distance calculation (GGDC) and a dotplot matrix comparison.

2.3.2 Average Nucleotide Identity

The distribution of nucleotide identity between the experimental genome (*F. pennivorans* strain T) and both the type and DYC strains was estimated through an Average Nucleotide Identity (ANI) test, which calculates the average nucleotide identity using both best hits (one-way ANI) and reciprocal best hits (two-way ANI) between two genomic datasets (Rodriguez-R and Konstantinidis, 2016, Goris et al., 2007).

2.3.3 Genome-to-genome distance analysis

Another genome comparison of the sequenced isolate with both *F. pennivorans* DYC and *F. pennivorans* DSM 9078 was performed, using the Genome-to-Genome distance calculator by

Leibniz Institute (DSMZ). This tool infers whole-genome distances mimicking the experimental genome-to-genome comparison by DNA hybridization (DDH) (Meier-Kolthoff et al., 2013).

2.3.4 Dot-plot matrix

A dotplot matrix comparison was made, to compare the experimental genome with the type and the DYC strains, using the Gepard tool (Krumsiek et al., 2007).

2.3.5 Genomic alignment

This alignment was conducted using Mauve software, which constructs multiple genome alignments in the presence of large-scale evolutionary events, such as rearrangement and inversion. Genomic alignments provide a basis for research into comparative genomics and the study of genome-wide evolutionary dynamics (Darling et al., 2004).

2.4 Conserved signature indels

The genomes of all the species in *Fervidobacterium* genus were aligned using BLAST Ring Image Generator (BRIG) v0.95 (Alikhan et al., 2011) to find candidates of genomic fragments horizontally transferred. The genbank file of the experimental genome was downloaded from the RAST website. The rest of genomes were obtained from the NCBI database (Tatusova et al., 2016, Benson et al., 2005). Then, the fragments suspected to have been horizontally transferred were searched in the sequenced genome, if present, using Artemis software (Rutherford et al., 2000). Those fragments having a GC content with at least a 2.5 Standard Deviation cut off were extracted to a FASTA file and translated with ExPaSy tool (Gasteiger et al.) if no annotation was available, and used to carry out a search with BLAST to find possible homologues using the blastp algorithm.

2.5 3D protein structure prediction

A 3D model structure was predicted using Swiss-model tool (Benkert et al., 2011, Bertoni et al., 2017, Biasini et al., 2014, Bienert et al., 2017, Guex et al., 2009).

2.5.1 Template Search

Template search with BLAST and HHBlits was performed against the SWISS-MODEL template library. The target sequence was searched with BLAST against the primary amino acid sequence contained in the SMTL. A total of 146 templates were found. An initial HHblits profile was then built using the procedure outlined in (Remmert et al., 2011) followed by 1 iteration of HHblits against NR20. The obtained profile was then searched against all profiles of the SMTL. A total of 235 templates were found. From those, the most suitable 50 templates for the query protein were automatically shown by the tool.

2.5.2 Template Selection

For each identified template, the template's quality has been predicted from features of the target-template alignment. The templates with the highest quality were then selected for building the model.

2.5.3 Model Building

The models were built based on the target-template alignment using ProMod3. Coordinates which are conserved between the target and the template are copied from the template to the model. Insertions and deletions are remodeled using a fragment library. Side chains are then rebuilt. Finally, the geometry of the resulting model is regularized by using a force field. In case loop modelling with ProMod3 fails, an alternative model is built with PROMOD-II (Guex et al., 2009)

2.5.4 Model Quality Estimation

The global and per-residue model quality has been assessed using the QMEAN scoring function (Benkert et al., 2011). For improved performance, weights of the individual QMEAN terms have been trained specifically for SWISS-MODEL.

2.5.5 Ligand Modelling

Ligands present in the template structure are transferred by homology to the model when the following criteria are met: (a) The ligands are annotated as biologically relevant in the template library, (b) the ligand is in contact with the model, (c) the ligand is not clashing with the protein, (d) the residues in contact with the ligand are conserved between the target and

the template. If any of these four criteria is not satisfied, a certain ligand will not be included in the model. The model summary includes information on why and which ligand has not been included.

2.5.6 Oligomeric State Conservation

The quaternary structure annotation of the template is used to model the target sequence in its oligomeric form. The method (Bertoni et al., 2017) is based on a supervised machine learning algorithm, Support Vector Machines (SVM), which combines interface conservation, structural clustering, and other template features to provide a quaternary structure quality estimate (QSQE). The QSQE score is a number between 0 and 1, reflecting the expected accuracy of the interchain contacts for a model built based a given alignment and template. Higher numbers indicate higher reliability. This complements the GMQE score which estimates the accuracy of the tertiary structure of the resulting model.

2.6 Multiple sequence alignment

A local multiple sequence alignment with all the found homologues for the expressed protein was built, to find conserved motives and catalytic sites. The amino acidic sequence of the expressed peptidase was used to perform a PSI-BLAST query, until convergence. After four iterations, no more sequences could be found. The found sequences were then downloaded and saved into a FASTA file. Following, and under JalView software, a Multiple Sequence Alignment was performed using Clustal Omega algorithm.

3. Cloning of a putative keratinase gene

3.1 Signal peptide prediction

The gene of interest to be expressed was selected based on its annotation in RAST, and was functionally annotated as a serine protease. The signal peptide of the protein was predicted using the Signal IP server (Petersen et al., 2011) and the primers were designed using the application available in the FX cloning website .

3.2 PCR and product cloning into sequencing vector

This procedure is known as FX or “fragment exchange” cloning, and is based on the method described by Geerstma & Dutzler (Geertsma and Dutzler, 2011). The FX cloning system uses

an initial cloning vector for the initial cloning of a PCR product. The sub-cloning of the gene from the initial vector into an expression vector utilizes the same cloning procedure; this expression vector has a different resistance marker. An overview of this method can be seen in Figure 7.

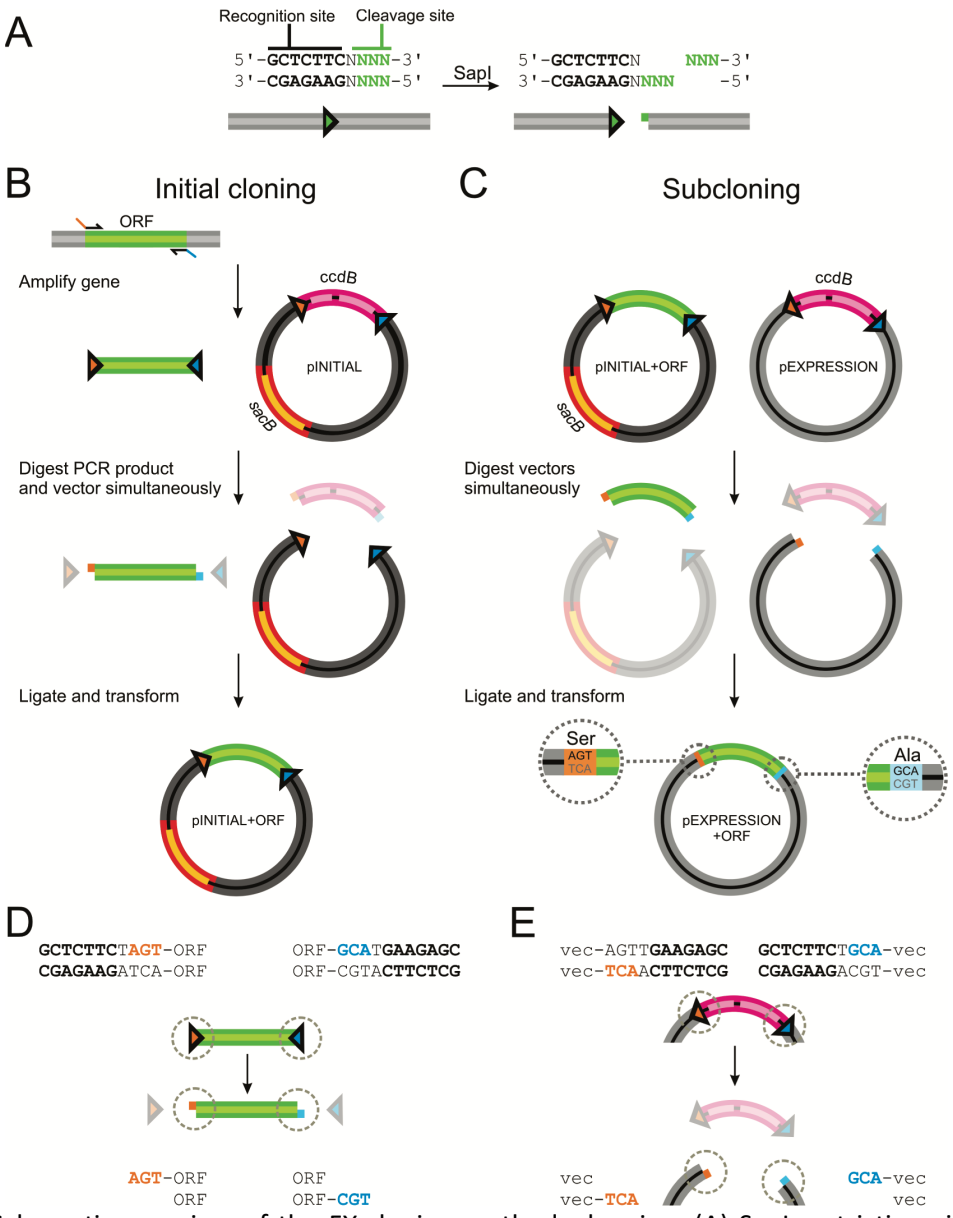


Figure 7. Schematic overview of the FX-cloning method, showing: (A) SapI restriction site, in bold letters, with any of the four nucleotides represented with N. Arrows indicate the direction of the restriction site. (B) Cloning of a PCR product into pINITIAL. The genes coding for the counterselection markers ccdB and sacB on pINITIAL are colored in magenta and orange, respectively. (C) Sub-cloning of an ORF into an expression vector. The three nucleotides added to either terminus of the ORF are shown as insets (circle). (D) Orientation of the SapI cleavage sites in the PCR product and pINITIAL and (E) in expression vectors. The single-stranded overhangs generated upon cleavage are shown in orange and magenta.

First, the gene of interest was amplified in a PCR reaction. The primers for this gene were designed using the FX Cloning website and their sequence is detailed in Table 7. Phusion High-Fidelity Polymerase by Thermo Fisher (catalog code: F530L) was used, and the extension temperature used for these primers was determined using OligoCalc (Buehler, 1996). The PCR program is detailed in Table 8. The amount of the DNA template to amplify was 120 ng. The PCR master mix without any DNA template was used as a negative control. Two replicas were used, for both the sample and the negative control.

Table 7. Primers used in the PCR amplification pointing to the gene of interest.

Primer name	Primer sequence
Forward	5'- ATATATGCTCTTCTAGTAACTCATTGGAGCCAAGATTTGAACCA
Reverse	5'- TATATAGCTCTTCATGCGTATGTCAAAGCCTTGTAAGCATCAA

Table 8. PCR program used to amplify the gene of interest.

Step	Temperature (Celsius)	Time (seconds)
Initial denaturation	98	30
30 Cycles	98	10
	52	30
	72	45
Final extension	72	420

The size and quality of the PCR products were checked in a 1.5% agarose gel electrophoresis, run for 45 minutes at 5 V/cm. The plasmid used for the initial cloning was pINIT with tetracycline resistance (Addgene plasmid # 46974) (Geertsma and Dutzler, 2011). The map of the pINIT_tet plasmid can be seen in Figure 8.

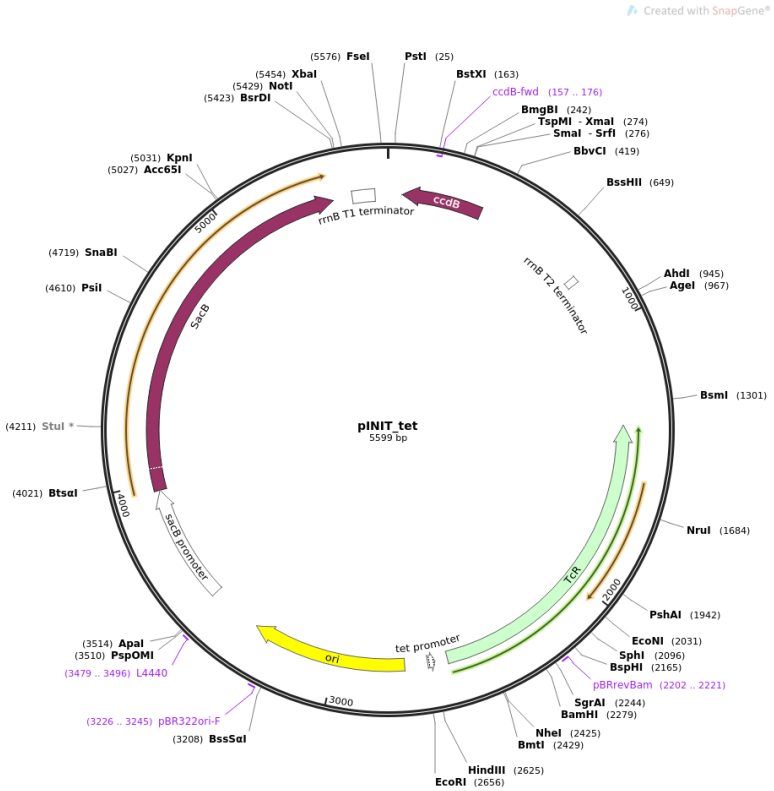


Figure 8. Map of the pINIT_tet plasmid (Addgene plasmid # 46974), used for the initial cloning of the gene of interest.

The *E. coli* strain containing this plasmid was grown overnight at 37 °C in LB with a final tetracycline concentration of 10 μg/mL. LB medium was obtained after combining 10 g of tryptone, 5 g of yeast extract, 10 g of NaCl, in 1 liter of distilled water; the mixture was autoclaved for 25 min at 120°C (Sezonov et al., 2007). The plasmid was extracted using the GeneJET Plasmid Miniprep Kit, by ThermoFisher Scientific (catalog number: K0502), and its size and quality was checked by running an 0.8 % agarose gel electrophoresis using Supercoiled DNA ladder by New England BioLabs (Sambrook, 1989) marker as size reference.

For initial cloning of the PCR products, 250 ng of these were mixed with 50 ng of the pINIT_tet plasmid, obtaining a final ratio of 1:5 (vector:insert). 1 μL of 10x buffer (200 mM Tris-acetate, 500 mM potassium acetate, 100 mM magnesium acetate, 10 mM dithiothreitol, pH 7.9) was added and the volume adjusted to 9 μL with MQ water. The digestion started by adding 1 μL of SapI (2U). After incubation of 1 hour at 37 °C, the enzyme was inactivated by heat, incubating the mixture at 65 °C for 20 minutes, and cooled to 25 °C. Following, a ligation of the consequent fragments was performed by adding 1.25 μL of ATP mix (10mM) and 1.25 μL

T4 DNA Ligase by New England BioLabs (catalogue number: M202L). The ligation mixture was incubated for 1 hour at 25 °C and inactivated by heating at 65 °C for 25 minutes.

The final mixture was used to transform One Shot™ TOP10 Chemically Competent *E. coli* cells, by ThermoFisher Scientific (Catalogue number: C404003). This was done by mixing a 5 μL aliquot with the competent cells and incubating them for 30 minutes on ice. A heat shock inhibition followed, by incubating in a 42 °C water bath for exactly 30 seconds. Then, 250 μL of S.O.C. medium by ThermoFisher Scientific (catalogue number: 15544034) was added and the cells incubated at 37 °C for 1 hour. Finally, 100 μL of the transformation mix was plated on agar with LB medium supplied with tetracycline (10 μg/mL) and incubated overnight at 37 °C. Then, six colonies were picked from the plate and incubated overnight again in 10 mL of LB medium supplied with tetracycline (10 μg/mL), and plasmid DNA isolation was conducted using the kit previously mentioned. The DNA was quantified as described before, and 200 ng of the construction were used to perform a PCR sequencing reaction according to the BigDye 3.1 protocol (http://www.uib.no/en/seqlab/55363/protocol-bigdye-v31) to verify the sequence of the insert. The primers and the program to run this reaction were chosen considering the pINIT_tet sequence, and are detailed in Table 9 and Table 10, respectively.

Table 9. Primers used in the sequence reaction conducted to verify the pINIT_tet + keratinase gene construction.

Primer name	Primer sequence
Forward	5'- GAGTAGGACAAATCCGC
Reverse	5'- TGCTTCGCAACGTTCAAATCCGC

Table 10. Sequencing program used to verify the pINIT_tet + keratinase gene construction.

Step	Temperature (Celsius)	Time (seconds)
Initial denaturation	96	300
25 Cycles	96	10
	46	5
	60	240

The sequences obtained were then aligned and merged using the EMBOSS merge tool (Rice et al., 2000) and the consensus sequence was likewise aligned with the expected gene sequence obtained in RAST; this alignment was performed with MEGA 7 software (Kumar et al., 2016).

3.3 Sub-cloning using FX Cloning

Then, sub-cloning into expression vectors followed. The vectors chosen were the plasmids p7xN3H (Addgene plasmid # 47064) and p7xC3H (Addgene plasmid # 47065) (Geertsma and Dutzler, 2011), which maps are detailed in Figure 9 and Figure 10, respectively. These vectors add a tag of 10 histidines to the translated protein, adding p7xNH3 these histidines to the N-terminus of the protein and p7xC3H to the C-terminus.

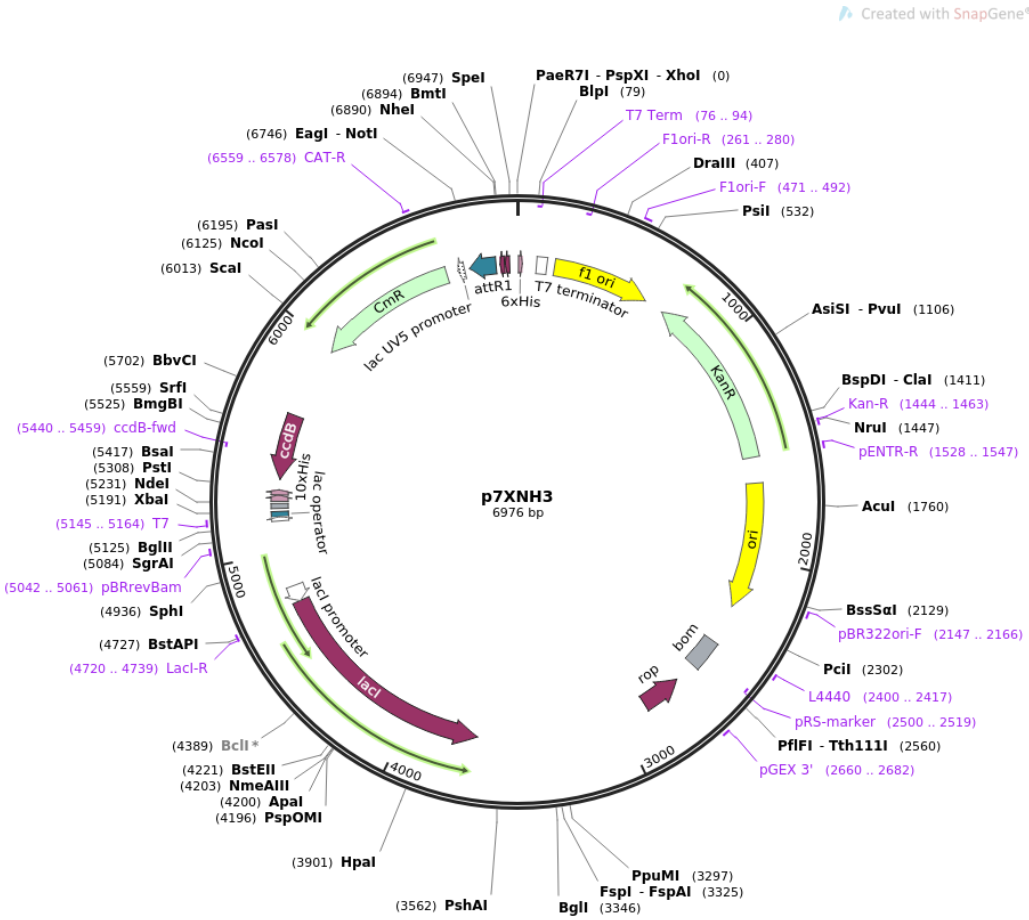


Figure 9. Map and sequence of the p7xN3H (Addgene plasmid # 47064), used for sub-cloning and expression of the gene of interest.

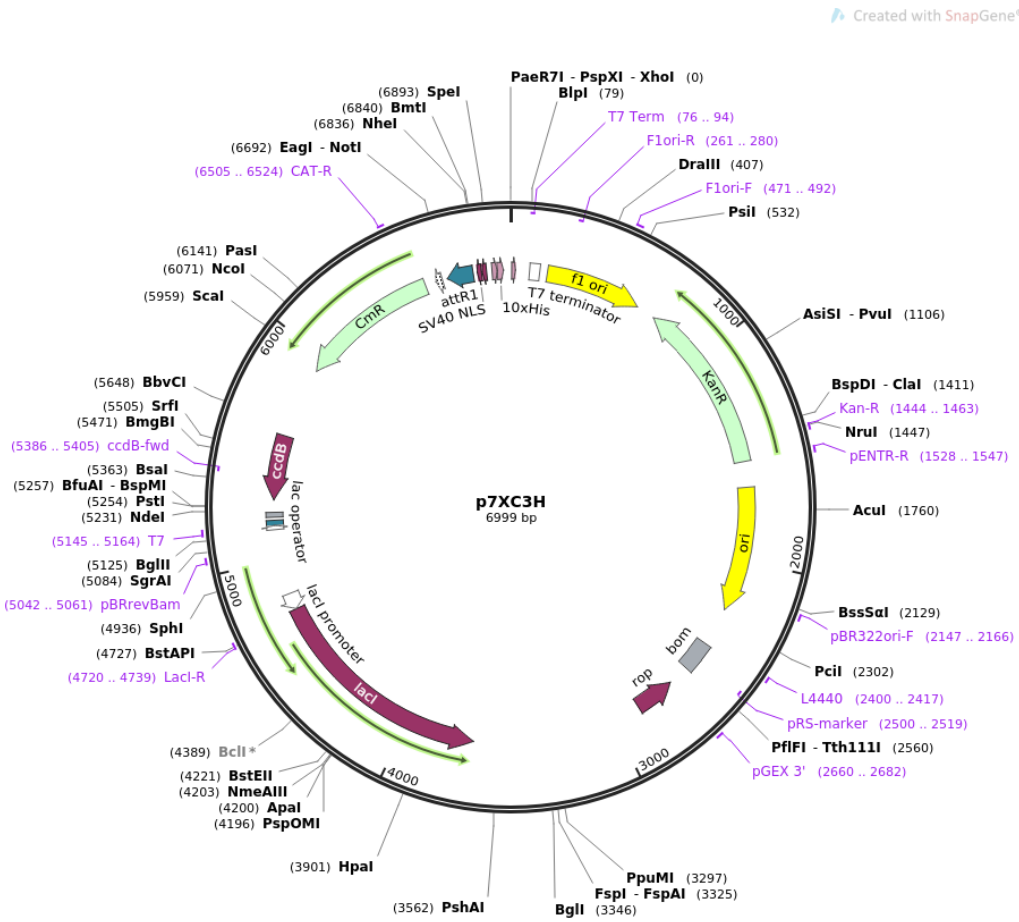


Figure 10. Map and sequence of the p7xCH3 (Addgene plasmid # 47065), used for sub-cloning and expression of the gene of interest.

This procedure started by growing *E. coli* strains containing these plasmids, incubating them overnight in 10 mL of LB medium with kanamycin (50 μg/mL). Then, the plasmids were isolated following the method already described, and its concentration and quality were determined using a NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer, and by running a 0.8% agarose gel electrophoresis, at 5 V/cm for 45 minutes. For sub-cloning, 50 ng of the chosen vectors were mixed with the pINIT_tet plasmid containing the keratinase gene to a final molar ratio of 1:4 (expression vector:pINIT_tet plasmid). The sub-cloning and transformation procedures were the same as described before, but the resultant cells were plated on LB with 50 μg/mL kanamycin and incubated overnight at 37 °C.

Three colonies per plasmid were picked from the plates and incubated again overnight at 37 °C in Falcon tubes with LB medium supplemented with kanamycin (50 μg/mL). Then, again, the plasmids were isolated as described before, and their quality and size checked in a 0.8 %

agarose gel electrophoresis, run at 5 V/cm for 45 minutes. The rest of the culture was stored to be used as pre-culture in the expression step. In addition, a PCR was conducted, using specific primers for p7xN3H and p7xC3H, to check their presence in these colonies, in duplicates. The sequence of the mentioned primers is detailed in Table 11. The program used to perform this reaction is described in Table 12.

Table 11. Sequence of the primers used to amplify the vectors p7xNH3 and p7xC3H.

Primer name	Primer sequence
T7 promoter	5'- TAATACGACTCACTATAGGG
T7 terminator	5'- GCTAGTTATTGCTCAGCGG

Table 12. PCR program used to verify the presence of p7xN3H and p7xC3H in the cells.

Step	Temperature (Celsius)	Time (seconds)
Initial denaturation	94	60
25 Cycles	94	15
	50	30
	68	60

3.4 Protein expression

Then, the transformation of the expression strain followed. The chosen strain was a derivative of *E. coli* BL21 named LOBSTR (Low Background Strain) (Andersen et al., 2013). First, 500 µL of the stored pre-culture was transferred to 25 mL of YT-2 medium (ThermoFisher, catalog number: 22712020) supplemented with kanamycin (25 µg/mL), previously warmed to 37 °C. This culture was incubated at 37 °C for 4 hours, shaking at 175 rpm, and then the O.D. was measured. Following, the cells were induced by adding IPTG (Isopropyl β-D-1-thiogalactopyranoside) until a final concentration of 0.4 mM; the cultures were incubated semi-anaerobically at 37 °C for 5 hours, while shaking at 175 rpm. The O.D. was measured again, the cultures transferred to new Falcon tubes and centrifuged for 15 minutes at 6000 rpm at 4 °C. The supernatant was discarded, and the pellet lysed by sonication: 250 µL of a buffer containing 50 mM TRIS, 50 mM NaCl and 10% Glycerol (pH = 8.0) was used to resuspend the pellet; lysozyme was added as well. After incubating on ice for 30 minutes, the cells were

given 5 cycles of 20 s on/5 s off pulses on ice at 80 % output power, using an Ultrasonic Homogenizer 4710 series sonicator (Cole-Parmer). Then, 150 µL of each sample were taken and centrifuged, to get rid of the insoluble fraction. The supernatant was heat treated at 75 °C for 20 min and centrifuged (5000 rpm for 20 minutes) to remove *E. coli* proteins (Godde et al., 2005). This lysate and the remaining 100 µL of the samples were used to run a 30% Mini-PROTEAN® TGX™ SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis), by Bio-Rad (catalog number 456-1093), for 40 minutes at 200 V. The samples were previously denaturized by adding 2-β-Mercaptoetanol to the SDS loading buffer and heating them at 95 °C for 10 min.

3.5 Enzyme activity assay

An EnzChek® Protease Assay Kit was used to assess the activity of the expressed protein. This test consists of a green direct fluorescence-based assay and detects serine, acid, and sulfhydryl proteases (Thompson et al., 2000, Bernard et al., 2003, Brzin et al., 2000). The procedure was followed as indicated in the kit manual. First, the substrate was reconstituted and the master mix prepared, containing 77,5 µL of the assay buffer and 12,5 µL of the 10 µg/mL BODIPY-FL casein per reaction. Then, 50 µL of the lysates were applied to the wells of a black microtiter plate, and 50 µL of the master mix was added to each well. As a positive control, Alcalase® by Sigma (catalogue number 126741) was used. The reaction master mix was used as a negative control for this assay. The samples were incubated in the dark for 1 hour at 37 °C. The fluorescence was determined in a microplate reader using standard fluorescein filters (excitation = 485, emission = 530).

Results

4. Enrichment and isolation

The sample containing the studied bacterium was enriched in MMF medium supplemented with peptone and yeast extract. After incubation for 48 hours at 65 °C, growth was checked under the contrast-phase microscope, showing very good growth. Then, the isolation by the dilution to extinction procedure followed. The microbe, named *Fervidobacterium pennivorans* strain T, was then studied in the contrast-phase microscope, showing the typical morphology found in the Thermotogales group: rods with a toga covering the cell. This bacterium can occur in couples or short chains. Some clusters were spotted as well, suggesting that this strain can form colonies when incubated using solid media. No spores were found. This culture was then assumed to be pure.

5. Phylogenetic identification

The phylogenetic analysis of the strain started with a PCR amplification of the 16S rRNA gene using the program described in the Material and Methods section. The PCR products were checked by a 1.5% agarose gel electrophoresis which results are shown in Figure 11. The gel showed bands of similar size for both control (wells 2 and 3) and *F. pennivorans* strain T (wells 4 and 5), and consistent with the bacterial 16S sRNA gene size (Woese et al., 1983).

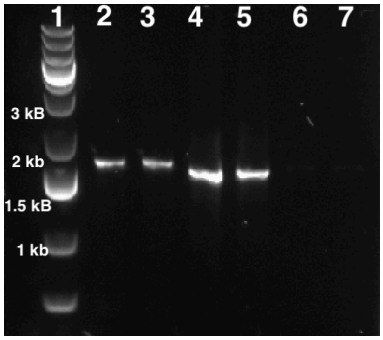


Figure 11. 1.5% agarose gel electrophoresis showing: 1Kb Ladder (well 1), 16S rRNA PCR product from a positive control (*Thermosipho africanus*) (wells 2 and 3), 16S rRNA PCR product from the *F. pennivorans* strain T (wells 4 and 5), and negative control (PCR master mix) (wells 6 and 7).

A cycle sequencing protocol using this PCR product was conducted. The DNA sequences obtained from each primer were merged and a consensus sequence was obtained corresponding to the nearly complete 16S rRNA gene. This sequence was compared in BLAST,

where it could be seen that the obtained 16S rRNA gene sequence is almost identical to *F. pennivorans* strains DYC (accession number CP011393.1) and DSM 9078 (the species type strain, with accession number CP003260.1), with an E-value of 0.0 and approximately a 99 % of sequence identity (1378 over 1394 nucleotides with 2 gaps for DYC strain, and 1376 over 1394 nucleotides with 2 gaps for DSM 9078).

6. Physiology and morphology

6.1 Carbon sources utilization

Different carbon sources were tested, for their potential use as growth substrates in biotechnological processes. 1 mL of culture was transferred from the master flask with 0.5 % peptone and 0.05 % yeast extract to flasks with 10 mL of MMF medium and the substrates shown in Table 13, with a concentration of 0.5% and 0.05% yeast extract. The flasks were incubated for 48h at 65 °C and then 0.5 mL was again transferred from these flasks to new ones with the same substrates and the same concentrations and incubated again for 48h. Peptone at 0.5 % was used as a positive control and only yeast extract at 0.05 % was used as a negative control. After this, growth was checked under the phase-contrast microscope. The flasks with galactose and glucose being the ones reporting the best growth. Following, dextrin was in the next step in preference, and then lactose, sorbitol and cellulose. Flasks with mannose, starch, sucrose and raffinose did not show good growth, but still were considered as positive, although using those substrates to grow this strain is not recommended. Finally, no growth was reported in flasks with arabinose and mannitol.

Table 13. Substrates used to test the carbon utilization by *F. pennivorans* strain T. Very good growth under the phase-contrast microscope is represented as +++, good growth as ++, weak growth as +, no clear growth as +/- and no growth as -.

Substrate	Growth
Peptone	+ + +
Arabinose	-
Cellulose	+
Dextrin	++
Galactose	+++
Glucose	+++
Lactose	+
Mannitol	-
Mannose	+/-
Sorbitol	+
Starch	+/-
Sucrose	+/-
Raffinose	+/-

6.2 pH test

The ability of this microbe to grow at different pH was tested by setting up flasks with 10 mL MMF medium supplemented with peptone (0.5 %) and yeast extract (0.1 %), incubating them for 48 h at 65 °C with 1 mL inoculum of a *F. pennivorans* culture. Two different pH values were tested: 5.5 and 8.5, and no growth could be reported in any of the flasks after checking by the phase-contrast microscope.

6.3 Osmotic stress

For testing the tolerance of *F. pennivorans* strain T to osmotic stress, a battery of flasks of 10 mL medium supplemented with peptone (0.5%) and yeast extract (0.1%) with 0.3 % (MMF without extra NaCl supplementation), 1 %, 2 %, 3 %, 4 % and 5 % of NaCl were set up and incubated at 65 °C for 48 h. Then, growth was checked under the phase-contrast microscope. The results can be seen in Table 14. Growth was positive and clear with NaCl concentration of 1 %, but when increasing the amount of NaCl, the organism grew slowly, with no clear growth in flasks with 2 % and 3 % NaCl. In 4 % NaCl flasks, although there was growth, some dead cells could be spotted, meaning that this concentration starts to be toxic for the organism. In flasks with 5 % NaCl, more dead cells could be seen under the phase-contrast microscope, being not clear the ability of *F. pennivorans* to grow in these conditions.

Table 14. Concentration of NaCl used to test the tolerance of *F. pennivorans* to osmotic stress. Clear growth is represented as ++, no clear or very little growth as +/- and uncertain as -/+

NaCl Concentration (%)	Growth	
	Flask 1	Flask 2
0.3	+++	+++
1	++	++
2	+/-	+/-
3	+/-	+/-
4	-	+/-
5	-/+	+/-

6.4 Temperature tolerance

Four different temperatures were used to test the thermo-tolerance of *F. pennivorans* strain T: 38 °C, 55 °C, 65 °C and 80 °C. Flasks were set up with 10 mL MMF medium containing peptone (0.5 %) and yeast extract (0.1 %), and incubated for 48h. Then, as no growth was seen in any of them under the phase-contrast microscope, all of them were again incubated for 48 extra hours. Results after these 96 hours of incubation can be seen in Table 15. *F. pennivorans* is not able to grow either at 38 °C nor at 80 °C, as any cell could be spotted under the phase-contrast microscope after the incubation time. It can thrive at 55.0 °C, though, needing more time to grow at this temperature.

Table 15. Temperatures used to test *F. pennivorans* thermo-tolerance. Very clear growtg is represented as +++, clear growth as ++, and no growth is represented as -.

Temperature (°C)	Growth	
	Flask 1	Flask 2
38	-	-
55	++	++
65	+++	+++
80	-	-

6.5 Keratinase activity

After four days, the start of the degradation of breast feathers was seen, and the same for wing feathers after one week of incubation. In Figure 12 can be seen that breast feathers were almost completely dissolved after one week, while wing feathers were almost fully disintegrated after ten days of incubation, as shown in Figure 13.

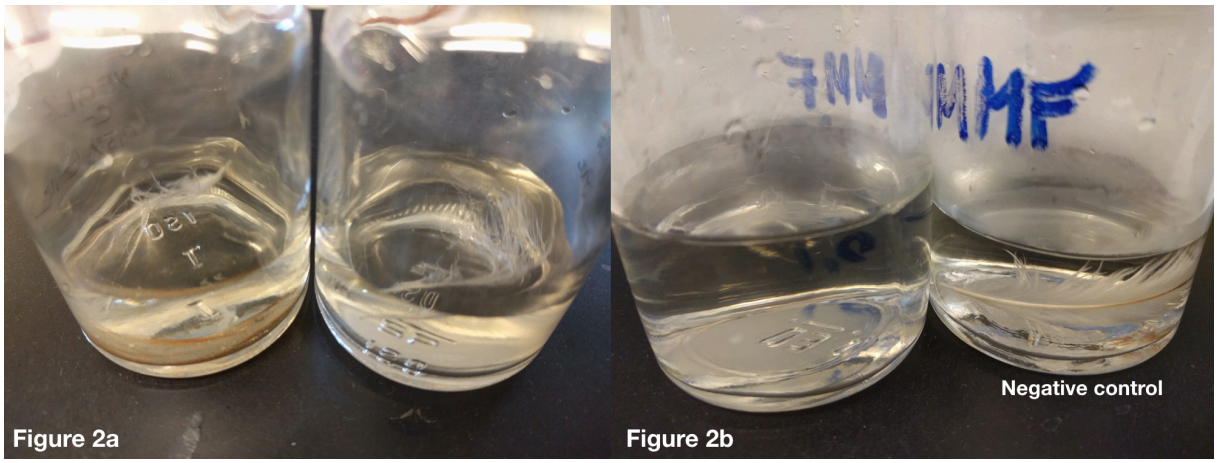


Figure 12. Breast chicken feathers set up to test the keratinolitic activity of *F. pennivorans* strain T. Figure 2a shows the feathers just before incubation at 65 °C. Figure 2b shows dissolved breast feather after one week of incubation at 65 °C (left) and negative control (right).

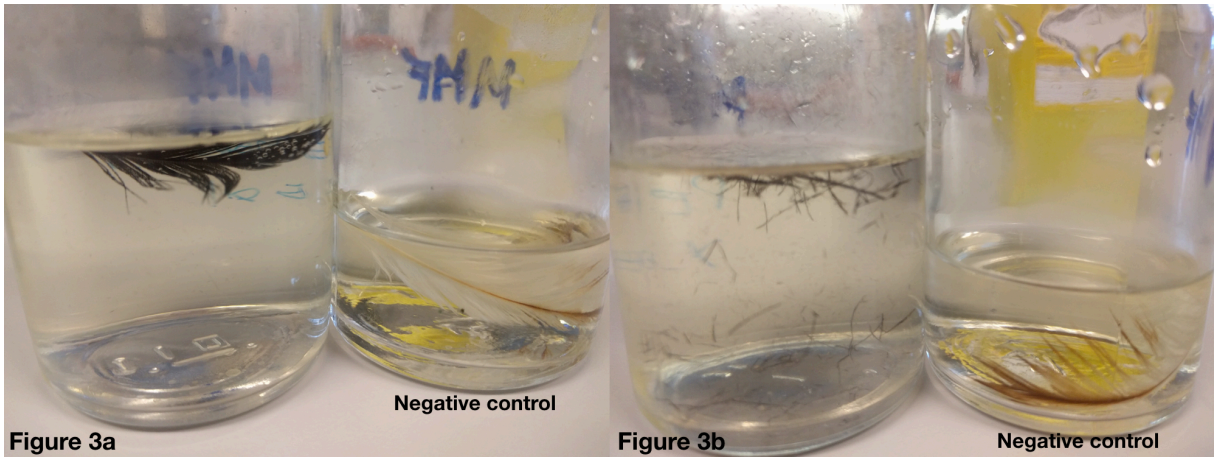


Figure 13. Wing chicken feathers set up to test the keratinolitic activity of *F. pennivorans* strain T. Figure 3a shows the feathers before incubation at 65 °C, with the negative control to the right. Figure 3b shows dissolved wing feather after ten days of incubation at 65 °C (left) and negative control (right).

6.6 Microscopy

F. pennivorans strain T was studied using contrast-phase microscopy and SEM. The obtained pictures of the isolated strain are shown in Figure 14 and Figure 15. This strain's morphology is typical for a bacterium belonging to the Thermotogae group: rod-shaped, straight to slightly curved cells with a sheath-like toga covering the bacterium. The cells occur singly, in pairs or in chains. Also, this strain presents pleomorphism, as some globular structures or terminal spheroids can be seen, as well as other different morphologies and shapes. No spores could be spotted.

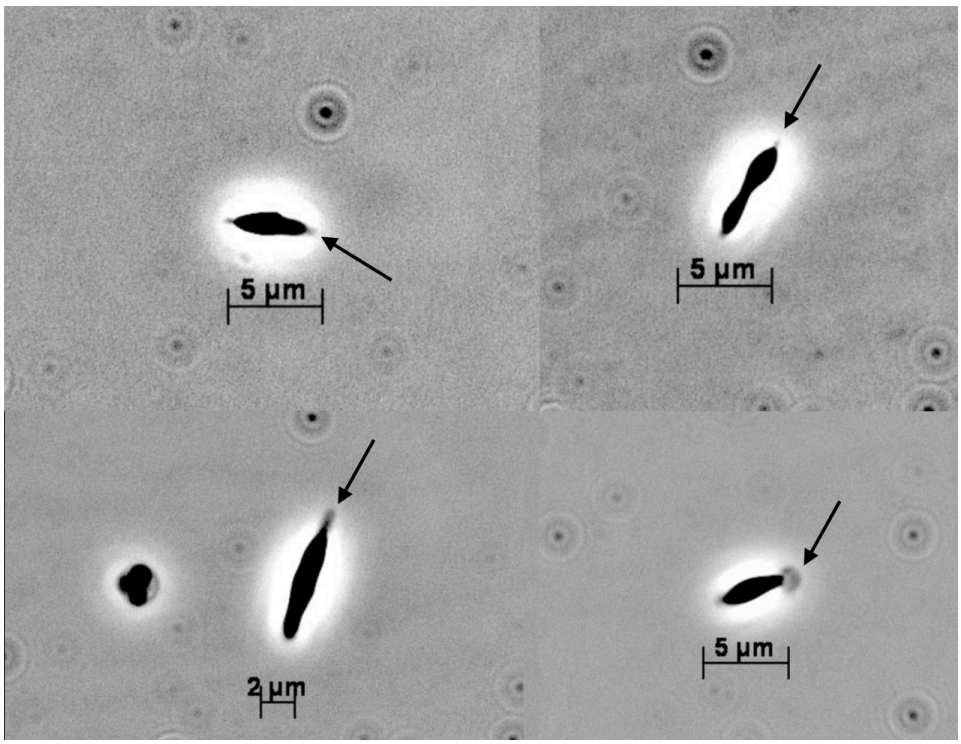


Figure 14. Phase-contrast microscope images of *F. pennivorans* strain T. The arrows point to the toga.

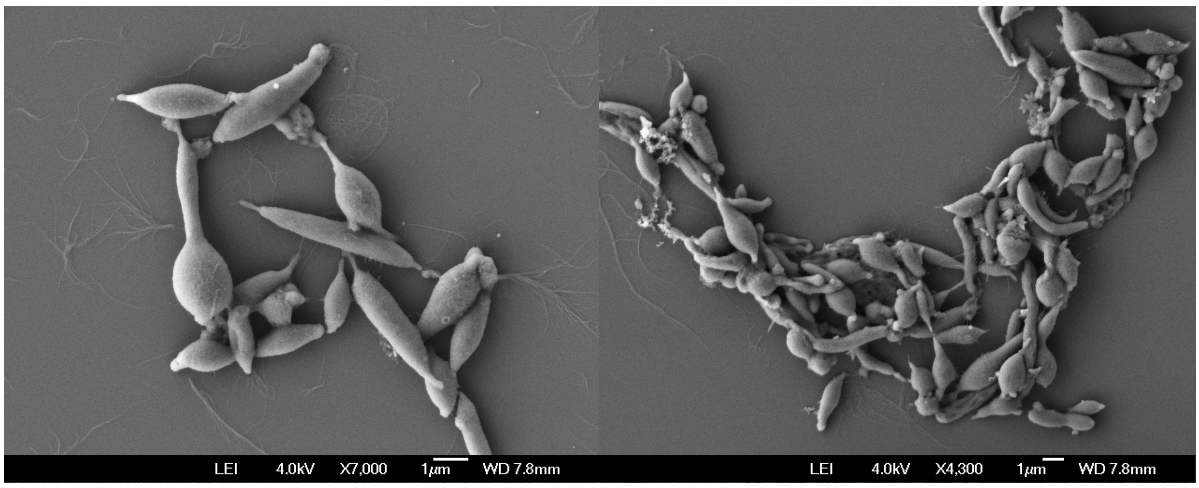


Figure 15. Scanning electron microscope images of the isolated strain. Bar 1 µm.

7. Genomic analyses

7.1 Genomic DNA isolation

After growing the bacterium in MMF medium as previously described and its growth checked, genomic DNA was isolated as mentioned. Then, a 0.8 % agarose gel electrophoresis was run,

loading approximately 150 ng of genomic DNA in two different wells. Figure 16 shows a picture of the mentioned gel, with the size marker used (left) and the two replicas of genomic DNA (right). The showed good quality and high molecular weight, being solid and without noticeable smearing, proving that the DNA had not suffered any disintegration during the extraction procedure.

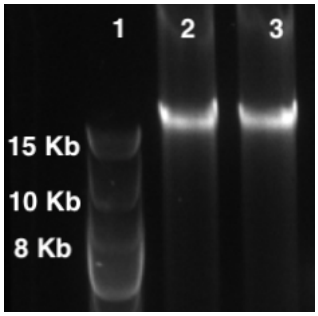


Figure 16. 0.8 % agarose gel electrophoresis showing: 1 kb Plus DNA Ladder (1) and *F. pennivorans* strain T genomic DNA (2 and 3).

8. Genomic and Bioinformatic analyses

8.1 Phylogeny

A phylogenetic tree based on the 16S rRNA gene sequence was reconstructed for the members of the *Fervidobacterium* genus and it is shown in Figure 17. According to the 16S RNA gene sequence, *F. pennivorans* strain T is placed next to the *F. pennivorans* type strain (DSM 9078) in the tree with a bootstrap value of 37 %, while *F. pennivorans* DYC is placed next to these two, being these three strains clustered together 89 times each 100 this tree was iterated (bootstrap value of 89 %).

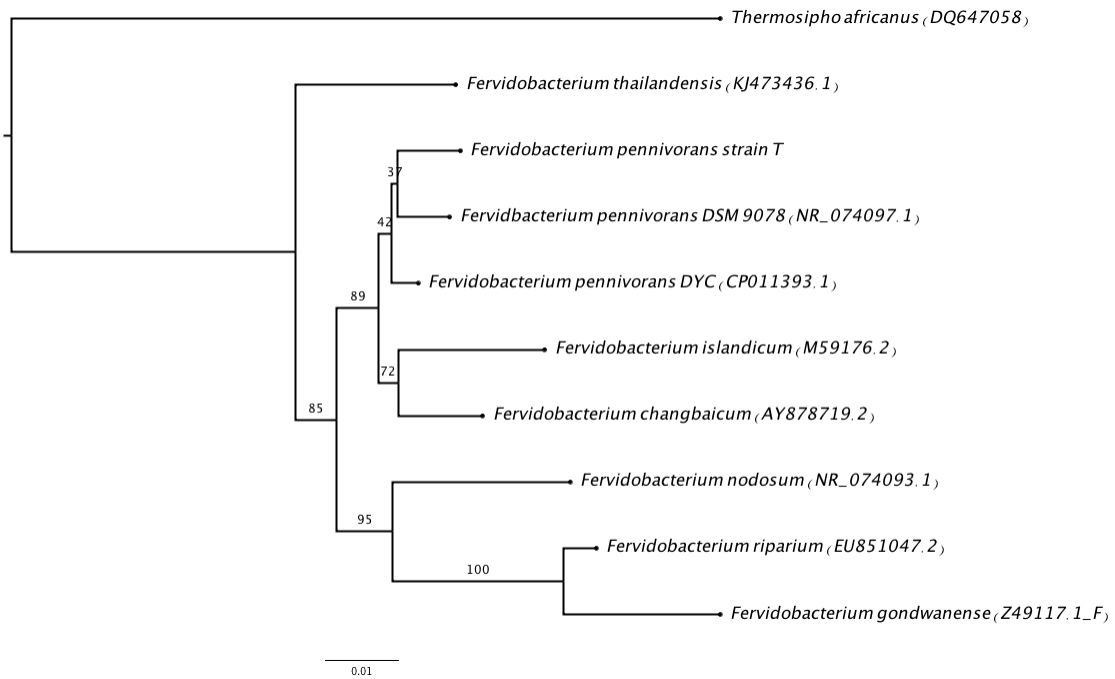


Figure 17. Neighbor-Joining phylogenetic tree of 16S rRNA gene sequences of members of the genus *Fervidobacterium* showing the relationship between the experimental strain and the rest of species of the genus. The 16S rRNA gene sequence of *Thermosipho africanus* was employed as an outgroup lineage. Bootstrap values as a percentage of 100 replications are presented. Bar, 0.01 changes per nucleotide position. Sequence accession numbers are shown in brackets

8.2 Genomic analyses

The genomic data of *F. pennivorans* strain T was downloaded and analysed. Prior to the genomic assembly the best k-mer length was determined using KmerGenie software, being the predicted best k value = 61, as shown in Figure 18.

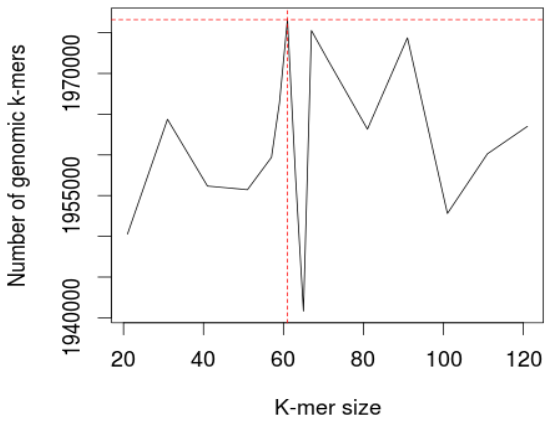


Figure 18. Representation of the number of genomic k-mers (y axis) related to the k-mer size (x axis). Optimal value is represented as the most represented k-mer size in the genome.

This k-mer size was used to assemble the genome with the CLC Genomic Workbench 11 suite. However, after a BLAST comparison, it was found that a certain number of contigs present in the sequence did not belong to *F. pennivorans* but to contaminations, most of them from human DNA. So, these contigs were manually filtered out from the sequence file. The genome was then uploaded to the RAST service for annotation, and analyzed using the SEED viewer, which summary output is shown in Table 16. The genomic size was 1,967.686 base pairs distributed in 28 contigs, with a GC content of 39.0 %. The genome was estimated to have 294 subsystems coding for 1896 sequences and 48 RNAs.

Table 16. Output of the summary of *F. pennivorans* draft genome analysis by the SEED viewer.

Genome	*Fervidobacterium pennivorans* strain T
Domain	Bacteria
Taxonomy	Bacteria; *Fervidobacterium pennivorans* strain T
Size (base pairs)	1,967,686
GC Content (%)	39.0
N50	155906
L50	5
Number of Contigs (with PEGs)	28
Number of Subsystems	294
Number of Coding Sequences	1896
Number of RNAs	48

The SEED overview also included a chart with this bacteria's subsystem information, where it could be seen the different categories and features of the subsystem, including metabolism and other features, according to the RAST automatic annotation. The mentioned chart is shown in Figure 19. It has been mentioned the ability of this group of bacteria to utilize a variety of carbon sources, and this can be seen in the SEED chart: up to 188 features are involved in carbohydrates metabolism, 35 in respiration, and 29 in protein degradation. In addition, there are remarkable features like those related with virulence, disease and defense, or the ones linked to dormancy and sporulation, although this bacterium belongs to a non-sporulating group and some of these features were annotated as «sporulation-associated proteins with broader functions».

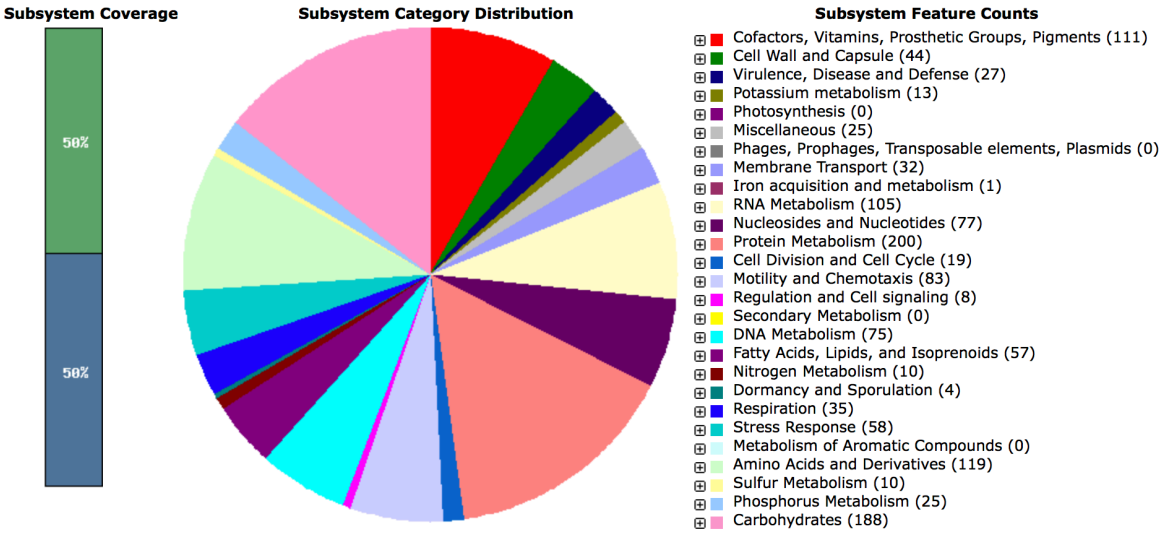


Figure 19. SEED viewer subsystem information which can be seen under the *F. pennivorans* overview in the browser.

The Seed viewer was also used to perform a sequence comparison between the type and DYC strains with strain T's genome. In Figure 20 can be seen the comparison of strain T with the type strain (DSM 9078), where the 85.2 % of the proteins show at least 95 % sequence identity with the type strain. When comparing with DYC strain the results were different, with 52.9 % of the sequences showing an identity of 95 % or more (Figure 21).

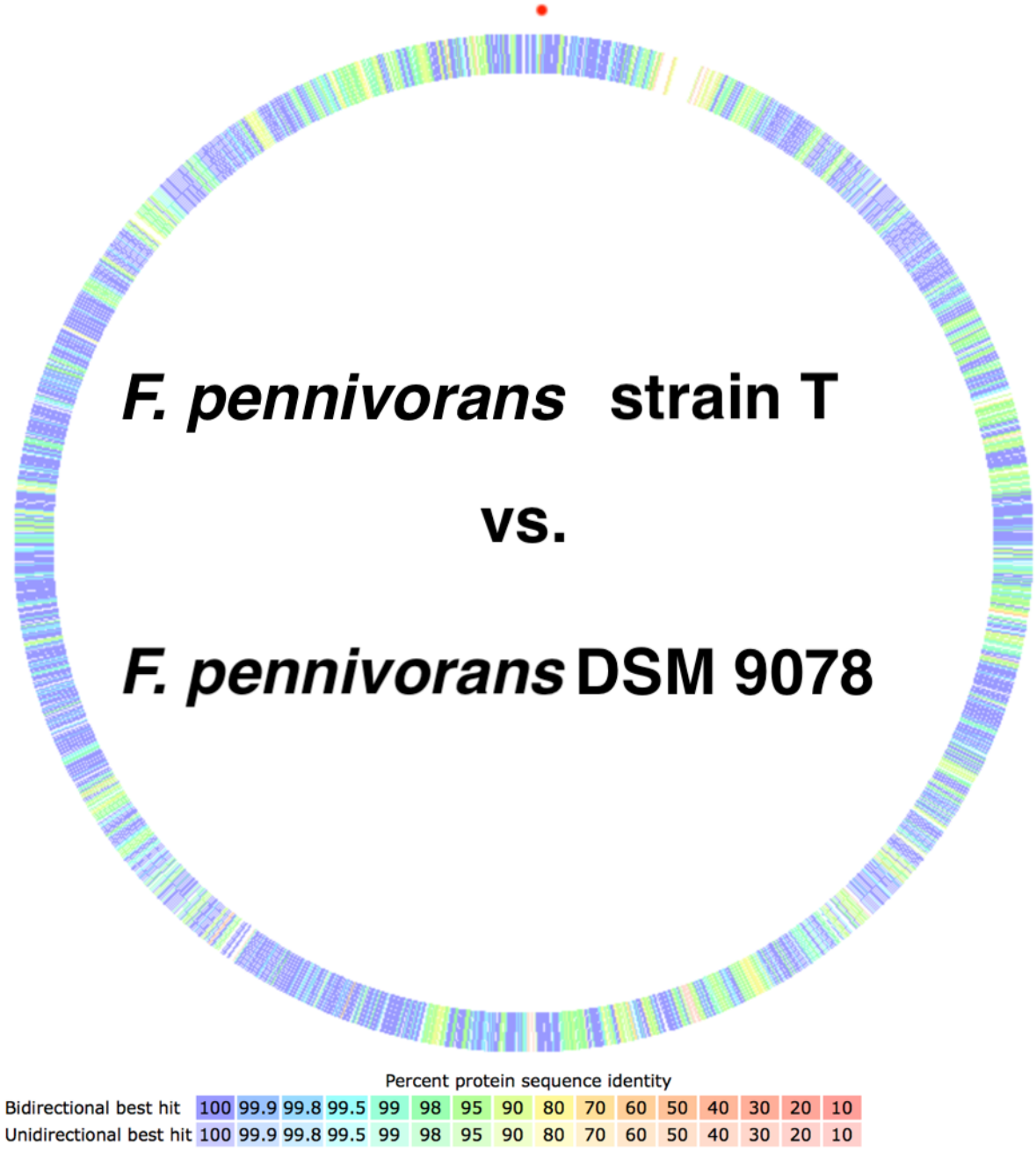


Figure 20. Percent protein sequence identity between *F. pennivorans* strain T genome, annotated in RAST, and *F. pennivorans* DSM 9078 strain.

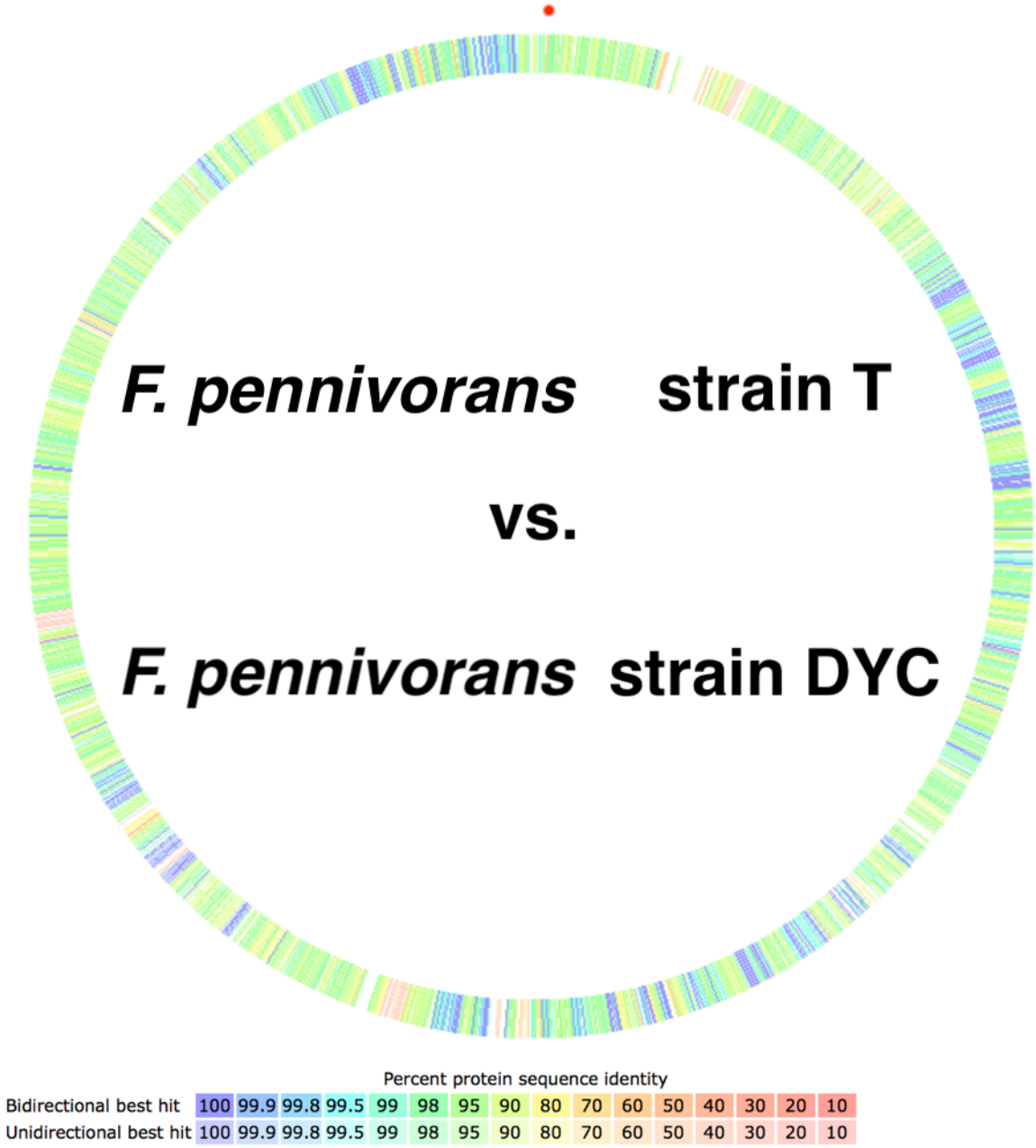


Figure 21. Percent protein sequence identity between *F. pennivorans* strain T, annotated in RAST, and *F. pennivorans* DYC strain.

Likewise, *F. pennivorans* DSM 9078 and *F. pennivorans* DYC strains were also compared using this tool. The representation of this comparison is shown in Figure 22. According to this annotation, these two strains have 46.2 % sequences that show a percent identity of 95 % or higher.

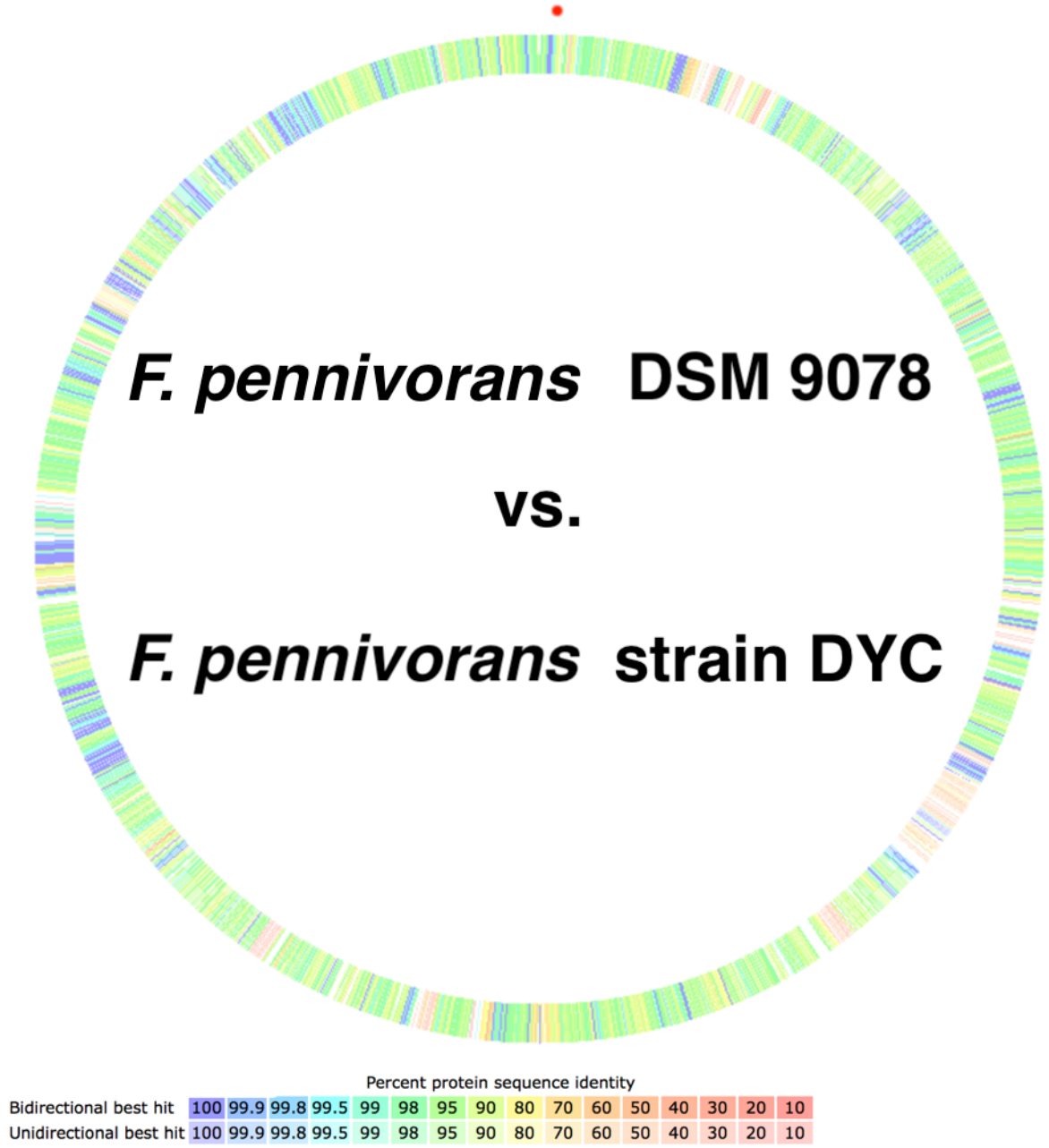


Figure 22. Percent protein sequence identity between *F. pennivorans* DSM 9078 strain and *F. pennivorans* DYC, according to RAST annotation.

8.2.1 Average Nucleotide Identity

F. pennivorans strain T was also compared to *F. pennivorans* type strain and DYC strain through an ANI estimation. In addition, DYC and DSM 9078 genomes were compared as well. The result of these comparisons is shown in Figure 23, Figure 24 and Figure 25, respectively. The comparison of the genome of *F. pennivorans* strain T with the type strain showed an average identity with the draft sequence of 97.28 % (SD = 4.39 %), with a median of 99.76. %. The average bit score was 1597, with a median of 1672.

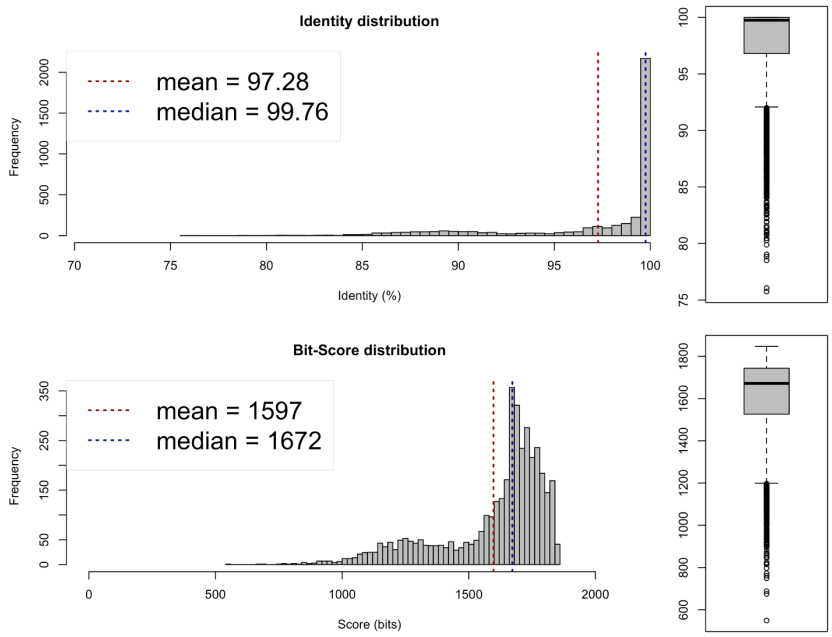


Figure 23. Average Nucleotide Identity comparison between *F. pennivorans* strain T genome and *F. pennivorans* DSM 9078, represented as the distribution of the percentage of identities between the compared organisms (above), and the distribution of the Score of the comparisons (below).

When comparing *F. pennivorans* strain T with *F. pennivorans* DYC strain, the average identity was 91.4 % (SD = 5.33 %), with a median of 89.97 %. Regarding the bit-score distribution, the ANI output showed an average score of 1335, with a median of 1279.

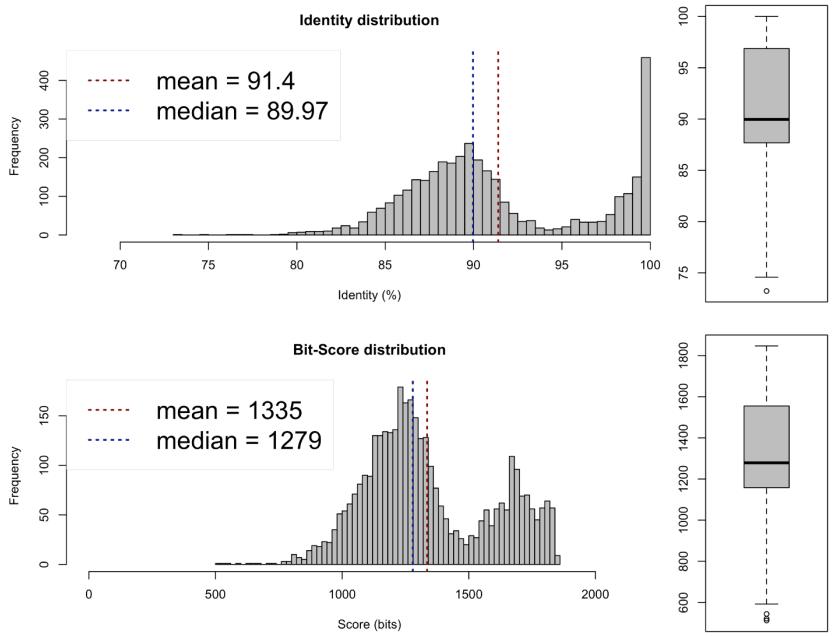


Figure 24. Average Nucleotide Identity comparison between *F. pennivorans* strain T genome and *F. pennivorans* DYC, represented as the distribution of the percentage of identities between the compared organisms (above), and the distribution of the Score of the comparisons (below).

Finally, *F. pennivorans* DSM 9078 and *F. pennivorans* DYC were also compared through an ANI analysis. In this case, the comparison showed an average identity of 90.99 % (SD = 5.17 %), with a median of 89.89. %. The average bit score was 1348, with a median of 1301.

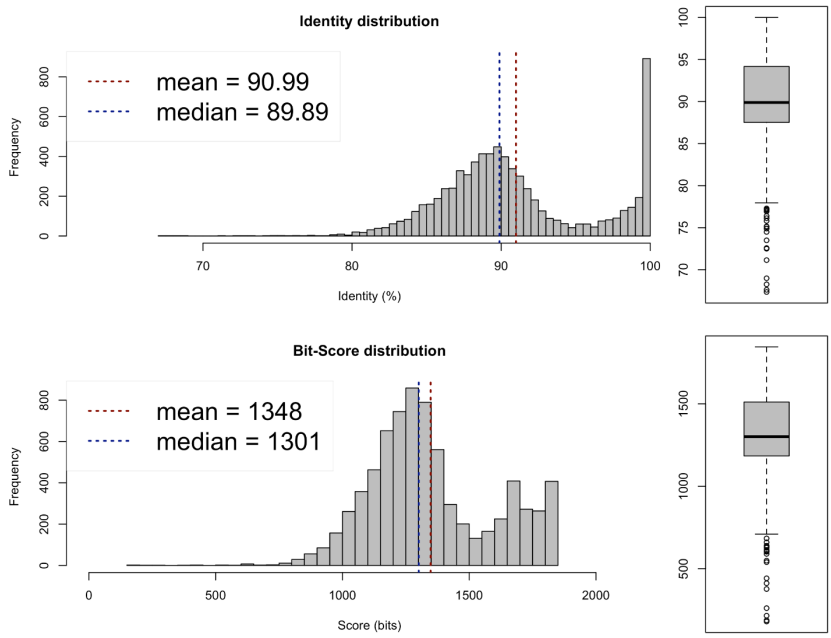


Figure 25. Average Nucleotide Identity comparison between *F. pennivorans* DYC and *F. pennivorans* DSM 9078 strains, represented as the distribution of the percentage of identities between the compared organisms (above), and the distribution of the Score of the comparisons (below).

So, according to the ANI estimation, *F. pennivorans* strain T and *F. pennivorans* strain DSM 9078 belong to the same species, as the ANI is above 95 %. Nevertheless, when comparing both of them with DYC strain, the ANI values are 91.4 % and 90.99 %, respectively, meaning that *F. pennivorans* DYC should not be included in the *F. pennivorans* species.

8.2.2 Genome-to-Genome distance calculation (GGDC)

Another genomic comparison between the experimental draft genome and these two strains was conducted using the Genome-to-Genome distance analysis tool by Leibniz Institute (DSMZ), and the result is detailed in Table 17. When comparing *F. pennivorans* strain T with *F. pennivorans* type strain, the recommended formula "Identities / HSP length" (High Scoring Segment Pair) estimated a probability of 91.41 % that the DNA-DNA Hybridization (DDH) between these two strains is higher than 70 %, representing this value the species threshold. The probability that the DDH between *F. pennivorans* strain T with *F. pennivorans* DYC strain

is higher than 70 % was estimated by this tool in 20.18 %. Likewise, the type strain was compared with DYC strain, as shown. According to the GGDC estimation and using its recommended formula (identities / HSP length), the probability that these two strains belong to the same species is 7.21 %.

Table 17. Genome-to-genome distance estimation between *F. pennivorans* strain T genome, the species type strain and DYC strain. Another comparison between type strain and DYC strain is also shown. DDH is the DNA-DNA hybridization value estimated by this calculator. The last row shows the estimation that the compared strains' genomes hybridation percentage is over 70 %.

Query genome	Reference genome	DDH	Prob. DDH >= 70%
F. pennivorans strain T	*F. pennivorans* DSM 9078	81.1	91.41
F. pennivorans strain T	*F. pennivorans* DYC	50.4	20.18
F. pennivorans DSM 9078	*F. pennivorans* DYC	44.3	7.21

So, according to these results, *F. pennivorans* strain T and *F. pennivorans* DSM 9078 should be considered as the same species, but different strains. Likewise, *F. pennivorans* DYC should not be considered as a *F. pennivorans*.

8.2.3 Genomic alignment

The genome of *F. pennivorans* strain T was then aligned with *F. pennivorans* type strain, DSM 9078, and with DYC strain, which got the highest score when comparing the 16S rRNA sequence of the studied bacterium with BLAST. These alignments were run with the Mauve software, using the progressive alignment algorithm, and reordered. Nevertheless, a high number of contigs of the sequenced genome did not find their correspondent sequence in the type strain. These contigs were compared with BLAST using the blastn algorithm and was found that they did not belong to *F. pennivorans* but to contaminations, proceeding most of them from human DNA.

So, these contigs were deleted from the FASTA file and new alignments with the type and DYC strains were conducted, following the same procedure as before. The output of these alignments is shown in Figure 26 and Figure 27. As can be seen, the type strain has a genome visibly larger than *F. pennivorans* strain T genome: 2,166.381 bp, according to the Genbank

database, while the sequenced genome is shorter than 2 Mb. The alignment shows an almost perfect match and synteny between both genomes, with small gaps representing those fragments which did not find their homologous regions.

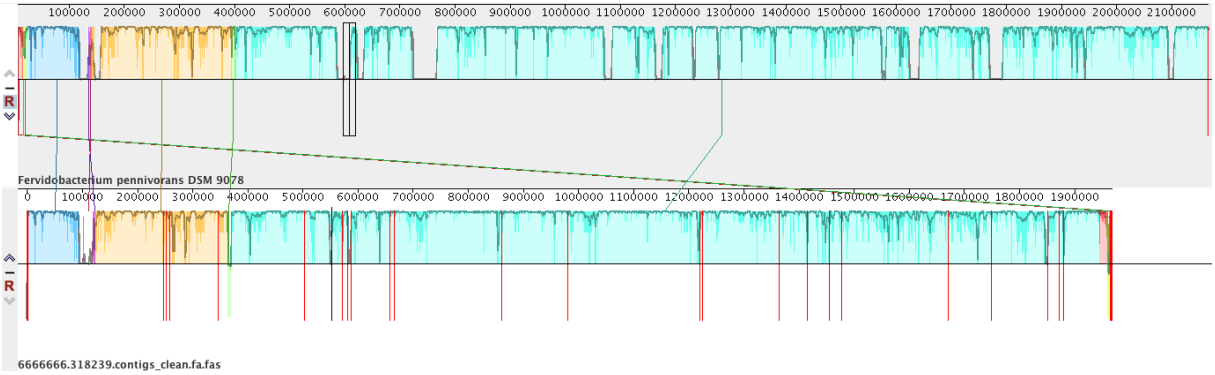


Figure 26. Mauve output after the alignment of the *F. pennivorans* strain T (below) and *F. pennivorans* type strain (DSM 9078) (above). The vertical red lines represent contig boundaries.

Likewise, the genome of the DYC strain is 2,061.852 bp long according to the Genbank database, visibly larger than *F. pennivorans* strain T genome, as can be seen in Figure 27. Both genomes aligned almost completely, with a small number of genomic gaps; nevertheless, according to Mauve, some parts of the genome of *F. pennivorans* strain T are inverted when comparing with DYC's genome, being these fragments in different order, even after running the "reorder contigs" option. This area corresponded to contigs 1, 5, 10, 16, 21 and part of contig 6.

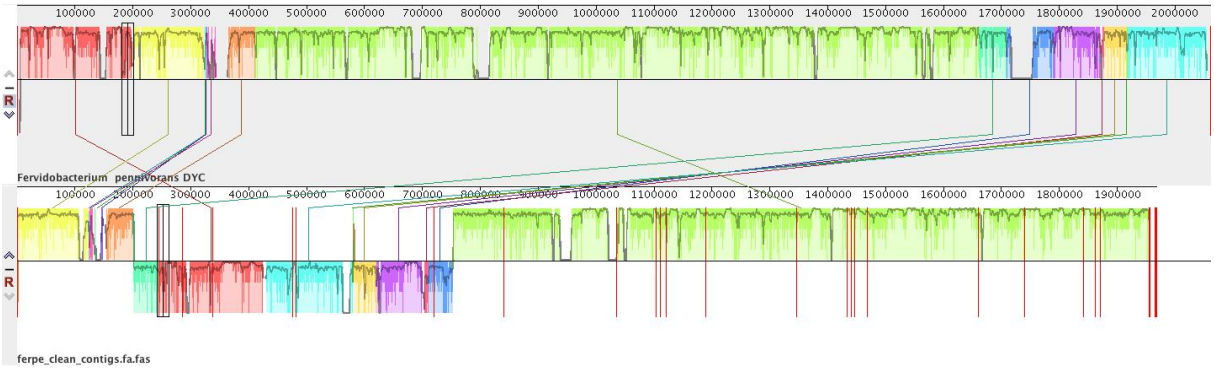


Figure 27. Mauve output after the alignment of *F. pennivorans* strain T (below) and *F. pennivorans* DYC strain (above). The vertical red lines represent contig boundaries.

8.2.4 Dot-plot analysis

Finally, these sequence comparisons were as well represented in a Dot-plot chart, using the Gepard tool and using the strain T contigs reordered, according to Figure 26, to spot the regions with high similarity and synteny. Figure 28 shows the representation of *F. pennivorans* DSM 9078 and *F. pennivorans* strain T. The contigs were previously reordered with the Mauve software, using the type strain as the template sequence. It can be seen as almost every contig from the experimental genome finds its match with the type strain genome, indicating that these two genomes are almost identical, with only a few small gaps appearing/showing up in the representation. Likewise, in Figure 29 can be seen the dot-plot matrix for the global alignment of *F. pennivorans* strain T genome and *F. pennivorans* DYC. In this case, according with the plot, the sequences are not identical, with different order of some fragments and even some large genomic inversions.

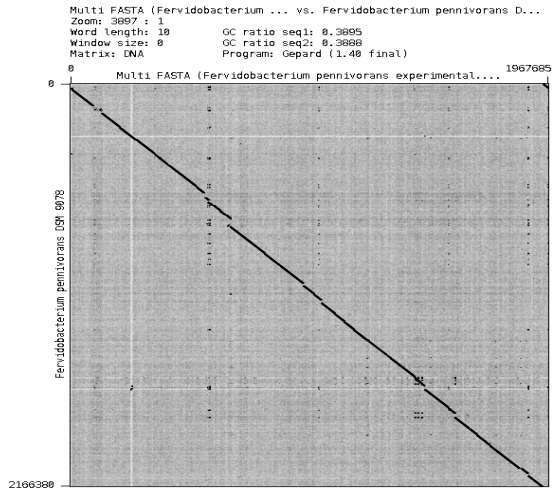


Figure 28. Dot-plot matrix comparison between the aligned genomes of *F. pennivorans* type strain, DSM 9078, (y axis) and the genome of *F. pennivorans* strain T (x axis). Gpard software with default parameters was used. The black line represents the matching sequences.

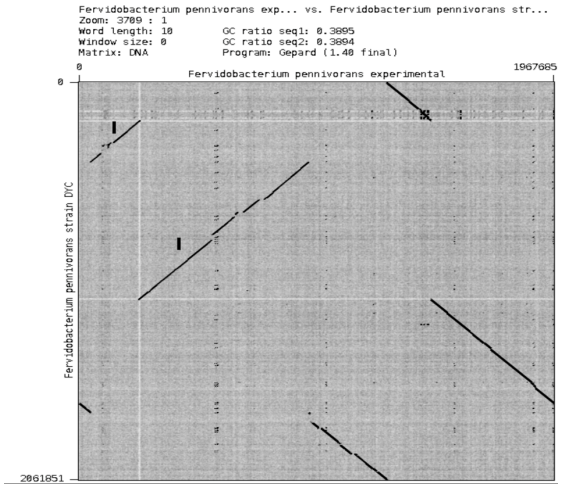


Figure 29. Dot-plot matrix comparison between the aligned genomes of *F. pennivorans* DYC (y axis) and the genome of *F. pennivorans* strain T (x axis). Gepard software with default parameters was used. The black lines descending from left to right represent the matching sequences, being also in the same order. The "I's" mark the inverted fragments.

Again, the species type strain, DSM 9078, was as well plotted against DYC strain to obtain their respective dot-plot matrix. As can be seen in Figure 30, the representation is almost identical to the one obtained after the plotting of the DYC strain against *F. pennivorans* strain T, with some differences between them and some inverted fragments.

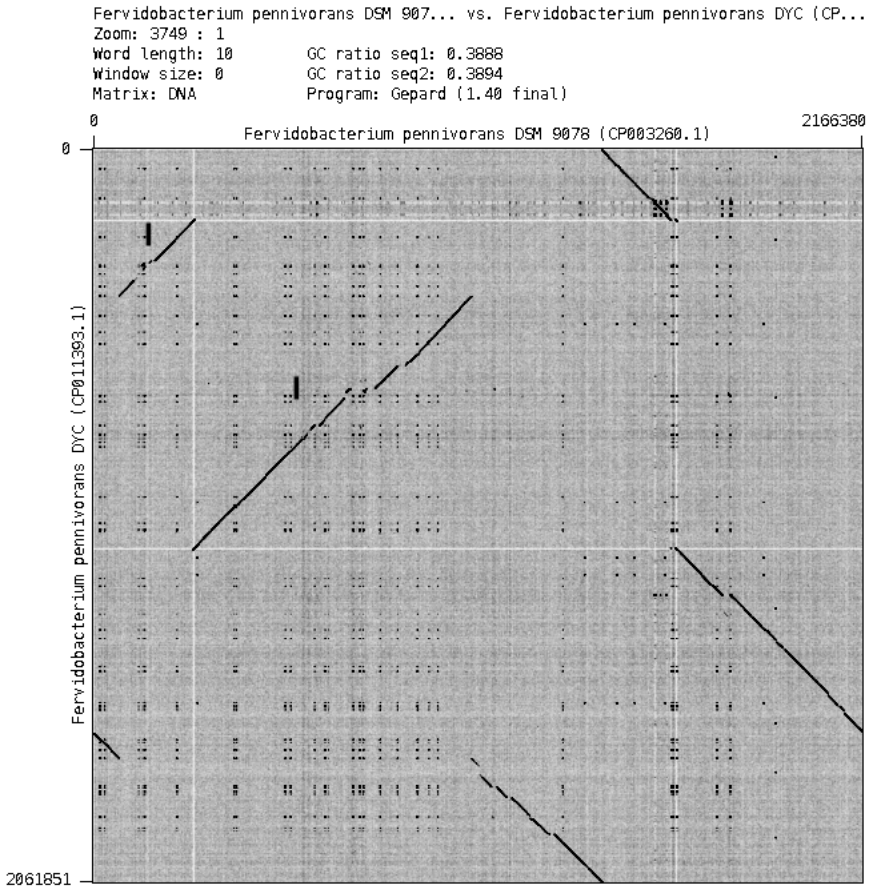


Figure 30. Dot-plot matrix comparison between the aligned genomes of *F. pennivorans* type strain, DSM 9078 (x axis) and *F. pennivorans* DYC strain (y axis). Gepard software with default parameters was used. The black lines descending from left to right represent the sequences which match and are in the same order. The "I's" mark the inverted fragments.

8.2.5 Conserved signature indels

The genomes of the bacteria belonging to the genus *Fervidobacterium* were aligned using BLAST Ring Generator (BRIG) v0.95, and an image with a ring representation of this alignment was generated (Alikhan et al., 2011). This image is shown in Figure 31. The inner ring belongs to *F. pennivorans* strain T genome, then the GC skew and content are represented. The external rings correspond to the other members of the *Fervidobacterium* genus. The more intense the color in each ring, the higher the similarity found with the experimental genome.

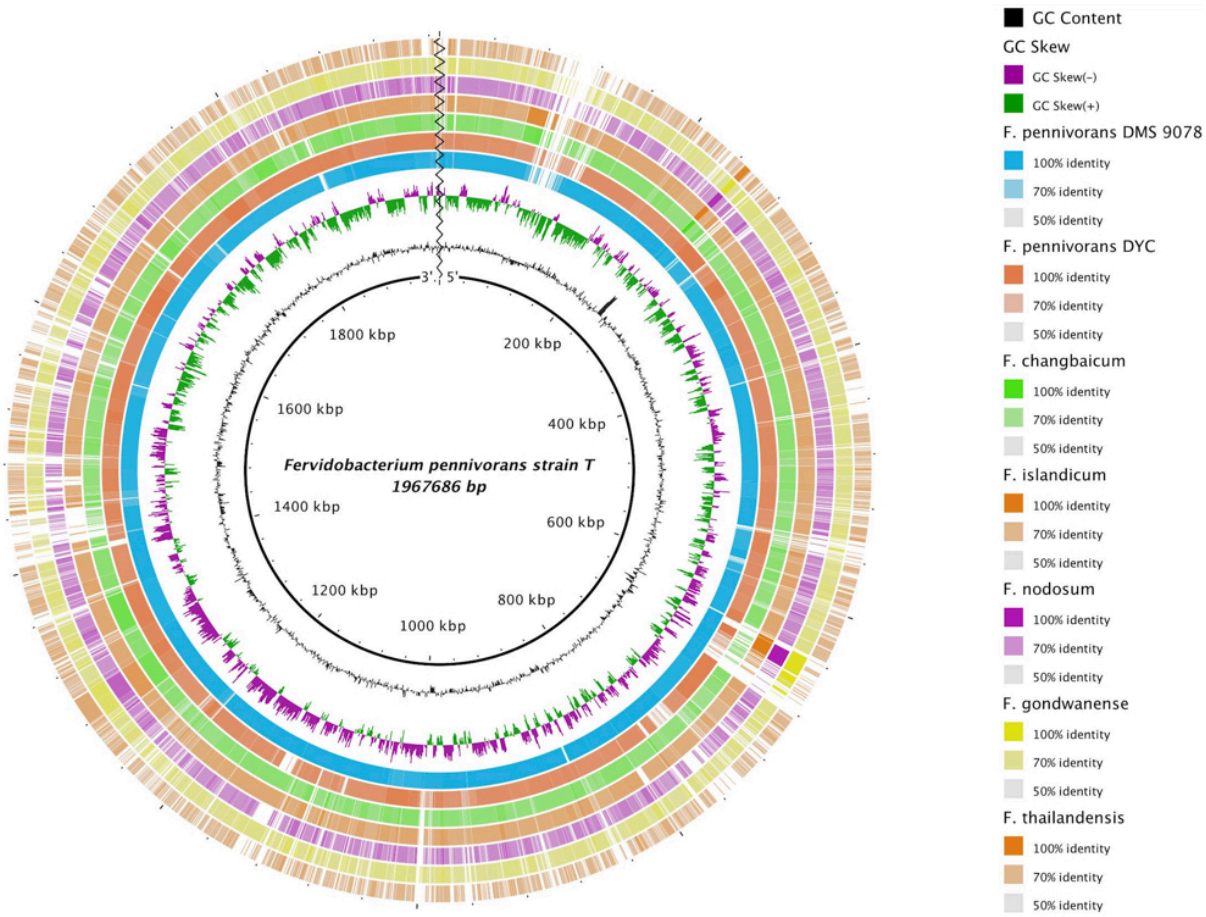


Figure 31. Blast Ring Generator (BRIG) representation of the alignments of the *Fervidobacterium* species' genomes. The internal circle corresponds to *F. pennivorans* strain T and the external ones belong to the different *Fervidobacterium* species. Sections with a GC stew are shown in green and purple, while the GC content is represented in black.

Analyzing this figure, some parts with high level of similarity and a high GC skew were found. These parts were investigated using Artemis software. An example of this can be seen in Figure 32, which shows a sequence in the strain T with a GC content significantly higher (values higher than 2.5 times the standard deviation) compared to the average content.

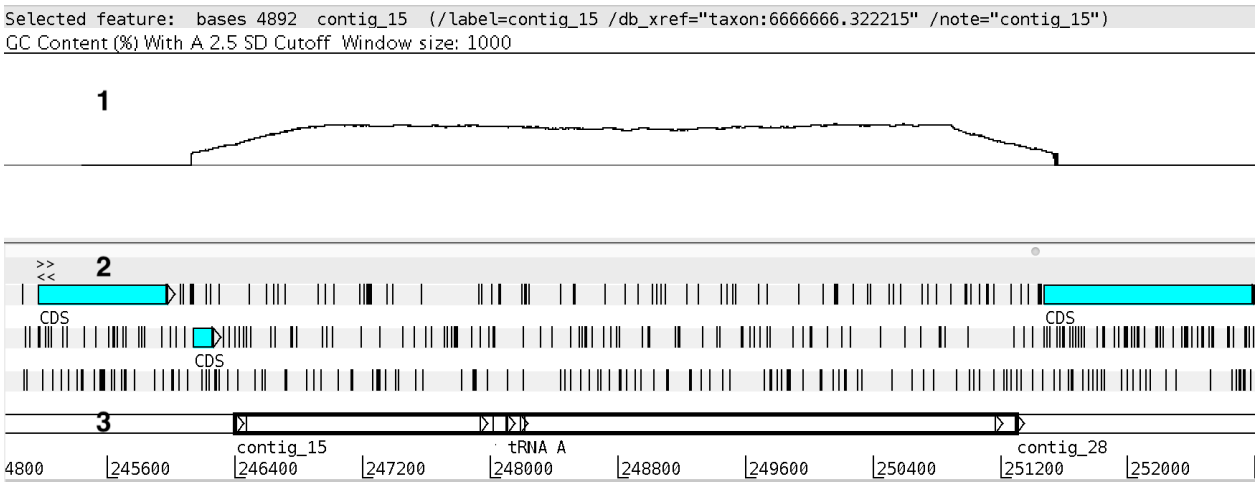


Figure 32. Artemis software snapshot showing a genome sequence region of *F. pennivorans* strain T with a GC content higher than 2.5 times the standard deviation (1), the three reading frames of the forward DNA chain, with the stop codons marked as a vertical black line and the annotated genes highlighted in blue (2). Also, the contig extension is represented (3).

These fragments were also compared with BLAST to find possible homologues for each fragment or gene product. Some of these fragments were annotated and unique for *F. pennivorans* species, but others are shared by some or all the members of the genus (Figure 33). A full table with all the Conserved signature indels (CSI) candidates is shown in the Appendix.

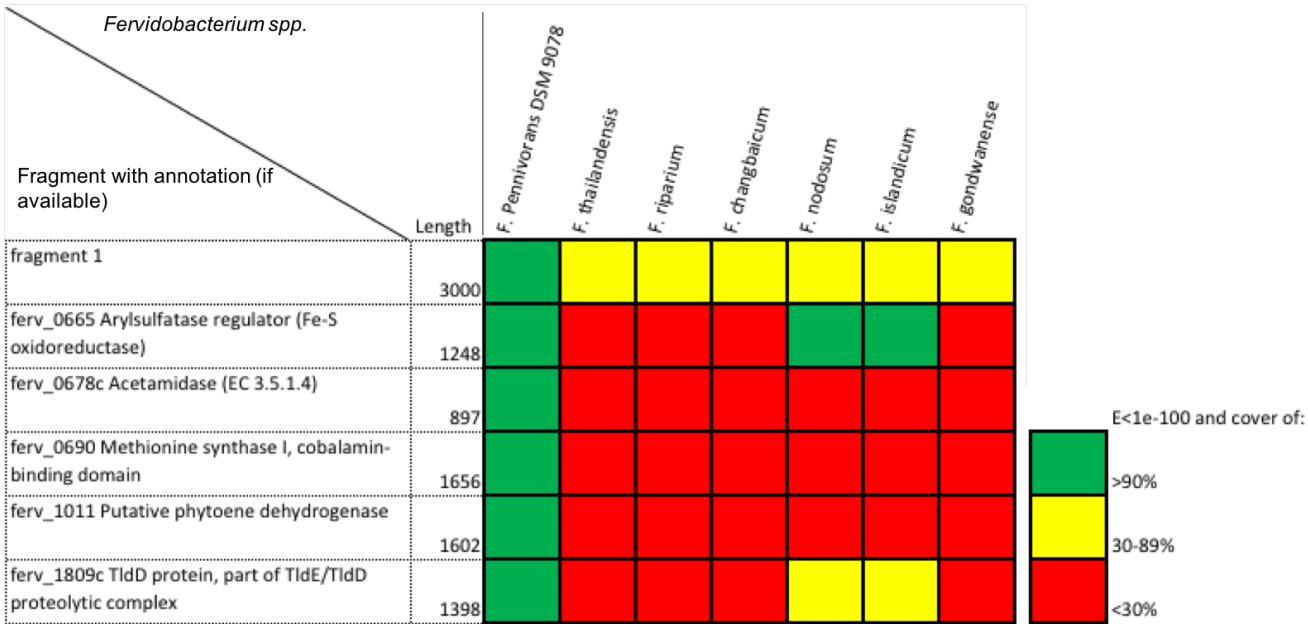


Figure 33. Annotated DNA sequences with a GC content 2.5 times higher (or more) the standard deviation. For each fragment or gene product the presence or absence in the other *Fervidobacterium* members is shown, with the identity level provided by BLAST, being this 90 % or higher (represented

in green), between 89 % and 30 % (represented in yellow), and below 30 % or absent (represented in red). Also, it is shown if the fragments were found in other species with an E value below $1 \cdot 10^{-100}$

9. Cloning of a putative keratinase gene

9.1 3D protein structure prediction

A candidate gene encoding a putative keratinase was chosen, based on the annotations made by RAST. It was a protein which activity had not been reported. A 3D model of this keratinase was predicted using the SWISS-MODEL tool. From the templates found by this tool, two were selected. The selection was based on both coverage of the protein and percentage of identity. These two templates corresponded to fervidolysin (ID: 1r6v) and Pro-F17H/S324A (ID: 4jp8), a subtilisin homolog from *Thermococcus kodakarensis* (Yuzaki et al., 2013), with a sequence identity of 37.16 % and 36.05 %, respectively. The structures of both templates were resolved through X-ray method, with a resolution of 1.7 Å and 2.2 Å, respectively.

The two models proposed by this tool had a coverage of 88 % for fervidolysin and an 82 % for Pro-F17H/S324A. The Global Model Quality Estimation was 0.63 for the one using the fervidolysin as a template and a GMQE of 0.61 for Pro-F17H/S324A, with a QMEAN of -5.09 and -4.60, respectively. So, the fervidolysin model was the chosen model to build a 3D structure for the query protein. The model and the template are shown in Figure 34 and Figure 35, respectively. Also, in Figure 36 can be seen the estimated QMEAN for the model in terms of the Z-score, supplied by SWISS-MODEL. QMEAN consists of four individual terms, listed in the mentioned figure, which compares: the interaction potential between Cβ atoms only, the interaction potential between all atoms, the solvation potential and the torsion angle potential. Values of Z-score equal to zero (white area of the plot) indicate that the value is similar to the expected one from experimental structures with similar size. Negative values of Z-score (red area of the plot) indicate that the model scores lower than experimental structures on average, and positive values (blue area of the plot) indicate that the model scores higher than experimental structures on average. The QMEAN is shown on top.

.

Figure 34. Three-dimensional model of the protein predicted by SWISS-MODEL tool using fervidolysin as a template.

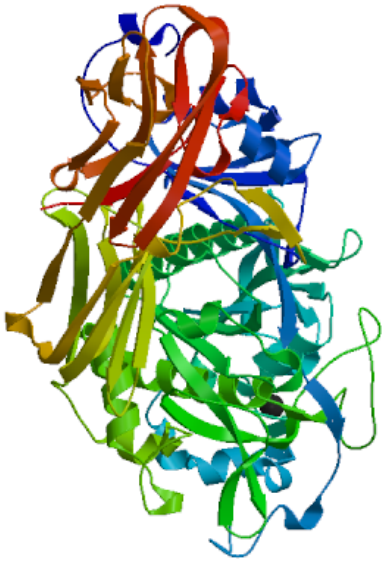


Figure 35. Three-dimensional structure of the fervidolysin, used as a template to predict the structure of the cloned keratinase.

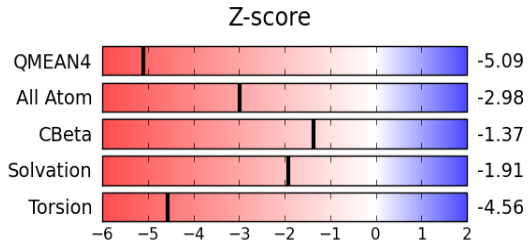


Figure 36. Estimated QMEAN for the model in terms of the Z-score, comparing: the interaction potential between Cβ atoms only, the interaction potential between all atoms, the solvation potential and the torsion angle potential.

In Figure 37 can be seen a plot with the Local Quality estimation provided by the tool. This plot shows, for each residue of the model (reported on the x-axis), the expected similarity to the native structure (y-axis). Finally, Figure 38 shows a comparison plot where it is shown the score of the model related to scores obtained for experimental structures of similar size. The x-axis shows protein length (number of residues). The y-axis is the normalized QMEAN score. Every dot represents one experimental protein structure, being the black dots structures within 1 standard deviation of the mean (Z-score between 0 and 1), and the grey dots those structures with a Z-score between 1 and 2, or above (within 1 and 2 standard deviation of the mean or above). The model is represented with a red star.

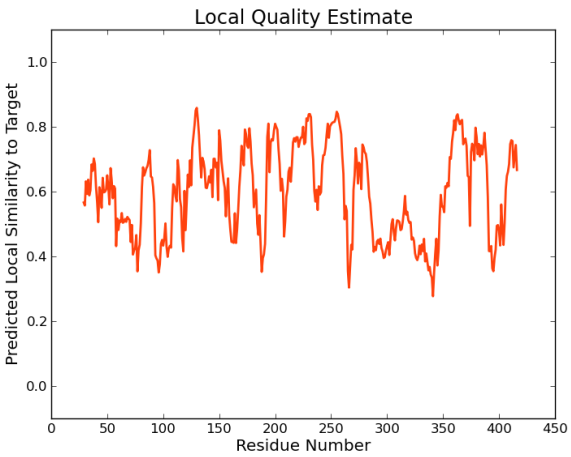


Figure 37. Local quality estimate of the model, showing, for each residue (x-axis), the expected similarity to the native structure (y-axis).

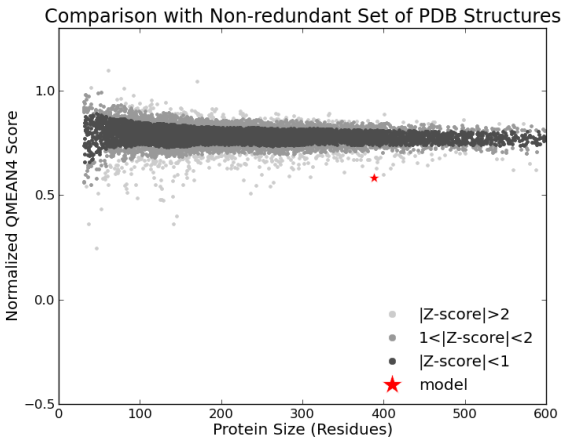


Figure 38. Comparison plot between size of proteins which structures have been experimentally resolved and their QMEAN score. Black dots represent proteins with a QMEAN within 1 standard deviation of the mean, and grey ones correspond to proteins with QMEAN between 1 and 2 or above 2 standard deviations of the mean. The model is represented as a red star

SWISS-MODEL also provided an alignment between the template used to resolve the model and the query sequence. In this alignment it can be seen that the catalytic triad of the enzyme is present in both the template and the model Figure 39 .

Figure 39. Sequence alignment of the putative keratinase of *F. pennivorans* strain T and the template used to resolve its three-dimensional model (fervidolysin).

9.2 Multiple sequence alignment

The conservation of the catalytic triad in the expressed keratinase (a peptidase S8) was assessed by comparing its sequence with all the homologues found after a PSI-BLAST query. 256 sequences where found after four iterations. The catalytic triad of the protein consists of three amino acids placed in specific points of the molecule (Figure 40).

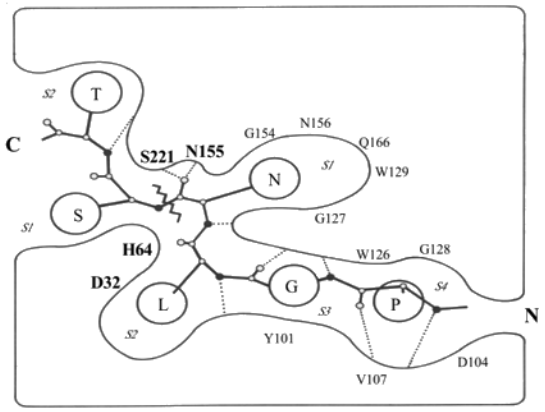


Figure 40. Schematic representation of the substrate-binding region of fervidolysin. The catalytic residues (D32, H64, and S221) and the oxyanion (N155) are marked in boldface. C and N denote the C-terminus and the N-terminus of the substrate, respectively (Kluskens et al., 2002)

These amino acids are: Aspartic acid, Histidine and Serine. Figure 41 shows the multiple sequence alignment made with JalView software, where it can be seen that the candidate protein had this mentioned triad: Asp_{154}, His_{190} and Ser_{377}. This catalytic center was also conserved in all the peptidase S8 sequences found after the PSI-BLAST query.

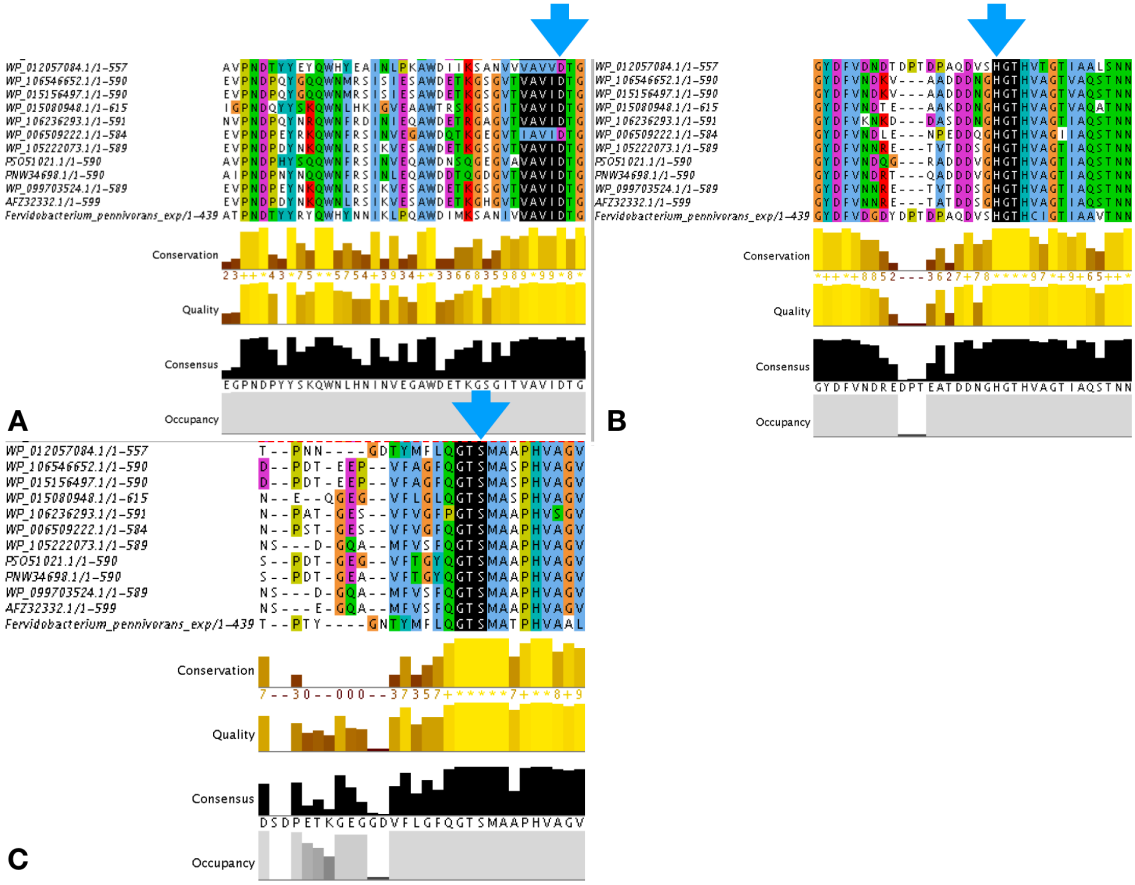


Figure 41. Multiple sequence alignment of 11 selected peptidase S8 found after a PSI-BLAST search. The catalytic triad of the peptidase S8 was found in all the sequences. In the expressed protein it was placed in Asp_{154}, His_{190} and Ser_{377} as can be seen in A, B and C, respectively.

9.3 Keratinase phylogeny

Another phylogenetic tree was built, based on the expressed keratinase, for the members of *Fervidobacterium* genus. *F. riparium* was excluded because its genome cannot be found in Genbank database. This tree is shown in Figure 42, and it can be seen that the expressed keratinase is placed next to the one from *F. pennivorans* type strain (DSM 9078) with a bootstrapping value of 100%, meaning that these strains were clustered together every time the tree was reconstructed. Likewise, *F. islandicum* and *F. changbaicum* were grouped in the same node with a bootstrapping value of 100%, too but these four species are clustered in a node which has been grouped together 69 times every 100 the tree was reconstructed.

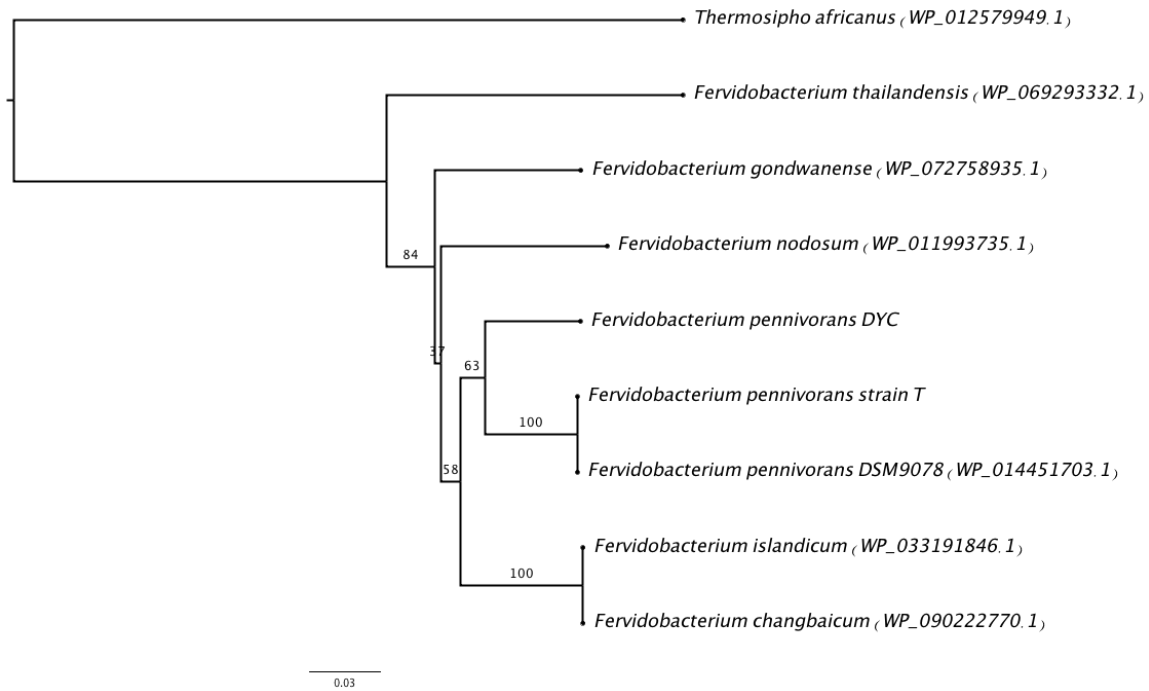


Figure 42. Bio Neighbour-Joining phylogenetic tree based on the expressed keratinase gene sequence with the members of the genus *Fervidobacterium*. The relationship between the experimental strain and the rest of species of the genus is shown. The homologue gene sequence of *Thermosipho africanus* was employed as an outgroup lineage. Bootstrap values as a percentage of 100 replications are presented. Bar, 0.03 changes per nucleotide position. Accession numbers, if available, are shown in brackets.

9.4 Signal peptide prediction

First, the signal peptide of the protein was predicted with the Signal IP server, using the algorithm optimized for gram negatives. The graphical output from SignalP shows the three different scores, C, S and Y, for each position in the sequence (Figure 43). C score shows signal peptide cleavage sites, S score is used to recognize positions within signal peptides, and the Y score predicts better the cleavage site than the raw C-score alone.

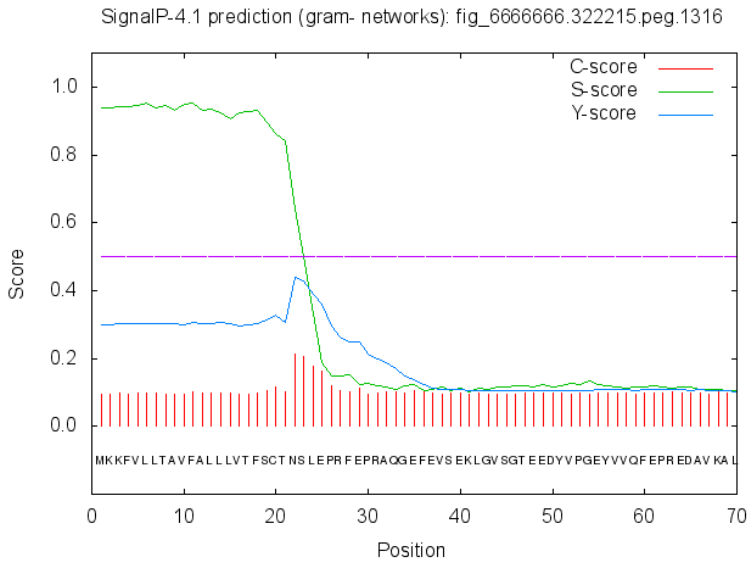


Figure 43. Graphical output from SignalP 4.1 server, showing C, S and Y scores for the first 70 amino acids of the query sequence. C score distinguish cleavage sites, S score distinguish positions with signal peptides and Y score predicts cleavage sites.

The summary supplied by SignalP 4.1 is shown in Table 18, and reports the maximal value for these three scores, and two extra values: the average S-score (Mean S) of the possible signal peptide (from position 1 to the position immediately before the maximal Y score) and the discrimination score (D-score), the value used to discriminate signal peptides from non-signal peptides, being a weighted average of the mean S and the maximal Y scores.

So, according to these calculations, a cleavage site was found between positions 21 and 22, suggesting a signal peptide from position 1 to 21.

Table 18. Summary output supplied by SignalP 4.1. Maximal C, Y and S values are shown. Also, Mean S (average S score from the positions 1 until the one immediately before the maximal Y score) and D score (weighted average of mean S and maximal Y scores).

Measure	Position	Value	Cutoff	Signal peptide
Maximal C	22	0.213		
Maximal Y	22	0.441		
Maximal S	11	0.953		
Mean S	1-21	0.926		
D Score	1-21	0.669	0.570	YES

9.5 PCR and product cloning

The designed primers were used to conduct a (PCR) targeting the gene encoding the protein we wanted to express, without the signal peptide. To check the success of the PCR and to assess the quality and size of the PCR product, an 1.5 % agarose gel electrophoresis was run (Figure 44). It can be seen that the reaction was positive only for the sample DNA, as no bands were found for the negative control. Furthermore, these positive bands belonged to a DNA fragment with a length consistent with the expected one: The length of the gene was estimated by RAST to be 1320 base pairs and 1257 after removing the signal peptide.

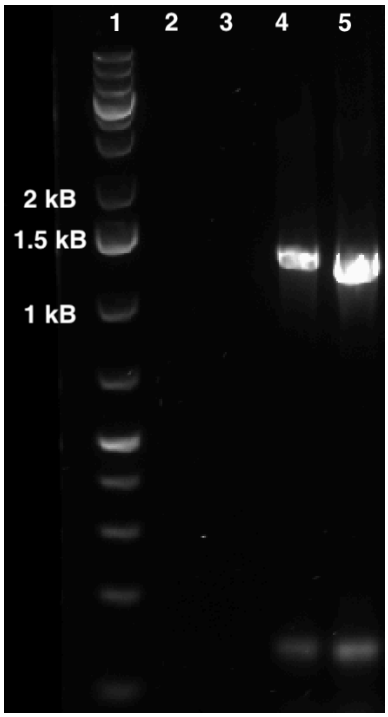


Figure 44. 1.5% agarose gel electrophoresis showing: 1Kb Ladder (well 1), negative control (PCR master mix) (wells 2 and 3), and PCR product pointing to the gene of interest from *F. pennivorans* strain T (wells 4 and 5).

The vector used for cloning was the plasmid pINIT_tet with a marker gene for resistance to tetracycline. After growing the host cells with vector plasmid, plasmids were isolated and a gel electrophoresis was run to check their quality and size. Figure 45 shows a picture of the mentioned gel: in the first well a Supercoiled plasmid marker was loaded, and the isolated plasmids were loaded in wells 2 and 3. Clear bands could be seen in wells 2 and 3, and their size were congruent with pINIT_tet length, which is 5599 base pairs.

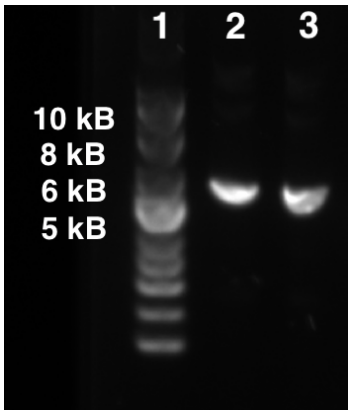


Figure 45. 0.8 % agarose gel electrophoresis showing: supercoiled DNA ladder (well 1) and pINIT_tet plasmid isolated from Escherichia coli (wells 2 and 3).

After ligation and transformation, a new plasmid DNA extraction was conducted to isolate the obtained recombinant constructions (pINIT_tet plasmids with the gene of interest inserted) from six colonies grown in LB agar supplemented with tetracycline (10 µg/mL). These constructions were used to run a 0.8 % agarose gel electrophoresis for evaluation (Figure 46). The empty pINIT_tet plasmid was also loaded as control in this gel. Bands 2 and 3 correspond to the empty pINIT plasmid, and 4 and 5 belong to 2 candidate colonies. It can be seen that both of them carried the plasmid, being the DNA loaded in well 5 larger.

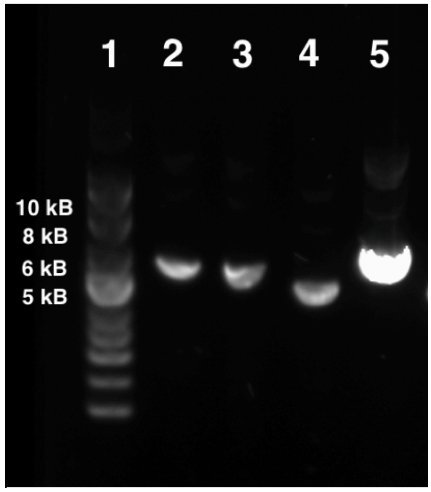


Figure 46. 0.8 % agarose gel electrophoresis showing: super coiled DNA ladder (well 1), empty pINIT_tet plasmid (wells 2 and 3) and two colonies candidates to carry pINIT plasmid with the insert (wells 4 and 5).

To verify the presence of the insert in the mentioned candidates, insert of the isolated plasmids were sequenced. The DNA fragments were then aligned and merged, and then

aligned with the peptidase gene using MEGA 7 software. It was found that one of the candidates (the one loaded in well 5 in the previous electrophoresis) matched 100 % with the expected sequence.

9.6 Cloning in expression vectors

Once the constructions were confirmed, the experiment followed as described, with the sub-cloning procedure of the keratinase gene in the expression vectors (p7xNH3 and p7xC3H) and the subsequent PCR to check the presence of the insert-containing plasmids in the cells after overnight incubation in LB medium with kanamycin. Three colonies per series were picked and analysed. Figure 47 shows a gel electrophoresis run after plasmid extraction to check its size and integrity. Two replicates of each plasmid were loaded, being found that both size and integrity were as expected. Lanes 2 and 3 correspond to the p7xC3H plasmid (6999 base pairs), and lanes 4 and 5 to the p7xNH3 plasmid (6976 base pairs). Figure 48 shows the 1.5 % gel electrophoresis run after the mentioned PCR. Lanes 2 to 7 show the bands belonging to the colonies expected to carry p7xC3H plasmids, and lanes 8 to 13 correspond, likewise, to the colonies candidate to carry p7xNH3. The expected size of the amplified fragment employing the described primers was approximately 1600 base pairs for both reactions. All colonies were verified as positive, except for one p7xNH3 candidate. The positive colonies showed good quality bands, consistent with the expected size.

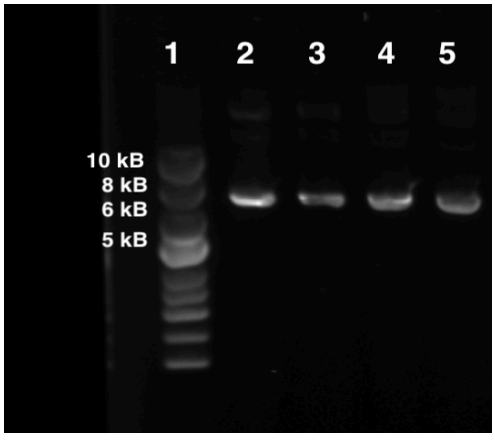


Figure 47. 0.8 % agarose gel electrophoresis showing: super coiled DNA ladder (lane 1), p7xC3H plasmid (lanes 2 and 3) and p7xNH3 plasmid (lanes 4 and 5).

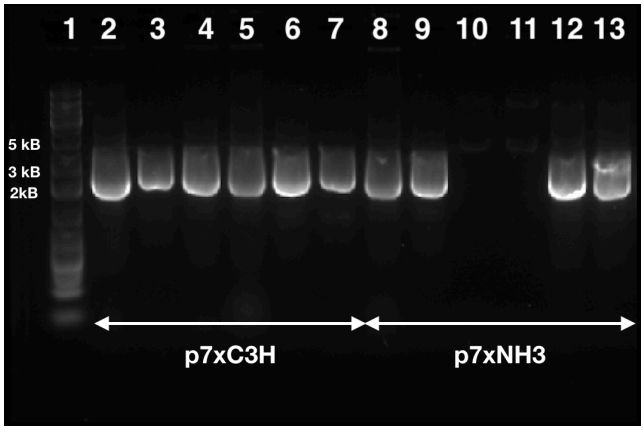


Figure 48. 1.5 % gel electrophoresis showing: 1 kB DNA ladder (lane 1), PCR products belonging to candidates to carry p7xC3H plasmid (lanes 2 to 7) and p7xNH3 (lanes 8 to 13).

9.7 Protein expression

The expression of the protein was tested by running a sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis, using two different fractions: the cell lysates on one side and the sonicated cells on another. In addition, cells without induction were used as negative control. Figure 49 shows the result of the mentioned gel. The first lane corresponds to the size marker, and lanes from 2 to 5 correspond to the cleared lysate samples, after heating and centrifugation. Lane 2 belong to the induced cells carrying p7xNH3 construction, and lane 4 belong to the induced cells carrying p7xC3H construction. Lanes 3 and 5 correspond to the uninduced cells. It can be seen a band in lanes number 2 and number 4, being this band congruent with the expected protein size, which is 45.76 kDa after removing the signal peptide, according to the molecular weight calculator available in the Sequence Manipulation Suite (Stothard, 2000). This band could not be seen in the uninduced cells. The presence of the bands in this fraction suggests that the expressed protein was soluble.

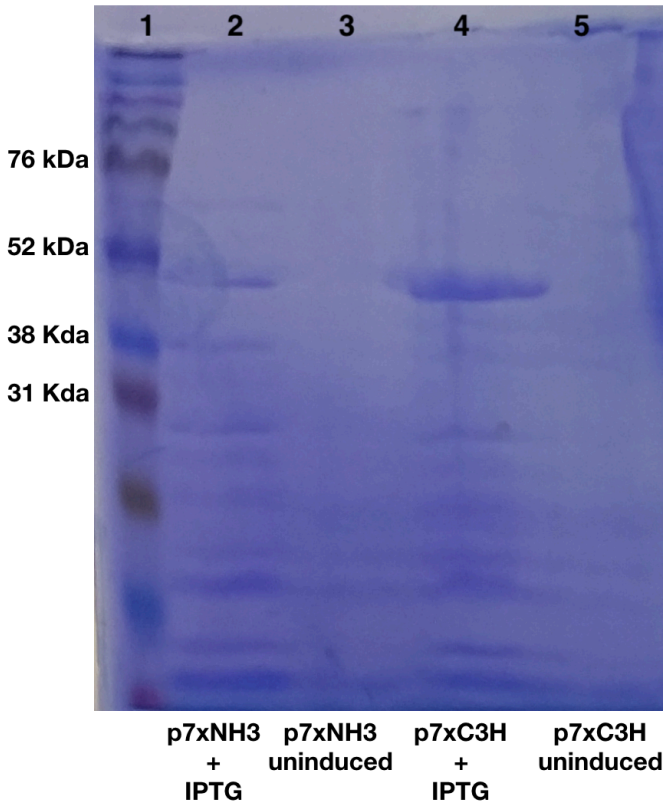


Figure 49. 30 % sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), showing: Rainbow size marker (lane 1), cell lysates (lanes 2, 3, 4 and 5), corresponding lanes 2 and 4 to the cells induced with Isopropyl β-D-1-thiogalactopyranoside (IPTG), carrying p7xNH3 and p7xC3H constructions, respectively.

9.8 Protease activity assay

The activity of the expressed protease was assessed by an activity assay based on a fluorometric measurement. After the master mix and sample preparations, the array was incubated for 1h at 37 °C, as described in the Materials and Methods section. After incubation, the fluorescence was read by a plate reader software (Hidex Sense, v0.5.11.1), and the results were plotted. This plot can be seen in Figure 50, where can be seen the Fluorescence Units (RFU) output by the software for each sample after 1 hour incubation at 37 °C.

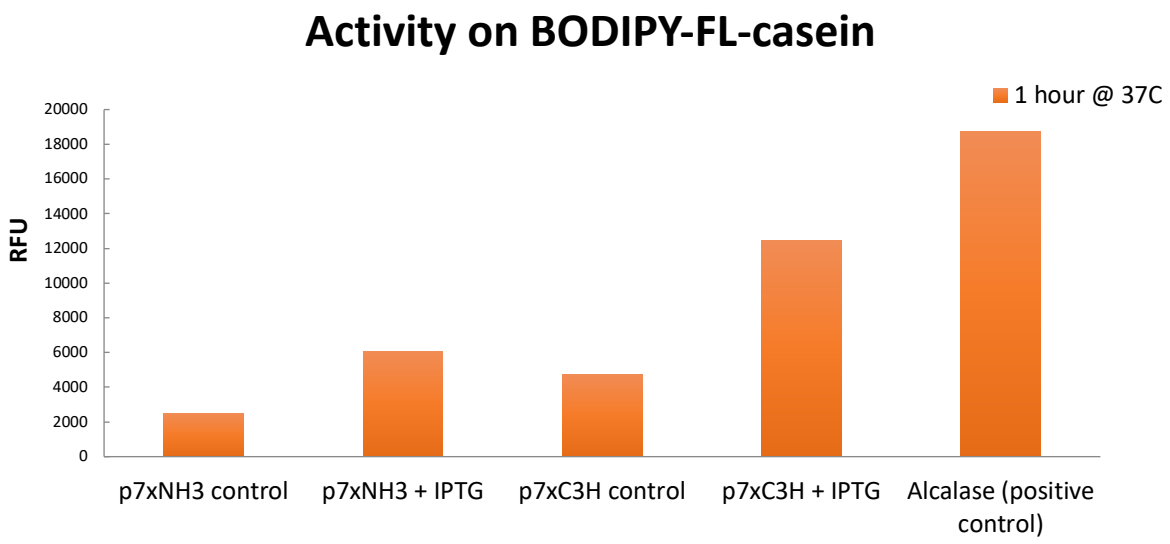


Figure 50. Activity of the different lysates tested in the protease activity assay, measured in RFU (Fluorescence Units). The chart shows the activity of the lysates without IPTG induction (negative controls), represented as NH3 control and C3H control, as well as the activity of the induced cells' lysates, represented as NH3 and C3H. Alcalase bars correspond to the positive control.

This assay was considered as a preliminary test to check the activity of the expressed keratinase. It can be seen that there the induced samples showed higher activity than the uninduced ones, suggesting that the protein might be active. However, as the enzyme is supposed to be thermos-stable, this assay was also conducted at 60 °C, but the obtained results were not conclusive.

Discussion

10. Discussion of the Material and Methods

10.1 Media preparation and enrichment

Although the water sample was stored at room temperature for about one year, it could still be used for enrichments when incubated at 65 °C in an appropriate medium. The medium recipe used to grow this strain was reported to work successfully for different Thermotogae species in the past. Its composition is similar to other recommended media for this species (DSMZ, 1969, Atlas, 2004). However, it had some differences, as the presence of a small amount of NaCl (3 g/L), or different concentration of other salts: for instance, in MMF the concentration of NH_4Cl is 0.25 g/L, but 0.5 g/L in the "*Fervidobacterium* medium" suggested by DSMZ, and 0.9 g/L in the one provided by Atlas. Still, the bacterium grew well when incubated in MMF supplemented with the proper carbon sources, so this was the medium used for the rest of experiments of this work.

Although the technique used to prepare the culture media was described in 1950 it is still used after some modifications and updates. The principle and the goal are the same than the ones described in the classic method: to take advantage of the low solubility of oxygen at high temperatures and flush the medium with the appropriate sterile gas (N_2, CO_2, etc.) while cooling the preparation in the presence of a reducing agent and a redox indicator (Hungate, 1950). Although this procedure requires some training and supervision the first times it is performed, and despite it also needs a relatively complex infrastructure, this is a highly effective method, that almost guarantees the absence of oxygen in the medium and the preservation of a reducing environment, perfect for anaerobic microorganisms.

10.2 Strain isolation

The dilution to extinction method was conducted to isolate the bacterium and obtain a pure culture. This method benefits the most abundant microbes present in a potentially heterogenous culture and it has been widely used, especially for oligotrophic, slow-growing microbes or, in general, those which are not easily culturable in the lab. The underlaying principle of this procedure is to progressively reduce the diversity of the culture by repeatedly

diluting the samples, ideally down to single cells, before their culture in isolation (R. et al., 2010, Button et al., 1993). After two rounds of dilution to extinction, the culture was assumed to be pure and the strain, thus, isolated. Furthermore, the incubation temperature was an extra selection factor, as only thermo or hyper-thermophiles can grow at 65 °C. These assumptions are congruent with the results of the further analysis conducted, such as the strain identification experiment or the genomic sequencing, as no contaminations could be spotted, neither as interferences in the sequence chromatograms after sequencing, nor under the microscope or in the form of genetic fragments from other microbes. However, this method is not appropriate for isolating the less represented species in a sample, as after several dilution rounds only the most abundant ones will remain/persist. This is an important drawback if it is needed to isolate a poorly represented species in the sample.

10.3 Carbon utilization

The Fervidobacteriaceae need yeast supplement to grow (Friedrich and Antranikian, 1996, Kanoksilapatham et al., 2016, Podosokorskaya et al., 2011, Cai et al., 2007, Andrews and Patel, 1996, Huber et al., Patel et al., 1985), and the recommended concentration for *F. pennivorans* is 2 g/L (DSMZ, 1969) or 3 g/L (Atlas, 2004), plus another organic carbon source. However, during these tests, growth was spotted in the negative control flasks unless the yeast extract concentration was 0.5 g/L or lower. So, all the negative control flasks contained only MMF basal medium and a yeast extract concentration of 0.5 g/L. Furthermore, an extra step was performed: 1 mL of the master medium was transferred to the flasks with MMF medium and the tested carbon source and, after 48 h, 1 mL from these flasks were transferred to new ones with, again, MMF medium and the respective carbon source to test. This extra step helped to avoid false positives, as the master flask contained a richer medium that might have stimulated the inoculum to grow in the new environment. The substrates were chosen based on its potential biotechnological interest and/or due to inconclusive results in the literature. Growth was verified by checking the presence of bacteria under the phase-contrast microscope, which is not the optimum way to assess the viability of the cells due to the risk of having false positives, as some of the bacteria spotted might have been optical artifacts or being dead. Nevertheless, direct verification under the phase-contrast microscope was considered a sufficient approximation to qualitatively assess the utilization of the nutrients. Furthermore, the microscope analysis allowed to check for changes in the morphology of the

cells, or spot any other visible changes in them: clustering, opacity changes, different growing patterns, etc.

10.4 Temperature tolerance

It has been reported that the type strain of *F. pennivorans* can grow between 50 °C and 80 °C (Friedrich and Antranikian, 1996). In this test, three different temperatures have been tested to try to assess the minimum and maximum temperatures this organism can thrive: 38 °C , as thermophiles cannot divide if temperature is lower than 40 °C, and 80 °C, as the limit between thermophiles and hyper-thermophiles (Madigan et al., 2014). Two extra temperatures were included: 65 °C, as the recommended incubation temperature for this organism, and 55 °C. The chosen medium for this test was MMF supplemented with peptone and a yeast extract concentration of 0.1 % (1 g/L) instead of 0.05 % (0.5 g/L) used in the carbon utilization test, to ensure the nutrients concentration were not a growth limiting factor. Again, growth was assessed by analyzing each sample under the phase-contrast microscope despite the objections exposed before, as the same benefits and reasons described also apply for this test.

10.5 Osmotic stress tolerance

The range of NaCl concentration *F. pennivorans* can tolerate has been established in 0-40 g/L, 4 g/L being the optimum concentration (Kanoksilapatham et al., 2016, Podosokorskaya et al., 2011, Cai et al., 2007). Thus, a battery of flasks, in duplicates, with NaCl concentrations ranging from 10 g/L until 50 g/L was set up. The basal MMF medium contains 3 g/L, close to the reported optimum, so this was the NaCl concentration chosen to grow the bacteria for biomass collecting when needed. This battery of flasks was directly inoculated from the master culture, without intermediate step, as the NaCl concentration in this culture was lower than the lowest one disposed in this test. Growth was determined after observation by phase-contrast microscope. In this case, some weird and unusual cellular shapes were spotted at higher NaCl concentrations, so checking the cultures under the phase-contrast microscope was especially useful for this test.

10.6 pH tolerance

The tolerance to pH variations for this *F. pennivorans* ranges from 5.5 to 8.0 (Kanoksilapatham et al., 2016, Podosokorskaya et al., 2011, Cai et al., 2007), and growth has been reported even

at pH 5.0, but in this case with doubling times of 24 h (Friedrich and Antranikian, 1996). So, for this test a battery of flasks with pH 5.5 and 8.0 were set up, in duplicates. The chosen method to assess the viability of the cells was again the analysis of each culture under the phase-contrast microscope, with the pitfalls and benefits previously reported. Again, the medium used in this test was MMF with 0.5 % peptone and 0.1 % yeast extract, to ensure the nutrients did not imply a growth limitation.

10.7 Keratinase activity

F. pennivorans is one of the Fervidobacterioaceae species able to grow on feathers, thus showing keratinolytic activity (Lee et al., 2015b, Friedrich and Antranikian, 1996). This ability was tested by inoculating 1 mL of a dense *F. pennivorans* culture into a flask with fresh MMF supplemented with 0.5 % peptone and 0.05 % of yeast extract. The feathers were sterilized by tindalization for 60 minutes at 100 °C (Friedrich and Antranikian, 1996). Beforehand, the feathers were cleaned by immersion in a methanol:ethanol solution (1:1) to wash out the dirt from the feathers, as well as to eliminate the lipases the feathers might have had, since they may have interfered with this experiment. Despite it has been reported that after an incubation of 48 hours this strain is able to completely degrade chicken feathers (Friedrich and Antranikian, 1996), in this experiment more incubation time was needed to obtain the same results; so, the cultures have been under a continuous monitoring, inspecting them every 24 hours, both by direct checking the feather inside the flask, and by analyzing the cells under the phase-contrast microscope until the feathers were degraded.

10.8 Discussion of the genomic methods

10.8.1 DNA isolation and strain identification

The genomic DNA was isolated using a *NA2100 SIGMA* GenElute™ Bacterial Genomic DNA extraction kit, by Sigma-Aldrich, since it was the fastest and most reliable method available. The DNA was quantified with a NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer. The NanoDrop is not the most reliable method to quantify DNA, but the quality of the DNA extraction can also be assessed, and it was considered that the quality check was more important than the quantification accuracy in this step. The DNA was run in

a 0.8 % gel electrophoresis, for 45 min at 5 V/cm, as ThermoFischer recommended this configuration for its DNA GeleRuler best performance.

The strain identification was based on PCR amplification of the 16S rRNA gene and following comparison of its sequence with BLAST searches (Woese et al., 1983). The PCR protocol (Peake, 1989) and program were designed following the indications of the polymerase supplier, to optimize its performance. It was decided to use genomic DNA from *Thermosipho africanus* as a positive control, since this bacterium was identified following this same procedure, so its behavior was known. The reaction Master Mix was used as a negative control, as it contained exactly the same components than the experimental PCR except for the DNA template. Furthermore, all reactions were run in duplicates, to minimize the possibility of errors.

For the sequencing reaction, four primers were chosen: two of them for the forward direction and the other two for the reverse. In addition, for each direction, the primers were chosen considering the gene sequence: so, two of them primed the elongation from the first positions of the gene (16S forward and 16S reverse), and the other two started feeding in middle positions (575 Forward and K517 Reverse).

So, four different nucleotide sequences were obtained, one per each primer. These were aligned, merged and a consensus sequence was generated. This consensus sequence corresponded to the 16S gene of the studied strain, and was used to perform a BLAST comparison to find its closest relatives. The studied strain was then identified as *Fervidobacterium pennivorans*, and, according to the score provided by the alignment calculated by blastn algorithm, closer to *F. pennivorans* DYC strain than to the species type strain, *F. pennivorans* DSM 9078. So, these two strains were the chosen ones to perform the following genetic and genomic comparisons.

10.9 Discussion of the Genomic and Bioinformatic analysis methods

10.9.1 Phylogenetic trees construction

The first step to build a phylogenetic tree is to perform a MSA with all the sequences of interest. In this case, two trees have been built: one based on the 16S genes of all the members of the *Fervidobacterium* genus, and another one based on the amino acidic sequence of the protein cloned and expressed in *Escherichia coli*, also taking in account all the species of the *Fervidobacterium* genus which had a homologue of the mentioned protein. These sequences were downloaded in FASTA format from the Genbank database. In addition, an extra species, external to the group, was included to root the tree. As all the sequences had the same length, the algorithm chosen to conduct the mentioned alignment was Clustal Omega, as it is fast, reliable and suitable to perform global alignments of sequences of similar length (Sievers et al., 2011). The tree was resolved using an improved Neighbor-Joining clustering algorithm, called Bio Neighbor-Joining, a model well adapted when the calculations are obtained from aligned sequences (Gascuel, 1997). The nucleotide distance was measured by the Kimura 80 method, which assumes different mutation rates for transversions and transitions, since the latter occur more frequently (Kimura, 1983). The tree quality was then tested by non-parametric bootstrapping. With this method, a multiple sequence alignment of the same length is generated a certain number of times with random duplicate of some of the sites, being the tree repeatedly sampled through this slightly perturbed dataset (Efron and Tibshirani, 1994).

10.10 Genomic analyses

10.10.1 Genome assembly and annotation

The experimentally obtained genome of *F. pennivorans* strain T was sequenced using a commercial Next Generation DNA sequence provider based on Illumina technology. DNA of good quality and in sufficient amount (at least 1 µg) was required. So, after the genomic DNA isolation these parameters were checked by analyzing and measuring the extracted DNA with NanoDrop and an agarose gel electrophoresis, as detailed before. Again, as the quality of the DNA was more important than the accuracy of the quantification, the NanoDrop was used. The genome was assembled with CLC Genomic Workbench 11.0. The optimal k-mer length for the genome assembly was determined using KmerGenie software. KmerGenie computes the k-mer abundance histogram for many values of k. Then, for each value of k, it predicts the

number of distinct genomic k-mers in the dataset, and returns the k-mer length which maximizes this number, estimating the best k-mer length for genome de novo assembly (Chikhi and Medvedev, 2014). As explained, the genome was then uploaded for annotation to the RAST website, which turned out to be fast, having the genome completely annotated after only a few hours. This annotation system, in combination with the SEED website, is useful and convenient, since it allows to easily find genes, clusters of operons, perform internal queries based on BLAST, or compare the query organism with external sequences or genes. It also permits comparing the annotated genome with other ones, either available in their database or uploaded and annotated by the researcher; these comparisons include function or sequence based genomic comparisons, or even dot-plot matrices representations. The main inconvenience of this tool, and others of its kind, is that it is very difficult or impossible to identify new gene functions if they are not annotated in the database. In addition, sometimes these annotations are not reliable and can be misleading.

10.10.2 Average Nucleotide Identity

Although a consensus has still not been reached, it is assumed that two bacterial strains belong to the same species if the similarity in their 16S rRNA gene sequence is higher than 97 % and their genomic DNA hybridization percentage is higher than 70 % (STACKEBRANDT and GOEBEL, 1994). Nevertheless, as the experimental DNA-DNA hybridizations has been criticized for being difficult to implement, bioinformatic estimations have become more important in the last years, the ANI calculation being one of them, as there are studies suggesting that it is an exceptionally robust and sensitive method for measuring evolutionary relatedness among closely related bacterial strains (Konstantinidis et al., 2006). The ANI test calculates the average percentage of DNA hybridization between two genomes to infer if they belong to the same species, based on a mathematical model that predicts the DNA hybridization between them. The output provided by the tool includes the Average Nucleotide Identity itself and two charts, representing the frequency of each identity percentage of the comparison, and the bit-score distribution of the compared genomes, respectively.

10.10.3 Genome-to-genome distance analysis

Similarly to ANI, this test estimates the distance between two genomes to infer if they belong to the same species or subspecies, based on a mathematical model that predicts the

percentage of DNA hybridization between them. The designers of this algorithm claim that the estimations by GGDC yield a better correlation with wet-lab DDH determinations than other approaches. They also state that these calculations can cope with repetitive sequence regions and that they have been proved very robust when estimating distances with incomplete or draft genomes (Meier-Kolthoff et al., 2013). The supplied output includes the probability that DNA-DNA Hybridization is higher than 70 %, meaning that the input genomes belong to the same species, and the probability that DNA-DNA Hybridization is higher than 79 %, meaning that the input genomes belong to the same sub-species.

10.10.4 Dot-plot matrix

A dot-plot matrix is a graphical representation of the global alignment two sequences. The comparison is done by scanning each residue of one sequence comparing with all residues in the other sequence. If a residue match is found, a dot is placed within the graph. When the two sequences have substantial regions of similarity, the sequence alignment is revealed by the formation of a diagonal line. The interruptions in the middle of the line indicate insertions or deletions. Parallel diagonal lines within the matrix represent repetitive regions of the sequences, and perpendicular lines represent fragment inversions (Xiong, 2006). Dot plot matrices allow to easily visualize in a graphic way the global alignment between two large sequences. It is easy to spot deletions, insertions or even inversions. The dot-plot matrix shown in this work was built using Gepard software, which builds the matrix based on heuristic calculations (Krumsiek et al., 2007).

10.10.5 Genomic alignment

The genomic alignments presented in this work were conducted using Mauve software. Mauve constructs genome alignments in the presence of large-scale evolutionary events such as rearrangement and inversion. These alignments allow the study of genome-wide evolutionary dynamics and provide a basis for comparative genomics. Genome alignments can also identify evolutionary changes in the DNA by aligning homologous regions of sequence. In addition, Mauve also identifies orthologous regions which may have been reordered or inverted. It also allows ordering contigs of a draft genome to compare it with a template genome (Darling et al., 2004).

10.10.6 Conserved signature indels

Conserved signature indels (CSI) can be used as molecular markers to easily identify a species or group of organisms (Gupta and Bhandari, 2011, Gupta, 1998) . The strategy followed to find CSI candidates was the analysis of the GC composition of the sequenced genome, and the following search of genomic areas with a GC skew in the genome (Che et al., 2014). To try to find CSI candidates in the different species of the genus *Fervidobacterium*, all the genomes were aligned and represented using BRIG, a tool based on BLAST capable of producing high resolution graphs with the aligned genomes and extra statistics, such as GC content, GC skew, etc. So, those fragments with a clear GC skew present in the experimental genome as well as in other of the members of the genus were traced. Then, these sequences were pasted into BLAST website to find homologues or similar sequences. Some of these fragments were annotated, but others were not. For the latter, the sequences were translated into their hypothetical products, and they were likewise compared to the BLAST database.

10.10.7 3D protein structure prediction

The 3D structure of the expressed protein was predicted based on the information contained in its amino acidic sequence, looking for homologues proteins with structure already resolved by experimental methods. When the similarity percentage is sufficient enough with a protein which structure has been experimentally described, then a reliable model for the query protein can be estimated. SWISS-MODEL was the chosen structure modelling server, based on homology and accessible via the ExPASy web server (Bienert et al., 2017). Thus, the first step was finding an experimentally resolved template with at least a 30 % identity with the peptidase of *F. pennivorans* (Xiong, 2006). The server found a total of 235 candidate templates combining different search strategies (BLAST and HHBlits) against the SWISS-MODEL library. From all of these, the most suitable 50 templates were selected and shown for manual inspection. The template selected to build the model was 1r6v.1.A, the fervidolysin of *F. pennivorans* DSM 9078, since both the identity level with the query protein and the resolution of the experimental structure were the highest ones among all the templates, with 37.16 % and 1.7 Å, respectively. In addition, a second template was selected: 4jp8.1, a peptidase from *Thermococcus kodakarensis*, due to the high sequence coverage with the query protein.

10.10.8 Multiple sequence alignment

A Multiple sequence alignment (MSA) of the expressed peptidase and all the homologues found was prepared, to assess the conservation and presence of the catalytic triad in the mentioned expressed protein. PSI-BLAST first performs a regular blastp search with the query protein, but then builds a position-specific scoring matrix (PSSM) or profile. Then, this PSSM is used in subsequent iterations to further search the database for new matches, being updated every iteration with the newly detected sequences. PSI-BLAST is a very sensitive and useful search strategy to detect distant related proteins, but still with biologically significant similarities with the query sequence. It has been estimated that the profile-based approach is able to identify three times more homologs than regular BLAST. (Altschul et al., 1997, Xiong, 2006). Four iterations were run, until no new sequences were found and added to the PSSM, then all the sequences were downloaded and saved in a FASTA file. The MSA was performed using ClustalO algorithm under JalView software.

10.11 Cloning of a keratinase gene

The method used to clone the gene of interest into *Escherichia coli* is called FX cloning. Its name is derived from “fragment exchange” cloning and is based on a class IIS restriction enzyme and negative selection markers. It allows sub-cloning either a sequence-verified open reading frame (ORF) or even PCR products to a variety of expression vectors, leaving only a single amino acid to either side of the protein (Geertsma and Dutzler, 2011).

10.11.1 Signal peptide prediction

The signal peptide of the protein was predicted with the Signal IP server. This tool uses neural networks which produce three output scores for each position of the query sequence to predict a possible signal peptide:

- C-score (raw cleavage site score): output from the CS networks, trained to distinguish signal peptide cleavage sites from everything else.
- S-score (signal peptide score): output from the SP networks, trained to distinguish positions within signal peptides from positions in the mature part of the proteins and from proteins without signal peptides.

- Y-score (combined cleavage site score): the geometric average of the C-score and the slope of the S-score, resulting in a better cleavage site prediction than the raw C-score alone.

The SignalP 4.1 tool provides also a summary, reporting the maximal value for the mentioned three scores, and two extra ones:

- Mean S: average S-score of the possible signal peptide (from position 1 to the position immediately before the maximal Y score).
- D score (discrimination score): a weighted average of the mean S and the maximal Y scores. This is the score used to discriminate signal peptides from non-signal peptides.

10.11.2 PCR and product cloning into sequencing vector

In this PCR Phusion Polymerase was used, instead of the most common OneTaq Polymerase, since high fidelity was needed to clone the gene of interest, and this enzyme's reliability and performance are higher. Thus, the PCR protocol (Peake, 1989) and program were designed according to the specifications of this polymerase, to optimize its performance. The reaction master mix was used as a negative control.

The cloning vector was a pINIT plasmid with a tetracycline resistance gene marker. In addition, it has a cassette with the ccdB gene, which encodes for a bacterial toxin that poisons the DNA gyrase, killing the cells. When cloning, this cassette is exchanged by the insert. So, only those cells carrying the plasmid with the gene inserted will grow when inoculated or spread on a plate with tetracycline: cells without the plasmid will be sensitive to tetracycline and will not survive, and those carrying the plasmid without the insert will be killed by the toxin produced by the ccdB gene.

The colonies able to grow in these conditions were candidates to carry the gene of interest inside the pINIT_tet plasmid. So, six were picked and the plasmids sequenced to verify the presence of the insert while ruling out possible mutations.

10.11.3 Sub-cloning using FX Cloning

The vectors chosen for the protein expression were the plasmids: p7xNH3 and p7xC3H. Both of them add a tag of 10 histidines to the translated protein, but the main difference between them is that p7xNH adds these histidines to the N-terminus of the protein while p7xC3H adds them to the end C-terminus of the protein. These histidine tags ease the protein purification since it has a high affinity to metal-ions, used in purification procedures. Just like pINIT_tet, these plasmids also carry the ccdB gene and a resistance gene marker, in this case, to kanamycin. So, after the transformation procedure, only those cells carrying both the plasmid and the insert will survive when spread on medium with kanamycin. Colonies able to grow in these conditions were picked and the plasmid presence checked by performing a PCR with specific primers.

10.11.4 Protein expression and activity assay

The expression *E. coli* strain chosen was a derivative of BL21 named LOBSTR (Low Background Strain). This strain has been engineered to eliminate the contaminants appearing in form of histidine-rich *E. coli* proteins which interfere in histidine-tag affinity purification (Andersen et al., 2013). To express the protein, the cells were incubated at 37 °C in 2-YT medium supplemented with kanamycin until the optical density of the culture was approximately 0.4, marking this absorbance the end of the early/mid lag phase (Madigan et al., 2014). The expression vectors also carry the lacI gene, which codes for the *lac* repressor, a transcription inhibitor. Then, to induce the expression of the protein, lactose or an analog had to be added to the medium. So, in the beginning of the exponential phase, IPTG was added to suppress the action of the transcription inhibitor and initiate the expression of the protein. After the incubation time, the cells were collected, pelleted and lysed by sonication. To assess if the expressed protein was soluble, the samples were centrifuged to pellet all the insoluble particles. As it was expected that the protein was in the soluble phase, before loading the SDS-PAGE this sub-sample was heat-treated to denature the proteins belonging to *E. coli* present in the sample. The activity of the expressed protein was determined in a fluorometric test based on a direct measurement able to detect the activity of several proteases. An enzyme of known activity was used as a positive control, while the reaction master mix was the negative control for this test.

11. Discussion of the results

11.1 Carbon utilization

Among all the different carbon sources tested, peptone gave the best results, with fast growth and high biomass amount. Peptone was then used during this work when fast growth was required, or when collecting sufficient amount of biomass was needed, for instance, to extract genomic DNA. Also, peptone was used in tests in which the nutrients where not the growth limitation factor, such as temperature, osmotic stress or pH tests. It was used as well as a positive control when needed. Galactose and glucose yielded efficient growth as well, following dextrin. All of them are recommended or suitable to use when growing this strain, since after 48 hours of incubation at 65 °C clear growth can be achieved. Still, although the biomass collected was enough for all the conducted analysis in this work, no high turbidity cultures were obtained in any case, being recommended to test different media in the future if higher cell density is required. It has been reported that some members of the Thermotogales group can reduce thiosulfate to sulfide, and that this process enhances cellular yields and growth rates, so it is recommended to add thiosulfate to the medium if higher yield is needed (Ravot et al., 1995). The rest of the tested nutrients are not recommended to grow this strain, either giving no clear or sufficient growth or not growth at all. Nevertheless, the method chosen to assess cell growth has the limitations previously discussed, since exist the risk of having false positives or negatives. False positives by mistaking an optical artifact or a dead cell for an alive cell, and false negatives by not spotting any bacterium in the slide checked because the culture density was too low. Alternative methods to check the growth are recommended if more accurate conclusions are needed, such as O.D. measurement, flow cytometry, etc. Nevertheless, when comparing with the available literature, the results obtained after this test are consistent con those published, being this experimental strain able to utilize the same substrates as the species type strain (Kanoksilapatham et al., 2016, Podosokorskaya et al., 2011, Cai et al., 2007, Friedrich and Antranikian, 1996).

11.2 pH tolerance

It has been reported that this species is able to grow in pH's ranging from 5.5 to 8.0 is (Kanoksilapatham et al., 2016, Podosokorskaya et al., 2011, Cai et al., 2007, Friedrich and Antranikian, 1996), and even at pH = 5.0 (Friedrich and Antranikian, 1996). However, no

growth was reported in any of the flasks set up in this test, neither at pH = 5.5, nor at pH = 8.0. Again, the same objections against the growth assessment method applies, and a different growth check system is required to extract more accurate conclusions: due to the long time this species need to divide when growing in these stressing conditions, low density cultures might have been mistaken for negative. Still, although this strain was able to grow in those conditions, the yield obtained would be so low that they are not recommended.

11.3 Osmotic stress

The optimal amount of NaCl reported for this species' type strain to grow is 0.4 % although it can tolerate until a NaCl concentration of 4 %, but with long division times (Kanoksilapatham et al., 2016, Podosokorskaya et al., 2011, Cai et al., 2007, Friedrich and Antranikian, 1996) . The results obtained are consistent with the literature: the culture medium used to optimally grow this bacterium had 0.3 % of NaCl, the reported optimal concentration. Furthermore, according to the results from this work, fewer cells were spotted under the phase-contrast microscope when increasing the NaCl concentration. In flasks with the highest amounts of salt (4 % and 5 %) some cells could still be spotted, but their viability was compromised as their morphology and structure appeared altered, with less dense and opaque cytoplasm and less defined cellular membrane.

11.4 Temperature tolerance

According to the literature, this species type strain can grow from 50 °C until 80 °C, but needing 12 hours to divide when incubated at 50 °C and 80 °C (Friedrich and Antranikian, 1996). However, in other works it has been reported that *F. pennivorans* is able to grow from 30 °C to 80 °C (Cai et al., 2007). The results presented here, though, show that this strain T cannot grow at 38 °C, as no growth could be seen after 96 hours of incubation. The results obtained from the flasks incubated at 55 °C are more congruent with the literature, being this strain able to divide slower than when incubated at 65 °C. So, according to these results, the strain can be catalogued as a thermophilic bacterium, with optimal growth temperatures similar to those required by the rest of species of the genus *Fervidobacterium*.

11.5 Keratinase activity

It is reported that *F. pennivorans* type strain (DSM 9078) and *F. islandicum* are able to grow on chicken feathers, completely degrading them after 48 hours of anaerobic incubation (Friedrich and Antranikian, 1996, Nam et al., 2002). However, as it has been shown, when incubating the studied strain with chicken feathers four days were needed start feather degradation in the flasks with breast feathers, and a week for the flasks with wing feathers. The total degradation time for the breast feathers was seven days, while the wing feathers needed ten days of incubation to be completely degraded. The difference in time between this experiment and those described in the literature is significant; so, trying a different culture medium is recommended if an accurate reproduction of those experiments is needed, or an improvement in the method is required.

11.6 Phase-contrast and Scanning Electron Microscopy

After checking different samples under the microscope, from culture media in very different conditions, it has been ascertained that the studied strain present some degree of pleiomorphism, showing variations in size, shape and morphology. Similarly to other members of the genus (Patel et al., 1985, Kanoksilapatham et al., 2016, Podosokorskaya et al., 2011, Huber et al.), this strain present terminal spheroids or blebs, and grows in pairs or short filaments. Clusters could also be spotted, suggesting the ability to form colonies when spread on the appropriate solid medium, although this was not investigated. Due to certain issues reported when processing the sample, the morphology of the cells shown in the SEM pictures might be altered.

12. Discussion of the Genomic results

12.1 DNA isolation and strain identification

As shown and described, the genome of *F. pennivorans* strain T had the expected size for a member of the Thermotogales group (Benson et al., 2005, Madigan et al., 2014). This DNA was then used to identify the strain relying on its 16S gene sequence. Thus, after sequencing this gene, the obtained sequence was compared with BLAST. It was found that the experimental strain belonged to the *Fervidobacterium* genus, specifically, to *F. pennivorans*

species. The most similar strains were F. pennivorans DYC and the species type strain (DSM 9078), with an E-value of 0.0 and a 99% identity, meaning that, according to the 16S gene sequence, the compared sequences were identical. However, as mentioned before, to assure that two bacteria belong to the same species, their 16S rRNA gene sequences have to be more than 97 % identical, and their genomic DNA hybridization percentage higher than 70 % (STACKEBRANDT and GOEBEL, 1994). Thus, the genomes of these three species were compared through an ANI calculation and a Genome-to-Genome distance analysis. Furthermore, they were compared with the SEED viewer. Also, using the Mauve algorithm their genomes were aligned and a dot-plot matrix was performed, as detailed. Finally, a phylogenetic tree comparing all the available species of the genus *Fervidobacterium* was reconstructed.

12.1.1 Genomic analyses

The summary provided by CLC Genomic Workbench after assembling the sequenced *F. pennivorans* genome stated that the mentioned genome had a total of 2,271,337 base pairs distributed over 505 contigs. As explained before, not all that DNA belonged to *P. pennivorans*, so it was removed. After removing all contaminant DNA, the genome was uploaded again to RAST, and this time its size was 1,967,686 distributed over 28 contigs. So, a total of 303.651 base pairs of the sequenced DNA was not *F. pennivorans* DNA, more than 13 %. This strain is similar in genomic size to other close species, such as *Thermosipho africanus*, *T. melanesiensis* or *Thermotoga petrophila*, being all of them around 2 mega bases. According to the RAST annotation, all these genomes show a remarkable variety of genetic versatility, since there could be seen a number of different subsystems covering a wide spectrum of metabolic activities, related with carbon utilization, protein degradation or respiration (Overbeek et al., 2014). Although neither of these organisms can sporulate, a small subsystem related with dormancy and sporulation was annotated by RAST. In this subsystem can be found sequences related with the Stage V sporulation protein (SpoVS) of *Bacillus subtilis*. No further tests regarding this topic were conducted, so it was not investigated whether these sequences are involved in some metabolic activity or represent a genetic ancestor of these sporulation proteins and systems.

The strain T genome sequence was compared with the type strain of *F. pennivorans* as well as with the DYC strain of the same species. Also, these two strains were compared to each other. The type strain showed higher level of protein sequence identity with the type strain: 98.6 % of the sequences were at least 70 % identical between these two strains, showing the high level of similarity they have. However, when comparing with DYC strain, both of these strains showed a lower level of identity, suggesting that DYC strain is a more distant strain, with large genomic inversions.

12.1.2 Average Nucleotide Identity

DDH values are the most important criterion in the delineation of bacterial species, being 70 % the threshold limiting the species identity level, corresponding this value to 95 % ANI (Goris et al., 2007, Rodriguez-R and Konstantinidis, 2016). According to this, since strain T shows an ANI of 97.28 % with the type strain (DSM 9078), it can be asserted that they clearly belong to the same species. When comparing strain T with DYC, ANI is 91.4 % with a median of 89.97 %; in this case, according the estimation they do not belong to the same species. Finally, the comparison between DYC and DSM 9870 strains gave an ANI of 90.99 % with a median of 89.89 %, which again indicates they should not be considered as members of the same species.

12.1.3 Genome-to-genome distance calculation (GGDC)

This calculation represents a different estimation of the DDH and, thus, it can be seen as a complementary test to ANI. When comparing *F. pennivorans* strain T with the type strain's genome, GGDC gave a probability that they belonged to the same species (DDH > 70 %) of 91.41 %, consistent with the one provided by ANI. However, the comparison between *F. pennivorans* strain T and strain DYC showed a different GGDC estimation, being 20.18 % the probability these two strains belong to the same species (probability that DDH > 70 %). So, according to this test, it is very unlikely that these strains belong to the same species. Finally, DSM 9078 and DYC comparison provide the result of 7.21 % probability they belong to the same species (probability that DDH > 70 %). So, again, according to GGDC, DYC strain does not belong to the *F. pennivorans* species.

12.1.4 Phylogenetic tree

The 16S rRNA gene phylogenetic tree for the *Fervidobacterium* genus is consistent with the one recently reported in (Kanoksilapatham et al., 2016) and shows bootstrapping values over 70 % for every branch except for those grouping the three strains of *F. pennivorans*: only 37 times every 100 the tree was built, strain T and the type strain were grouped together. When it comes to the DYC strain, the given bootstrapping value was 42 %. However, these three strains were grouped together with a bootstrapping value of 89 %, meaning that their phylogenetic relationships are unclear.

12.1.5 Genomic alignment and dot-plot matrix

The genomic alignment between the sequenced genome with *F. pennivorans* DSM 9078 using Mauve's algorithm showed an almost perfect match and synteny for every contig. Only a small number of gaps between these genomes was estimated. The genomic size was also similar, both of them being close to 2 mega bases. When comparing *F. pennivorans* strain T genome with DYC strain the match was not so exact, with some gaps and differences arising, although again both genomes had similar size, around 2 mega bases. The most remarkable difference between these genomes, according to Mauve, is the presence of a large inverted fragment, belonging to contigs 1, 5, 10, 16, 21 and part of contig 6. This strain was also aligned with the species type strain, and the result was very similar to the previously discussed, with a large part of the genome inverted.

The dot-plot matrices were based on the Mauve alignment and represented an alternative visualization of the same estimation. The matrix between the sequenced genome and the type strain showed an almost complete diagonal line, from the upper left corner of the chart until the lower right corner, only broken by small points representing those parts of the genomes which did not show sequence match. Likewise, the experimental genome and the type strain showed very similar dot-matrices when their genomes were plotted compared with DYC strain's genome: long parts matched, represented by descending lines from left to right, while others appeared perpendicular to the before mentioned lines, meaning that these sequences were in reverse directions.

12.1.6 Conserved signature indels

Different Conserved Signature Indels (CSI) have been described and in Thermotogae genomes, uniquely present in the phylum or its different groups, representing specific molecular markers (Bhandari and Gupta, 2014). After analyzing *F. pennivorans* strain T genome using Artemis software, several genomic fragments only found either in this strain, *Fervidobacterium* genus or Thermotogales group were spotted. Some of these segments were annotated as recombinases, transposases or mobile elements, suggesting the occurrence of horizontal genetic transfer in this group. Other segments were annotated as functional or hypothetical proteins, so their products were obtained using ExPaSy and compared with BLAST database. An Acetamidase and a Methionine synthase were found only in the studied strain and in other members of the Thermotogales group, being candidates to carry CSI in their sequences and become molecular markers for the group. However, further investigations should be conducted regarding this topic.

13. Discussion of the peptidase cloning and expression

13.1 3D peptidase structure prediction

Among all the candidate templates found by SWISS-MODEL, two were selected: fervidolysin (ID: 1r6v) and Pro-F17H/S324A (ID: 4jp8). Since the GMQE value was higher when using the fervidolysin as a template, this protein was chosen to build the model. GMQE estimation combines properties from the target–template alignment and the template search method. The resulting GMQE score is expressed as a number between 0 and 1, reflecting the expected accuracy of a model built with that alignment and template and the coverage of the target. Higher numbers indicate higher reliability. The QMEAN4 of the estimated model was -5.09, indicating a low quality model, since values close to zero represent models comparable to what would be expected from experimental structures of similar size. The local quality plot provided by SWISS-MODEL shows, for each residue of the model, the expected similarity to the native structure. According to the tool, values below 0.6 are considered to be of low quality. Again, this local plot shows a significant number of residues which estimation lays under this threshold of 0.6, considered then as low-quality estimations. Finally, another plot is provided, comparing the model with proteins of similar size which structure has been experimentally resolved. The model estimated for the peptidase of the sequenced strain

shows a score over two times the Z-score (score over 2 times the standard deviation of the mean), resulting, again, in an estimation of low quality for the model. So, in conclusion, the estimated model is not reliable because the templates found do not have a sufficient identity level to build a high-quality model. If that was needed, it is recommended using alternative methods, such as threading or ab-initio modelling (Xiong, 2006).

13.2 Multiple sequence alignment

As explained, the catalytic triad of the sequenced protein was assessed by performing an MSA with it and all the homologues found after a PSI-BLAST search. When analyzing it could be seen that the catalytic triad of the protein, present in all the homologues found, was also present in the target protein, suggesting that it may have proteolytic activity.

13.3 Peptidase cloning and expression in *Escherichia coli*

13.3.1 Signal peptide prediction

Signal sequences are N-terminal extensions of 16 to 30 amino acid residues in length present in newly synthesized secretory and membrane proteins (Martoglio, 2003). The Signal IP server estimates the presence and length of signal peptides of query proteins. For the peptidase S8 cloned in this work, a signal peptide of 21 amino acids in length was predicted, which is the same length as fervidolysin, and similar size to islandisin, a serine protease from *F. islandicum*, with a signal peptide 33 amino acids length (Kluskens et al., 2002, Godde et al., 2005).

13.3.2 PCR and cloning

PCR amplification of the peptidase gene was successful, as could be seen in the agarose gel run afterwards: a band congruent with the size of the gene could be seen, with good quality. After cloning the gene in pINIT_tet plasmid, another electrophoresis was run, as well as sequencing of the candidate plasmids, to assess the presence of the gene in the plasmid and investigate possible mutations. The presence of the insert was verified by this process, and no mutations were found in the sequence.

13.3.3 Cloning in expression vectors and protein expression

Again, after cloning the gene in the expression vectors, its presence was verified by a PCR. The gel showed that all the picked colonies, but one, carried the plasmid and, thus, the insert, since clear bands of the right size were seen. In the SDS-PAGE run to verify the expression of the protein after the cloning experiment, bands congruent with the protease size (45.76 kDa.) were seen. These bands appeared in lysate samples induced with IPTG. So, the protein was successfully expressed in *E. coli* BL21 LOBSTR.

13.3.4 Enzyme activity assay

The chart obtained after this test shows that there were differences between the samples induced with IPTG and the uninduced negative controls. These differences suggest the protein was active, especially obvious for p7xC3H. However, the activity of the positive control was almost two times higher than for p7xC3H. So, no clear conclusions can be extracted from this test. As the enzyme is assumed to be thermo-tolerant, this assay was also conducted at 60 °C. However, the obtained results were not conclusive. The protocol supplied by (Thompson et al., 2000) in the kit did not specify the range of temperatures in which the substrate is stable, so perhaps after incubation at 60 °C it was altered or partially denatured. So, it is recommended to repeat the experiment, optimizing the conditions. Due to lack of time it was not possible to repeat the experiment modifying the experimental conditions. Furthermore, due to the insufficient amount of lysate collected after the whole procedure, no replicates could be prepared, and thus no statistical test was conducted.

14. General conclusions

The new isolate belongs to *F. pennivorans* species, although some minor differences between this strain and the species type strain could be found and described, suggesting these two strains may be classified as different strains of the same species. These two strains showed the same metabolic abilities in terms of carbon sources utilization. However, when comparing these two strains with *F. pennivorans* DYC, the highest scoring organism when comparing the studied strain's 16S rRNA gene sequence with BLAST, differences are more obvious. In fact, the analysis conducted during this work suggest this strain should be classified as a separate species within the *Fervidobacterium* genus.

Similarly to other members of the Thermotogales, the genome of *F. pennivorans* showed the presence of genomic parts only found within the Thermotogales group, being some of them found only among the Fervidobacteriaceae or being only unique of *F. pennivorans*. These fragments could be useful as genetic markers to specifically identify bacteria from this groups, and may represent also a valuable insight into the Evolution of thermophilic bacteria. Furthermore, they could lead to discovery of novel properties that are unique to these bacteria.

F. pennivorans shows keratinolytic activity, and it is able to completely degrade chicken feathers after several days of incubation at 65 °C in MMF medium supplemented with peptone and yeast extract. This ability could be very useful in biotechnological processes, as feathers cause a serious waste problem. Furthermore, *F. pennivorans* is an anaerobic and thermophilic bacterium, making it especially convenient for industrial processes, as neither no oxygen supply is needed, nor refrigeration systems. On the other hand, as *F. pennivorans* is a thermophilic bacterium, it would require a heating system, which is an inconvenience for using this organism for biotechnological processes. However, this is not an absolute drawback, as high temperatures make easier the degradation of substrates, and avoids contamination of mesophiles.

The analysis of the genome of *F. pennivorans* strain T allowed identification of different keratinase enzyme candidates. One of these contained an amino acidic catalytic triad, previously shown to be part of the active site. This candidate was successfully expressed in *E. coli*, but whether the recombinant protein is active remains unclear after the tests conducted.

However, which features are required for keratin degradation is not fully understood, and most research in this field centers on screening novel microorganisms with keratinolytic activity. Furthermore, for other complex substrates such as cellulose, it has been suggested that degradation of keratin does not occur by the action of a single enzyme, but requires a mixture of different enzymes with protease ability. Thus, the production of recombinant keratinase by heterologous remains relatively limited (Wu et al., 2017, Lange et al., 2016).

15. Further directions

Thermotogae is a relatively new group, placed in a deep branch of the Tree of Life, meaning they are closely related to the first living organisms on Earth, being their study under an evolutionary approach of great importance to understand how life did appear. Their metabolic versatility and thermophilic features makes them interesting for biotechnological processes.

During this project, a member of Thermotogae, *F. pennivorans* strain T, has been identified, characterized and studied, and a gene encoding a keratinase has been cloned and expressed. We aim for following further research in the future:

- Carry out more in-depth phylogenetic analyses studying housekeeping genes, to give new insights into the evolutionary relationships of the Thermotogae group.
- Use genomic, metagenomic and bioinformatic approaches to understand how the metabolic processes of *F. pennivorans* work, interact and are regulated, building a metabolic model of the organism.
- Further genomic investigations, to be able to close the genome of F. *pennivorans* strain T and obtain a complete map of the chromosome.
- Optimization of growth of this strain to get denser cultures, subsequently improving feather degradation.
- Carry out further assays to assess the activity of the expressed keratinase, optimizing the procedure and the conditions.
- Investigation of the bacterial secretome, which may lead to the discovery of new enzymes and give new insights into keratinases and proteases action (Lange et al., 2016), and discover other keratinase candidates.
- Application of *F. pennivorans* in industrial processes, due to its thermophilic condition and metabolic versatility.
- Conduct more microbiological prospections aiming to find new isolates with keratinolitic ability.

Appendix

16S rRNA gene merged sequence:

ATGCAGTCGAGCGGTGCAGCTGGAGGCTTCGGCCGAAGGCTGCATAGCGGCGGACGGGTGCGTAA
CGCGTAGGAACGTGCCCCCTGGAGGCGGATAGCCGCGGGAAACTGCGGGTAAACCGCCATAGACTC
GGGAGAGTAAAGGCCGAAAGGCGCCAGGGGAGCGGCCTGCGTCCCATCAGGTAGTTGGTAGGGTA
ATGGCCTACCAAGCCTACGACGGGTAGCCGGTCTGAGAGGATGGCCGGCCACAAGGGCACTGAGA
CACGGGCCCTACTCCTACGGGAGGCAGCAGTGGGGGATATTGGACAATGGGCGAAAGCCTGATCCA
GCGACGCCGCGTGGAGGACGAAGCCCTTCGGGGTGTAAACTCCTTTTGTCGGGGAAAAAGGACTGA
TGGTACCCGACGAATAAGCCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCGAG
CGTTACCCGGAATCACTGGGCGTAAAGGGTGCGTAGGCGGCCGCGTGTGTCTGGCGTAAAATACCA
CGGCTCAACCGTGGGAATGCGCTGGAAACTACGTGGCTTGGGTGCGGCAGAGGCAGACGGAACTG
CTGGTGTAGGGGTGAAATCCGTAGATATCAGCAGGAACGCCGGTGGAGAAGTCGGTCTGCTGGGC
CGTAACCGACGCTGAGGCACGAAAGCTAGGGGAGCGAACCGGATTAGATACCCGGGTAGTCCTAGC
CGTAAACGATGCTCACTAGGTGTGGGGGCGGAAGCCTCCGTGCTGAAGCTAACGCGATAAGTGAGC
CGCCTGGGGAGTACGCCCGCAAGGGTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTG
GAGCGTGTGGTTTAATTGGAAGCTAAGCCAAGAACCTTACCAGGGCTTGACATGCTGGTGGTACCG
AGCCGAAAGGTGAGGGACCCTGCCTTATGGCAGGGAGCCAGCACAGGTGGTGCACGGTCGTCGTC
AGCTCGTGCCGTGAGGTGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTGCCCTTAGTTGCCAGCG
GTTAGGCCGGGCACTCTAAGGGGACTGCCGGCGACGAGCCGGAGGAAGGAGGGGATGACGTCAG
ATACTCGTGCCCCTTATGTCCTGGGCGACACACGCGCTACAATGGGGTGGACAGCGGGAAGCGAGG
CAGCGATGCTGAGCAAATCCCTGAAACCACCCCCCAGTTCAGATTGTGGGCTGAAACCCGCCCACAT
GAAGCCGGAATCGCTAGTAATCGCGGATCAGCCATGCCGCGGTGAATACGTTCCCGGGCCTTGTACA
CACCGCCCGTCAAGCCACCCGAGTCGCGGGCACCCGAAGACGGTGACCCTTAGGGGCGCCGTTGAG
GGTGAACGCTGGCGGCGTGCCTAACACATGCAAGTCGAGCGGTGCAGCTGGAGGCTTCGGCCGAA
GGCTGCATAGCGGCGGACGGGTGCGTAACGCGTAGGAACGTGCCCCCTGGAGGCGGATAGCCGCG
GGAAACTGCGGGTAAACCGCCATAGACTCGGGAGAGTAAAGGCCGAAAGGCGCCAGGGGAGCGG
CCTGCGTCCCATCAGGTAGTTGGTAGGGTAATGGCCTACCAAGCCTACGACGGGTAGCCGGTCTGA
GAGGATGGCCGGCCACAAGGGCACTGAGACACGGGCCCTACTCCTACGGGAGGCAGCAGTGGGGG
ATATTGGACAATGGGCGAAAGCCTGATCCAGCGACGCCGCGTGGAGGACGAAGCCCTTCGGGGTGT
AAACTCCTTTTGTCGGGGAAAAAGGACTGA

Keratinase sequence. The signal peptide is highlighted in yello. Hyphen represent the signal peptide cleavage position. The catalytic triad is bolded and marked in red:

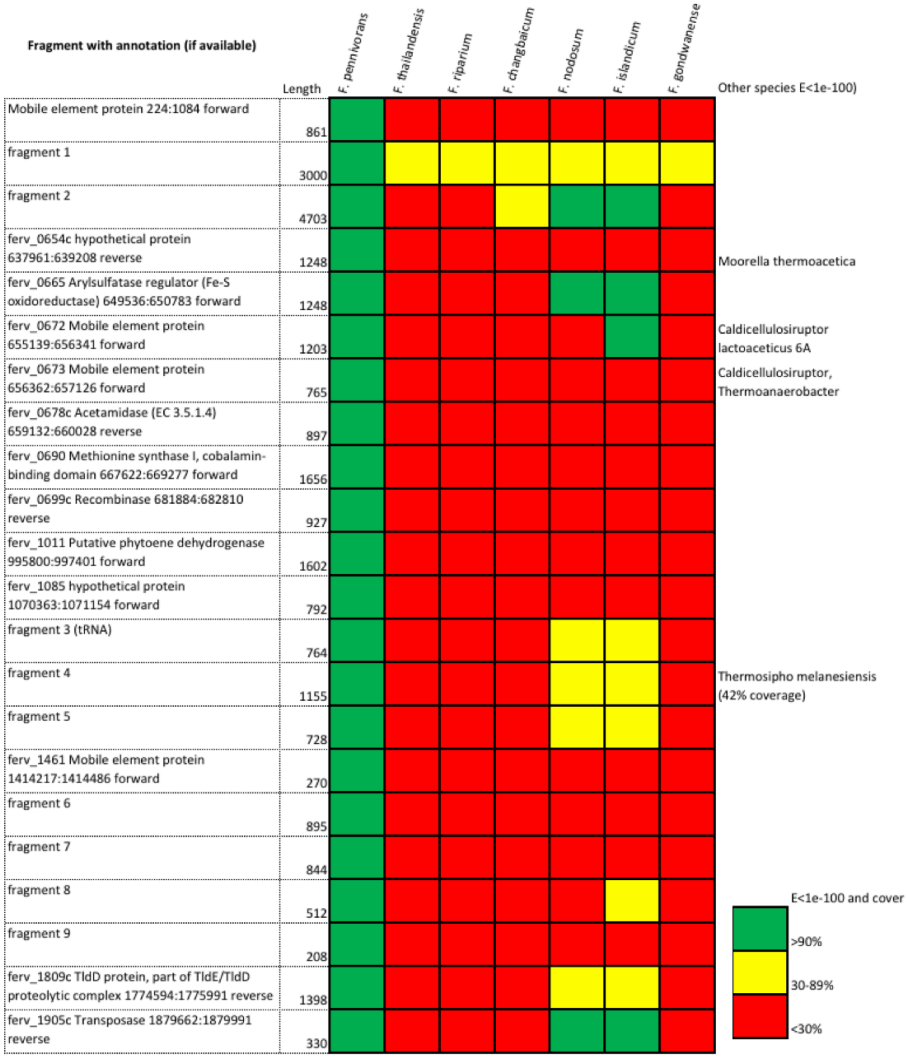
MKKFVLLTAVFALLLVTFSCT-NSLEPRFEPRAQGEFEVSEKLGVSGTEEDYVPGEYVVQF
EPREDAVKALSSVGAEVVRAYSFSDVQIVTVRTEKPELLNSLPGVKSVDKNYIYRALATP
NDTYYRYQWHYNNIKLPQAWDIMKSANIVVAVIDTGVSFTHPDLQGIFVQGYDFVDGDYD
PTDPAQDVSHGTHCIGTIAAVTNNSLGVAGVNWGGYGIKIMPIRVLGADGSGTLDNVAAG
IRWAVDNGAKIVSMSLGGSGAQVLMDAVKYAYSRNVTLICAAGNESRPSLSYPAAYVETI
AVGATRYDNTRARYSNYNYTRYYDPYRKAYVYHYLDVVAPGGDTSVDQNGDGYADGVLST
TWTPTYGNTYMFLQGTSMATPHVAALAAMLYAKGYTTPEAIRSRLIKTAYKIPGYTYNSS
GWNKYVGYGLIDAYKALTY

Keratinase gene sequence. Red hyphen represents the signal peptide cleavage position.

ATGAAGAAGTTTGTTTTACTTACAGCAGTTTTCGCACTTTTGCTGGTAACATTCAGCTGT
ACC-AACTCATTAGAGCCAAGATTTGAACCACGCGCACAAGGTGAGTTTGAGGTTTCAGAG
AAACTTGGCGTATCCGGAACAGAAGAAGATTACGTTCCTGGAGAATATGTTGTCCAGTTC
GAGCCAAGAGAAGATGCGGTTAAAGCATTATCAAGTGTTGGTGCAGAAGTAGTCAGAGCA

TATTCATTCAGCGATGTTCAAATCGTAACAGTAAGAACGGAAAAACCAGAACTTCTTAAT
TCTCTTCCAGGTGTCAAGTCAGTTGATAAGAACTACATTTACAGAGCACTCGCAACACCA
AACGACACATACTACCGATACCAGTGGCACTACAACAATATCAAACTGCCACAGGCATGG
GATATCATGAAATCTGCTAATATCGTTGTAGCAGTTATTGATACAGGAGTTAGCTTTACA
CATCCAGACCTGCAAGGCATATTCGTTCAAGGCTATGACTTTGTCGATGGAGATTACGAT
CCGACAGACCCGGCACAGGATGTGAGCCATGGAACACATTGTATAGGAACAATAGCCGCT
GTTACAAACAACAGCCTTGGTGTTGCCGGAGTTAATTGGGGAGGATATGGAATAAAGATA
ATGCCTATCAGGGTTCTTGGCGCAGACGGTTCCGGAACACTCGATAATGTCGCAGCTGGT
ATCAGATGGGCAGTTGACAACGGTGCAAAAATAGTGAGTATGAGTCTTGGTGGTAGCGGT
GCACAAGTTCTTATGGATGCCGTTAAATATGCTTACAGCAGAAATGTAACACTTATCTGC
GCAGCAGGAAATGAGAGTAGACCTTCGCTATCCTATCCAGCAGCATATGTTGAAACGATC
GCAGTAGGTGCAACAAGATACGACAACACACGCGCTCGGTATTCTAACTACAATTACACA
AGATACTACGATCCTTACAGAAAAGCGTATGTATACCATTACCTTGACGTTGTTGCTCCT
GGTGGAGATACAAGTGTTGACCAAAACGGTGATGGATACGCAGATGGTGTGCTCAGCACA
ACCTGGACACCGACATACGGAAATACATATATGTTCTTGCAAGGTACATCGATGGCAACA
CCACATGTTGCAGCGCTTGCAGCTATGCTTTACGCAAAAGGTTACACAACACCAGAGGCG
ATTAGAAGCAGACTTATCAAAACAGCTTATAAGATTCCTGGATACACATATAATTCGAGC
GGATGGAACAAATACGTTGGCTACGGTTTAATTGATGCTTACAAGGCATTGACATACTAA

CSI candidates for *F. pennivorans* strain T found:



References

Basic Local Alignment Search Tool [Online]. Available: https://blast.ncbi.nlm.nih.gov/.

CLC Genomic Workbench 11.01.

FX Cloning [Online]. Available: http://www.fxcloning.org.

ACHENBACH-RICHTER, L., GUPTA, R., STETTER, K. O. & WOESE, C. R. 1987. Were the original eubacteria thermophiles? *Syst Appl Microbiol,* 9**,** 34-9.

ALIKHAN, N. F., PETTY, N. K., BEN ZAKOUR, N. L. & BEATSON, S. A. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics,* 12**,** 402.

ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res,* 25**,** 3389-402.

ANDERSEN, K. R., LEKSA, N. C. & SCHWARTZ, T. U. 2013. Optimized *E. coli* expression strain LOBSTR eliminates common contaminants from His-tag purification. *Proteins,* 81**,** 1857-1861.

ANDREWS, K. T. & PATEL, B. K. 1996. *Fervidobacterium gondwanense* sp. nov., a new thermophilic anaerobic bacterium isolated from nonvolcanically heated geothermal waters of the Great Artesian Basin of Australia. *Int J Syst Bacteriol,* 46**,** 265-9.

ATLAS, R. M. 2004. *Handbook of Microbiological Media, Third Edition*, CRC Press.

AZIZ, R. K., BARTELS, D., BEST, A. A., DEJONGH, M., DISZ, T., EDWARDS, R. A., FORMSMA, K., GERDES, S., GLASS, E. M., KUBAL, M., MEYER, F., OLSEN, G. J., OLSON, R., OSTERMAN, A. L., OVERBEEK, R. A., MCNEIL, L. K., PAARMANN, D., PACZIAN, T., PARRELLO, B., PUSCH, G. D., REICH, C., STEVENS, R., VASSIEVA, O., VONSTEIN, V., WILKE, A. & ZAGNITKO, O. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics,* 9**,** 75.

BARRETT, A. J. & RAWLINGS, N. D. 1995. Families and clans of serine peptidases. *Arch Biochem Biophys,* 318**,** 247-50.

BENKERT, P., BIASINI, M. & SCHWEDE, T. 2011. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics,* 27**,** 343-50.

BENSON, D. A., KARSCH-MIZRACHI, I., LIPMAN, D. J., OSTELL, J. & WHEELER, D. L. 2005. GenBank. *Nucleic Acids Research,* 33**,** D34-D38.

BERNARD, D., MEHUL, B., THOMAS-COLLIGNON, A., SIMONETTI, L., REMY, V., BERNARD, M. A. & SCHMIDT, R. 2003. Analysis of proteins with caseinolytic activity in a human stratum corneum extract revealed a yet unidentified cysteine protease and identified the so-called "stratum corneum thiol protease" as cathepsin l2. *J Invest Dermatol,* 120**,** 592-600.

BERTONI, M., KIEFER, F., BIASINI, M., BORDOLI, L. & SCHWEDE, T. 2017. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Sci Rep,* 7**,** 10480.

BHANDARI, V. & GUPTA, R. S. 2014. Molecular signatures for the phylum (class) Thermotogae and a proposal for its division into three orders (Thermotogales, Kosmotogales ord. nov. and Petrotogales ord. nov.) containing four families (Thermotogaceae, Fervidobacteriaceae fam. nov., Kosmotogaceae fam. nov. and Petrotogaceae fam. nov.) and a new genus Pseudothermotoga gen. nov. with five new combinations. *Antonie Van Leeuwenhoek,* 105**,** 143-68.

BIASINI, M., BIENERT, S., WATERHOUSE, A., ARNOLD, K., STUDER, G., SCHMIDT, T., KIEFER, F., GALLO CASSARINO, T., BERTONI, M., BORDOLI, L. & SCHWEDE, T. 2014. SWISS-MODEL:

modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res,* 42**,** W252-8.

BIENERT, S., WATERHOUSE, A., DE BEER, T. A., TAURIELLO, G., STUDER, G., BORDOLI, L. & SCHWEDE, T. 2017. The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Res,* 45**,** D313-d319.

BRZIN, J., ROGELJ, B., POPOVIC, T., STRUKELJ, B. & RITONJA, A. 2000. Clitocypin, a new type of cysteine proteinase inhibitor from fruit bodies of mushroom clitocybe nebularis. *J Biol Chem,* 275**,** 20104-9.

BUEHLER, E. 1996. *OligoCalc* [Online]. University of Pittsburgh School of Medicine. Available: http://www.basic.northwestern.edu/biotools/OligoCalc.html [Accessed].

BUTTON, D. K., SCHUT, F., QUANG, P., MARTIN, R. & ROBERTSON, B. R. 1993. Viability and Isolation of Marine Bacteria by Dilution Culture: Theory, Procedures, and Initial Results. *Applied and Environmental Microbiology,* 59**,** 881-891.

CAI, J., WANG, Y., LIU, D., ZENG, Y., XUE, Y., MA, Y. & FENG, Y. 2007. *Fervidobacterium changbaicum* sp. nov., a novel thermophilic anaerobic bacterium isolated from a hot spring of the Changbai Mountains, China. *Int J Syst Evol Microbiol,* 57**,** 2333-6.

CHE, D., HASAN, M. S. & CHEN, B. 2014. Identifying Pathogenicity Islands in Bacterial Pathogenomics Using Computational Approaches. *Pathogens,* 3**,** 36-56.

CHEN, X. G., STABNIKOVA, O., TAY, J. H., WANG, J. Y. & TAY, S. T. 2004. Thermoactive extracellular proteases of *Geobacillus caldoproteolyticus*, sp. nov., from sewage sludge. *Extremophiles,* 8**,** 489-98.

CHIKHI, R. & MEDVEDEV, P. 2014. Informed and automated k-mer size selection for genome assembly. *Bioinformatics,* 30**,** 31-37.

CONNERS, S. B., MONGODIN, E. F., JOHNSON, M. R., MONTERO, C. I., NELSON, K. E. & KELLY, R. M. Microbial biochemistry, physiology, and biotechnology of hyperthermophilic Thermotoga species.

CUECAS, A., KANOKSILAPATHAM, W. & GONZALEZ, J. M. 2017. Evidence of horizontal gene transfer by transposase gene analyses in *Fervidobacterium* species. *PLoS One,* 12**,** e0173961.

DARLING, A. C., MAU, B., BLATTNER, F. R. & PERNA, N. T. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res,* 14**,** 1394-403.

DSMZ, L.-I. 1969. *Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH* [Online]. Available: https://www.dsmz.de [Accessed].

EFRON, B. & TIBSHIRANI, R. J. 1994. *An Introduction to the Bootstrap*, Taylor & Francis.

FRIEDRICH, A. B. & ANTRANIKIAN, G. 1996. Keratin Degradation by *Fervidobacterium pennavorans*, a Novel Thermophilic Anaerobic Species of the Order Thermotogales. *Appl Environ Microbiol,* 62**,** 2875-82.

FROCK, A. D., NOTEY, J. S. & KELLY, R. M. 2010. The genus Thermotoga: recent developments. *Environ Technol,* 31**,** 1169-81.

GASCUEL, O. 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol Biol Evol,* 14**,** 685-95.

GASTEIGER, E., GATTIKER A FAU - HOOGLAND, C., HOOGLAND C FAU - IVANYI, I., IVANYI I FAU - APPEL, R. D., APPEL RD FAU - BAIROCH, A. & BAIROCH, A. ExPASy: The proteomics server for in-depth protein knowledge and analysis.

GEERTSMA, E. R. & DUTZLER, R. 2011. A versatile and efficient high-throughput cloning tool for structural biology. *Biochemistry,* 50**,** 3272-8.

GODDE, C., SAHM, K., BROUNS, S. J., KLUSKENS, L. D., VAN DER OOST, J., DE VOS, W. M. & ANTRANIKIAN, G. 2005. Cloning and expression of islandisin, a new thermostable subtilisin from *Fervidobacterium islandicum*, in *Escherichia coli*. *Appl Environ Microbiol,* 71**,** 3951-8.

GORIS, J., KONSTANTINIDIS, K. T., KLAPPENBACH, J. A., COENYE, T., VANDAMME, P. & TIEDJE, J. M. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol,* 57**,** 81-91.

GOUY, M., GUINDON, S. & GASCUEL, O. 2010. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular Biology and Evolution,* 27**,** 221-224.

GRIEKSPOOR, A. & GROOTHUIS, T. 2015. *4Peaks* [Online]. Available: www.nucleobytes.com.

GUEX, N., PEITSCH, M. C. & SCHWEDE, T. 2009. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis,* 30 Suppl 1**,** S162-73.

GUPTA, R. S. 1998. Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiol Mol Biol Rev,* 62**,** 1435-91.

GUPTA, R. S. & BHANDARI, V. 2011. Phylogeny and molecular signatures for the phylum Thermotogae and its subgroups. *Antonie Van Leeuwenhoek,* 100**,** 1-34.

HUBER, R., LANGWORTHY, T. A., KÖNIG, H., THOMM, M., WOESE, C. R., SLEYTR, U. B. & STETTER, K. O. 1986. *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. *Archives of Microbiology,* 144**,** 324-333.

HUBER, R., WOESE, C. R., LANGWORTHY, T. A., KRISTJANSSON, J. K. & STETTER, K. O. *Fervidobacterium islandicum* sp. nov., a new extremely thermophilic eubacterium belonging to the “Thermotogales”.

HUNGATE, R. E. 1950. The anaerobic mesophilic cellulolytic bacteria. *Bacteriol Rev,* 14**,** 1-49.

INOUYE, M. 1991. Intramolecular chaperone: the role of the pro-peptide in protein folding. *Enzyme,* 45**,** 314-21.

JACOBS M FAU - ELIASSON, M., ELIASSON M FAU - UHLEN, M., UHLEN M FAU - FLOCK, J. I. & FLOCK, J. I. Cloning, sequencing and expression of subtilisin Carlsberg from *Bacillus licheniformis*.

KALISZ, H. M. 1988. Microbial proteinases. *Adv Biochem Eng Biotechnol,* 36**,** 1-65.

KANOKSILAPATHAM, W., PASOMSUP, P., KEAWRAM, P., CUECAS, A., PORTILLO, M. C. & GONZALEZ, J. M. 2016. *Fervidobacterium thailandense* sp. nov., an extremely thermophilic bacterium isolated from a hot spring. *Int J Syst Evol Microbiol,* 66**,** 5023-5027.

KIM, J. S., KLUSKENS, L. D., DE VOS, W. M., HUBER, R. & VAN DER OOST, J. 2004. Crystal structure of fervidolysin from *Fervidobacterium pennivorans*, a keratinolytic enzyme related to subtilisin. *J Mol Biol,* 335**,** 787-97.

KIMURA, M. 1983. *The neutral theory of molecular evolution,* Cambridge, Cambridge University Press.

KLUSKENS, L. D., VOORHORST, W. G., SIEZEN, R. J., SCHWERDTFEGER, R. M., ANTRANIKIAN, G., VAN DER OOST, J. & DE VOS, W. M. 2002. Molecular characterization of fervidolysin, a subtilisin-like serine protease from the thermophilic bacterium *Fervidobacterium pennivorans*. *Extremophiles,* 6**,** 185-94.

KONSTANTINIDIS, K. T., RAMETTE, A. & TIEDJE, J. M. 2006. The bacterial species definition in the genomic era. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 361**,** 1929-1940.

KRUMSIEK, J., ARNOLD, R. & RATTEI, T. 2007. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics,* 23**,** 1026-8.

KUMAR, S., SHARMA, N. S., SAHARAN, M. R. & SINGH, R. 2005. Extracellular acid protease from *Rhizopus oryzae*: purification and characterization. *Process Biochemistry,* 40**,** 1701-1705.

KUMAR, S., STECHER, G. & TAMURA, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol,* 33**,** 1870-4.

LANGE, L., HUANG, Y. & BUSK, P. K. 2016. Microbial decomposition of keratin in nature—a new hypothesis of industrial relevance. *Applied Microbiology and Biotechnology,* 100**,** 2083-2096.

LEE, Y. J., DHANASINGH, I., AHN, J. S., JIN, H. S., CHOI, J. M., LEE, S. H. & LEE, D. W. 2015a. Biochemical and structural characterization of a keratin-degrading M32 carboxypeptidase from *Fervidobacterium islandicum* AW-1. *Biochem Biophys Res Commun,* 468**,** 927-33.

LEE, Y. J., JEONG, H., PARK, G. S., KWAK, Y., LEE, S. J., LEE, S. J., PARK, M. K., KIM, J. Y., KANG, H. K., SHIN, J. H. & LEE, D. W. 2015b. Genome sequence of a native-feather degrading extremely thermophilic Eubacterium, *Fervidobacterium islandicum* AW-1. *Stand Genomic Sci,* 10**,** 71.

MACY, J. M., SNELLEN, J. E. & HUNGATE, R. E. 1972. Use of syringe methods for anaerobiosis. *Am J Clin Nutr,* 25**,** 1318-23.

MADIGAN, M. T., MARTINKO, J. M., BENDER, K. S., BUCKLEY, D. H. & STAHL, D. A. 2014. *Brock biology of microorganisms*.

MADIGAN, M. T., MARTINKO, J. M. & BROCK, T. D. 2006. *Brock biology of microorganisms,* Upper Saddle River, NJ, Pearson Prentice Hall.

MARTOGLIO, B. 2003. Intramembrane proteolysis and post-targeting functions of signal peptides. *Biochem Soc Trans,* 31**,** 1243-7.

MEIER-KOLTHOFF, J. P., AUCH, A. F., KLENK, H. P. & GOKER, M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics,* 14**,** 60.

MIRANDA-TELLO, E., FARDEAU, M. L., THOMAS, P., RAMIREZ, F., CASALOT, L., CAYOL, J. L., GARCIA, J. L. & OLLIVIER, B. 2004. *Petrotoga mexicana* sp. nov., a novel thermophilic, anaerobic and xylanolytic bacterium isolated from an oil-producing well in the Gulf of Mexico. *Int J Syst Evol Microbiol,* 54**,** 169-74.

NAM, G. W., LEE, D. W., LEE, H. S., LEE, N. J., KIM, B. C., CHOE, E. A., HWANG, J. K., SUHARTONO, M. T. & PYUN, Y. R. 2002. Native-feather degradation by *Fervidobacterium islandicum* AW-1, a newly isolated keratinase-producing thermophilic anaerobe. *Arch Microbiol,* 178**,** 538-47.

NIEHAUS, F., BERTOLDO, C., KAHLER, M. & ANTRANIKIAN, G. 1999. Extremophiles as a source of novel enzymes for industrial application. *Appl Microbiol Biotechnol,* 51**,** 711-29.

OVERBEEK, R., OLSON, R., PUSCH, G. D., OLSEN, G. J., DAVIS, J. J., DISZ, T., EDWARDS, R. A., GERDES, S., PARRELLO, B., SHUKLA, M., VONSTEIN, V., WATTAM, A. R., XIA, F. & STEVENS, R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res,* 42**,** D206-14.

PAPADOPOULOS, M. C. 1989. Effect of processing on high-protein feedstuffs: A review. *Biological Wastes,* 29**,** 123-138.

PARRY, D. A., CREWTHER, W. G., FRASER, R. D. & MACRAE, T. P. 1977. Structure of alpha-keratin: structural implication of the amino acid sequences of the type I and type II chain segments. *J Mol Biol,* 113**,** 449-54.

PATEL, B. K. C., MORGAN, H. W. & DANIEL, R. M. 1985. *Fervidobacterium nodosum* gen. nov. and spec. nov., a new chemoorganotrophic, caldoactive, anaerobic bacterium.

PEAKE, I. 1989. The polymerase chain reaction. *Journal of Clinical Pathology,* 42**,** 673-676.

PETERSEN, T. N., BRUNAK, S., VON HEIJNE, G. & NIELSEN, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods,* 8**,** 785-6.

PODOSOKORSKAYA, O. A., MERKEL, A. Y., KOLGANOVA, T. V., CHERNYH, N. A., MIROSHNICHENKO, M. L., BONCH-OSMOLOVSKAYA, E. A. & KUBLANOV, I. V. 2011. *Fervidobacterium riparium* sp. nov., a thermophilic anaerobic cellulolytic bacterium isolated from a hot spring. *Int J Syst Evol Microbiol,* 61**,** 2697-701.

R., V. S., M., P. R. & G., W. W. 2010. Strategies for culture of 'unculturable' bacteria. *FEMS Microbiology Letters,* 309**,** 1-7.

RAINEY, F. A., DORSCH, M., MORGAN, H. W. & STACKEBRANDT, E. 1992. 16S rDNA Analysis of *Spirochaeta thermophila*: Its Phylogenetic Position and Implications for the Systematics of the Order Spirochaetales. *Systematic and Applied Microbiology,* 15**,** 197-202.

RAVOT, G., OLLIVIER, B., MAGOT, M., PATEL, B., CROLET, J., FARDEAU, M. & GARCIA, J. 1995. Thiosulfate Reduction, an Important Physiological Feature Shared by Members of the Order Thermotogales. *Applied and Environmental Microbiology,* 61**,** 2053-2055.

RAWLINGS, N. D. & BARRETT, A. J. 1994. Families of serine peptidases. *Methods Enzymol,* 244**,** 19-61.

REMMERT, M., BIEGERT, A., HAUSER, A. & SODING, J. 2011. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods,* 9**,** 173-5.

RICE, P., LONGDEN, I. & BLEASBY, A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet,* 16**,** 276-7.

RODRIGUEZ-R, L. M. & KONSTANTINIDIS, K. T. 2016. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Preprints,* 4**,** e1900v1.

RUTHERFORD, K., PARKHILL, J., CROOK, J., HORSNELL, T., RICE, P., RAJANDREAM, M. A. & BARRELL, B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics (Oxford, England),* 16**,** 944-945.

SAMBROOK, J., FRITSCH, E.F. AND MANIATIS, T. 1989. *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press.

SEQUENCING FACILITY, U. *Big Dye v3.1 protocol* [Online]. Available: http://www.uib.no/en/seqlab/55363/protocol-bigdye-v31.

SEZONOV, G., JOSELEAU-PETIT, D. & D'ARI, R. 2007. *Escherichia coli* Physiology in Luria-Bertani Broth. *Journal of Bacteriology,* 189**,** 8746-8749.

SHAVANDI, A., SILVA, T. H., BEKHIT, A. A. & BEKHIT, A. E.-D. A. 2017. Keratin: dissolution, extraction and biomedical application. *Biomaterials Science,* 5**,** 1699-1735.

SHAW, A. K. & PAL, S. K. 2007. Activity of Subtilisin Carlsberg in macromolecular crowding. *Journal of Photochemistry and Photobiology B: Biology,* 86**,** 199-206.

SIEVERS, F., WILM, A., DINEEN, D., GIBSON, T. J., KARPLUS, K., LI, W., LOPEZ, R., MCWILLIAM, H., REMMERT, M., SODING, J., THOMPSON, J. D. & HIGGINS, D. G. 2011. Fast, scalable

generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol,* 7**,** 539.

STACKEBRANDT, E. & GOEBEL, B. M. 1994. Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *International Journal of Systematic and Evolutionary Microbiology,* 44**,** 846-849.

STOTHARD, P. 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques,* 28**,** 1102, 1104.

SZYBALSKI, W., KIM, S. C., HASAN, N. & PODHAJSKA, A. J. 1991. Class-IIS restriction enzymes--a review. *Gene,* 100**,** 13-26.

TATUSOVA, T., DICUCCIO, M., BADRETDIN, A., CHETVERNIN, V., NAWROCKI, E. P., ZASLAVSKY, L., LOMSADZE, A., PRUITT, K. D., BORODOVSKY, M. & OSTELL, J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res,* 44**,** 6614-24.

THOMPSON, V. F., SALDANA, S., CONG, J. & GOLL, D. E. 2000. A BODIPY fluorescent microplate assay for measuring activity of calpains and other proteases. *Anal Biochem,* 279**,** 170-8.

WANDERSMAN, C. 1989. Secretion, processing and activation of bacterial extracellular proteases. *Mol Microbiol,* 3**,** 1825-31.

WILLIAMS, C. M., LEE, C. G., GARLICH, J. D. & SHIH, J. C. H. 1991. Evaluation of a Bacterial Feather Fermentation Product, Feather-Lysate, as a Feed Protein.

WOESE, C. R., GUTELL, R., GUPTA, R. & NOLLER, H. F. 1983. Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol Rev,* 47**,** 621-69.

WU, W.-L., CHEN, M.-Y., TU, I. F., LIN, Y.-C., ESWARKUMAR, N., CHEN, M.-Y., HO, M.-C. & WU, S.-H. 2017. The discovery of novel heat-stable keratinases from *Meiothermus taiwanensis* WR-220 and other extremophiles. *Scientific Reports,* 7**,** 4658.

XIONG, J. 2006. *Essential Bioinformatics,* Cambridge, Cambridge University Press.

YUZAKI, K., SANDA, Y., YOU, D. J., UEHARA, R., KOGA, Y. & KANAYA, S. 2013. Increase in activation rate of Pro-Tk-subtilisin by a single nonpolar-to-polar amino acid substitution at the hydrophobic core of the propeptide domain. *Protein Sci,* 22**,** 1711-21.