People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

Abdelhamid Ibn Badis University–Mostaganem

Faculty of Natural and Life Sciences

Department of Biology

Laboratory of Microbiology and Plant Biology

A Thesis submitted in partial fulfillments of the requirements for admittance to the degree of Doctor in Biology

Filière: Biology

Option: Molecular Biotechnology

Presented by: **FELLAHI Soltana**

Keratinolytic protease genes in *Bacillus pumilus* strain C_4 - Enhancement of extracellular protease production and molecular screening.

For public examination on 27/06/2018

Examiners' Committee

Pr. DJIBAOUI Rachid	Chairman	University of Mostaganem
Pr. DALACHE Fatiha	Examiner	University of Mostaganem
Pr. BENSOLTANE Ahmed	Examiner	University of Oran
Dr. RACHEDI Abdelkarim	Examiner	University of Saida
Pr. BELAHCENE Miloud	Examiner	University Center of Ain Temouchent
Pr. CHIBANI Abdelwaheb	Supervisor	University of Mostaganem

2017-2018

Aknowledgment

In the name of Allah, the most merciful and the most compassionate.

All praise and thanks to Allah for the successful and completion of this research work. He gave me the health, patience, and the strength to reach this far.

This thesis work was mainly carried out at Swedish Centre for Resource Recovery, University of Borås, with the financial support from the centre and Sparbanksstiftelsen Sjuhärad in Borås, Sweden. Also a part of the practical work was done at the laboratory of Biotechnology, Institute of Graduate Studies and Research, Alexandria University, in Egypt.

I am grateful to everyone involved in this process. I am most grateful to my thesis supervisor Professor CHIBANI Abdelwaheb for all the help and guidance he has provided over the years.

I would also like to express my sincere gratitude to Dr. Elisabeth Feuk Lagerstedt, Doctor of molecular biology at the Swedish Centre for Resource Recovery, University of Borås, for her great and valuable help, continuous support and advice.

Also, I am greatly indebted to Dr.Mohammad J. Taherzadeh, Professor of chemical engineering at the Swedish Centre for Resource Recovery, University of Borås and to Dr. Johan O. Westman working at Gothenberg University, for their great help and advice.

I thank my thesis advisory committee members Professors BENSOLTANE Ahmed, DJIBAOUI Rachid, DALACHE Fatiha, BELLAHCENE Miloud and RACHEDI Abdelkarim for their helpful discussions and valuable advice regarding this thesis project.

I owe my deepest gratitude to my family and specially my brother Mohammed for his support and encouragement.

Finally, I am grateful to my husband for his endless support and unconditional love and to my son Ilyes.

SoltanaFellahi,2018

List of Tables

List of Figures

Summary

Optimization of extracellular protease production by *Bacillus* sp. $C_4$was carried out using a two-step strategy One Variable at A Time (OVAT). On the basis of the results obtained from OVAT experiments, 1XSG medium supplemented with 0.2% yeast extract was the best medium formula for maximum protease production. An overall of 1.8-fold increase in protease production was achieved in the optimized medium as compared to the non-optimized one. The optimum predicted production conditions were as follows: beef extract (0.3%), peptone (0.5%), yeast extract (0.2%), glucose (0.1%), KCl (0.2%), $FeSO_4$ (1mM), $MgSO_4$ (2mM), $MnCl_2$ (0.1mM), and $CaCl_2$ (1 mM) and an incubation period of 24 hours. It was also found out that under the optimal conditions the extracellular protease production started four hours (04 hours) earlier than that produced in the non-optimized conditions.

A keratinolytic protease gene (ker1 gene) was isolated from a keratin-degrading *Bacillus* sp. C_4 using PCR amplification and degenerate primers. Ker1 gene consists of an open reading frame of 1,152 bps, encoding a prepropeptide (polypeptide) consists of 383 amino acidres residues. The BLASTp analysis showed that the deduced amino acid sequence of *ker1* gene has identity to many serine proteases from different *Bacillus pumilis*. This indicated that ker1 gene encodes one of the serine protease from *Bacillus* sp. C_4. The three residues, found in all subtilisins, at the active site that form the catalytic triad (Asp-140, His-172 and Ser-329) were conserved in the deduced protein sequence of the ker1 gene. The *ker1* gene was subsequently submitted to the NCBI GenBank (accession number: KX184831). Genome sequencing and analysis of the *Bacillus* sp. C_4 was performed as an alternative approach to investigate further keratinase genes involved in keratin degradation. *Bacillus* sp. C_4 genome was constructed as a single circular chromosome containing 3,659,360 bps with an average G+C content of 41.4% and no plasmids were detected. Annotation results indicated that *Bacillus sp* C_4 belongs to the *B. pumilus* group of organisms. NCBI Prokaryotic genome annotation pipeline predicted a total of 3,698 genes, from which 3,596 protein-coding sequences (CDS), 71 RNA genes and 31 pseudogenes. The RNA coding genes predicted include 60 tRNAs, 10 rRNAs, and 1 non-coding RNA (ncRNA). The genome was found to be 99.6 % complete when analyzed. This whole-genome *Bacillus pumilus* C_4 sequence has been deposited at NCBI GenBank under the accession number CP011109. The bioproject number was PRJNA278012. Eighty one candidate genes for peptidase were found overall in the draft genome. Out of them six (06) genes were annotated as putative peptidase S8 in which two genes were

predicted also as subtilisins based on BASys annotation. The first one is the *Ker1* gene (accession number: KX184831) previously isolated in this study. The second one is a homolog to *Ker1* gene and was named *Ker2* gene. This gene was subsequently submitted to NCBI GenBank (Accession Number: KX184832). The ORF of this gene has an equal amount of base pairs as *Ker1* gene. The two genes had a gene sequence similarity of 67% using BLASTn at NCBI. The deduced amino acid sequence of *ker2* gene of *B. pumilus* C_4 contained 383 amino acid residues and the three residues at the active site that form the catalytic triad in all subtilisins are Asp-140, His-172 and Ser-329. Results of PCR amplification, *de novo* sequencing, and genome annotation indicated that the protease producer and keratin degrading *Bacillus pumilus* C_4 has at least two genes encoding subtilisin-like serine proteases. The broad substrate specificity and keratinolytic activity towards both type of keratin of this strain could originate from the presence of this high number of putative peptidases and proteases.

Identification of some protease enzymes produced upon growing *Bacillus pumilus* C_4 on optimized medium was done using nLC-ESI-MS/MS followed by Mascot database search. Mascot Search results indicated that the partial purified protease fraction in fact consists of two enzymes, corresponding to the previously identified genes *Ker1* and *Ker2* of the *B. pumilus* strain C_4. This confirmed that *Bacillus pumilus* C_4 produced at least two protease enzymes that could be related to its remarkable keratinilytic activity towards both type of keratins.

Key words: *Bacillus pumilus* C_4, extracellular protease, subtilisins, keratins, whole genome sequencing.

الملخص

لتحسين إنتاج إنزيم البروتياز خارج الخلية ، تم استخدام طريقة متكونة من مرحلتين: طريقة المتغير الواحد في المرة الواحدة (OVAT). أوضحت نتائج الخطوة الأولى أن الوسط (1XSG) المزود بـ 0.2% مستخلص الخميرة كان التركيبة الجيدة لأحسن إنتاج للبروتياز بمقدار 419 وحدة/مجم بروتين بعد 24 ساعة عند درجة حرارة 37 ْم والأس هيدروجيني 7.0، و بحجم لقاح قدره 2%. لقد تم تحسين إنتاج إنزيم البروتياز في هذا الوسط بمعدل قدره اكثر من مرة و نصف مقارنة بالوسط الأساسي. لقد تم ايضا ايجاد ان النشاط البروتيازى لهده السلالة البكتيرية يبدا مبكرا باربع ساعات تحت الظروف المثلى مقارنة بالظروف الاساسية للنمو.

تم عزل المورثة المحللة للكيراتين (جين ker1) من *Bacillus* sp. C_4 المكسرة للكيراتين باستخدام تفاعل البلمرة المتسلسل (PCR) وبادئات البلمرة. المورثة ker1 تتكون من اطار قراءة مفتوح يضم 1152 زوج قاعدى، المشفر لمتعدد بيبتيد يتكون من 383 حمض امينى. اظهر تحليل BLASTP لسلسلة الاحماض الامينية المستنتجة لمورثة ker1 أن لديها تشابه مع الكثير من السيرين بروتياز المنتج من عدد من سلالات *Bacillus pumilus*. هدا يشيرالى ان المورثة ker1 مسؤولة عن انتاج واحد من السيرين بروتياز من السلالة *Bacillus* sp. C_4. الاحماض الامينية الثلاث الموجودة فى جميع انزيمات السبتليزين فى الموقع الفعال المكون لثلاثى التحفيز (Asp-140, His-172 and Ser329) **كانت** محفوظة فى سلسلة الاحماض الامينية الناتجة عن المورثة ker1. **تم ادراج المورثة** ker1 ضمن قاعدة البيانات العالمية (NCBI) تحت رقم القبول KX184831. تم تحديد التتابع النكليوتيدى لجينوم *Bacillus* sp. C_4 وتحليله كطريقة بديلة للبحث عن المزيد من المورثات المسؤولة عن تحليل الكيراتين. وجد ان جينوم *Bacillus* sp. C_4 يتكون من كروموزوم حلقى واحد مؤلف من 3,659,360 **زوج قاعدى بنسبة** G+C **بحدود** % 41.4 و لا وجود للبلازميدات. تم القيام بدراسة الناحية الوظيفية والتركيبية لمسودة تسلسل جينوم البكتيريا *Bacillus* sp. C_4 بواسطة النظامين الاليين BASys و PGAP فى موقع NCBI. اظهرت النتائج ان السلالة *Bacillus sp* C_4 تنتمى الى مجموعة *B. pumilus*. وقد استطاع نظام الكشف على المورثات التعرف على ما مجموعه 3692 مورثة منها 3596 تتابع مشفر للبروتين، 71 مورثة خاصة بـ RNA و 31 مورثة كاذبة. المورثات المتنبأ بها والمشفرة لل RNAتضم 60 مورثة خاصة بـ RNA الناقل (tRNA)، 10 مورثات خاصة بـ RNA الريبزومي (rRNA) ومورثة واحدة للRNA غير معبر عنها. تم تحديد 99.6% من التتابع النكليوتيدى لجينوم *Bacillus* sp. C_4. تم ايداع هدا الجينوم فى قاعدة البيانات GenBank تحت رقم الدخول CP011109 **ورقم المشروع** PRJNA278012**.** واحد وثمانون مورثة مرشحة لان تشفر للبيبتيداز تم ايجادها فى مسودة الجينوم. ستة من هده المورثات تم تصنيفها على انها بيبتيداز S8 مفترضة، من بينها جينين تم التنبؤ بهما كسبتليزين اعتمادا على نظام BASys**.** أولهما المورثة *Ker1* الدى تم عزلها فى هده الدراسة (**رقم الدخول**: KX184831). ثانيهما هى مورثة مشابهة للمورثة *Ker1* وسميت **مورثة ل** *Ker2***.** هده الاخيرة تم ايداعها ضمن قاعدة البيانات GenBank تحت رقم الدخول KX184832. اطار القراءة المفتوح لمورثة *Ker2* لها نفس العدد من القواعد للمورثة Ker1. باستعمال BLASTN على موقع NCBI، وجد ان لهاتين المورثتين 67% نسبة تشابه. سلسلة الاحماض

الامينية المستنتجة للمورثة Ker2 وجد انها تتكون من 383 حمض أميني والثلاث أحماض امينية المكونة لثلاثي التحفيز في الموقع الفعال المميزة لجميع انزيمات السبتليزين كانت كالتالي: Asp-140، His-172 و Ser-329. أظهرت نتائج كل من تفاعل البلمرة المتسلسل، تحديد التتابع النكليوتيدي من جديد واضافة الوظائف لجينات الجينوم ان *Bacillus pumilus* C_4 المنتجة للبروتياز والمحللة للكيراتين تحتوي على الاقل على مورثتين مشفرتان للسيرين بروتياز مثل السبتليزين. المجال الواسع لهذه البكتيريا تجاه عدة ركائز وقدرتها على تحليل كلا النوعين من الكيراتين الفا وبيتا يمكن ارجاعه الى هدا العدد الكبير من المورثات المشفرة لانزيمات البروتياز والبيبتيداز المفترضة.

تم التعرف على بعض انزيمات البروتياز المنتجة من طرف *Bacillus pumilus* C_4 و المنماة على وسط الانتاج بواسطة تقنية كروماتوغرافيا السائلة – التاين بالرداد الالكتروني- قياس الطيف الكتلي المترادف nLC-ESI-MS/MS المتبوعة بالبحث في قاعدة البيانلت باستعمال Mascot. اشارت نتائج البحث باستخدام Mascot ان عينة الانزيم المنقى جزئيا في الحقيقة تتكون من انزيمين موافقين للجينين Ker1 و Ker2 السابق عزلهما. هدا يؤكد مرة اخرى ان *Bacillus pumilus* C_4 تنتج على الاقل انزيمين من البروتياز والدى من الممكن ان يكون لهما علاقة بقدرتها على كسر نوعي الكيراتين.

الكلمات المفتاحية: *Bacillus pumilus* C_4، بروتياز خارج خلوي، سبتليزين، كيراتين، قراءة التتابع النيكليوتيدي للجينوم.

Résumé

L’optimisation de la production de la protéase extracellulaire par la souche*Bacillus*sp. C_4 a été réalisée en utilisant une stratégie à deux étapes; une approche à une variable à la fois (OVAT). Basant sur les résultats d’OVAT, le milieu 1XSG additionné d'extrait de levure à 0,2% était la meilleure formule du milieu pour une production maximale de protéase par *Bacillus* sp. C_4. Une augmentation globale de 1,8 fois la production de la protéase extracellulaire par *Bacillus* sp. C_4 a été obtenue dans le milieu optimisé par rapport à celui non optimisé. Il a été également découvert que dans les conditions optimales, la production de protéase extracellulaire a commencé quatre heures (04 heures) plus tôt que celle produite dans les conditions non optimisées.

Le gène de la protéase kératinolytique (gène *ker1*) a été isolé à partir de *Bacillus* sp. C_4; une souche protéolytique possédant une activité kératinolytique; en utilisant l’amplification par PCR et des amorces spécifiques. Le gène *Ker1* se compose d'un cadre de lecture ouvert de 1, 152 bps, codant pour un polypeptide constitué de 383 résidus d'acides aminés. L'analyse par BLASTP a montré que la séquence d'acides aminés déduite du gène *ker1* a une similarité avec de nombreuses serine protéases provenant de différents *Bacillus pumilus*. Ceci indique que le gène *ker1*code pour l'une des serine protéases de *Bacillus* sp. C_4. Les trois résidus présents dans toutes les subtilisines au site actif formant la triade catalytique (Asp-140, His-172 et Ser-329) ont été conservés dans la séquence d'acides aminés déduite du gène ker1. Le gène *ker1*est ensuite déposédans la banque de données internationale GenBank à NCBI (numéro d'accession: KX184831). Le séquençage et l'analyse du génome de *Bacillus* sp. C_4 a été réalisée comme une approche alternative pour isoler d'autres gènes impliqués dans la dégradation de la kératine par cette souche. Le génome de *Bacillus* sp. C_4 a été construit sous la forme d'un seul chromosome circulaire contenant 3,659,360 pb avec une teneur moyenne en GC de 41,4% et aucun plasmide n'a été détecté.Les résultats de l'annotation indiquent que *Bacillus* sp C_4 appartient au groupe d'organismes *B. pumilus*. L’annotation par le NCBI prédit un total de 3 698 gènes dont 3 596 séquences codant pour protéine (CDS), 71 gènes d’ARN et 31 pseudogènes. Les gènes d'ARN prédits comprennent 60 ARNt, 10 ARNr et 1 ARN non codant (ARNnc). Le génome s'est révélé être complet à 99,6% lorsqu'il a été analysé. Le génome entier de *Bacillus pumilus* C_4 a été déposé à la base de données GenBank sous le numéro d'accession CP011109 et le nombre de bioprojets était PRJNA278012. Quatre-vingt-un gènes candidats pour la peptidase ont été trouvés globalement dans ce génome. Six (06) gènes entre eux ont été annotés comme peptidase S8 dans laquelle deux gènes ont été prédits aussi comme

subtilisines basées sur l'annotation par BASys. Le premier est le gène *Ker1* (numéro d'accession: KX184831) préalablement isolé dans cette étude. Le second est un homologue du gène *Ker1* et a été nommé gène *Ker2*. Ce gène a ensuite été soumis à NCBI GenBank (Numéro d'accession: KX184832).Le cadre de lecture ouvert de ce gène a une quantité de paires de base égale à celledugène *Ker1*. Les deux gènes ont une similarité de séquence de 67% en utilisant BLASTN à NCBI. La séquence d'acides aminés déduite du gène *ker2* de *B. pumilus* C_4 contenait 383 résidus d'acides aminés et les trois résidus au site actif qui forment la triade catalytique dans toutes les subtilisines sont Asp-140, His-172 et Ser-329. Les résultats de l'amplification par PCR, du séquençage *de novo* et de l'annotation du génome de cette bactérie indiquent que *Bacillus pumilus* C_4possède au moins deux gènes codant pour des sérine-protéases de type subtilisine. La large spécificité du substrat et l'activité kératinolytique vis-à-vis des deux types de kératine de cette souche pourraient provenir de la présence de ce nombre élevé de peptidases et de protéases putatives.

L'identification de certaines protéases produites lors de la croissance de *Bacillus pumilus* C_4 sur le milieu optimisé a été effectuée en utilisant nLC-ESI-MS/MS suivie d'une recherche de base de données par Mascot. Les résultats de Mascot indiquent que la fraction de protéase partiellement purifiée se compose en fait de deux enzymes, correspondant aux gènes *Ker1* et *Ker2* précédemment identifiés de la souche de *B. pumilus*C_4. Ceci confirme que *Bacillus pumilus* C_4 produit au moins deux protéases qui pourraient être liées à son activité kératinilytique remarquable envers les deux types de kératines.

Mots clés: *Bacillus pumilus* C_4, protéase extracellulaire, subtilisins, kératins, le séquençage du génome entier.

Table of Contents

Introduction

Introduction

Proteasesalsoknown as peptidyl-peptides hydrolases (E.C. 3.4.21-24 and 99) are one of the most valuable industrial enzyme swhich catalyze the hydrolysis of peptide bonds in a protein molecule. The currentestimated value of the worldwide sales of industrial enzymes is $4.2 billion **(Singh *et al,* 2016).** Proteases represent one of the three largest groups of industrial enzymes and their global market is projected to reach approximately $2.21 billion in terms of value by 2021 at a compound annual growth rate (CAGR) of 6% from 2016 to 2021 (**Industrial Enzyme Market, 2015**). Proteases from microbial origin are accounted for about 60% of the total worldwide enzyme sales, followed by the animal source. Microbial proteases, especially from *Bacillus* sp. have traditionally held the predominant share of the industrial enzyme market of the worldwide enzyme sales with major applications in food and feed, detergent, pharmaceuticals, leather, diagnostics, waste management, silver recovery and molecular biology (**Anisworth, 1994; Outtrup *et al,* 1995; Inhs *et al,* 1999; Saxena and Singh, 2010**).

The overall of enzyme production and downstream processingis a major obstacle against the successful application of any technology in the enzyme industry. In particular, in biotechnical production processes in which even small improvements can be decisive for the commercial success; optimizations on everybody's agenda at present (**Reddy *et al,* 2008; Potumarthi *et al,* 2008**). Researchers and process engineers have used several methods to increase the yields of proteases with respect to their industrial requirements. Selection of appropriate carbon, nitrogen and other nutrients is one of the most critical stages in the development of an efficient and economic process. Conventional experimental approach used for media optimization employs ‘change-one-factor-at-a-time’ method. This method is time consuming and requires large number of experiments to study the effect of individual factors. Response surface methodology (RSM) is one of the popularly used optimization procedures. It is a collection of statistical techniques for designing experiments, building models(**Oh *et al,* 1995**), identifying the effective factors, study interactions, select optimum conditions and quantify the relationships between one or more measured responses and the vital input factors in a limited number of experiments (**Gupta *et al,* 2002a;Ahuja*et al,* 2004**).

Keratinolytic peptidases or keratinases (EC 3.4.21/24/99.11) encompass a distinctive group of proteolytic enzymes; show their complete involvement in hydrolysis of insoluble keratin into simple polypeptides and amino acids **(Gupta and Singh, 2013 and 2014)**. Keratin is the third most

abundant polymer in nature after cellulose and chitin. It is a structural component of skin, hair, feather, wool, horns, hooves, cloves, nails, beaks, reptilian osteoderm, and fish teeth and slime (**McKittrick *et al,* 2012; Lange *et al,* 2016**). Keratin belongs to the intermediate filament proteins that are packed tightly either as a α-helix (α-keratin) or β-sheet (β-keratin) in a super coiled polypeptide chain with a high degree of cross-linking by disulfide bonds, hydrophobic interactions, and hydrogen bonds (**Fraser and Parry 2008; Riffel *et al,* 2011**). Thus, keratin is a very strong and insoluble protein with high resistance to mechanical stress and recalcitrance to common proteolytic enzymes like pepsin, trypsin, and papain **(Riffel *et al,* 2011; Huang *et al,* 2015**).

A large number of microorganisms have been reported to produce keratinases (**Brandelli *et al,* 2010; Gupta and Ramnani 2006; Onifade *et al,* 1998**), and among bacteria, the best studied are organisms from the genus *Bacillus* (**Gobinath *et al,* 2015**). Microbial keratinases have attracted a great deal of attention in the recent decade as special proteases with their extended substrate spectrum as they act on insoluble structural proteins. Today, due to their wide substrate specificity, they are considered as better catalyst for their application in conventional protease sectors like feed, fertilizer, detergent, leather, and pharmaceutical/biomedical industries. Apart from these, they are indispensible for their application in the non-polluting processes such as the treatment of keratin-containing wastes from poultry and leather industry (**Sangali and Brandelli, 2000**). After hydrolysis, the feathers can be converted to feed-stuffs, fertilizer, glues and films or selected amino acids such as serine, cysteine and proline (**Riffel *et al,* 2003**).

Keratinases are mostly serine proteases(**Gupta and Ramnani 2006**), and many bacterial keratinase genes have been sequenced, cloned, and characterized indicating a sequence similarity with the subtilisin family, Family S8, of serine proteases. According to MEROPS, subtilases are grouped under the S8 family of SB clan and are characterized by a catalytic triad of three amino acids, namely aspartate, histidine, and serine. They are further classified into two subfamilies namely subtilisins (S8A) and kexins (S8B) (**Rawlings *et al,* 2008; Jasmin *et al,* 2010**).

The whole Genome sequencing is becoming one of the most widely used applications of Next-Generation Sequencing (NGS) technologies (**Goodwin *et al,* 2016**). It has played a central role in the advancement of the field of life sciences disciplines and has become as common and as straightforward as PCR technique. The complete genome sequence of important industrial organisms provide opportunities for human health, industry, and the environment. *Bacillus* species are the dominant workhorses in industrial fermentations. Today, genome sequences of several

Bacillus species are available, and comparative genomics of this genus helps in understanding their physiology, biochemistry, and genetics **(Sharma and Satyanarayana, 2013**)

This study was an attempt to enhance the extracellular protease production by *Bacillus* sp. C_4 through screening for some keratinolytic protease genes that may contribute to the keratinolytic ability of this bacterium towads feather and wool keratins using *de novo* genome sequencing and protein analysis.

Research Objectives

II. Research Objectives

The objectives of this study are:

i. To enhance the extracellular protease production by *Bacillus* sp. C4 through optimization of medium composition,

ii. To screen for some protease genes related to keratinolytic activity of the bacterium under study using PCR amplification and the whole genome sequencing.

iii. Moreover, to identify the secreted proteases in the optimized medium and to investigate their relationship to the isolated protease genes from *Bacillus* sp. C4 in this study.

Review of Literature

Chapter One

1. Micobial Proteases

1.1. Defenition

Proteases (peptidases or proteolytic enzymes) have been studied extensively since the advent of enzymologist (**Gupta *et al,* 2002b**).They are widely distributed in nature and play a vital role in life processes. Proteases are essential constituents of all forms of life on the earth, including prokaryotes, fungi, plants and animals. They are particularly known for their capacity to hydrolyze peptide bonds in aqueous environments and to synthesize peptide bonds in non-aqueous biocatalysis (**Jaouadi *et al,* 2011**).Proteases from microorganisms are the largest group of industrial enzymes and account for greater than 60% of total global sale of enzymes **(De Souza *et al,* 2015).**Bacterial proteases are the most widely exploited when compared with plant, animal, and fungal proteases **(Breithaupt, 2001; Olajuyigbe and Falade 2014)**.

1.2.Sources of microbial proteases

Although protease-producing microorganisms, plants and animals have cosmopolitan distribution in nature; microbial community is preferred over the others for the large scale production of proteases due to their fast growth, broad biochemical diversity and simplicity of life for the generation of new recombinant enzymes with desired properties (**Jisha *et al,* 2013**). Furthermore, microbial proteins have a longer shelf life and can be stored under less than ideal conditions for weeks without significant loss of activity (**Gupta *et al,* 2002c**).

In general, microbial proteases are extracellular in nature and are directly secreted into the fermentation broth byte producer, thus simplifying downstream processing of the enzyme as compared to proteases obtained from plants and animals. Despite the long list of protease-producing microorganisms, only a few are considered as appropriate producers for commercial exploitation, being ‘generally regarded as safe’ (GRAS), non-toxic and non-pathogenic. A large number of microbes belonging to bacteria, fungi, yeast and actinomycetes are known to produce alkaline proteases of the serine type (**Kumar and Takagi 1999**).

Bacteria are the most dominant group of alkaline protease producers with the genus *Bacillus* being the most prominent source. A myriad of *Bacillus* species from many different exotic environments have been explored and exploited for alkaline protease production but most potential alkaline protease producing bacilli are strains of *B. licheniformis*, *B. subtilis*, *B amyloliquifaciens*, and. *mojavensis* (**Gupta *et al,* 2002b; Kalisz 1988; Kumar and Takagi 1999; Rao *et al,* 1998**).

Another bacterial source known as a potential producer is *Pseudomonas sp*. (**Bayoudh *et al,* 2000; Ogino *et al,* 1999**). Among actinomycetes, strains of *Streptomyces*are the preferred source (**Petinate *et al,* 1999**). In fungi, Aspergilli (**Chakrabarti *et al,* 2000; Rajamani and Hilda 1987**) is the most exploited group, and *Conidiobolus sp.* (**Bhosale *et al,* 1995**), and *Rhizopus sp.* (**Banerjee and Bhattacharyya1993**) also produce alkaline protease. Among yeasts, *Candida sp* .has been studied in detail as a potential alkaline protease producer (**Poza *et al,* 2001**). Despite this interest in other sources, survey of the literature conclusively shows that *Bacillus spp.* are by far the most popular source of commercial alkaline proteases (**Gupta *et al,* 2002b**).

1.3. Classification of proteases

According to the Enzyme Commission (EC) classification, proteases belong to group 3 (hydrolases), and sub-group 4 (which hydrolyse peptide bonds, peptide hydrolasesor peptidases E.C 3.4).However, proteases do not comply easily with the general system of enzyme nomenclature due to their huge diversity of action and structure (**Jisha *et al,* 2013**). Protease scan be separated into two major groups based on their ability to cleave N- or C- terminal peptide bonds (exopeptidases, (E.C 3.4.11-19))or internal peptide bonds (endopeptidases, (E.C 3.4.21-99)).While aminopeptidases cleave the N-terminal peptide linkage, carboxypeptidases cleave the C-terminal peptide bond.

Proteases are also distinguished by the presence or absence of charged groups in positions relative to the susceptible bond and are classified on a number of bases: their pH optima (as acidic (pH 2.0 to 6.0),neutral(pH 6.0 to8.0) or alkaline (pH 8.0 to 13.0)) **(Gupta *et al,* 2002b; Roa *et al,* 1998 ; Sabotic and Kos, 2012, De Souza1 *et al,* 2015**); substrate specificity (collagenase, keratinase, elastase, *etc*.); or their homology to well-studied proteins such as trypsin, pepsin, etc. (trypsin-like,pepsin-like, etc.). Hartley(**Hartley 1960**) classified end proteases into four groups on the basis of their active site and sensitivity to various inhibitors; serine proteases, aspartic proteases, cysteine proteases, and metalloproteases, Table1 (**Sumantha *et al,* 2006**). There are a few miscellaneous proteases which do not precisely fit into the standard classification; ATP-dependent proteases which require ATP for activity (**Menon and Goldberg, 1987; Rao *et al,* 1998**).

Based on their amino acid sequences, Rawlings and Barrett have classified proteases into different families (**Argos, 1987**) which are further subdivided into “clans” to accommodate sets of peptidases that have diverged from a common ancestor (**Rawlings and Barett, 1993**). They are available in the MEROPS database (http://merops.sanger.ac.uk/).Each family of peptidases has been assigned a code letter denoting the type of catalysis, i.e., S, C, A, M, or U for serine, cysteine, aspartic, metallo-, or unknown type, respectively (**Rao *et al,* 1998**).

Table 1: General classification of proteases with their enzyme commission (EC) code, coupled with specific mechanism of action of each subgroup (**Jisha *et al,* 2013**).

Protease	E.C code	Mechanism
Exopeptidases	3, 4, 11-19	cleave the peptide bond proximal to the amino or carboxy termini of the substrate
Aminopeptidases	3, 4, 11	Those acting at a free N-terminus liberate a single amino acid residue
Dipeptidyl peptidase	3, 4, 14	Releases of an N-terminal dipeptide from a polypeptide
Tripeptidyl peptidase	3, 4, 14	Releases of an N-terminal tripeptide from a polypeptide
Carboxypeptidases	3, 4, 16-18	Release of a single residue C-terminal from a polypeptide
Serine type protease	3, 4, 16	Carboxypeptidase have an active centre serine involved
Metalloprotease	3, 4, 17	Carboxypeptidase use a metal ion in the catalytic mechanism
Cysteine type protease	3, 4, 18	Carboxypeptidase have a cysteine in the active centre
Peptidyldipeptidase	3, 4, 15	Release of free C-terminus liberate a dipeptide
Dipeptidases	3, 4, 13	Exopeptidases specific for dipeptides
Omega peptidases	3, 4, 19	Remove terminal residues that are linked by isopeptide bonds
Endopeptidases	3, 4, 21-24	Cleave internal bonds in polypeptide chains
Serine protease	3, 4, 21	Endopeptidases have an active centre serine involved
Cysteine protease	3, 4, 22	Possesses a cysteine in the active centre
Aspartic protease	3, 4, 23	An aspartic acid residue for their catalytic activity
Metalloprotease	3, 4, 24	Usesa metal ion (often, but not always, Zn^{2+}) in the catalytic mechanism
Endopeptidases of unknown catalytic mechanism	3, 4, 99	Acting on peptide bonds

1.3.1. Serine Proteases

Serine proteases are of considerable interest, in view of their activity and stability at alkaline pH; and thus they have applications in a number of industries. These serine proteases have applications detergent industry, in tannery, waste water treatment, silver recovery and resolution of amino acids mixtures **(Sundus *et al,* 2016)**. They have a nucleophilic serine residue located in their active site. Apart from this, these proteases are also distinguished by having essential aspartate and histidine residues which, along with the serine, form the catalytic triad. They are generally active at neutral and alkaline pH, with optima at pH 7–11, although higher pH optima (10.0–12.5) for proteases produced from *Bacillus* sp. YaB have also been reported (**Shimogaki *et al,* 1991, Gupta *et al,* 2002b**).Serine proteases contribute to one third of the share in the enzyme market and are readily inactivated by Phenyl Methane Sulfonyl Fluoride (PMSF).The prototypical group of bacterial serine proteases are sublitisins (subtilases) (**Gupta *et al,* 2002b**). The name of subtilisin derives from the name of the bacterial species, *Bacillus subtilis*, from which the protease was first isolated and the term also now covers proteases isolated from related *Bacillus* species. *Bacillus* spp. are the best known producers of subtilisin (**Rao *et al,* 1998)**. Their amino acid sequence and three dimensional structures can be clearly differentiated from other seine proteases, such as chymotrypsin, carboxypeptidase and peptidase A from *E. coli* (**Kumar *et al,* 2008b**).

1.4.Protease assays

Proteolytic activity can be measured by either qualitative or quantitative methods. Both methods based on the measurement of either the products of protein hydrolysis or of residual protein itself. The available methods for the proteolytic activity measurement vary in their simplicity, rapidity, range of detection and sensitivity.

1.4.1. Qualitative assays rely on the formation of a clear zone of proteolysis on agar plates as a result of protease production. They include:

Protein agar plate assay is commonly used for the initial screening of proteolytic activity. Different protein substrates can be used; skim milk (**Rajamani and Hilda, 1987**), casein, gelatin, bovine serum albumin (BSA) (**Vermelho *et al,* 1996**) and keratin (**Friedrich *et al,* 1999; Gupta *et al,* 2002c**)

Radial (zone) diffusion assay (**Wikstrom *et al,* 1981**) is used for semi-quantitative assessment of protease activity. Protease is detected by observing the zone of hydrolysis around small wells

cut in agar plates containing the appropriate protein substrate. In another method, the Coomassie prestained substrate agarose gel allows direct assessment of protease activity (**Hagen *et al,* 1997**).

Thin-layer enzyme assay (**Wikstrom, 1983**),

The inner surface of a Perti dish is coated with protein and then covered with the agar culture medium. The enzymes, produced during growth of the microorganisms, reach the protein-coated surface. Degradation of the protein was visualized by condensation of water vapor on the surface after removal of the agar medium. The wet ability of the enzyme-affected protein-coated polystyrene surface was decreased compared with the unaffected protein surface. The main advantage of this assay is that it is a sensitive, convenient and inexpensive method, which can be used to select specific microbial protease producers in a mixed sample.

1.4.2. Quantitative assays

The commonly used methods employ natural or synthetic substrates using techniques such as spectrophotometry, fluorimetry, radiometry and enzyme-linked immunosorbent assay-based assays (ELISA).

Spectrophotometric methods

The commonly accepted procedure for estimating the activity of proteases is to determine the quantity of peptides in acid-soluble hydrolyzed product fractions after proteolytic action on a protein substrate (BSA, casein, hammerstein casein, hemoglobin). These peptide residues are estimated either by absorption at 280 nm (direct estimation method) or using conventional Folin's reagent (colorimetric method), (**Gupta *et al,* 2002c**).

Both naturally occurring insoluble proteins, e.g., fibrin, elastin, gelatin, keratin, collagen, or soluble proteins rendered insoluble either by cross-linking with bifunctional agents (**Safarik, 1987a, b, 1989**) or entrapment into appropriate polymermatrix (**Safarik 1988a**), thermally modified substrates (**Safarik 1987c, 1988b**), or synthesized chromogenic substrates using 3,5-dinitro-salicylic acid (**Gallegos *et al,* 1996**) can be used.

Fluorescent oligopeptide energy transfer assay (**Ng and Auld, 1989**),

A fluorescent peptide substrate is designed to explore the protease specificity for the amino acids in the region of the cleavage site (C- and N-terminal). This assay is based on intramolecular quenching of indole fluorescence by an N-terminal dansyl group separated by six amino acid

residues. Although the sensitivity of this method for detection and quantification of specific endoproteases, it is not used much because of the high cost of the assay components (**Gupta *et al,* 2002c**).

Enzyme-linked immunosorbent assay (**Bedouet *et al,* 1998**)

In this essay biotinylated-BSA was used as substrate in polystyrene-coated microtiter plates, and the absorbance was recorded at 405 nm using a microtiter ELISA reader. It is used for detection and quantification of low levels (~0.24 nag/ml) of proteases. ELISA-based assay methods have the limitation that details of the complete three-dimensional structure of the test enzyme must be known before using an antibody against it.

X-ray based method **(Cheung *et al,* 1991)**

It is based on utilizing gelatin on the surface of an unprocessed Kodak X-Omat AR film as a proteolytic substrate. The assay is convenient, rapid and simple. It can be used for a variety of proteases under a wide pH range of 5 to 8.5. The method is useful as a general laboratory procedure.

1.5. Regulation of protease biosynthesis(Gupta *et al,* 2002c)

Protease production is an inherent property of all organisms and these enzymes are generally constitutive; however, at times they are partially inducible (**Beg *et al,* 2002, Kalisz, 1988**). Proteases are largely produced during stationary phase and thus are generally regulated by carbon and nitrogen stress. Further, proteases are known to be associated with the onset of stationary phase, which is marked by the transition from vegetative growth to sporulation stage in spore-formers. Therefore, protease production is often related to the sporulation stage in many bacilli, such as *B. subtilis* (**O'Hara and Hageman, 1990**), and *B. licheniformis* (**Hanlon and Hodges, 1981**).

1.6. Protease fermentation and yield improvement

The overall cost of enzyme production and downstream processing is the major obstacle against the successful application of any technology in the enzyme industry. Researchers and process engineers have used several methods to increase the yields of alkaline proteases with respect to their industrial requirements. Recent approaches for increasing protease yield include screening for hyper-producing strains, cloning and over-expression, fed-batch, chemostat fermentations, and

optimization of the fermentation medium through a statistical approach, such as response surface methodology **(Gupta et al, 2002b)**.

1.7. Fermentation methods

Generally, microbial proteases are constitutive or partially inducible in nature. Extracellular microbial protease production is also strongly influenced by media components, e.g. variation in C/N ratio, presence of some easily metabolizable sugars, such as glucose (**Beg *et al,* 2002**), and metal ions (**Varela *et al,* 1996**). Protease synthesis is also affected by rapidly metabolizable nitrogen sources, such as amino acids in the medium. Besides these, several other physical factors, such as aeration, inoculum density, pH, temperature and incubation, also affect the amount of protease produced (**Hameed *et al,* 1999, Puri *et al,* 2002**). In order to scale up protease production from microorganisms at the industrial level, biochemical and process engineers use several strategies to obtain high yields of protease in a fermentor. Controlled batch and fed-batch fermentations using simultaneous control of glucose, ammonium ion concentration, oxygen tension, pH and salt availability (**Hameed *et al,* 1999; Hubner *et al,* 1993; Mao *et al,* 1992; Van Putten *et al,* 1996**) and chemostat cultures (**Frankena *et al,* 1985, 1986**) have been successfully used for improving protease production for long-term incubations, using a number of microorganisms. In a study, the overall alkaline protease yield from *B. mojavensis* (**Beg *et al,* 2002**) was improved up to 4-fold under semi-batch and fed-batch operations by separating biomass and protease production phases, using intermittent de-repression and induction during the growth of the organism.

1.8. Protease production using statistical design

Researchers and process engineers have used several methods to increase the yields of proteases with respect toothier industrial requirements. Selection of appropriate carbon, nitrogen and other nutrients is one of the most critical stages in the development of an efficient and economic process.

The methodologies used for screening the nutrients fall into two major categories: classical and statistical. The "one variable at-time-approach" is the most frequently used operation to obtain high yields of protease in a microbial system. However, the statistical methodologies are preferred because of various advantages in their use in terms of rapid and reliable short-listing of nutrients, understanding the interactions among the nutrients at varying concentrations and a tremendous reduction in total number of experiments resulting in saving of time, glassware, chemicals and manpower (**Reddy *et al,* 1999**). Response surface methodology (RSM) is one of the popularly used optimization procedures. RSM helps identify the effective factors, study interactions, select

optimum conditions and quantify the relationships between one or more measured responses and the vital input factors in limited number of experiments (**Gupta *et al,* 2002b, Ahuja *et al,* 2004**).

Response Surface Methodology (RSM) is a collection of mathematical and statistical techniques useful for the modeling and analysis of problems in which a response of interest is influenced by several variables and the objective is to optimize this response (**Montgomery, 2005**).

It is possible to separate an optimization study using RSM into three stages. The first stage is the preliminary work in which the determination of the independent parameters and their levels are carried out. The second stage is the selection of the experimental design and the prediction and verification of the model equation. The last one is obtaining the response surface plot and contour plot of the response as a function of the independent parameters and determination of optimum points (**Das and Boyaci, 2007**).

Central composite design (CCD) is the most popular design for fitting a second-order model. A second-order model is useful in approximating a portion of the true response surface with parabolic curvature. This design was introduced by Box and Wilson (**Box and Wilson, 1951**, **Box and Hunter, 1957; Venil and Lakshmanaperumalsamy, 2009**).

For statistical calculations, the variables X_i were coded as x_i according to the following transformation,

$$x_i = (Xi - X_0)/ \Delta X$$

where x_i is the dimensionless coded value of the variable X_i,
X_0 the value of the X_i at the center point,
and ΔX the step change.

The behavior of the system was explained by the following quadratic model:

$$Y = \beta_0 + \sum \beta_i \text{ x } x_i + \sum \beta_{ii} \text{ x } x_i^2 + \sum \beta_{ij} \text{ x } x_i x_j$$

Where Y is the predicted response, β_0 the intercept term, β_i the linear effect, β_{ii} the squared effect, β_{ij} the interaction effect.

Microsoft Excel 97 was used for the regression analysis of the experimental data obtained. Several statistical software packages were also used for the regression analysis of the experimental data, and also to plot the response graphs. The statistical significance of the model equation and the model terms was evaluated *via* the Fisher's test .The quality of fitting of the

polynomial model equation was expressed by the coefficient of determination (R^2) and the adjusted R^2. The multiple linear regression analysis estimate *t*-value, *p*-value and confidence

Table 2: Statistical methods used to improve protease production from some *Bacillus spp.*

Microorganism	Design	Software	Yield improvement	Reference
Bacillus licheniformis	Central composite design	Design-Expert Software	1.4-fold	Lakshmi and Hemalatha, 2016
Bacillusamyloliquefaciens 35s	Plackett-Burman + face-centered central composite design (FCCCD)	Design Expert 8.0.7.1 Stat-Ease	1.6-fold	Nassar *et al,* 2015
Bacillus sp.		n.s.	2-fold	Saxena and Singh, 2010
Bacillus cereus	Plackett-Burman + Central composite design	Minitab 15	n.s.	Rathakrishnan and Nagarajan, 2013
	Plackett-Burman + Central composite design			Kshetri1*et al*, 2016
Bacillus sp. KW2		Design-Expert	4.8-fold	Chandra *et al,* 2015
Bacillus subtilis SHmIIIa	OVAT +Plackett-Burman + Central composite design	Statistical analysis system SAS software v.	37.2-fold	Cheng *et al,* 2012

		9.0		
Bacillus amyloliquefaciens B7	OVAT + Plackett-Burmen	Design Expert 7.1.6	3.92-fold	Khusro, 2015
Bacillus licheniformis strain BIHPUR 0104	OVAT + Plackett-Burman + Central composite design	Design Expert9.0.0	2-fold	Mardina *et al,* 2015
Bacillus licheniformis ATCC12759	OVAT + Box-Behenken	Design Expert 7.1.6	9-fold	
	Plackett-Burman +Face Centered Central Composite Design			

n.s. not specified.

level. The significant level (*p*-value) was determined using the *t*-test. The *t*-test for any individual effect allows an evaluation of the probability of finding the observed effect purely by chance. If this probability is sufficient small, the idea that the effect was caused by varying the level of the variable under test is accepted. Confidence level is an expression of the *p*-value in percent (**Reddy *et al,* 2008; Ghaemi *et al,* 2008)**.

The yield of a number of proteases from different spieces of *Bacillus* have been improved using RSM and a summary of various production strategies adopted for enhanced protease production is presented in Table 2.

1.9. Microbial protease applications

Generally microbial proteases have a large variety of applications, in various industries. These include food industries, detergent, pharmaceutical, cosmetic industries and have been widely commercialized by various companies throughout the world. The applications of these enzymes vary considerably.

1.9.1. Food and feed industry

The use of proteases in the food industry dates back to antiquity (**Rao *et al,* 1998**).The principal applications of proteases in food processing are in brewing, cereal mashing, and beer haze clarification, in the coagulation step in cheese making, in altering the viscoelastic properties of dough in baking and in production of protein hydrolysates (**Ward, 2011**).The hydrolytic quality of proteases is exploited for degradation of the turbidity complex resulting from protein in fruit juices and alcoholic liquors, the improvement of quality of protein-rich foods, soy protein hydrolysis, gelatin hydrolysis, casein and whey protein hydrolysis, meat protein recovery, and meat tenderization (**Kumari *et al.,* 2012; Tomar *et al.,* 2008**).

The major application of proteases in dairy industry is in the cheese manufacturing, where the primary function of enzymes is to hydrolyze the specific peptide bond to generate casein and macropeptides (**Rao *et al,* 1998**). Protein hydrolysates commonly generated from casein, whey protein and soy protein have applications as constituents of dietetic and health products, in infant formulae, clinical nutrition supplements, beverages targeted at pregnant/lactating women and people allergic to milk proteins, and as flavoring agents (**Kumar and Takagi, 1999; Ramamurthy *et al.,* 1991; Rao *et al.,* 1998, De Souza *et al,* 2015**).

Keratinolytic activity of alkaline protease has also been exploited in the production of proteinaceous fodder from waste feathers or keratin-containing materials. Dalev, 1990 and1994 and Cheng *et al,* 1995 reported the use of alkaline proteases (B72 from *B. subtitles* and *B. licheniformis* PWD-1) for the hydrolysis of feather keratin, to obtain a protein concentrate for fodder production (**Gupta *et al,* 2002b**).

1.9.2. Detergents industry

The use of enzymes as detergent additives represents the largest application of industrial enzymes. Proteases in laundry detergents account for approximately 25% of the total worldwide sales of enzymes (**Demain and Adrio, 2008**). The use of enzymes in detergents formulations enhances the detergents ability to remove tough stains and making the detergent environmentally safe. Nowadays, many laundry-detergent products contain cocktails of enzymes including

proteases, amylases, cellulases, and lipases (**Hmidet *et al,* 2009**). Alkaline proteases added to laundry detergents enable the release of proteinaceous material from stains (**De Souza *et al,* 2015**).

1.9.3. Leather industry

The major building blocks of skin and hair are proteinaceous. The conventional methods of leather processing involve hazardous chemicals such as sodium sulfide, which create problems of pollution and effluent disposal(**Rao *et al,* 1998**).Alkaline proteases with elastolytic and keratinolytic activity can be used in leather-processing industries (**Gupta *et al,* 2002b**).Increased usage of enzymes for dehairing and bating not only prevents pollution problems, but also is effective in saving time with better quality leather (**Zambare *et al,* 2011**).In addition, studies carried out by different workers have demonstrated the successful use of alkaline proteases in leather tanning from *Aspergillus flavus*, *Streptomyces* sp.,*B. amyloliquefaciens* and *B. subtilis* (**Schechler and Berger, 1967)**.Novo Nordisk manufactures three different proteases, Aquaderm, NUE, and Pyrase, for use in soaking, dehairing, and bating, respectively (**Rao *et al,* 1998**).

1.9.4. Silver recovery

Alkaline proteases are used in silver recovery from used X-ray films. Used X-ray film contains approximately 1.5% to 2.0% (by weight) silver in its gelatin layers. The silver recovery by burning film causes a major environmental pollution problem; hence the enzymatic hydrolysis of the gelatin layers on the X-ray film enables the recycling of both silver and poly-ester film base (**Debette, 1991)**.

1.9.5. Silk degumming

Sericin protein, which is about 25% of the total weight of raw silk, covers the periphery of the raw silk fibers, thus providing the rough texture of the silk fibers (**Beg *et al,* 2002b**).Silk processing from cocoons to the finished clothing articles consists of a series of steps which include: reeling, weaving, degumming, dyeing or printing, and finishing. Degumming is a key process during which sericin is totally removed and silk fibers gain the typical shiny aspect and soft handle (**Freddi *et al,* 2003**).This sericin is conventionally removed from the inner core of fibroin by conducting shrink-proofing and twist-setting for the silk yarns (**Gupta *et al,* 2002b**). The process is generally expensive and therefore an alternative method suggested is the use of enzyme preparations, such as protease, for degumming the silk prior to dyeing (**Freddi *et al,* 2003**).

1.9.6. Pharmaceutical and medical usage

The wide diversity and specificity of proteases are used to great advantage in developing effective therapeutic agents. Oral administration of proteases from *Aspergillus oryzae* (Luizymand Nortase) has been used as a digestive aid to correct certain lytic enzyme deficiency syndromes. Clostridial collagenase or subtilisin is used in combination with broad-spectrum antibiotics in the treatment of burns and wounds. An asparginase isolated from *Escherichia coli* is used to eliminate asparagines from the bloodstream in the various forms of lymphocytic leukemia (**Rao *et al,* 1998**).

Collagenolytic proteases have been directly employed in clinical therapy, includes wound healing, treatment of sciatica in herniated intervertebral discs, treatment of retained placenta, and as a pretreatment for enhancing adenovirus mediated cancer gene therapy (**Watanabe, 2004; De Souza *et al,* 2015**).

Chapter Two

Microbial keratinases

2.1. Definition

Keratinases/keratinolytic enzymes are special proteases which attack hard to degrade, insoluble keratin substrates. They are robust enzymes displaying a great diversity in their biochemical characteristics. They stand apart from the conventional proteases due to their broad substrate specificity towards a variety of insoluble, keratin rich substrates (**Gupta and Ramnani, 2006; Gupta *et al,* 2013a**).

2.2. Keratins, structure and sources

Keratin is a fibrous and recalcitrant structural protein and is the third most abundant polymer in nature after cellulose and chitin. A wide spectrum of animals (mammals, fish, birds, and reptiles) has developed diversified keratin as a structural part of their outer protection. Keratin is a structural component of skin, hair, feather, horns, hooves, cloves, nails, beaks, reptilian osteoderm, and fish teeth and slime (**McKittrick *et al,* 2012; Lange *et al,* 2016**).

In general, keratins can be classified as Type I (small; 40-56.5 KDa and acidic) or Type II (53-67 KDa and more basic) (**Kopan and Fuchs, 1989**). The structural keratinaceous proteins are recalcitrant polymers. The recalcitrance is due to properties such as a high degree of cross-linking by disulfide bonds, hydrogen bonds, and hydrophobic interactions. Based on their secondary structure, keratins are classified into two major types, α-keratin and β-keratin. β-Keratin is rich in β-pleated sheets (**Meyers *et al,* 2008**) and is found in nails, claws, shells and beaks of animals (**Wang *et al,* 2016b**). α-Keratin consists of α-helical-coil structured strands which are self-assembled into intermediate filaments (**McKittrick *et al,* 2012; Meyers *et al,* 2008**), Figure 1 and 2.

Post-translational modifications of keratin, such as the formation of disulfide bonds, phosphorylation, and glycosylation, can result in diverse types of modified keratin filaments (**Yamada *et al,* 2002**). The different keratin characteristics give different degrees of bioaccessibility. Almost all keratinaceous materials (such as feathers, hair, bristles, and wool)

possess a mixture of keratins including both α-keratin and β-keratin. Ng *et al*, 2014 reported that α- and β-keratins are preferentially expressed in different feather parts. It was found that feathers are composed of 41–67% α-keratins, 33–38% β-keratin, and also amorphous keratin (**Barone *et al,* 2005; Fraser and Parry, 2008**). Other keratinaceous materials, such as hair, bristle, and wool, consist mostly of α-keratins (50–60%), matrix proteins (keratin-associated proteins located in the amorphous space around the intermediate filaments) (20–30%), and also minor amounts of β-keratins (**Daroit and Brandelli, 2014**). β-Keratin is more accessible for degradation by some keratinases than α-keratin because β-keratin has less disulfide bonds and exhibits the fibril and porosity structure(**Gupta and Ramnani, 2006**). Hairs and feathers belong to what is called hard keratins, based on their function, regulation, and high content of cysteine (**Daroit and Brandelli, 2014**). These hard keratins have diversified morphological structures and numerous disulfide bonds. This makes them insoluble in water, in weak acid and alkaline solutions and in organic solvents, and hard keratins are also resistant to degradation by most protease treatments.

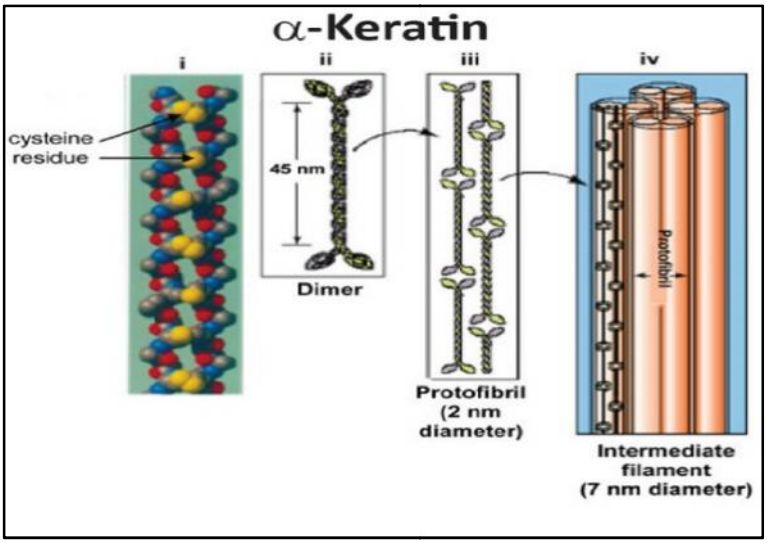


Figure1: Hierarchy of α-keratin showing the assembly from two polypeptide chains (i) to a fibrous structure (iv), (**Naleway *et al,* 2016**).

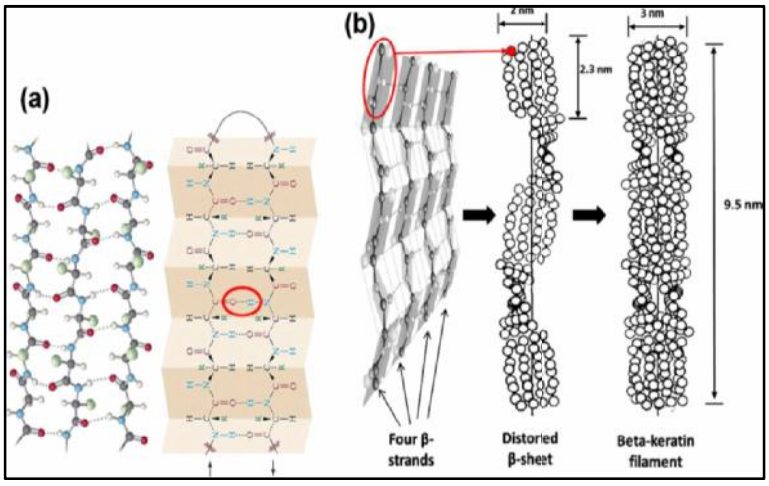


Figure2: Structure of the beta-keratin filaments. b) one polypeptide chain folds to form four b-strands which twist to form the distorted b-sheet. Two sheets assemble to form a beta-keratin filament (**Wang *et al,* 2016)**.

2.3. Sources of Microbial Keratinases

Microbial keratinases are elaborated by a compendium of microorganisms like bacteria, actinomycetes, and fungi. Keratinase producers are mainly isolated from keratinous wastes or keratin-rich areas such as poultry wastes (**Brandelli *et al,* 2010**), slaughterhouse wastes (**Fakh-fakh *et al,* 2010**), tannery wastes (**Khardenavis *et al,* 2009**), and other sources **(Gupta *et al,* 2013b)**, but they are widely distributed in nature and have been isolated from a variety of habitats, such as, antarctic soils, soybean wastes, Mediterranean Sea, solfataric muds, polluted rivers, and hot springs. Diverse source of isolation like Amazon Basin has also been explored for keratinolytic microbes by Giongo *et al,*2007; Correa *et al,* 2010; Daroit *et al,*2009.Among these, keratinolytic fungi fall in the category of dermatophytic microorganisms as they attack α-keratin and are not regarded as safe (**Gupta and Ramnani, 2006**). Therefore, here only diversity of keratinolytic bacteria and actinomycetes of biotechnological relevance is considered.

From the literature, it can be observed that among bacteria, strains of *Bacillus* species are the most prominent feather degraders and, hence, keratinase producers (**Gupta and Ramnani 2006; Zhang *et al,* 2009; Brandelli *et al,* 2010; Prasad *et al,* 2010**). Diverse strains of *Bacillus licheniformis* and *Bacillus subtilis* have been described for feather degradation and keratinase production (**Brutt and Ichida, 1999;Evans *et al,* 2000; Suh and Lee, 2001; Macedo *et al,* 2005; Suntornsuk *et al* ,2005;Hossain *et al,* 2007; Kumar *et al,*2008a; Cai and Zheng, 2009; Fakhfakh *et al,* 2009; Matikeviciene *et al,* 2009; Rai *et al,* 2009; Kumar *et al,* 2010; Desai *et al,* 2010; Linand Yin, 2010; Mazotto *et al,* 2010**). Other *Bacillus* sp., like *B. pumilus*, *B. cereus*, *B. amyloliquefaciens*, *B. halodurans*, *B. thuringiensis*, *B. megaterium*, and *B. pseudofirmus*, have also been reported for keratinase production (**Kojima *et al,* 2006; Adigüzel *et al,* 2009; Cortezi *et al,* 2008; Park and Son, 2009; Fakhfakh *et al,* 2010; Infante *et al,* 2010; Lateef *et al,* 2010; Nagal and Jain, 2010**).

Besides these *Bacillus* sp., many other bacteria like *Thermoanaerobic sp.*(**Riessen and Antranikian 2001; Kublanov *et al,* 2009**), *Lysobacter* (**Allpress *et al,* 2002**), *Kocuria* (**Vidal *et al,* 2000; Bernal *et al,* 2003**), *Microbacterium* (**Thys et al, 2004**), *Nesterenkonia* (**Bakhtiar *et al,* 2005**), *Serratia* (**Khardenavis *et al,* 2009**), *Chryseobacterium* (**Wang *et al,* 2008; Silveira *et al,* 2010**), *Stenotrophomonas*(**Yamamura *et al,* 2002; Jeong *et al,* 2010**), *Pseudomonas* (**Tork *et al,* 2010**), and *Vibrio* (**Sangali and Brandelli, 2000**) have also been reported for feather degradation and keratinase production.

Among actinomycetes, mainly *Streptomyces* spp. has been reported as a keratinolytic microbes. These have been largely isolated from poultry wastes, and few have also been reported from various habitats like antarctic soil, Mediterranean Sea, and Lonar Lake (**Chitte *et al,* 1999; Szabo *et al,* 2000; Gushterova *et al,* 2005; Esway, 2007; Jaoudi *et al,* 2010**). Highly keratinolytic actinomycetes strains, *Streptomyces flavis* 2BG and *Microspora aerate* IMBAS-11A, were isolated from antarctic soil (**Gushterova *et al,* 2005**). Thermophilic sp. like *Streptomyces gulbargensis* (**Syed *et al,* 2009**), *Streptomyces thermoviolaceus* (**Chitte *et al,* 1999**), and *Streptomyces thermonitrificans* (**Mohamedin, 1999**) were isolated from soil.

It can, thus, be put forth that keratinolytic microorganisms are widespread to a variety of environment and habitat; however, generally they are selected on the basis of their capability to degrade keratin in the form of chicken feather, wool, horn, or hove among them chicken feather is the most preferred substrate. Till date, strains of *Bacillus* species are predominant among bacteria and *Streptomyces* sp. among actinobacteria (**Gupta *et al,* 2013c**).

Most published work on industrial processes for degradation of keratin has focused on commercializing and using bacterial keratinases for decomposition of keratinaceous materials (**Table 3**). However, it appears to be very difficult or maybe impossible to find single-component bacterial enzymes that can do the job alone (**Lange *et al,* 2016**). All bacterial enzymes in Table 3except the last one are from *Bacillus* species and classified as serine end peptidases of the subtilisin-type belonging to the S8 protease family with a preference to cleave after hydrophobic residues.

Table 3: Current commercial keratinolytic proteases (**Lange *et al,* 2016**).

Product name	**Source of enzyme**	**Enzyme function EC number**	**Merops protein family**	**Substrates of enzyme product**	**Example of trade name and provider**
Protease P4860	*Bacillus licheniformis*	3.4.21.62	S8	Chicken (leg) bone protein	Alcalase, Novozymes A/S
Protease P5860	*Bacillus sp.*	3.4.21.62	S8	Keratin	Esperase, Novozymes A/S
Protease P3111.	*Bacillus sp.*	3.4.21.62	S8	Keratin	Savinase, Novozymes A/S.

Versazyme	*Bacillus licheniformis*	3.4.21.62	S8	Simple and complex vegetable and animal proteins, feather	Versazyme, BioResource International, Inc.
Prionzyme	*Bacillus licheniformis*	3.4.21.62	S8	Prion, keratin	Prionzyme, Genencor International, Inc.
Proteinase k	*Tritirachium album*	3.4.21.64	S8	Prion, keratin	Proteinase K, New England Biolabs

2.4. Mechanism of Keratin Degradation

Keratin hydrolysis by microorganisms is supposed to be assisted by keratinases and requires additional involvement of reducing agent for breakdown of disulfide bonds. Release of thiol groups has been often observed during the microbial growth on keratinous substrate which supports the essential role of reduction of disulfide bonds during keratin hydrolysis (**Daroit *et al,* 2009**).The definitive mechanism(s) and exact nature of keratinolysis are highly complex and not yet fully understood. Production of intracellular and/or extracellular disulfide reductase (**Yamamura *et al,* 2002; Kumar *et al,* 2008a**), sulfite and thiosulfate release, and cell-bound redox system (**Gupta *et al,* 2013c**) also leads to sulfitolysis.

In case of fungi, mechanical pressure occurs at keratinous substrate during mycelia penetration (**Gupta *et al,* 2013c**). Thus, it can be inferred that sulfitolysis and proteolysis may be involved in keratin degradation. Till date, none of the purified keratinase could degrade keratin completely.

Keratin hydrolysis using purified keratinase was observed only in the presence of reducing agents, which promote sulfitolysis (**Gradisar *et al,* 2005**). Hydrolysis of disulfide bonds changes the confirmation of keratin and exposes more sites for keratinase action leading to keratin degradation. Thus, it can be concluded that multiplicity of catalytic mechanisms are observed involving a variety of microbial keratinases, i.e., serine, thiol, or metalloproteases.

However, keratin hydrolysis in nature is still a result of synergistic action of keratinolytic microbes and their enzymes and metabolites which hydrolyze recalcitrant protein in comparison to pure enzyme (**Gupta *et al,* 2013c**).

2.5. Biochemical Properties of Keratinases

The properties of microbial keratinases may be diverse depending on the producer microorganism. These enzymes are predominantly extracellular, although cell-bound and intracellular enzymes have been described (**Gupta and Ramnani, 2006; Brandelli *et al,* 2010**).

The reaction conditions, especially optimum pH and temperature, varies considerably. Keratinases are usually active in alkaline conditions and reported pH optima ranges from pH 8.0 to 10.0 **(Khardenavis *et al,* 2009; Allpress *et al,* 2002; Schrooyen *et al,* 2001; Ulfig, 2003)** with

exception of acidic pH optima which is usually found in case of keratinases from fungal sources and extremely basic pH optima also exists **(Tiwary and Gupta, 2010; Hassan *et al,* 2013)**. Nam and coworkers **(Wang *et al,* 2004)** reported that keratinase from *Fervidobacterium islandicum* AW-1 with a temperature optima of 100°C, is one of the most thermostable keratinases reported to date. Temperature optima of keratinases range from 40 to 80°C, as reported by various workers **(Son *et al,* 2008; Allpress *et al,* 2002; Schrooyen *et al,* 2001; Kannahi and Ancy, 2012; Ulfig *et al,* 2003; Wang *et al,* 2015)**.

Keratinases are mostly serine type proteases with a serine residue at their catalytic site and showing characteristic PMSF inhibition **(Bressollier *et al,* 1999; Tiwary and Gupta, 2015; Kannahi and Ancy, 2012; Ulfig, 2003)**. Metalloproteases and thiol proteaseshave also been reported by various workers (**Tiwary *et al,* 2010; Sivakumar*et al,* 2013; Grazziotin*et al,* 2008; Park and Son, 2009)**. Metal ions are usually required by the enzymes for their structure stabilization as well as proper substrate binding at the active site along with transition state stabilization (**Verma *et al,* 2016**).

The molecular masses of several keratinases have been determined. Despite they range from 18 to 240 kDa for *Streptomyces albidoflavus* (**Bressollier *et al,* 1999**) and *Kocuria rosea* (**Bernal *et al,* 2006**), respectively, most keratinases have less than50 kDa **(Brandelli *et al,* 2010**).

Keratinases show broad substrate specificity and are reported to be active against both soluble and insoluble proteinaceous substrates. They have the unique ability to cleave complex, insoluble substrates such as feather keratin, collagen, elastin, fibrin, and nail which are otherwise resistant to degradation by conventional proteases (**Evans *et al,* 2000**). Keratinases efficiently hydrolyze soluble protein substrates like casein, azocasein, bovine serum albumin, gelatin, and insoluble proteins like azokeratin, collagen, elastin, feather, horn, hair, nail, silk, stratum corneum, wool, and skin. Study on synthetic *p*- nitroanilide substrates indicated that keratinase preferably cleaves at hydrophobic and aromatic amino acids at P1 position **(Bressollier *et al,* 1999; Evans *et al,* 2000, Gradisar *et al,* 2005; Brandelli *et al,* 2010; Gupta *et al,* 2013c)**.

Biochemical properties of keratinases have been extensively studied by many workers (**Gupta and Ramnani, 2006; Brandelli *et al,* 2010)** highlighting the industrial potential related to keratinase research (**Verma *et al,* 2016**).

2.6. Molecular characterization of keratinases

Till date, many keratinase genes from different bacteria such as *Bacillus licheniformis* PWD-1*kerA*, *Bacillus licheniformis* OWU 1411T (Accession number; Acc. no AF282893), *Bacillus licheniformis* RG1 (Acc no. AY590140), *Bacillus licheniformis* RG2 (Acc no. AY817143), *Bacillus licheniformis* MKU2 (Acc no. DQ071569), *Bacillus licheniformis* MKU3 (Acc no. DQ071570), *Bacillus subtilis aprA* (**Zaghloul, 1998**), *Bacillus mojavensis* (Acc no. AY665611), *Bacillus pumilus* (ACM477351), *Bacillus pumilus* KS12 (HM219183), *Bacillus* sp. JB99 (ADD644651), *Bacillus licheniformis* RPk (ACA979911), *Bacillus cereus* (AAR192201), *Fervidobacterium pennivorans* (**Kluskens *et al,* 2002**), actinomycetes *Nocardiopsis* sp. TOA-1 NAPase (**Mitsuiki *et al,* 2004**), *Streptomyces fradiae* (**Li *et al,*2007**), and *Pseudomonas aeruginosa* (**Lin *et al,* 2009**) have been cloned and sequenced (**Gupta *et al,* 2013c**) and many of them have been successfully expressed in various heterologous hosts (**Gupta *et al,* 2013a**).The nucleotide sequences of enzymes serve as a prelude to their phylogenetic analysis and assist in deciphering structure-function relationships.

The N-terminal sequences of a number of keratinases have been comprehensively analysed and reviewed by Gupta and Ramnani (2006) and Brandelli *et al,* 2010. Depending on the microbial source, keratinases produced by each class and group share a high degree of similarity in their N-terminal sequences. Most of characterized keratinases isolated from the *Bacillus* sp. found to be subtilisin-like proteases, belonging to the serine proteases(S8 family) and have very high (>90%) N-terminal sequence homology with the subtilisin Carlsberg produced by *B. licheniformis* (**Brandelli *et al,* 2010; Lange *et al,* 2016**).Further analysis of some keratinases by multiple sequence alignment revealed that The catalytictriad of Asp191, His 252 and Ser 441 are totally conserved throughout all the diverse keratinases along with high level of conservation in the oxyanion hole region (GVVVVAAAGN), (**Gupta *et al,* 2013a**).

2.7. Applications of microbial keratinases

Keratinases from microorganisms have attracted a great deal of attention in the recent decade as special proteases with their extended substrate spectrum as they act on insoluble structural proteins. Today, due to their wide substrate specificity, they are considered as better catalyst for their application in conventional protease sectors like feed, fertilizer, detergent, leather, and pharmaceutical/biomedical industries. Apart from these, they are indispensible for their application in keratin waste management especially chicken feather wastes. Keratinases are special proteases which act on structural proteins.

2.7.1. Recycling of keratinous wastes

Keratinous wastes are a major by-product of poultry, slaughterhouse, leather- and fur-processing industries, and are abundantly generated in various forms such as feather, hair, horn, hoof, nails, claws, wool, and bristles (**Kornillowicz-Kowalska and Bohacz 2011, Karthikeyan *et al,* 2007; Gupta *et al,* 2013b**).Among these, feather constitutes a major portion as poultry wastes with8.5 million metric tonnes of poultry waste was produced worldwide annually; India contributes about 3.5 million tonnes (**Gupta *et al,* 2013a**), the United State 1.8 million tonnes and the United Kingdom 1.5 million tonnes(**Okoroma *et al,* 2012;Purchase, 2016**).

These feathers are generally land filled or burnt which cause environmental problems (**Vasileva-Tonkova *et al,* 2009**). Feathers are made up of more than 90% protein and rich in hydrophobic amino acids and essential amino acids like cysteine, arginine, and threonine (**Coward-Kelly *et al,* 2006**). Nowadays, feathers are converted into feather meal and used as poultry feed, cattle feed, fish feed, etc. (**Brandelli *et al,* 2010**). Most popular method of conversion of feathers is by hydrothermal process where feather is cooked under high pressure at high temperature (**Onifade *et al,* 1998; Gousterova *et al,* 2003; Gupta and Ramnani, 2006**). This hydrothermal treatments result in the destruction of essential amino acids like methionine, lysine, tyrosine, and tryptophan and produced feather meal that has poor digestibility and low nutritional value (**Papadopolous *et al,* 1986; Wang and Parsons, 1997**). In the last decade, bioconversion of feather into feather meal using feather degraders has gained importance. These microbes offer considerable opportunities for bioconversion of poultry feathers from a potent pollutant to a nutritionally upgraded protein-rich feedstuff for livestock (**Onifade *et al,* 1998**). Bioconversion of feather into feather meal and improvement as poultry feed has been studied extensively (**Onifade *et al,* 1998; Grazziotin *et al,* 2006**). Biodegradation of feather can be achieved either by crude culture filtrate containing keratinases or by cultivation of keratin-degrading microorganism (**Gupta and Ramnani, 2006**).

2.7.2. Production of Fertilizer for organic farming

Composting of keratinous wastes utilizing keratinolytic microorganisms results in slow release of a nitrogen fertilizer and hence can be employed in organic farming. This has emerged as one of the safest and most cost effective technology to employ the keratinous waste. Being nitrogen rich, keratinous wastes can be used as an attractive and economically feasible co-substrate for compost. This can be used as rich nitrogen supplement to support plant growth, promote plant microbe

interactions, improve soil texture, enhance soil health status, and prevent soil erosion (**Verma *et al,* 2016; Gupta *et al,* 2013c**).

2.7.3. Leather and Textile Industry

Keratinolytic enzymes not showing collagenolytic and having mild elastolytic activities are being increasingly explored by the leather industry, particularly for dehairing processes, representing a suitable alternative to the conventional tannery processes which utilize sulfide (**Friedrich *et al,*2005; Anbu *et al,* 2005; Macedo *et al,* 2005; Giongo *et al,* 2007; Wang *et al,* 2007; Prakash *et al,* 2009**). In the traditional process, hair is gelatinized and converted into a pulp, whereas in the enzymatic process the hair remains intact (**Pillai and Archana, 2008**). Therefore, the utilization of keratinolytic enzymes might contribute to the production of high-quality leather, also resulting in improvement of wastewater quality and reduced pollution (**Riffel *et al,* 2003; Giongo *et al,* 2007**).In this respect, many keratinases from *Bacillus halodurans* PPKS2 (**Prakash *et al,* 2009**), *Bacillus halodurans* JB99 (**Shrinivas and Nayak, 2011**),*Bacillus cereus*MCM-B-326 (**Nilegaonkar *et al,* 2007**), and *Paenibacillus woosongensis* TKB2 (**Paul *et al,* 2013**) were demonstrated for dehairing of goat and buffalo skin (**Gupta *et al,* 2013c**).

Raw silk requires degumming to remove a fibrous protein, sericin that cements the fibroin fibers together, so as to provide the fibers luster and soft feel. This process is also important for subsequent dyeing. Enzymatic treatments by way of proteases such as Degummase, Papain, Trysin, Pepsin, Alcalase, Savinase, Protease A Amano, Protease N Amano, Protease M Amano, and Palkobate are now in focus over the conventional methods (**More *et al,* 2013, Sumana *et al,* 2013**). However, most of the proteases are characterized by a low degree of specificity towards sericin (**Freddi *et al,* 2003**) (**Gupta *et al,* 2013b**). Thus, enzymes with better specificity are the need of the hour and keratinases with their wide substrate range can be exploited for degumming of silk and finishing textile fibers like wool. They finish the fibers and increase smoothening, shining, and dying capacity (**Riessen and Antranikian, 2001; Cao *et al,* 2009**).

2.7.4. X-ray film reutilization

X-ray films, which are widely utilized in research and medical services, are made up of polyterephthalate (PET) and has emulsion based coating on both sides which is responsive to light. In the coating of the X-ray films are metallic silver impregnated in a gelatin layer. After

radiography the used photographic/X-ray films are the cause of solid waste problems because most of them still contain the metallic silver and the plastic PET base.

Traditional ways of silver recovery from used X-ray films involves incineration and acid treatment. In the conventional method of incineration, silver is recovered but the PET base does not and the process is costly and forms secondary pollutants such as smoke and soot. Similarly, the acid treatment involves the leaching of a gelatin layer with strong acid solution which leads to the problem of the acid effluent produced after silver recovery. Contrarily, keratinolytic enzymes offer an eco-friendly method of silver recovery from X-ray/photographic films because they can use gelatinas a substrate thus resulting in release of metallic silver in the reaction solution and the undistracted PET base **(Gradisar *et al,* 2005; Han *et al,* 2012; Verma *et al,* 2016)**.

2.7.5. Detergent Industry

Proteases are widely utilized in the detergent industry as safe alternative replacing harmful chemicals such as caustic soda. Nowadays, many of the top brands in the detergent industry contain proteases as a key intermediate (**Nescient and martin, 2006**). Keratinolytic proteases have an upper handover general proteases as they can act on stained surfaces such as shirt collars and sleeve heads more efficiently (**Jeong *et al,* 2010**) Various studies demonstrated the ability of keratinases for detergent stability and compatibility as well as their cleaning potential for different stains (**Gradisar *et al,* 2005; Park *et al,* 2009; Paul *et al,* 2014; Zarai Jaouadi *et al,* 2015)**. Apart from use in washing detergents, keratinases are also used as an additive in lens cleaning solutions and for drain cleaning etc (**Verma *et al,* 2016**).

2.7.6. Biological control

The potential for keratinases to act as a biological control agent has been explored by several research groups recently. Keratinase produced by *Stenotrophomonas maltophilia* R13 is effective against several fungal pathogens including *Fusarium solani*, *F. oxysprum*, *Mucor* sp. And *Aspergillus niger* that cause diseases in valuable plants and crops (**Jeong *et al,* 2010**). Similarly, keratinase produced by *Thermoactinomyces* also showed antifungal properties (**Gousterova *et al,* 2012**). Yue *et al*, 2011 reported that the keratinase produced by *Bacillus* sp. 50-3 has the ability to work effectively against agricultural pests such as root-knot nematodes (*Meloidogyne incognita*).

2.7.7. Biofuel Production

Poultry waste can be utilized as a source of bioenergy. Keratinolytic microorganisms and their keratinases cane utilized for generation of natural gas **(Puhl *et al,* 2009)** methane gas fuel pellets (**Gushterova *et al,* 2005)**, and biohydrogen production (**Saha *et al,* 2013**). Conversion of keratinous byproducts into fuels may address the emerging interest for energy conservation and recycling. Poultry waste can be utilized for biodiesel production and the remaining protein rich feather meal makes an excellent animal feed **(Anbu *et al,* 2007**). This will result in the creation of an extra economic benefit to poultry farmers. This work can be much more elaborated by utilizing keratinolytic microorganisms for fermentation of the resultant feather meal after biodiesel production for the production of keratinase in a cost effective manner and the further spent fermented material can act as slow release nitrogenous fertilizer (**Verma *etal,* 2016**).

Chicken feather waste pre-treated with a recombinant *B. megaterium* strain showing keratinase activity prior to biogas production, was able to produce methane in the order of 0.35 Nm^3/kg dry feathers, corresponding to 80% of the theoretical value on proteins (**Forgács *et al,* 2011, 2013**).

2.7.8. Cosmetics/Personal Care Products

Hair comprises mainly of keratin protein (90%) and a small amount of lipid (1–9%). Keratin hydrolysates are efficient restorers in hair care processes, they contain active peptides that repair and condition the hair (**Villa *et al,* 2013**). Most keratin hydrolysates for hair care products are obtained from nails, horns and wool via chemical hydrolysis and hydrothermal methods (**Barba *et al,* 2008**). However, using microbial keratinases to obtain keratin hydrolysis is also gaining popularity. Crude chicken feather hydrolysate produced by *S. maltophilia* is found to be protective to hair, as evidenced by the improved flexibility and strength for both normal and damaged hair (**Cao *et al,* 2012**). Villa *et al,* 2013successfully formulated a mild shampoo and a rinse off conditioner with the enzymatic hydrolysate which appeared to increase the brightness and softness of hair.

Keratinases also found applications in other personal care products (**Gupta *et al,* 2013b**) including: cosmetic skin whitening and bleaching (**Yang, 2012**), exfoliation and removal of stratum corneum (**Ding and Sun, 2009**), removal of corns and calluses (**Encarna and Elena 2011**), treatment of acne (**Spyros, 2003**) due to the build-up of sebum caused by blockage of hair-shafts by excess keratin, and anti-dandruff shampoo (**Selvam and Vishnupriya, 2012**). Proteos Biotech produces two types of commercial products: *Keratoclean® Hydra PB* and *Pure100 Keratinase,* for

the removal of corns and calluses; and *Keratoclean Sensitive PB* and *Keatopeel PB* for the treatment of acne (**Gupta *et al,* 2013b; Lange *et al,* 2016**).

2.7.9. Medical/Pharmaceutical Applications

2.7.9.1. Enhanced Drug Delivery

The two most common diseases affecting the nail unit are onychomycosis (fungal infections of the nail plate and/or nail bed) and psoriasis (an immune-mediated disease causing nail pitting and onycholysis detachment of the nail from the nail bed) (**Murdan, 2002**). The nail plate consists mainly of 80% 'hard' keratin and 20% soft keratin (**Lynch *et al,* 1986**). In order to deliver an effective topical treatment for nail disease, it is necessary for the hard keratin of nail plate to be weakened or compromised. Keratinases are effective instruments to hydrolyze the nail keratins as they cleave the disulphide linkage to increase the access of drug treatment, thus they can act as ungula enhancers (**Gupta *et al,* 2013a**). Commercial products involving keratinases for the treatment of nail disorders include FixaFungus™ by FixaFungus and Kernail-Soft PB by PreteosBiotech (**Gupta *et al,* 2013b**).

The ability of keratinases to hydrolyze keratin can also be applied in wound healing. In third degree burns, the avascular nature of the wound eschar may prevent effective diffusion of systemic antimicrobial agents to the wound where the amount of microorganisms is usually very high (**Manafi *et al,* 2008**). Enzymatic debridement of the wound will enhance penetration of the topically administered antibiotics and encourage wound healing (**Krieger *et al,* 2012**).

2.7.9.2. Prion protein management

Prions are the whole protein particles without nucleic acids with an unknown function in brain cells (**Verma *et al,* 2016**). Prion proteins (PrP) are causative agents of transmissible spongiform encephalopathies (TSE)**(Suzuki *et al,* 2006; Hui *et al,* 2004)**, which are a unique group of fatal neurodegenerative diseases of humans and animals that are both inheritable and infectious (**Wang *et al,* 2005**). Bovine Spongiform Encephalopathy (BSE) in cow popularly known as ''Mad Cow Disease'', scrapie in sheep and goat, chronic wasting disease (CWD) in elk and deer, transmissible mink encephalopathy (TME) in mink and kuru and Creutzfeldt-Jakob Disease (CJD) in human are types of TSE (**Suzuki *et al,* 2006)**. Infectivity by prions is accompanied by the conversion of harmless PrP^{c} to infectious PrP^{Sc}, facilitated by PrP^{Sc} itself (**Gupta and Ramnani, 2006; Hui *et al,* 2004**). In contrast to PrP^{c}, PrP^{Sc} is insoluble (**Wang *et al,* 2005**) and resistant against proteolytic

digestion (**Muller-Hellwiga *et al,* 2006)**.The structure of PrP^{Sc} has high β-sheet content (**Hui *et al,* 2004; Wang *et al,* 2005**). Due to its structural stability, PrP^{Sc} is resistant to common disinfection processes such as heat and autoclaving. It can be transmitted by contaminated food and feed prepared from infected animal tissues (**Wang *et al,* 2005)**.

Shih and coworkers at BioResource International (BRI, North Carolina) have reported that the broad-spectrum keratinase PWD1 (Versazyme) is capable of completely degrading prions from brain tissue of bovine spongiform encephalopathy (BSE) - and scrapie-infected animals in the presence of detergents and heat treatment. Furthermore, proteinase K, Alcalase, and subtilisin Carlsburg, all serine proteinases, were also able to fully degrade PrP^{Sc} under the same conditions, including pre-heating and the presence of the detergent (**Suzuki *et al,* 2006**). The enzymatic breakdown of prions would most importantly help revive the use of animal meal as feed, which faced much criticism by the European Union despite its high nutritive value due to risk of TSE **(Gupta and Ramnani, 2006)**. A thermophilic keratinase from *Bacillus* sp.WF146 disintegrates prion protein at 80°C (**Liang *et al,* 2010**). Three commercial keratinase-based enzymes are marketed for degradation of infectious prion proteins: Versazyme® is manufactured by BRI, Pure100 Keratinase™ is produced by Proteos Biotech and Prionzyme™ produced by Genencor International (**Gupta *et al,* 2013b**).

These enzymatic methods have also been used in the decontamination of precision instruments that are susceptible to prion contamination (**Yoshioka *et al,* 2007**). Prionzyme™ is currently the only effective enzyme-based decontamination technology that demonstrates significant removal of prion from medical and dental instruments (**Gupta *et al,* 2013b**).

2.7.10. Other Applications

Keratinases could be used in wastewater treatment to remove clogs in bathroom drain pipes and to clear obstructions in sewage system (**Takami *et al,* 1992; Chitte *et al,* 1999; Tapia and Contiero, 2008**).Nowadays, the use of biodegradable films, coatings, and glues is gaining importance. In this respect, processed keratinous wastes have been used for the formation of biodegradable products (**Schrooyen *et al,* 2001; Schrooyen and Radulf, 2004; Gupta and Ramnani, 2006**). Presently, one of the emerging applications of the keratinous wastes, such as hairs, feathers, fur, hooves, horns etc. is in the preparation of agro-textiles i.e. fruit covers, crop covers, edible films having biodegradable feature (**Abdel-Fattah, 2013; Chitte *et al,* 1999)**. Thus, keratinases have many potential applications which require extensive exploration.

Chapter Three

Genome Sequencing and Annotation

3.1. Genome Sequencing Technologies

The ability to sequence the DNA of an organism has become one of the most important tools in modern biological research (**Imelfort and Edwards, 2009**).Determining the genome sequence of an organism opens new avenues in deciphering the phenotypic characteristics, physiology, and ecology of the organism in question. The Whole Genome sequencing (WGS) has allowed for detailed phylogenetic analysis of many important pathogens. Studies of their spread on a global and local scale, as well as their response to pressure from therapeutics and the human immune response, are made possible by these analyses.DNA sequencing technologies have revolutionized biology **(Pop, 2009**)**.**

After the first bacterial genome was completely sequenced and published in 1995 (**Fleischmann *et al,* 1995; Fraser *et al,* 1995**), the number of sequenced genomes has continued to increase dramatically in the last 10 years (**Figure 3**). In 2006 there were about 300 sequenced bacterial genomes and only two published metagenomic projects; this represented growth of more than 100-fold from the mere two genomes sequenced in 1995. (**Binnewies *et al,* 2006)**. There are more than 30, 000 sequenced bacterial genomes publically available in 2014 (National Center for biotechnology Information: NCBI, 2014) (**Land *et al,* 2015**).

This wealth of data has resulted from the introduction of Next-Generation Sequencing technologies (NGS) or massively parallel sequencing technologies generating more data rapidly and at thousands of times lower cost(**Wetterstrand, 2011**).For almost 30 years, the Sanger method waste leading technology in genome sequencing. This method generates low-throughput long reads (800–1000 bp) with high costs. Since the emergence of next-generation sequencing technology, sequencers can produce vast volumes of data (up togigabases) during a single run with low costs

(**El-Metwally *et al,* 2013**).The NGS technologies are different from the Sanger method in aspects of massively parallel analysis, high throughput, and reduced cost (**Liu *et al,*2012;Voelkerding*et al,* 2009)**. They can produce over 100 times more data compared to the most sophisticated capillary sequencers based on the Sanger method. With the ongoing developments of high throughput sequencers and advancement of modern bioinformatics tools at unprecedented pace, the target goal of sequencing individual genomes of living organism at a cost of $1,000 each seemed to be realistically feasible in the near future (**Pareek *et al,* 2011**).

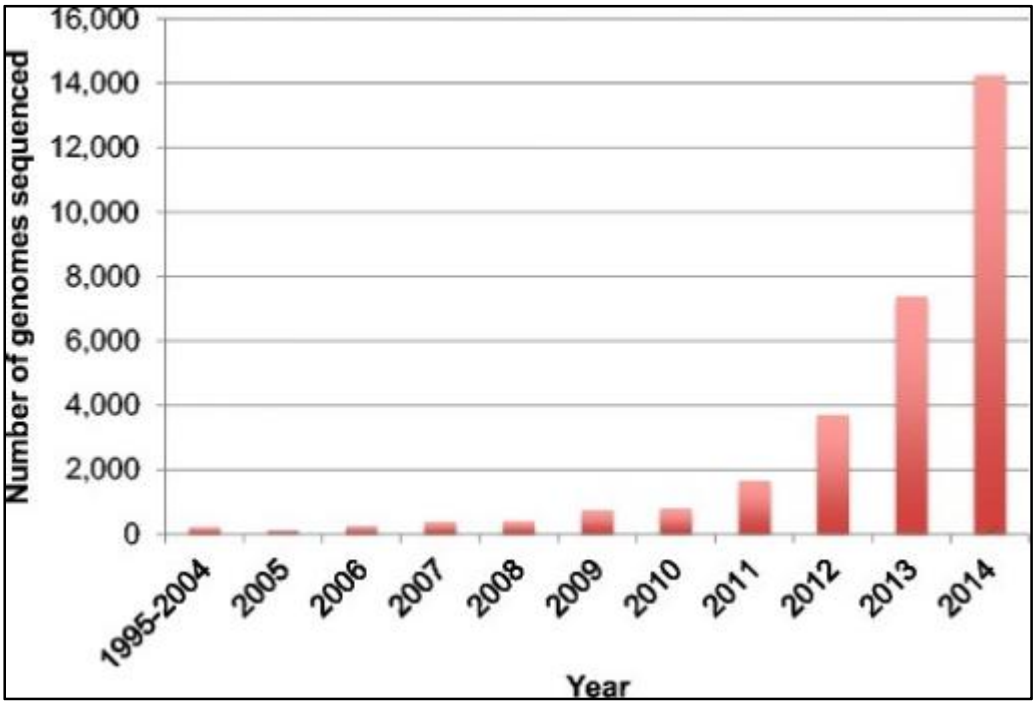


Figure3: Number of bacterial and archaeal genomes sequenced each year and submitted to NCBI. Source: GenBank prokaryotes.txt file downloaded 4 February 2015

3.2. Development of DNA Sequencing Technologies

There are a number of different NGS platforms using different sequencing technologies. However, all NGS platforms perform sequencing of millions of small fragments of DNA in parallel. Bioinformatics analyses are used to piece back together these short fragments or reads into longer continuous stretches of sequence (contigs) and contigs are typically joined to form longer stretches of sequence (known as scaffolds), (**Behjati and Tarpey, 2013; Ekblom and Wolf, 2014**).

For ease of discussion, we can categorize the progress of sequencing technology into three generations.

3.2.1. First Generation Technologies or platforms

First Generation Technologies encompasse the chain termination method pioneered by Sanger and Coulson in 1975 and the chemical method of Maxam and Gilbert in 1976–1977 (**Niedringhaus *et al,* 2011**). In 1977, Frederick Sanger developed DNA sequencing technology which was based on chain-termination method (also known as Sanger sequencing), and Walter Gilbert developed another sequencing technology based on chemical modification of DNA and subsequent cleavage at specific bases. Because of its high efficiency and low radioactivity, Sanger sequencing was adopted as the primary technology in the "first generation" of laboratory and commercial sequencing applications (**Liu *et al,* 2012**). At that time, DNA sequencing was laborious and radioactive materials were required. After years of improvement, Applied Biosystems introduced the first automatic sequencing machine (namely ABI370) in 1987, adopting capillary electrophoresis which made the sequencing faster and more accurate. ABI370 could detect 96 bases at one time, 500Kb a day, and the read length could reach 600 bases. The current model ABI3730xl can output 2.88Megabases per day and read length could reach 900 bases since 1995. Emerged in 1998, the automatic sequencing instruments and associated software using the capillary sequencing machines (**Collins *et al,* 2003**). Sanger technology was used in the sequencing of the first human genome, which was completed in 2003 through the Human Genome Project, a 13-year effort with an estimated cost of $2.7 billion (**Wheeler *et al,* 2008**).

This project greatly stimulated the development of powerful novel sequencing instrument to increase speed and accuracy, while simultaneously reducing cost and manpower. Not only this, X-prize also accelerated the development of NGS technologies **(http://genomics.xprize.org/).**

3.2.2. Next Generation Sequencing or more precisely second generation sequencing platforms

NGS platforms share a common technological feature massively parallel sequencing of clonally amplified or single DNA molecules that are spatially separated in a flow cell. This design is a paradigm shift from that of Sanger sequencing, which is based on the electrophoretic separation of chain-termination products produced in individual sequencing reactions. In NGS, sequencing is performed by repeated cycles of polymerase-mediated nucleotide extensions or, in one format, by iterative cycles of oligonucleotide ligation. As a massively parallel process, NGS generates hundreds of megabases to gigabases of nucleotide sequence output in a single instrument run, depending on the platform (**Voelkerding *et al,* 2009**).

3.2.2.1. Roche 454 system

Roche 454 was the first commercial second-generation sequencing system, produced by 454 Technologies and commercialized by Roche (**Imelfort and Edwards, 2009; Liuet *al,* 2012; Margulies *et al,* 2005; Zhou *et al,* 2010**). The GS FLX system based on sequencing-by-synthesis with pyrophosphate chemistry (**Zhou *et al,* 2010**).

In this system, the DNA sample is first sheared into fragments via nebulization or sonication. Fragments several hundred base pairs in lengthier end-repaired and ligated to two adapter oligonucleotides, an A-adaptor and a B-adaptor (**Voelkerding *et al,* 2009**). The adaptors provide priming sites for amplification and sequencing, as well as a special key sequence. The B-adaptor also contains a 5'-biotin tag that enables the immobilization of library fragments onto streptavidin-coated magnetic beads. The double-stranded products bound to the beads are then denatured to release the complementary non-biotinylated strands containing both an A- adaptor sequence and a B-adaptor sequence. These denatured strands form the single-stranded template DNA library (**Figure4A**). For DNA amplification, the Genome Sequencer system employs emulsion-based clonal amplification, called emPCR (**Dressman *et al,* 2003**). The single-stranded DNA library is immobilized by hybridization onto primer-coated capture beads. The process is optimized to produce beads where a single library fragment is bound to each bead. The bead-bound library is emulsified along with the amplification reagents in a water-in-oil mixture. Each bead with a single library fragment is captured within its own emulsion micro reactor, where the independent clonal

amplification takes place. After amplification, the microreactors are broken, releasing the DNA-positive beads for further enrichment (**Figure4B**).For sequencing, the DNA beads are layered onto a PicoTiter-Plate device, depositing the beads into the wells, followed by enzyme beads and packing beads. The enzyme beads contain sulfurylase and luciferase, which are key components of the sequencing reaction, while the packing beads ensure that the DNA beads remain positioned in the wells during that sequencing reaction (**Figure4C**). The fluidics sub-system delivers sequencing reagents that contain buffers and nucleotides by flowing them across the wells of the plate. Nucleotides are flowed sequentially in a specific order over the PicoTiter-Plate device. When a nucleotide is complementary to the next base of the template strand, it is incorporated into the growing DNA strand by the polymerase. The incorporation of a nucleotide releases a pyrophosphate moiety. The sulfurylase enzyme converts the pyrophosphate molecule into ATP using adenosine phosphosulfate. The ATP is hydrolyzed by the luciferase enzyme using luciferin to produce oxyluciferin and give off light. The light emission is detected by a charge-coupled device CCD camera, which is coupled to the PicoTiter-Plate device **(Zhou *et al,* 2010**). With the flow of each dNTP reagent, wells are imaged, analyzed for their signal-to-noise ratio, filtered according to quality criteria, and subsequently algorithmically translated into a linear sequence output (**Voelkerding *et al,* 2009**).The light emitted from a particular well indicates the incorporation of nucleotides (**Figure4D**), (**Zhou *et al,* 2010a**) and the intensity is proportional to the number of incorporated nucleotides (**Pettersson *et al,* 2009**).

With the newest chemistry, termed "Titanium" a single GS FLX run generates approximately1x10^6 sequence reads, with read lengths of≥400 bases yielding up to 500 million base pairs (Mb)of sequence per 10 h instrument run (**Voelkerding *et al,*2009;Zhou*et al,* 2010**).A major limitation of the 454 technology relates to resolution of homopolymer-containing DNA segments, such as AAA and GGG (**Rothberg and Leamon, 2008**).Because pyrosequencing relies on the magnitude of light emitted to determine the number of repetitive bases. Another disadvantage of 454 sequencing platform is that the per-base cost of sequencing is much higher than that of other nextgeneration platforms (**Zhou *et al,* 2010b**)

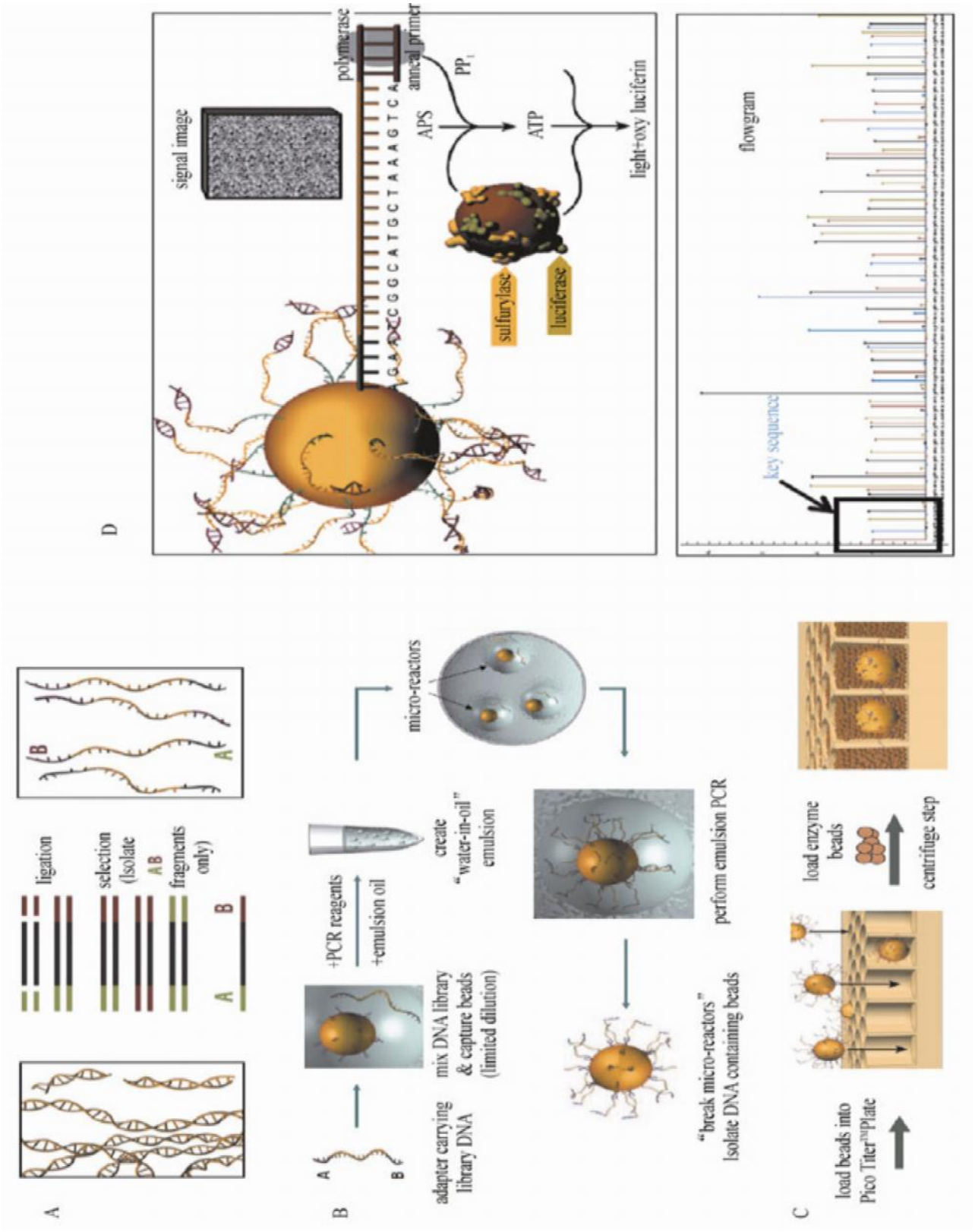


Figure 4: The GS FLX system working principle. **(A)** Prepare adapter ligated ssDNA library (A-[insert]-B). **(B)** Emulsion-based clonal amplification. **(C)** Depositing DNA beads into the PicoTiter™ plate. **(D)** Sequencing and base calling, (**Zhou *et al*,2010**).

3.2.2.2. Illumina Sequencing

The Illumina sequencing technology is based on the ideas of the British chemists Shankar Balasubramanianand David Klenerman. They founded Solexa in 1998, and their goal during early development of sequencing single DNA molecules was not achieved. By 2006, the Solexa Genome Analyzer (GA), the first "short read" sequencing platform, was commercially launched. Acquired by Illumina (http://www.Illumina.com) a year later (**Voelkerding *et al,* 2009).**

The working principle (**Figure 5**) is sequencing by synthesis chemistry using fluorescently labeled nucleotides with reversible termination chemistry and modified polymerases for improved incorporation of nucleotide analogues. (**Liu *et al,* 2012; Zhou *et al,* 2010a; Zhou *et al,* 2010b; Niedringhaus *et al,* 2011**).Template DNA is fragmented into lengths of several hundred base pairs and end-repaired to generate 5'-phosphorylated blunt ends. The polymerase activity of Klenow fragment is used to add a single A base to the 3' end of the blunt phosphorylated DNA fragments. This addition prepares the DNA fragments for ligation to oligonucleotide adapters, which have an overhang of a single T base at their 3' end to increase ligation efficiency. The adapter oligonucleotides are complementary to the flow-cell anchors. Under limiting-dilution conditions, adapter-modified, single-stranded template DNA is added to the flow cell and immobilized by hybridization to the anchors (**Voelkerding *et al,* 2009**). In contrast to emulsion PCR, DNA templates are amplified in the flow cell by "bridge PCR amplification" (**Adessi *et al,* 2000; Fedurco *et al,* 2006**), which relies on captured DNAstrands "arching" over and hybridizing to an adjacent anchor oligonucleotide. Multiple amplification cycles convert the single-molecule DNA template to a clonally amplified arching "cluster," with each cluster containing approximately 1000 clonal molecules.

Approximately $50x10^6$ separate clusters can be generated per flow cell. For sequencing, the clusters are denatured, and a subsequent chemical cleavage reaction and wash leave only forward strands for single-end sequencing. Sequencing of the forward strands is initiated by hybridizing a primer complementary to the adapter sequences, which is followed by addition of polymerase and a mixture of four different terminator nucleotides. These nucleotides contain different cleavable fluorescent dye and a removable blocking group. The nucleotides are incorporated according to sequence complementarity in each strand in a clonal cluster. Each nucleotide is 3′-OH blocked to prevent further addition (**Voelkerding *et al,* 2009**).After incorporation into the DNA strand, the terminator nucleotide as well as its position on the support surface are detected and identified via its fluorescent dye by the CCD camera. The terminator group at the 3'endof the base and the

fluorescent dye are then removed from the base and the synthesis cycle is repeated. This series of steps continues for a specific number of cycles, asdetermined by user-defined instrument settings (**Zhou *et al,* 2010a**).

Incomplete incorporation of nucleotides and insufficient removal of reverse terminators or fluorophores may be the explanation for the relatively short read length of 35 bases (**Pettersson *et al,* 2009**).

Although shorter read lengths than the454 system, the throughput is much higher. At first, Illumina GA output was 1 giga basepairs (Gbp)/run. Through improvements in polymerase, buffer, flow cell, software, and the use of paired-end (PE) sequencing in 2009 the output of GA increased to 20 Gbp/run in August (75PE), 30Gbp/run in October (100PE), and 50 Gbp/run in December (Truseq V3, 150PE), and the latest GAIIx series can attain 85 Gbp/run.

In early 2010, Illumina launched HiSeq 2000, which adopts the same sequencing strategy with GA. Its output was 200Gbp per run initially, improved to 600Gbp per run currently which could be finished in 8 days. MiSeq, a benchtop sequencer launched in 2011which shared most technologies with HiSeq, is especially convenient for amplicon and bacterial sample sequencing. It could sequence 150PE and generate 1.5Gbp/run in about 10 hours including sample and library preparation time (**Liu *et al,* 2012**).

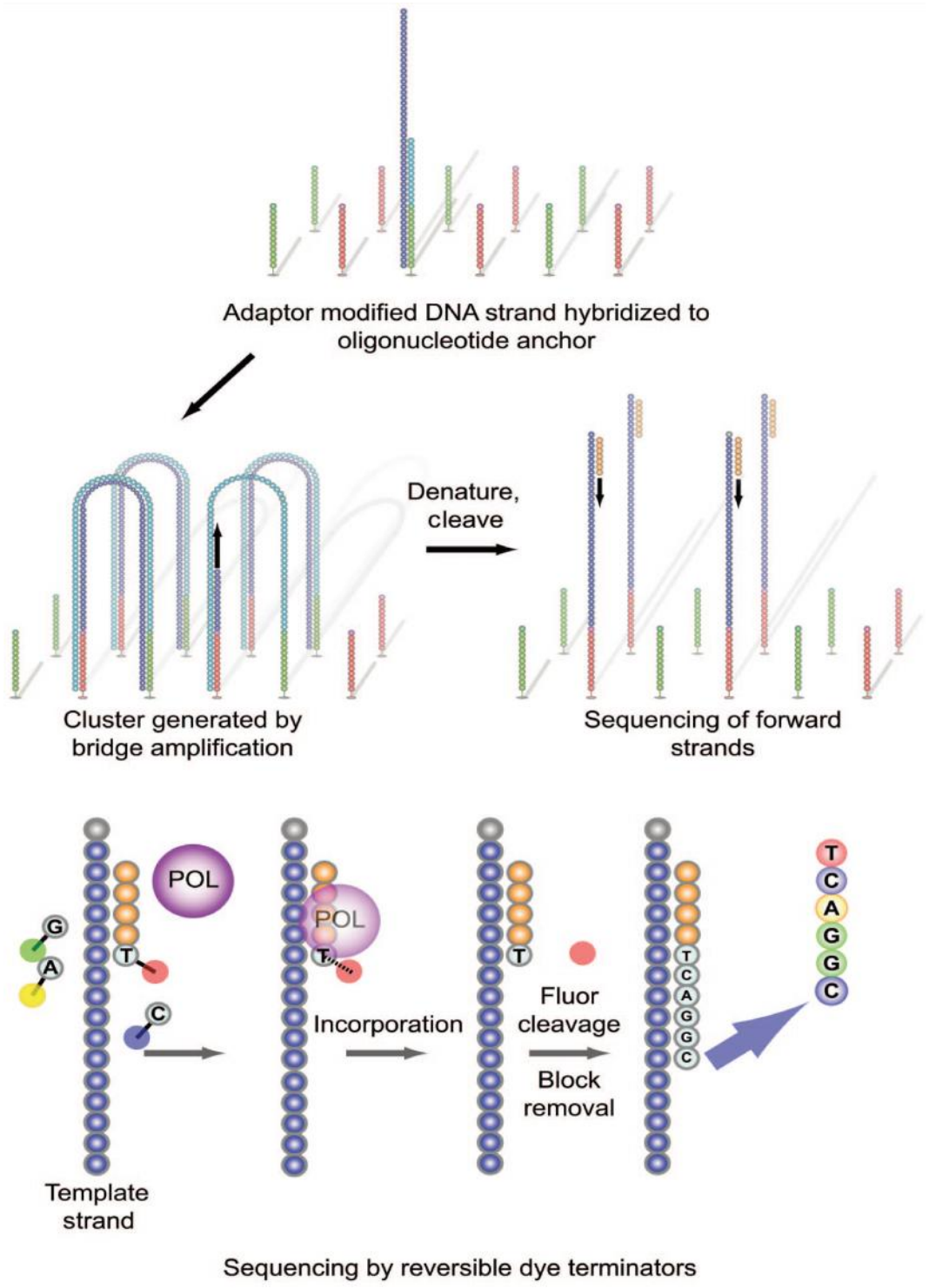


Figure 5: The working principle of the Illumina sequencing platform.

Adapter-modified, single-stranded DNA is added to the flow cell and immobilized by hybridization. Bridge amplificationgenerates clonally amplified clusters. Clusters are denatured and cleaved; sequencing is initiated with addition of primer, polymerase (POL) and 4 reversible dye terminators. Post-incorporation fluorescence is recorded. The fluorescent dye and block are removed before the next synthesis cycle (**Voelkerding*et al,* 2009**).

3.2.2.3. SOLiD™ Sequencing

Sequencing by Oligo Ligation Detection (SOLiD) was purchased by Applied Biosystems in 2006. The sequencer adopts the technology of two-base sequencing based on ligation sequencing (**Liu *et al,*2012**).This platform has its origins in the system described by Shendure *et al,* 2005and in work by McKernan and colleagues 2006 at AgencourtPersonal Genomics (acquired by Applied Biosystems in2006).

The generation of a DNA fragment library and the sequencing process by subsequent ligation steps are shown in **Figure 6**. Sample preparation shares similarities with the 454 Technology in that DNA fragments are ligated to oligonucleotide adapters, attached to beads, and clonally amplified by emulsion PCR. Beads with clonally amplified template are immobilized onto a derivitized-glass flow-cell surface, and sequencing is begun by annealing a primer oligonucleotide complementary to the adapter (**Figure 6**). Instead of providing a 3' hydroxyl group for polymerase-mediated extension, the primer is oriented to provide a 5' phosphate group for ligation to interrogation probes during the first "ligation sequencing" step. Each interrogation probe is an octamer, which consists of (in the 3' to 5' direction) 2 probe-specific bases followed by 6 degenerate bases with one of 4 fluorescent labels linked to the5' end. The 2 probe-specific bases consist of one of 16possible 2-base combinations (for example TT, GT, and so forth). In the first ligation-sequencing step, thermostable ligase and interrogation probes representing the 16 possible 2-base combinations are present. The probes compete for annealing to the template sequences immediately adjacent to the primer. After annealing, a ligation step is performed, followed by wash removal of unbound probe. Fluorescence signals are optically collected before cleavage of the ligated probes, and a wash is performed to remove the fluor and regenerate the 5' phosphate group. In the subsequent sequencing steps, interrogation probes are ligated to the 5' phosphate group of the preceding pentamer. Seven cycles of ligation, referred to as a "round, «are performed to extend the first primer. The synthesized strand is then denatured, and a new sequencing primer offset by 1 base in the adapter sequence (n-1) is annealed. Five rounds total are performed, each time with a new primer with a successive offset (n-2, n-3, and so on). By this approach, each template nucleotide is sequenced twice (**Voelkerding *et al,* 2009**).

The read length of SOLiD was initially 35 bp reads and the output was 3G data per run with an accuracy of 99.85% after filtering. In late 2010, the SOLiD 5500xl sequencing system was released. From SOLiD to SOLiD 5500xl, five upgrades were released by ABI in just three years. The SOLiD 5500xl realized improved read length, accuracy, and data output of 85 bp, 99.99%, and 30G per run,

respectively. But the short read length and resequencing only in applications is still its major shortcoming (**Liu *et al,* 2012**).

In conclusion, of the three NGS systems described before, the Illumina features the biggest output and lowest reagent cost, the SOLiD system has the highest accuracy (**Huse *et al,* 2007**), and the Roche 454 system has the longest read length (**Liu *et al,* 2012**).

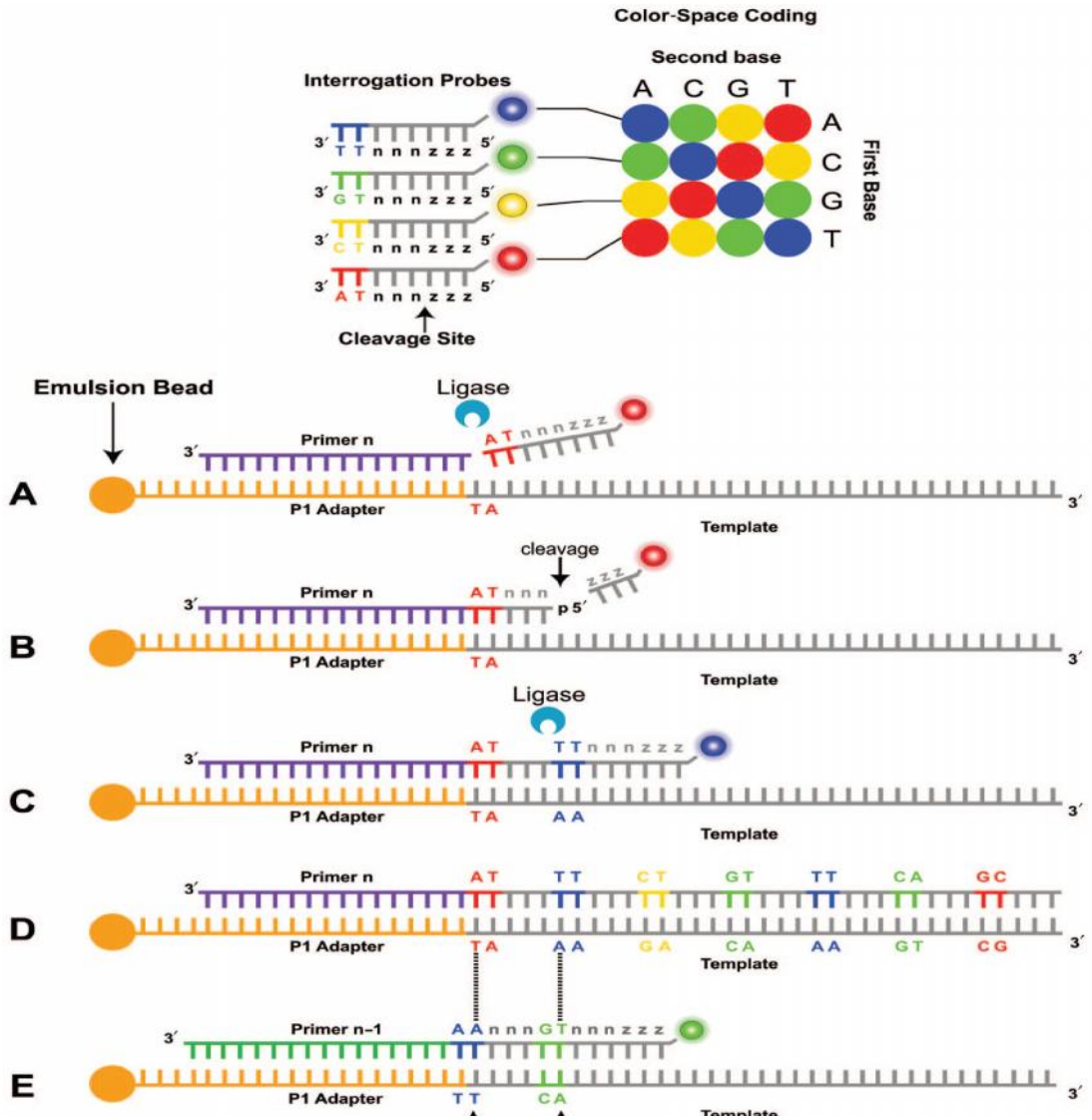


Figure 6: Applied Biosystems SOLiD Technology.

Top: SOLiD color-space coding. Each interrogation probe is an octamer, which consists of (3'-to-5' direction) 2 probe-specific bases followed by 6 degenerate bases (nnnzzz) with one of 4 fluorescent labels linked to the 5' end. The 2 probe-specific bases consist of one of 16 possible 2-base combinations. **Bottom:(A)**, The P1 adapter and template with annealed primer (n) is interrogated by probes representing the 16 possible 2-base combinations. In this example, the 2 specific bases complementary to the template are AT. **(B)**, After annealing and ligation of the probe, fluorescence is recorded before cleavage of the last 3 degenerate probe bases. The 5' end of the cleaved probe is phosphorylated (not shown) before the second sequencing step. **(C)**, Annealing and ligation of the next probe. **(D)**, Complete extension of primer (n) through the first round consisting of 7 cycles of ligation. **(E)**, The product extended from primer (n) is denatured from the adapter/template, and the second round of sequencing is performed with primer (n-1), (**Voelkerding*et al,*2009**).

3.2.2.4. Paired-end sequencing

A major advance in NGS technology occurred with the development of paired-end (PE) sequencing (**Figure 7**). PE sequencing involves sequencing both ends of the fragments in the sequencing library and aligning the forward and reverse reads as read pairs (**Nakazato *et al,* 2013, Illumina 2016**).Paired-end sequencing was already described in 1981 by Hong (**Hong, 1981**), and the first use of paired-end sequencing was described by Ewards and Caskey in 1990 (**Edwards and Caskey, 1991)**. Read pairs can be obtained by one of two mechanisms: paired ends or mate pairs.

In paired-end sequencing, a linear fragment with a length of less than 1 kb has adapter sequences at each end with different priming sites on each adapter. The sequencing instrument is designed to sequence from one adapter priming site then, in a subsequent reaction, the opposite adapter is primed and sequence data are obtained. These reads are paired with one another during the alignment step in data analysis, which provides higher overall certainty of placement than does a single end read of the same length (**Mardis, 2013**).Mate pair library preparation is designed to generate short fragments consisting of two segments that originally had a separation of several kilobases in the genome. Fragments of sample genomic DNA is end-biotinylated to tag the eventual mate pair segments. Self-circularization and refragmentation of these large fragments generates a population of small fragments, some of which contain both mate pair segments with no intervening sequence. These mate pair fragments are enriched using their biotin tag. Mate pairs are sequenced using a similar two-adaptor strategy as described for paired-end sequencing (**Illumina, 2010)**.

Often, mate-pair and paired-end reads are used in combination to achieve genome coverage when attempting longer-range assemblies through difficult regions of a genome or when attempting to assemble a genome for the first time (*de novo* sequencing). In this combined coverage approach, the mate-pair reads provide longer-range order and orientation (a separation of up to 20 kb is possible), and the paired ends provide the ability to assemble, in a localized way, difficult-to-sequence regions that can then be layered on top of the scaffold provided by an assembly of mate-pair reads (**Mardis, 2013; Illumina, 2010)**.Mate-pair library sequencing is significant beneficial for *de novo* sequencing, because the method could decrease gap region and extend scaffold length (**Liu *et al,* 2012**).

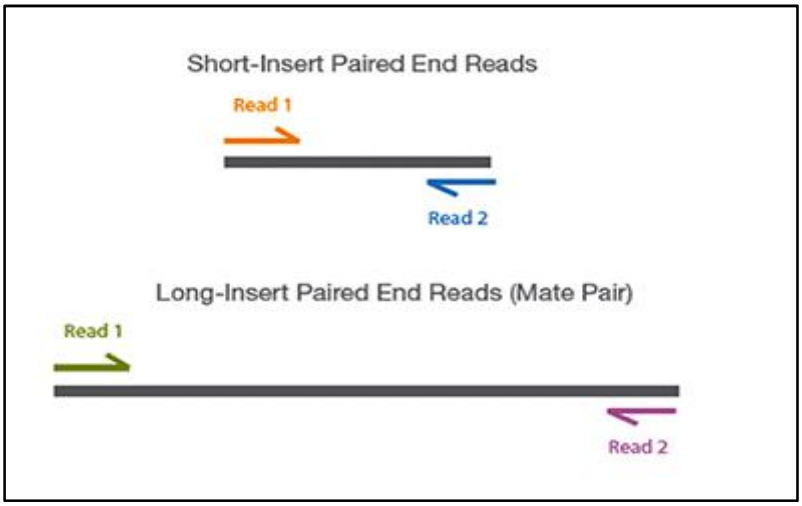


Figure 7: Paired-end sequencing.

3.2.2.5. Other NGS technologies

3.2.2.5.1. Ion Personal Genome Machine (PGM) from Ion Torrent

Ion PGM was released at the end of 2010by Ion Torrent, a company that was later purchased by Life Technologies™ Corp (**Mardis, 2013**). PGM uses semiconductor sequencing technology (**Liu *et al,* 2012**) detecting the protons released as nucleotides are incorporated during synthesis (**Quail *et al,* 2012**). PGM is the first commercial sequencing machine that does not require fluorescence and camera scanning (**Liu *et al,* 2012**).

DNA fragments with specific adapter sequences are linked to and then clonally amplified by emulsion PCR on the surface of 3-micron diameter beads, known as Ion Sphere Particles (**Quail *et al,* 2012**). The templated beads are loaded into a highly dense microwell array (Ion Chip) in which each well acts as an individual DNA polymerization reaction chamber. Just below this layer of microwells is an ionsensitive layer followed by a sublayer composed of a highly dense field-effect transistors (FET) array. Following the pyrosequencing scheme, sequential cycling of the four nucleotides into the microwells enables primary sequence resolution since the FET detector senses the change in pH created during nucleotide incorporation and converts this signal to recordable voltage change. The change in voltage scales with the number of nucleotides incorporated at each step, **Figure 8** (**Niedringhaus *et al,* 2011)**.

Ion Torrent has already released Ion 314, 316 and Ion 318 chips in late 2011. The chips are different in the number of wells resulting in higher production within the same sequencing time. The Ion 318chip enables the production of >1Gb data in 2 hours. Read length is >400 bp (**Liu *et al,* 2012**). There are several limitations with regards to sequencing complete genomes. Currently, the short read lengths and sequencing through smaller repetitive regions of the same nucleotide(homopolymer regions) on the order of 5 to 10 bases can prove challenging (**Niedringhaus *et al,* 2011)**.For a more complete list of sequencing technologies see review by Zhang *et al*,2011.

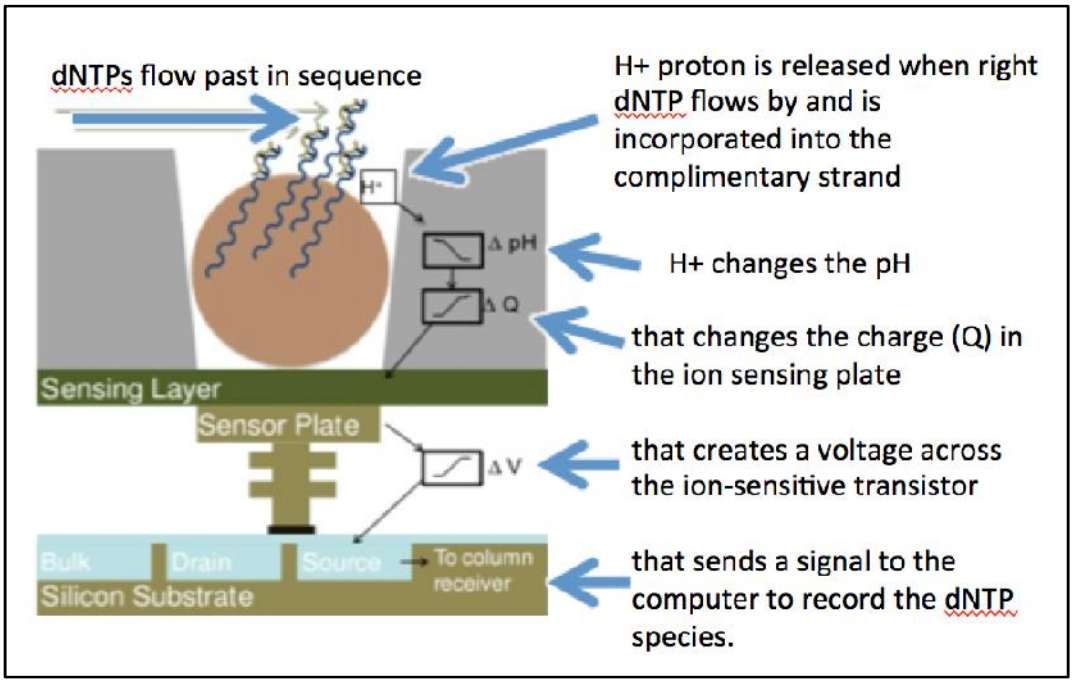


Figure 8: Structure of the Ion Torrent Ion Chip used in pH-based sequencing. **(Mardis, 2013)**

3.2.2.5.2. Heliscope™ Single Molecule Sequencer

The HeliScope Genetic Analysis System is one of the first techniques for sequencing from a single DNA molecule was introduced by Braslavsky *et al*,2003 and licensed by Helicos Biosciences as the first commercial single-molecule DNA sequencing system in 2007 (**Pareek *et al,* 2011**). The principle of Heliscope sequencer relies on sequencing by synthesis without a prior amplification

step; single-molecule sequencing (**Pettersson *et al,* 2009**). Single molecule sequencing technology begins with DNA library preparation through DNA shearing and addition of poly(A) tail to generated DNA fragments (**Ozsolak *et al,* 2010**), followed by hybridization of DNA fragments to the poly (T) oligonucleotides which are attached to the flow cell and simultaneously sequenced in parallel reactions. Helicos has developed proprietary labeled dNTPs, termed "virtual terminators," so that only single bases are added (**Voelkerding *et al,* 2009**). Sequencing is performed in a similar fashion as Illumina sequencing, with single nucleotide extension followed by detection and cleavage of fluorophores.

This Single Molecule Sequencer is comparable with other current NGS sequencing systems with regards to accuracy. However, the Helicos system is currently limited by shorter read length (maximal length of 55 bases), and lower throughput compared to the Illumina and SOLiD sequencing systems (up to 28 Gb in a single sequencing run and takes about 8 days (**Pareek *et al,* 2011**).

3.2.3. Third Generation Sequencing Technologies

Although the advents of next or second generation sequencing platforms had and still have a huge impact on life Sciences, there are difficulties associated with sequencing from polymerase amplified fragment populations. These include polymerase errors during library construction that may appear to be variant bases in the original genome but are not, as wells preferential amplification of certain fragments in the library population that cause them to be differentially overabundant relative to others. Furthermore, DNA modifications, such as different types of methylation, are diluted out during the amplification process because this step only copies the DNA bases (**Mardis, 2013**).On the other hand, the National Human Genome Research Institute (NHGRI) has funded several groups developing alternative approaches to improving second generation technologies, as well as novel approaches with the final goal of bringing the cost of a human genome to under $1000 (**Niedringhaus *et al,* 2011**). A goal not that far away, since the price in 2013 was less than $5,000 in average (**Mardis, 2013**).

Therefore, advantages would exist for a platform that could obtain sequence data from individual molecules of a DNA isolate, eliminate many of the aforementioned challenges and effective read length increased as a result. Many of the so called ''third generation'' or ''next next generation'' sequencing technologies started to appear before the second generation sequencing platforms peaked such as the PacBio RS system and a portable single-molecule sequencer, MinION.

3.2.3.1. The PacBio RS system

The PacBio RS sequencer is commercialized in 2010 by Pacific Biosciences, Inc. The principle of this third-generation sequencing platform relies on a real-time, single-molecule sequencing technology by synthesis method provided on the sequencing chip containing thousands of zero-mode waveguides (ZMWs) and uses phospho-linked fluorescently labeled dNTPs. The sequencing reaction of a DNA fragment is performed by a single DNA polymerase molecule attached to the bottom of each ZMW hole (**Pareek *et al,* 2011**).

Here, DNA polymerase molecules, bound to a DNA template, are attached to the bottom of 50 nm-wide wells termed zero-mode waveguides (ZMWs). Each polymerase is allowed to carry out second strand DNA synthesis in the presence of γ-phosphate fluorescently labeled nucleotides. The width of the ZMW is such that light cannot propagate through the waveguide, but energy can penetrate a short distance and excite the fluorophores attached to those nucleotides that are in the vicinity of the polymerase at the bottom of the well. As each base is incorporated, a distinctive pulse of fluorescence is detected from the bottom in real time (**Quail *et al,* 2012**).

This will give out not only the fluorescent signal but also the signal difference a long time, which may be useful for the prediction of structural variance in the sequence such as DNA methlyation (**Liu *et al,* 2012**). Currently, the most widely used instrument in long-read sequencing is the PacBio RS II instrument. This device is capable of generating single polymerase reads in excess of 50 kb with average read lengths of 10–15 kb for a long-insert library (**Goodwin *et al,* 2016**),raw accuracies ranging from 81% to 83% and with reagent cost no more than $60 per run (**Zhou et al, 2010a**, **Mardis,2013**).

3.2.3.2. Nanopore Sequencing Technologies

Nano-technologies have long been considered a cutting-edge technology for single-molecule DNA sequencing (**Iqbal *et al,* 2007; Branton *et al,* 2008**). One concept is based on the observation that when a DNA strand is pulled through a nanopore by an electrical current, each nucleotide base (A, T, C, and G) creates a unique pattern in the electrical current. This unique nanopore electrical current fingerprint can be used for nanopore sequencing (**Zhang *et al,* 2011**). Various groups and organizations around the globe are exploring this idea: Agilent, DNA Electronics, IBM, NabSys, Oxford Nanopore Technologies, Sequenom, etc (**Zhou *et al,* 2010b**).

In June 2014OxfordNanopore Technologies, collaborating with University of Oxford, launched a portable single-molecule sequencer, MinION (**Ip *et al,* 2015**). This device, the MinION, is a nanopore based device in which pores are embedded in a lipid-bilayer membrane placed over an electrical detection grid. To carry out sequencing, DNA is passed through a protein pore (alpha hemolysin) as current is passed through the pore. As the DNA translocates through the action of a secondary motor protein, a voltage blockade occurs that modulates the current passing through the pore (**Goodwin *et al,* 2016**).These fluctuations are sequence dependent and thus can be used by a base-calling algorithm to infer the sequence of nucleotides in each molecule. Although it is still in development, MinION promised read lengths orders of magnitude longer than existing technologies together with low per-base costs (**Goodwin *et al,* 2015**), but the data are still of insufficient quality owing to systematic errors and further improvement will therefore be required (**Van Dijk *et al,* 2014**).

3.3. Comparison of Next Generation Sequencing technologies

Table 4summarizes the features of the different NGS technologies described above.

Table 4: Comparison of selected high-throughput NGS technologies(This table is compiled from the data presented in **Goodwin *et al,* 2016; Niedringhaus *et al,* 2011; Reuter *et al,* 2015; Liu *et al,* 2012; Loman *et al,* 2012 and Van Dijk*et al,* 2014**).

Platform	**Roche/454, GS FLX Titanium+**	**ABI/SOLiD, 5500xl**	**Illumina, MiSeq**	**Ion PGM System, Ion 318**	**Helicos Biosciences Heliscope**	**PacBio, RSII**	**Nanopore, MinION**
Release of the first version	2005	2006	2006	2010	2007	2010	2014
Amplification	Emulsion PCR	Isothermal template walking	Bridge PCR	Emulsion PCR	--	--	--
Sequencing chemistry	pyrosequencing	Sequencing by oligonucleotide probe ligation	Reversible dye termination	pH-based sequencing	Single-molecule sequencing by synthesis	Real time single molecule sequencing by synthesis	Nanopore-based sequencing
Max read length (bp)	Up to 1000 (SE, PE)	75 (SE)	150 (PE)	400 (SE)	25-55	20000	Not yet quantified
Advantages	Longest read lengths among NGS platforms	high throughput accuracy	Highest throughput Best cost - effectiveness	Fast run times	High throughput, lowest reagent cost	Very long read lengths, fast run time	No fluorescent labeling or optics necessary
Limitations	Errors inhomopolymer regions > 6,high cost, low throughput, planned	Very short read length, long run time	Short read length, substitution errors	Errors in homopolymer repeats	Low accuracy, high cost per base, short read length	High cost, high overall error rate,low throughput	High error rate, insertion/deletion errors, low throughput

	discontinuation in 2016						
Suitable microbiological applications	*De novo* genome sequencing, rapid draft *de novo* genome sequencing	Transcriptome sequencing, SNP detection	Draft *de novo* genome sequencing, genome re-sequencing, transcriptome sequencing, metagenomics	Rapid draft *de novo* genome sequencing	Direct RNA sequencing	*De novo* genome sequencing, rapid draft *de novo* genome sequencing, transcriptomics	*De novo* genome sequencing, genome scaffolding

3.4. Whole genome sequencing

Currently, most genome projects use a shotgun sequencing strategy for whole genome sequencing. In a first step, genomic DNA is sheared into small random fragments. Two types of fragmentation are widely used: mechanical and enzymatic (**Loman *et al,* 2012**). Fragments can consist of single reads (typically50–1000 bp) or of paired-end reads of varying insert size. Mate-pair libraries span larger genomic regions (~2–20 kb inserts). Depending on the technology, these fragments are sequenced independently to a given length. Powerful computer algorithms are then utilized to piece the resulting sequence reads back together into longer continuous stretches of sequence (contigs), a process known as *de novo* assembly. After the initial assembly, contigs are typically joined to form longer stretches of sequence (known as scaffolds) using the so-called paired-end, mate-pair or jump libraries, **Figure 9** (**Ekblom and Wolf, 2014**).

Ideally, a draft genome would represent the complete nucleotide base sequence for all chromosomes in the species of interest, a ‘physical map’ of its genetic content (as opposed to the ‘genetic or linkage map’ which establishes the order and recombination distances among genetic markers).

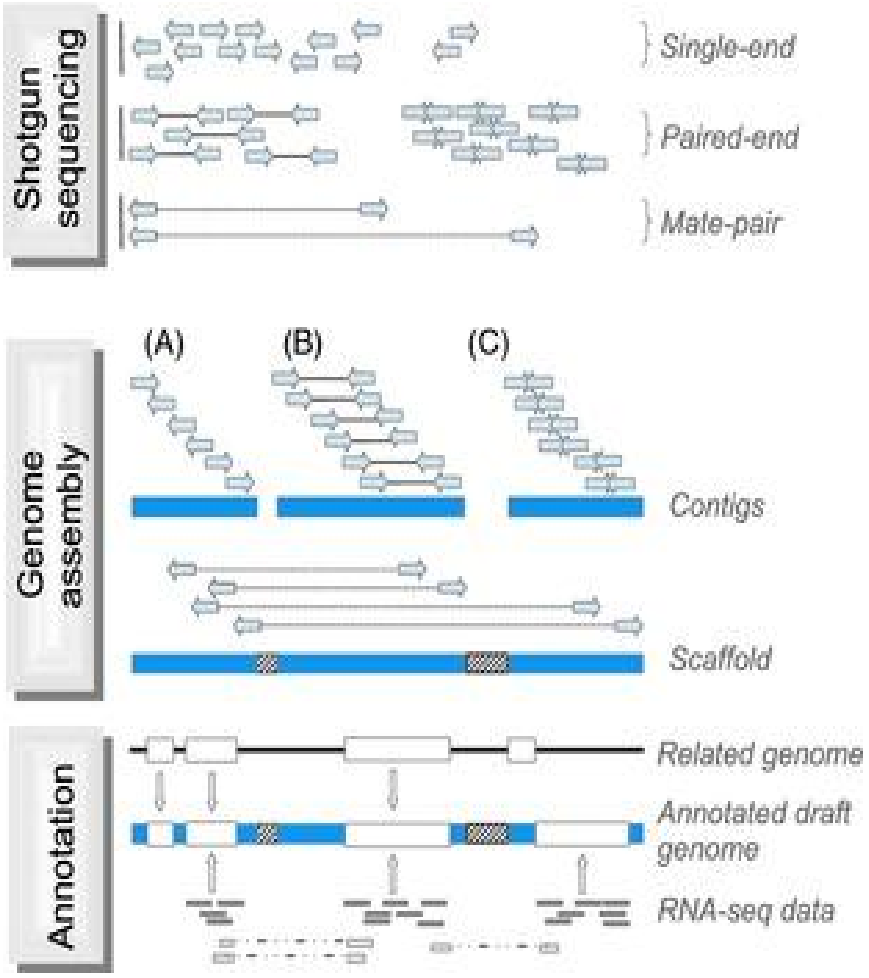


Figure 9: Simplified illustration of the assembly process and terminology.
Shotgun sequencing: short fragments of DNA from the target organism are sequenced at random positions across the genome to a given depth of coverage. **Genome assembly:** (A) short-read *de novo* assemblers extend the disperse sequence information from the reads into continuous stretches called contigs. (B) Paired-end reads provide additional information on whether a read is supported for a given contig. (C) Some assemblers work with overlapping read pairs that are joined into a virtual longer read prior to the assembly. Read pairs from mate-pair libraries can be used to order and orient contigs into scaffolds. Gap size between contigs is estimated from the expected length of mate-pairs and marked with 'N's. **Annotation:** gene models can be inferred *in silico* by prediction algorithms, by lifting over information from genomes of related organisms and by using transcriptome data (RNA-seq, expressed sequence tag) from the target organism itself (**Ekblom and Wolf, 2014**).

3.5. Genome assembly

Genome sequencing samples the chromosomes that make up one genome. Genome assembly is the reconstruction of sequences up to chromosome length. It groups reads into longer contiguous reads known as contigs by a computer program known as an assembler. These contigs are joined together to form longer contigs known as scaffolds. Contigs provide multiple sequence alignment of reads plus the consensus sequence. The scaffolds, sometimes called super contigs or metacontigs, define the contig order and orientation and the sizes of the gaps between contigs (**Figure 9**).The assembly task is relegated to computer software. The assembly process is often compared to solving a jigsaw puzzle-metaphor.

Assemblies are measured by the size and accuracy of their contigs and scaffolds. Assembly size is usually given by statistics including maximum length, average length, combined total length, and N50.The contig N50 is the length of the smallest contig in the set that contains the fewest (largest) contigs whose combined length represents at least 50% of the assembly (**Miller *et al,* 2010**).

There are two approaches for genome assembly; the comparative approach and the *de novo* approach. During comparative assembly, also known as reference-based assembly, a reference genome from the same organism or a closely related species is used as a map to guide the assembly process by aligning the fragments being assembled. This approach is used in resequencing applications, for example (**Pop *et al,* 2004**). During *de novo* assembly, no map or guidance is available for assembling the genome, so this approach represents assembly in the strict sense. Therefore, *de novo* assembly is used to reconstruct genomes that are not similar to previously sequenced genomes (**Martin and Wang, 2011; El-Metwally*et al,* 2013**).

Genome assembly combines four basic stages: preprocessing filtering, a graph construction process, a graph simplification process, and post processing filtering. Preprocessing filtering is responsible for detecting and correcting erroneous reads before the assembly begins. The graph construction process is responsible for creating a graph model, which is used to organize short-read sequences into a compact form and to create longer reads during assembly. The graph simplification process is used to simplify the graph by reducing the number of graph nodes and edges, and removing erroneous ones. Post processing filtering builds contigs, detects misassembled ones, and extends them into scaffolds. Mate pair reads are used in this stage as a guide map to order and orient contigs during the scaffolding process. Appropriate contigs are joined together to form scaffolds depending on the positions

of the mate pair reads in the contigs, their orientation, and expected insert size (**Dayarian *et al,* 2010)**. A mate pair that spans two contigs is taken as evidence for the juxtaposition of those two contigs within the genome (**Paszkiewicz and Studholme, 2010**).If the gaps between the contigs are not filled with other contigs, they are filled with N characters that denote unknown bases between them and the total number of N can be estimated easily using mate pair constraints **Dayarian *et al,* 2010; El-Metwally *et al,* 2013**).

The NGS assemblers are organized into three categories, all based on graphs. The Overlap/Layout/Consensus (OLC) methods rely on an overlap graph. The de Bruijn Graph (DBG) methods use some form of K-mer graph. The greedy graph algorithms may use OLC or DBG (**Miller *et al,* 2010**).The graph is an abstract data structure, which describes the similarity relations within a set of reads. Mathematically, a graph is represented as a set of vertices (nodes) and edges. In the assembly graph, the nodes represent strings or substrings of reads, while the edges represent the suffix-to-prefix overlaps between reads (**Medvedev and Brudno, 2009; Medvedev *et al,* 2007**).

3.5.1. Overlap/Layout/Consensus Assemblers

A classical overlap-based approach for *de novo* assembly consists of three stages: overlap, layout, and consensus (OLC) **(Peltola *et al,* 1984)**. Assemblers following this paradigm start by detecting the overlaps among the set of unassembled reads through pair-wise read comparison. Then, the overlap information is organized into a graph where nodes correspond to reads and edges encode the (suffix-to-prefix) overlaps among them. The goal of the layout step is to find a shortest path that visits each node in the graph exactly once and hence this path represents a solution to the assembly problem. Finally, the overlaps between the reads (nodes) are combined in the consensus step. This paradigm is implemented in several short-read assemblers such as Newbler (**Margulies *et al,* 2005**), CABOG (**Miller *et al,* 2008**) Shorty (**Hossain *et al,* 2009)**, Forge (**DiGuistini *et al,* 2009**), Edena (**Hernandez*et al,* 2008)**, SGA (**Simpson and Durbin, 2012**), Fermi (**Li, 2012**), and Readjoiner (**Gonnella and Kurtz, 2012**). The OLC assemblers target variable-length reads in the 100-800 bp range (**Miller *et al,* 2010**).

Newbler Assembler

Newbler software has been developed by 454 life sciences and it has been successfully applied for the assembly of a bacterial genome (**Poly *et al,* 2007**).As described in 2005, Newbler implements OLC twice. The first-phase OLC generates unitigs from reads.

Unitigs are mini-assemblies that are, ideally, uncontested by overlaps to reads in other unitigs (**Myers *et al,* 2000**). The unitigs serve as preliminary, high-confidence, conservative contigs that seed the rest of the assembly pipeline. The second-phase OLC generates larger contigs from the unitigs. (**Miller *et al,* 2010**)Although originally intended as assembler for Roche 454 and Sanger data only, new versions of this software are able to efficiently process smaller amounts of Illumina data. Especially for hybrid assemblies and smaller genome sizes, the Newbler pipeline is a well-performing alternative for the standard in-house Velvet pipeline (**Reinhardt *et al,* 2009**).

3.6. Bacterial Genome Annotation

Annotation of prokaryotic sequences is a multilevel process that strives to define both the structural and functional properties of a given sequence (**Ouyang *et al,* 2009**). Structural annotation is dependent on algorithmic interrogation of experimental evidence to discover the physical characteristics of a gene; identifies the location(s) of genes, promoters, pseudogenes, RNA genes (rRNAs, tRNAs, and other small RNAs), and untranslated regions. Functional annotation is dependent on sequence similarity to other known genes or proteins in an effort to defines the role of the aforementioned genetic structures encoded in the DNA sequence (**Ouyang *et al,* 2009**).Combining structural and functional annotation across genomes in a comparative manner promotes higher levels of accurate annotation as well as an advanced understanding of genome evolution (**Beckloff *et al,* 2012**).

Annotation is a multistep process, requiring the integration of a large number of tools for the identification of specific genomic features and the interoperability of tools for gene finding and assignment of function.

Figure10describes a very general process used for bacterial genome annotation (**Richardson and Watson, 2013**).Usually the first step in interpretation of a genome is to use gene-prediction programs, which scan the sequence for regions that are likely to encode proteins or functional RNA products, depending on the particular program. The identified genes are then compared to databases of DNA or protein sequences in an attempt to identify related sequences. If hits of a certain similarity are identified, information about their function is transferred to the new sequence (**Stothard and Wishart, 2006**).

Most annotation pipelines use a combination of multiple extrinsic and intrinsic gene prediction tools in tandem for complete genomes. First, extrinsic tools will query databases to identify genes based on homology searches of the sequence of interest. Hits to these genes are considered real and are added to the training set for the intrinsic software. The intrinsic

software uses the profile of the training set to predict and discover more genes that were missed in the previous step. In addition to identifying protein-coding genes, ribosomal RNA operons are generally found via similarity to rRNA sequences, while transfer RNAs can be reliably identified using tRNA scan-SE (**Delcher *et al,* 1999**).Annotation pipelines can also add several other types of information to genes and genomes, such as protein chemical properties, protein structural properties, predicted operons, gene ontologies, evolutionary relationships and metabolic pathways. These annotations are obtained using sequence similarity searches, calculated from the predicted gene and protein sequences directly, or derived from comparisons of gene order between species. The final and most frequently overlooked step of the annotation process is to organize and present the results in a useful manner (**Stothard and Wishart, 2006**).

There are a range of online automatic bacterial annotation pipelines, including RAST (Rapid Annotation using Subsystem Technology) (**Aziz *et al,* 2008)**, BASys (Bacterial Annotation System) (**Van Domselaar *et al,* 2005)**, IMG (Integrated Microbial Genomes) (**Markowitz *et al,* 2009**), JCVI annotator (J. Craig VenterInstitute, **Orvis *et al,* 2010**), PGAP (Prokaryotic Genomes Annotation Pipeline at NCBI) (**Tatusova *et al,* 2016**), and others (**Beckloff *et al,* 2102**).

BASys is a web server that performs automated, in-depth annotation of bacterial genomic (chromosomal and plasmid) sequences. It accepts raw DNA sequence data and an optional list of gene identification information and provides extensive textual and hyperlinked image output. BASys uses more than 30 programs to determine nearly 60 annotation subfields for each gene, including gene/protein name, Gene Onthology (GO) function, Cluster of Orthologous Groups (COG) function, possible paralogues and orthologues, molecular weight, isoelectric point, operon structure, subcellular localization, signal peptides, transmembrane regions, reactions, and pathways. The textual annotations and images that are provided by BASys can be generated in approximately 16 hours for an average bacterial chromosome (5 Megabases. 5000 genes), or approximately 350 coding regions per hour (**Van Domselaar *et al,* 2005**)

Various programs used by the BASys pipeline to generate genome annotations including, GLIMMer (Gene Locator and Interpolated Markov ModelER), NCBI BLAST, HMMER, and databases such as : NCBI non-redundant database of bacterial protein sequences, the PDB database of 3D biological macromolecular structure data (**Berman *et al,* 2000**), and the COG database of orthologous groups of proteins (**Tatusov *et al,* 1997**).

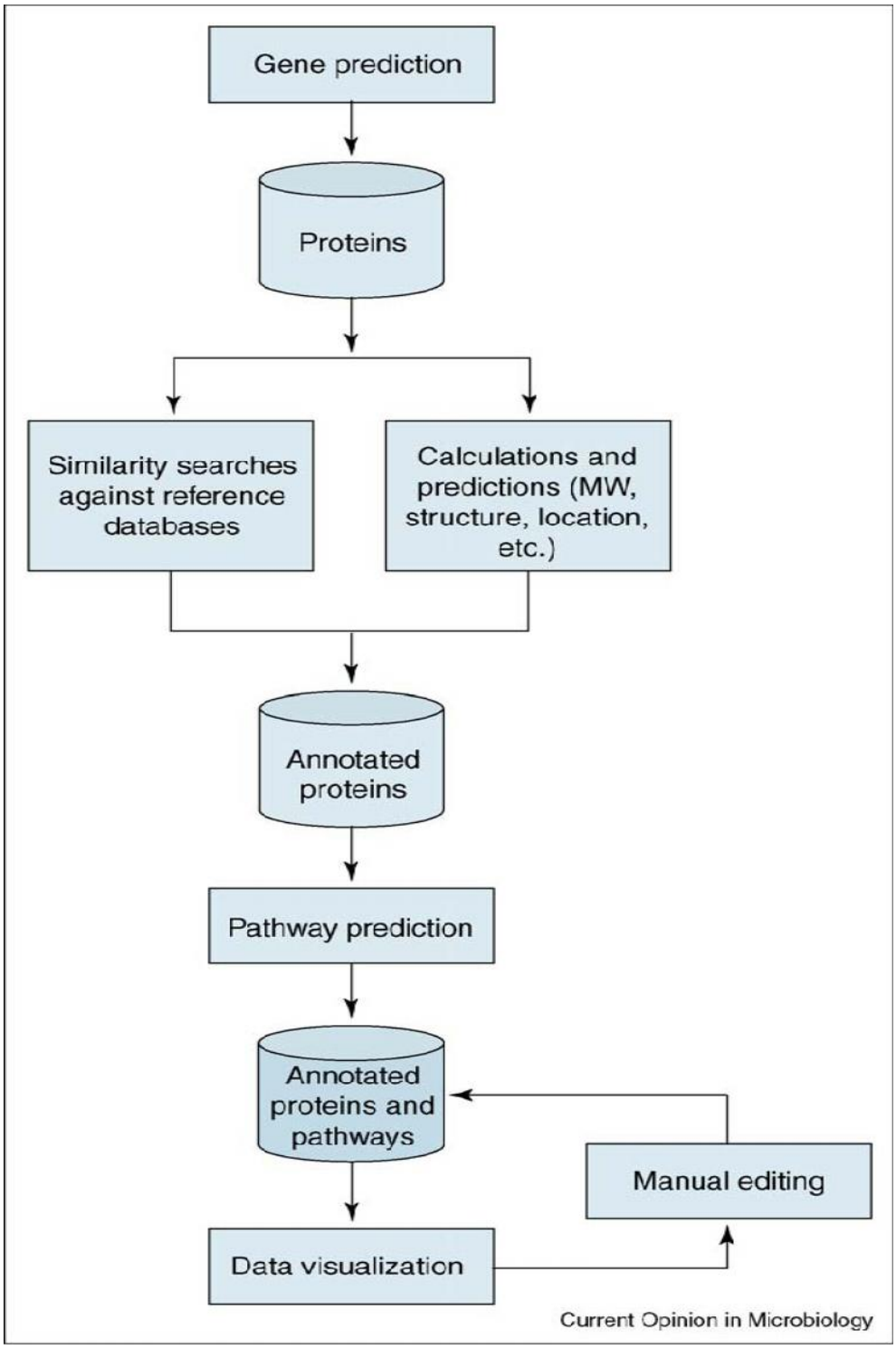
The first version of NCBI's Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) combiningHMM-based gene prediction algorithm with a protein sequence similarity search methods was developed in 2000-2001. The pipeline predicts protein-coding genes, structural RNAs (5S, 16S, and 23S), tRNAs, and small non-coding RNAs. The NCBI PGAP is available as a service for GenBank submitters. The pipeline is capable of annotating both complete genomes and draft WGS genomes consisting of multiple contigs. The submitters can request PGAP annotation during submission of the genome to GenBank. The initial pipeline used a combination of automatic protein coding gene model prediction via two prediction methods, Gene Marks (**Besemer *et al,* 2001**) and Glimmer (**Delcher *et al,* 1999**).Genes of ribosomal RNAs were predicted either by the BLASTn sequence similarity search using the entries from the RNA sequence database as queries or by running specialized tools, such as Infernal (**Nawrocki and Eddy, 2013; Nawrocki, 2014**) and Rfam (**Nawrocki *et al,* 2015**). Genes of transfer RNAs were predicted using tRNAscan-SE (**Lowe and Eddy, 1997**).The new PGAP version employs a robust automatic system optimized for high throughput. PGAP is able to annotate more than 1200 genomes a day. NCBI's PGAP is an evolving system. The main difference between the old and new version is the use of GeneMarkS+ tool for gene finding (**Tatusova *et al,* 2016**).

For additional information on PGAP see

https://www.ncbi.nlm.nih.gov/genome/annotation_prok/

and the NCBI handbook;

https://www.ncbi.nlm.nih.gov/books/NBK174280.


Gene prediction
Proteins
Similarity searches against reference databases
Calculations and predictions (MW, structure, location, etc.)
Annotated proteins
Pathway prediction
Annotated proteins and pathways
Manual editing
Data visualization
Current Opinion in Microbiology

Material & Methods

Material and Methods

4.1. Microbial strain and media

The microorganism used in this study was *Bacillus* sp. C_4; CCUG 66887. It has been isolated from the compost and identified using biochemical tests and 16S rDNA technique (GenBank accession: FJ214667).This bacterium is able to secrete extracellular protease enzyme upon growing on several proteinaceous substrates. It has broad substrate specificity; it can degrade chicken feather and sheep wool keratins as well. It was able to degrade native chicken feather within 48 hours (**Fellahi *et al,* 2014**).

Several culture media were used to cultivate the strain including: peptone yeast extract (PY) medium (10 g bactopeptone, 5g Difco yeast extract and 5g sodium chloride per liter) (**Bernhardt *et al,* 1978**) and PA medium which is PY supplemented with 1.5 % agar; used for the activation of *Bacillus* sp. C_4. Skimmed milk agar (SMA) medium contained: 0.5g yeast extract, 10g skimmed milk and 15g agar per liter. This medium was used for testing the proteolytic activity of the strain. Sporulation medium (modified Schaeffer's medium, 1XSG) (**Leighton and Doi, 1971**) contained 3g beef extract, 5g bactopeotone, 2g KCl per liter, after autoclaving, several separately sterilized compounds were added as follow: $MgSO_4.7HO_2$ was added to a final concentration of 2 mM, $CaCl_2$ to 1mM, $MnCl_2$ to 0.1mM, $FeSO_4$ to 1 mM, and glucose to 0.1% (w/v). Luria-Bertani broth (LB) (**Luria *et al,* 1960**) contained per liter 10g treptone, 5g yeast extract, and 10g NaCl. Nutrient broth (NB) composed of 5g peptone and 3g beef extract per liter. The pH of the above media was adjusted to pH 7.0 using either NaOH as a base and HCl as an acid.

4.2. Proteolytic activity

Proteolytic activity was measured as described by Cliffe and Law (1982) using hide powder azure (HPA, Sigma) as a substrate. For a standard assay, 1 ml 0.1M Tris-HCl buffer, pH 8 was added to 5 mg of HPA (insoluble material) in 5 ml reaction tube. A hundred µl of an appropriately diluted enzyme preparation was added to the reaction mixture. The mixture was

incubated at 37°C until a measurable blue color was developed. The reaction was stopped by chilling the assay tubes on ice, after which the tubes were centrifuged at 7000 rpm for 5 minutes and the absorbance of the clear supernatant was measured at 595 nm. A control reaction was prepared by adding 0.1M Tris-HCl buffer instead of enzyme solution. One unit of enzyme was equal to the amount of enzyme that develops a change in absorbance (0.1) against the control at 595 nm per 30 minutes at 37°C using HPA as substrate.

4.3. Determination of protein concentration

Soluble protein concentration was determined by the method of Lowry *et al,* 1951. Protein sample (Up to 0.5 ml) was added to 2.5 ml of protein assay mixture [49 ml of 4% (w/v) Na_2CO_3, 49 ml of 0.2 N NaOH, 1 ml of 1% (w/v) $CuSO_4$. $5HO_2$, and 1 ml of 2% sodium potassium tartrate]. Reaction tubes were mixed well and were allowed to stand at room temperature for 10 minutes, after which, 0.25 ml of diluted Folin reagent (1X) was added. The developed color was measured at 750 nm after 30 minutes. A standard curve was established using bovine serum albumin (BSA).

4.4. Activation of the *Bacillus* sp. $C_4$and inoculum preparation

Several recently growing colonies of *Bacillus* sp. $C_4$from overnight PA plate were used to inoculate 20 ml PY medium. This mother culture was grown at 37°C with shaking at 160 rpm for 4 hours to reach absorbance 3 at 420nm (corresponding to approximately 3.0 $x10^9$colony forming unit per ml). One milliliter (1 ml) of the growing culture was centrifuged at 7,000 rpm for 3 min using a micro-centrifuge. The bacterial pellet was washed with saline [0.85% (w/v) NaCl] and was used to inoculate culture media. All experiments were carried out in250ml Erlenmeyer flasks containing 50 mL of medium.

4.5. Optimization of extracellular protease production

In an attempt to enhance the extracellular protease production of *Bacillus* sp. C_4 a two-step procedure was adopted. One variable at a time (O.V.A.T.) approach was first applied to identify the most effecting factors controlling the production of this enzyme followed by response surface methodology (RSM) to optimize these significant factors. In each step and for each tested parameter, protease activity in the cell-free supernatant and protein content were determined after 24 hours.

4.5.1. Selection of significant factors affecting protease production through (O.V.A.T.) approach

A variety of culture media were used and many physical factors (inoculum size, initial pH, temperature, and agitation rate) were assessed for their effect. Several nutritional parameters including carbon sources (starch, glucose, glycerol, Tween-20 and Tween-80), several nitrogen sources (peptone, beef extract, yeast extract and sodium nitrate), and some inorganic salts (KCl, $CaCl_2$, $CaCO_3$, $FeSO_4$, $MnCl_2$, and $MgSO_4$) were tested separately. For each tested factor the protease production in terms of specific protease activity (U/mg protein) was determined.

4.5.1.1. Effect of Culture medium

The bacterial strain was grown on several commonly used laboratory media including peptone-yeast extract (PY and 1/2 PY), sporulation medium (1XSG and 1/2XSG), nutrient broth (NB and 1/2 NB) and Lauria-Bertani medium (LB and 1/2 LB). The pH of these media was adjusted to pH 7. The pellet of one ml of mother culture was used to inoculate 50 ml of each medium in 250 ml Erlenmeyer flask. Cultures were incubated at 37°C with shaking at 160rpm for 24 hours. For each culture medium the specific protease activity was measured.

4.5.1.2. Effect of inoculum size

The effect of inoculum size on protease production was determined by inoculating 50 mL 1XSG medium, pH 7 with several inoculum sizes (1, 2, 4, 6, and 8%). Each milliliter contains approximately 3.0 $x10^9$colony forming unit. The inoculated flasks were incubated at 37^oC with shaking at 160 rpm for 24 hours. For each tested inoculum size the specific protease activity was determined.

4.5.1.3. Effect of initial pH

To evaluate the influence of initial pH, 1XSG medium prepared at different pHs (6, 7, 8, and 9) was inoculated with mother culture and cultures were grown at 37^oC with shaking at 160 rpm for 24 hours. Specific protease activity was determined for each assessed phi These pHs levels were chosen based on literature data used for this type of enzyme (**Sharma *et al,* 2017**).

4.5.1.4. Effect of incubation temperature

Incubation temperature effect on the production of extracellular protease activity by *Bacillus* sp. C_4 was evaluated by incubating the inoculated cultures (1XSG, pH7) at different temperatures (30, 37, and 45^{o}C) under shaking for 24 hours (**Sharma *et al,* 2017**). Specific protease activity was determined for each assessed temperature.

4.5.1.5. Effect of agitation rate

Inoculated culture media (1XSG, pH7) was grown at 37°C for 24 hours and under several agitation speeds (100, 160, 200, and 250 rpm). Specific protease activity was determined as above.

4.5.1.6. Effect of different carbon and nitrogen sources

To evaluate the effect of different carbon sources on the protease production, the liquid medium 1XSG was used as a basic production medium and supplemented separately with various concentrations of glucose (0.2%), starch (0.3%), glycerol (0.2%), Tween-20 and Tween-80 (0.2%). Likewise the effect of different nitrogen sources on protease production was studied by supplementing the 1XSG medium with various concentrations of complex nitrogen sources such as beef extract (0.6% and 0.9%), peptone (1%), yeast extract (0.2%), and inorganic sources (0.5%, 1.5%, and 2% of sodium nitrate).

4.5.1.7. Effect of metal ions

The effect of the following metal ions KCl (0.4%), $FeSO_4$ (2 mM), $MnCl_2$ (0.5 mM), $CaCl_2$ (2mM), and $MgSO_4$ (4mM) was tested.

4.6. Isolation of keratinase gene (s) from *Bacillus* sp. C_4

As stated above the proteolytic activity of *Bacillus* sp. $C_4$was broadly specific, and the bacterium could grow and produced a significant level of keratinase when using sheep wool and chicken feather as substrates. A total hydrolysis of the keratinous waste was obtained in less than three days (**Fellahi *et al,* 2014**).In an attempt to increase our understanding of the *Bacillus* strain's ability to hydrolyze both alpha- and beta-keratins, we in this study aimed to isolate and identify the keratinolytic gene (s).

4.6.1. Genomic DNA isolation

An overnight culture of *Bacillus* sp. C_4 grown on the optimized production medium at 37°C was used for the isolation of total genomic DNA. Genomic DNA was isolated using Master Pure™ Gram Positive DNA Purification Kit (Epi centre, Germany) following the manufacturer's standard protocol. The concentration and the purity of the prepared DNA were determined using NanoDropND-2000 UV-VIS spectrophotometer as well as by running small aliquots on agarose gel (1 %) electrophoresis.

4.6.2. Agarose gel electrophoresis

Agarose gel electrophoresis was carried out essentially as reported earlier (**Sambrook *et al.*, 1989**). Agarose gel used to analyze DNA was prepared by adding 1 g of agarose to 100 ml (1X) Tris-borate (TBE) buffer. The mixture was completely dissolved by heating. The solution was allowed to cool to 50ºC, after which 5 µl of 10 mg/ml ethidium bromide stock solution was added. The solution was poured into a suitable gel tray (10×8 cm). The gel was left to cool at room temperature for approximately 15 minutes, after which the gel was placed appropriately in an electrophoresis apparatus (Hoefer Scientific Instrument, USA). After removing the well-forming comb, the gel was covered with electrophoresis buffer (1X TBE buffer). DNA samples were mixed with the loading dye (6X) on 6:1 volume bases, after which the DNA samples were loaded into the gel and electrophoresis was carried out at 80 volts for about 45 minutes. The gel was visualized under UV-trans-illuminator (Ultra-Violet Products Pvt Ltd, Cambridge UK) and photographed. The size of DNA fragments of unknown size, migrating in the agarose gel, was estimated by comparing their migration with those of GeneRuler^{TM}1Kb DNA Ladder (Invitrogen, USA).

4.6.3. PCR amplification of Keratinase gene (ker gene)

Multiple sequence alignement program CLUSTALW2 (**Chenna *et al.*, 2003**) (http://www.ebi.ac.uk/Tools/msa/clustalw2/), was used to align keratinase gene sequences from different *Bacillus pumilus* strains to investigate the resemblance among the sequences to be able to design the first primer set (Primer set A, Table 5). All primers used in this study were purchased from Eurofins Genomics, Germany. Standard PCR was performed using 100ng bacterial genomic DNA in a 50µl reaction mixture containing 1X amplification buffer (Dream Taq^{TM} Green Buffer Fermentas, Germany), 0.2 mmol L^{-1} dNTPs, and 0.2µmol L^{-1} of each primer with 5 U of Taq polymerase (Dream Taq^{TM}Green DNA polymerase Fermentas, Germany). PCR was performed in a Thermocycler (Progene, England) programmed for 30 cycles, with an initial denaturation at 95°C for2 minutes and each cycle with denaturation at 95°C for 30 seconds (sec), annealing at 51°C for 30 secondes, extension at 72°C for 75

secondes; final extension at 72°C for10 minutes. The PCR product was then sent to Eurofins Genomics, Germany for sequencing and from the retrieved DNA sequence, a two new primer sets (Primers sets B and C, Table 5) were designed to get the full open reading frame (ORF) of the gene. The PCR conditions for them were similar, except the annealing condition of 55˚C for 30secondes.PCR products in each time were resolved by electrophoresis on 1% agarose gel, eluted with gel elution kit (Jetquick, Germany) and sequenced at Eurofins Genomics, Germany.

Table 5: Oligonucleotide primers used in PCR reactions.

Primer	**Primer sequence 5' - 3'**
FwA	ATGTGCGTGAAAAAGAAAAATGTG
RvA	TTAGAAGCCGCTTGAACGTTA
FwB	AAGTATTAGATCGTTACGGCGATGGAC
RvB	CCAAGAACACCAATCGTGTTATCAAGG
RvC	TCAGGGCTTTCCCAAGTCAAT

4.6.4. Characterization of the nucleic acid sequence and deduced amino acid sequence

Sequence similarity analysis was done on the web server of the National Centre for Biotechnology and Information (NCBI, http://www.ncbi.nlm.nih.gov/) in the non-redundant nucleotide (nr-nt) or amino acid (nr-aa) databases using BLASTN, BLASTP and BLASTX programmes. The ker gene sequence was subsequently submitted to the NCBI GenBank. Motif search for amino acid patterns in the deduced protein sequence was performed using the Scan Prosite tool in the ExPASy proteomics server.

4.7. Library preparation and genome sequencing of *Bacillus* sp. C_4

To screen the genome of feather-degrading *Bacillus* C_4 for additional putative keratinase genes, sequencing of the whole genome using the bench-top sequencer Illumina MiSeq and the MiSeq Control Software 2.3.0.3 was performed by Eurofins Genomics, Germany.

Combining short-insert, paired-end and long-insert, mate pair sequences is the ideal way to maximize coverage of the genome for *de novo* assembly (**Mardis, 2013**). In this regard, the *Bacillus* sp. C_4genomic DNA was isolated as stated in section 4.6.1 and used to prepare two

types of sequencing libraries: one genomic shotgun (SG) paired-end library with a fragment size approximately between 70 and 150 bp and one proprietary long jumping distance (LJD) library (with a jumping distance of 8 kbp and library insert size of approximate 400-700 bp) with a jumping with an average insert size of 7865 bp. Before sequencing a phiX library were added to estimate the sequencing quality. The libraries were placed in the reagent cartridge and loaded on the instrument along with the flow cell. All subsequent steps were performed on Illumina MiSeq v2 including, cluster generation and 2x150 bp paired-end sequencing.

Illumina MiSeq, a bench-top sequencer, is especially convenient for amplicon and small bacterial genome sequencing. It could sequence 150 pair-ends and generate 1.5Gbp/run in about 10 hours including sample and library preparation time (**Liu *et al,* 2012**).

4.8. *De novo* assembly of the *Bacillus* sp. C_4 genome

The resulting sequence reads were joined into longer continuous stretches of sequence (contigs) which were then joined to form longer stretches of sequence (known as scaffolds). *De novo* sequence assembly and scaffolding of trimmed reads was done by Eurofins Genomics in Germany using Illumina's in-house pipeline based on the Newbler assembler software v2.9 (20130529_1641). Although originally intended as assembler for Roche 454 and Sanger data only, new versions of this software are able to efficiently process smaller amounts of Illumina data. Especially for hybrid assemblies and smaller genome sizes, the Newbler pipeline is a well-performing alternative for the standard in-house Velvet pipeline.

Prior to assembly read sequences were analyzed and clipped using the software Trimmomatic (version 0.32, http://www.usadellab.org/cms/index.php?page=trimmomatic). Low quality bases with a Phred score <20 were removed from the start as well as from the end of each single read. In addition, reads that did not fulfill the average quality threshold with a window size of 4 bp and Phred score <20 were removed as well as reads shorter than a length threshold of 70 pb. To obtain high quality long jumping distance pairs without any adapter or spacer sequences a sequence length cut-off of 30 base pairs is applied for paired sequences as well as for singleton (single end) sequences and length filtering of processed LJD read pairs (minimum length 75 bp) was done by Newbler pipeline. Paired reads become singleton reads if the other read of a pair is shorter than the length cut-off after the processing. Various filtering steps were also applied to the illumina sequences in order to remove phiX related reads.

4.9. Annotation of *Bacillus* sp. C_4 genome

The draft genome sequence of *Bacillus* sp. $C_4$was annotated by submitting it to the web based automated Bacterial Annotaion System (BASys) available on https://www.basys.ca/ (**Van Domselaar *et al,* 2005**) and the NCBI Prokaryotic Genomes Annotation Pipeline (PGAP) version 2.10.The annotated genome sequence was then deposited in the NCBI GenBank database (http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html).

4.10. Identification of active proteins produced by *Bacillus* sp. C_4 and involved in protein degradation

Nano-scale liquid chromatography-electro-spray ionization tandem mass spectrometry (nLC-ESI-MS/MS) is a well-established technique for identification of peptides from complex mixtures (**Yang *et al,* 2007**).In a HPLC-MS/MS based peptide sequencing, a complex protein mixture is often separated via gel electrophoresis first to simplify the sample. Subsequently, proteins are digested with a specific enzyme such as trypsin. To further reduce sample complexity, peptides are separated by liquid chromatographic systems (LC): eluents are ionised, separated by their mass over charge ratios and subsequently registered by the detector. Each MS/MS spectrum (corresponding to a specific peptide sequence) is then used to search protein database for matched peptides. A Protein hit is identified often by multiple independently sequenced peptides from the same protein.

nLC-ESI-MS/MS followed by Mascot database search is used to identify the secreted enzyme (s) involved in protein degradation by *Bacillus* sp. C_4. Mascot is a software package from Matrix Science (http://www.matrixscience.com/search_form_select.html) that interprets mass spectral data into protein identities. Mascot compares the observed spectra to a database of known proteins and determines the most likely matches.

4.10.1. Extracellular protease fraction preparation from *Bacillus* sp. C_4

Bacillus sp. $C_4$was grown on the production medium for 24 hours at 37^oC with shaking. Cell-free supernatant was obtained by pelleting bacterial cells through centrifugation at 8,000 rpm for 15 minutes at 4°C (Optima Max-XP, Beckman Coulter). The enzyme fraction was precipitated by adding solid ammonium sulfate (NH_3SO_4) to 65% saturation at 4°C and slowly mixing for 1hour using an ice bath. The precipitate was then collected by centrifugation at 12,000 rpm for 30 minutes using the above rotor at 4^oC. Pellet was resuspended in a small volume of 20 mM Tris-HCl buffer, pH8 and dialyzed overnight against the same buffer. Protein content and proteolytic activity were determined in the crude and dialyzed preparation as described earlier.

Dialysis tubing was prepared as follow. Dialysis tubing was cut into pieces of convenient length (10 – 30 cm), soaked in distilled water for 10 minutes, after which the pieces were

soaked in 50 % ethanol for about 10 minutes. The dialysis tubing was boiled in a solution of 2% sodium bicarbonate and 1 mM Ethylenediaminetetraacetic acid (EDTA) for 15 minutes. The tubing was allowed to cool then stored in 50 % ethanol at 4°C. Before use, the tubing was carefully washed from inside and outside with distilled water then with the buffer to be used.

4.10.2. SDS polyacrylamide gel electrophoresis

The dialyzed fraction wasrunon12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) as described earlier (**Laemmli, 1970**). SDS-PAGE (12% resolving gel) was prepared by mixing 1.725 ml of distilled water; 2 ml of 30% acrylamide-bisacrylamide mixture [29.2 g acrylamide, 0.8 g bisacrylamide in 100 ml distilled water]; 1.25 ml of 1.5 M Tris-HCl, pH 8.8 (lower Tris-HCl buffer); 0.05 ml 10% of SDS; 0.05 ml of 10% ammonium persulfate; and 0.004 ml of N, N, N, N, tetrametylenediamine (TEMED). The above mixture was poured in an assembly unit (10 x 10 cm) (Mighty small ^{TM} SE 245, Hoefer, USA) and overlayed carefully with isopropanol. The gel was allowed to be polymerized at room temperature for about 15 minutes. Stacking gel (5%) was prepared by mixing 2.7 ml distilled water; 0.67 ml of 30% acrylamide-bisacrylamide mixture; 0.5 ml of 1.0 M Tris-HCl , pH 6.8 (upper Tris-HCl buffer); 0.04 ml of 10 % SDS; 0.05 ml of 10% ammonium persulphate and 0.004 ml of TEMED. After the lower running gel was set, the isopropanol was decanted and the gel surface was washed with distilled water after which the stacking gel mixture was poured to the top of the small plate. Wells forming comb was placed in the gel. After polymerization of the gel, the comb was removed carefully and the wells were washed with the running buffer. The assembled gel plates were transferred to the electrophoresis tank and covered with SDS-running buffer [0.05 M tris-HCl, pH 8.3, 0.384 M glycine, and 0.1% (w/v) SDS]. Protein sample was prepared by mixing 50 µg proteins with 5X sample application buffer (SAB) [0.6 M Tris-HCl, pH 6.8, 1% (w/v) SDS, 10 % β-mercaptoethanol, 10 % sucrose and 0.05 % bromophenol blue] to give a final concentration 1X (SAB), boiled in water bath for 3 minutes at 95°C and was applied to the slab gel along with a molecular weight protein marker (All BlueProtein Precision Standard, Bio-Rad). Electrophoresis was carried out at a constant voltage 150 volts for about 1.5 hours. The gel was then stained with Coomassie Brilliant Blue R-250 [0.1 % Coomassie Brilliant Blue R-250 in 50 % Methanol, 10 % acetic acid] for 3 hours with gentle agitation at room temperature. The gel was destained overnight using a destain solution (100 ml methanol, 70 ml acetic acid and 830 ml distilled water).

4.10.3. NanoHPLC-ESI-MS/MS strategy

Protein identification of the enzyme fraction produced upon growing *Bacillus* sp. C_4 on production medium was performed by Proteome Factory AG, Germany. Two protein spots from a 12% SDS-PAGE-gel were cut out and digested in-gel by trypsin (Promega, Mannheim, Germany) and analyzed by nano HPLC-ESI-MS/MS.

The LC-MS/MS system consisted of an Agilent 1100 nanoHPLC system (Agilent, Waldbronn, Germany), PicoTip electrospray emitter (New Objective, Woburn, MA), and an Orbitrap XL or LTQFT mass spectrometer. The retrieved peptides were analyzed using an ultra-mass spectrometer (ThermoFisher Scientific, Bremen, Germany). Peptides were first trapped and de-salted on the enrichment column Zorbax 300SB-C_{18}, 0.3 x 5 mm (Agilent) for five minutes (solvent: 2.5% acetonitrile/0.5% formic acid), then separated on a Zorbax 300SB-C_{18}, 75 µm x 150 mm column (Agilent) using a linear gradient from 10% to 32% B (Solvent A: 5% acetonitrile in water, Solvent B:acetonitrile. Both solvents contained 0.1 % formic acid). Ions of interest were data-dependently subjected to tandem MS according to the expected charge state distribution of the peptide ions.

4.10.4. Data Analysis and Protein Identification

Protein identification of the crude enzyme fraction involved in protein utilization by *Bacillus* sp. $C_4$was performed by Proteome Factory AG, Germany. MS/MS data were identified by database search against the *B. pumilus* entries from the RefSeq protein database available at NCBI, which was appended to an existing bacterial database. The search parameters were set to: 3 ppm peptide mass tolerance (MS accuracy), 0.6 KDa fragment mass tolerance (MS/MS accuracy), maximum one missed cleavage by trypsin, fixed modification of cysteine and variable modification of oxidized methionine. MS/MS ion search of the Mascot search engine (Matrix Science, London, England) was performed, and only peptide matches with a score of 26 or above were accepted. The search results were also run against the amino acid sequences retrieved from the *Ker1* and *Ker2* genes.

Results & Discussion

V. Results and Discussion

5.1. Optimization of extracellular protease production from *Bacillus* sp. C_4

There is no general medium for protease production and each strain has its own conditions for maximum enzyme productivity (**Reddy *et al,* 2008**). Process optimization is important in industrial production processes, in which even small improvement scan be decisive for commercial success (**Cheng *et al,* 2012**). Therefore, the required constituents and their concentration shave to be optimized accordingly.

The optimization of culture medium was carried out by a combination of non-statistical methodology and statistical methodology-based experimental design. A series of experiments were first carried out to study the effects of culture media as well as the physico-chemical and nutritional factors on the extracellular protease production by *Bacillus* sp. C_4 with one-factor-at-a-time experiments. The second approach deals with optimizing the significant factors that selected in the first step by response surface methodology.

5.1.1. Selection of the most important physico-chemical determinants and the most suitable carbon & nitrogen sources via OVAT

5.1.1.1. Effect of culture medium

The ability of *Bacillus* sp. C_4 to produce extracellular protease was monitored in various commonly used laboratory media. These media include PY and 1/2 PY medium, 1XSG and 1/2 XSG medium, NB and 1/2 NB medium, LB and 1/2 LB medium, as stated in the Materials and Methods.

Bacillus sp. C_4 cells produced protease enzyme when grown in all the above media. However, the amount of the secreted enzyme varied extensively. The medium 1XSG was the most suitable medium for the production of active protease giving 213.69 Units/mg protein after 24 hours of incubation at 37^{o}C, while the medium 1/2 XSG was the worst, Table 6. This could be due to the fact that sporulation medium (1XSG) normally gives a

short logarithmic phase and an extended stationary one which is convenient for the protease production. These results are in agreement with those reported earlier (**El-Dadaly and Zaid, 1999**). Therefore, 1XSG medium was used as the reference culture medium in the following studies.

5.1.1.2. Effect of inoculum size

The optimization of inoculum size is quite important, as high inoculum density can attenuate enzyme production due to competition for available nutrients. In a similar manner, low density can result in a delay of enzyme secretion, due to the drop in cellular number. Different inoculum sizes were used to inoculate the 1XSG medium at the level of 1, 2, 4, 6, and 8%. Proteolytic activity was optimally produced at the inoculum size of 2% (about 31.6 x 10^8 CFU/mL), (Table 6).

5.1.1.3. Effect of pH

The effect of initial pH of the medium on the production of protease by *Bacillus* sp. $C_4$was also tested. Cells produced extracellular protease optimally at pH 7, (Table 7). Proteases from *B*. *mojavensis* (**Beg *et al*, 2003)** and *Bacillus* sp. RGR-14 (**Chauhan and Gupta, 2004**) were generally produced at a medium with an initial pH around 7. However, protease production in a medium of starting pH higher than 7 (pH 8, 9, 10, and 12) has been also reported (**Tari *et al*, 2006; Khusro, 2015; Lakshmi and Hemalatha, 2016; Kshetri1, 2016; Lakshmi *et al*, 2014**).

5.1.1.4. Effect of temperature

Temperature is another critical parameter that needs to be controlled. *Bacillus* sp. C_4 was incubated at 30°, 37° and 45°C in 1XSG medium for 24 hours to determine which temperature is suitable for maximal protease production. At 37°C *Bacillus* sp. C_4 produced the maximal enzyme activity (268.37 U/mg) after 24 hours of incubation period. A sharp decline in proteolytic enzyme production was observed, when the incubation temperature was increased from 37° to 45°C. However, proteolytic activity produced by this bacterium at 30°C was closer to that of 37°C, (Table 6).In contrast, an optimum temperature of 45°C and 55°C for protease production by *B*. *subtilis* and *Saccharomonospora viridis* SJ-21 has been reported (**Pant *et al*,2015; Jani *et al*, 2012**).

5.1.1.5. Effect of agitation rate

As shown in Table 6, the agitation rate of 160 rpm favor maximum proteolytic activity and its increase drastically lowered the enzyme yield. In some cases, higher agitation rates have been shown to decrease the enzyme production which can be due to the denaturation of the enzyme caused by higher agitation rates (**Lee *et al,* 2002**). Moreover, excessive agitation and aeration may lead to cell lysis (**Darah and Ibrahim, 1996; Kumar *et al,* 2016**). However, higher agitation rates have been reported to be good for protease production from *B. amyloliquefaciens* (**Cheng *et al,* 2012**), *B. licheniformis* (**Suganthi *et al,* 2013**), *B. clausii* (**Oskouie *et al,* 2008**) and isolate AP11 (**Kumar *et al,* 2016**).

5.1.1.6. Effect of some carbon and nitrogen sources as well as metal ions on extracellular protease production

In an attempt to optimize the 1XSG medium as a basic and a reference medium for protease production by *Bacillus* sp. C_4, it was supplemented with several carbon and nitrogen sources as well as some salt ions, (Table 8).Carbon sources were 0.2% of Tween-20, Tween80, glycerol, and glucose and 0.3% starch. For nitrogen sources, peptone (1%), beef extract (0.6 and 0.9%), yeast extract (0.2%), and $NaNO_3$ (0.5, 1.5, and 2%) were used. The medium was also supplemented with the following metal ions KCl (0.4%), $FeSO_4$ (2mM), $MnCl_2$ (0.5mM), $CaCl_2$ (2mM), $CaCO_3$ (0.3%), and $MgSO_4$ (4mM), (Table 7).

Data from Table 7 indicated that among the different carbon sources tested, starch had slightly increased the protease production over the reference medium (1XSG), while Tween-20, Tween-80 and glycerol did not support the production. Strong catabolite repression was observed when glucose concentration was higher than that cited in the reference medium (Table 6). This phenomenon is well known in the control of protease production in several microorganisms as well (**Gupta *et al,* 2002b, Gupta *et al,* 2002c, Beg *et al,* 2003**).Our results are in accordance with those obtained by other investigators, where glycerol did not act as an inducer for protease production (**Liu *et al,* 2005**). Interestingly, according to several reports many surfactants acted as inducers for proteolytic synthesis in other cases (**Kanekar *et al,* 2002; Jani *et al,* 2012**).

Table 6: Effect of different laboratory media, inoculum size, initial pH, incubation temperature, and agitation rate on the production of protease by *Bacillus* sp. C_4 cells.

Variable or Determinant	Proteolytic activity (Units/ml)	Protein content (mg/ml)	Specific proteolytic activity (Units/mg protein)
Media			
PY	244.28	2.99	81.69
½ PY	270.00	1.87	144.38
1XSG	312.00	1.46	**213.69**
½ XSG	11.22	0.82	13.68
NB	95.71	1.32	72.50
½ NB	99.28	0.85	166.80
LB	35.06	1.35	25.97
½ LB	230.71	2.00	115.25
Inoculum size (%)			
1	352.28	1.64	214.80
2	389.14	1.45	**268.37**
4	282.85	1.55	182.84
6	298.57	1.57	190.17
8	297.14	1.50	198.09
Ph			
6	11.19	2.07	5.40
7	389.14	1.45	**268.37**
8	361.71	1.55	233.36
9	302.14	1.52	198.77
Temperature ^oC			
30	345.71	1.36	254.19

37	389.14	1.45	**268.37**
45	14.83	1.99	7.43
Agitation rate (rpm)			
100	257.50	1.44	178.81
160	389.14	1.45	**268.37**
200	69.10	1.80	38.38
250	10.04	1.58	6.35

Concerning the nitrogen sources tested (peptone, beef extract, yeast extract, and $NaNO_3$), yeast extract exerted the most significant effect on the production of extracellular protease by *Bacillus* sp. C_4 and higher specific protease activity was achieved (419.06 U/mg), (Table 7). Similar reports described an enhanced protease production in the presence of yeast extract (**Reddy *et al,* 2008; Krishnan *et al,* 1998; Cheng *et al,* 2012**). However, peptone and beef extract in amounts beyond those used in the reference medium (1XSG) caused a significant reduction in the enzyme production. On the other hand, sodium nitrate ($NaNO_3$) did not have any significant effect on enzyme production, which was almost similar. However, $NaNO_3$was a stimulatory chemical for protease production by many microorganisms (**Beg *et al,* 2003; Chi *et al,* 2007; Rathakrishnan and Nagarajan, 2013**).

The effects of increase in the level of the existing metal ions had adverse effects on enzyme production except for $MnCl_2$, where the protease activity was almost constant (Table 7). However, the deletion of metal ions from the reference medium caused decrease in the proteolytic activity over the control, and this may be attributed to the important role of Mn, Mg, and Fe metals in the sporulation process and in other spore properties such as heat resistance and dormancy (**Kihm *et al,* 1988**).

Table 7: Protease production by *Bacillus* sp. C_4 in 1XSG medium supplemented with different concentrations of carbon, nitrogen sources, and metal ions. Cells were grown at 37^{o}C for 24 hours.

Substrates supplied	**Enzyme activity (mg mL^{-1}protein**	**Specific activity (**
Reference medium	389.14	1.45	268.37

(1XSG)			
Carbon sources			
Tween-20 (0.2 %)	0.0	N.D	0
Tween-80 (0.2 %)	0.0	N.D	0
Glycerol (0.2 %)	0.0	N.D	0
Starch (0.3%)	318.0	1.01	314.85
Glucose (0.2 %)	0.0	N.D	0
Nitrogen sources			
Peptone (1%)	186.90	1.93	96.83
Beef extract (0.6%)	220.32	1.36	162.0
Beef extract (0.9%)	284.28	1.99	142.85
Yeast extract (0.2%)	519.64	1.24	**419.06**
$NaNO_3$ (0.5%)	316.11	1.26	250.88
$NaNO_3$ (1.5%)	366.85	1.24	295.84
$NaNO_3$ (2%)	314.57	1.25	251.65
Metal ions			
KCl (0.4%)	186.90	1.93	96.83
$FeSO_4$ (2mM)	186.90	1.93	96.83
$MnCl_2$ (0.5mM)	383.81	1.28	299.85
$CaCl_2$ (2mM)	186.90	1.93	96.83
$MgSO_4$ (4mM)	186.90	1.93	96.83
Without metal ions	331.71	1.70	195.12

N.D: Not detected.

5.2. Isolation of keratinase gene (s) from keratin-degrading *Bacillus* sp. C_4

As reported earlier (**Fellahi *et al,* 2014**) protein-degrading *Bacillus* sp. C_4 could grow and produced a significant level of keratinase when using wool and chicken feather as substrates. A total hydrolysis of the

feather keratin was obtained in less than three days. Therefore, a keratinolytic gene was isolated using three sets of designed primers.

Polymerase Chain Reaction was carried out to isolate keratinase gene from strain C_4. The PCR product was found to be less than 1.5 Kb as shown in Figure 10. Sanger sequencing revealed one major open reading frame of 1152 bps (Figure 11), which encodes a polypeptide of 383 amino acid residues. The analysis of the deduced amino acid sequence of *ker* gene of the proteolytic *Bacillus* sp. C_4 showed 100% identity with alkaline serine proteinase from *B. pumilus* (ACO94164.1) and Lehensis serine protease from *B. lehensis* (AFP23380.1) and 99% identity to alkaline serine protease from *B. pumilus* (BAE79641.1). It exhibited also significant similarity to some known keratinases indicating its keratinolytic activity. It shares 99% identity with dehairing protease precursor from *B. pumilus* (AAR19220.1), 96% similarity to keratinase precursor (ACM47735.1) and keratinase (ADK11996.1) from *B. pumilus,* Table 8. Keratinase precursor has been reported to degrade chicken feather.

The three residues at the active site that form the catalytic triad and which have been found in all serine proteases are Asp-140, His-172 and Ser-329 in polypeptide sequence of *ker* gene. They are predicted using the ScanProsite tool in the ExPASy proteomics server. This result confirmed that the isolated DNA fragment (*Ker* gene) encodes one of the serine proteases produced from the keratin-degrading *Bacillus* sp. C_4 strain. The gene was subsequently submitted to the NCBI GenBank (accession number: KX184831).

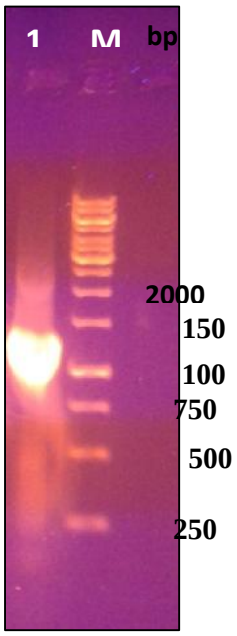


Figure 10:1% agarose gel electrophoresis of PCR product (lane 1) generated from *Bacillus* sp. C_4 using keratinase gene specific primers. Lane M represents 1Kb DNA Ladder.

```
1ATG TGC GTG AAA AAG AAA AAT GTG ATG ACA AGT GTT TTA TTG GCT GTC CCT CTT CTG TTT TCA
1 M   C   V   K   K   K   N   V   M   T   S   V   L   L   A   V   P   L   L   F   S
64   GCA GGG TTT GGA GGC TCG ATA GCA AAT GCC GAG ACT GCC TCA AAG TCA GAA AGC GAA AAA
22  A   G   F   G   G   S   I   A   N   A   E   T   A   S   K   S   E   S   E   K
124 AGC TAT ATC GTT GGC TTT AAA GCT TCT GCC ACC ACA AAC AGC TCT AAG AAA CAA GCA GTG
42  S   Y   I   V   G   F   K   A   S   A   T   T   N   S   S   K   K   Q   A   V
184 ACA CAA AAT GGT GGG AAA TTA GAA AAA CAA TAC CGC CTC ATT AAT GCT GCA CAA GTG AAG
62  T   Q   N   G   G   K   L   E   K   Q   Y   R   L   I   N   A   A   Q   V   K
244 ATG TCT GAA CAA GCC GCA AAA AAA CTT GAA CAC GAC CCT AGC ATT GCT TAT GTA GAA GAA
82   M   S   E   Q   A   A   K   K   L   E   H   D   P   S   I   A   Y   V   E   E
304 GAC CAC AAA GCA GAA GCA TAT GCC CAA ACC GTC CCT TAT GGT ATC CCT CAA ATC AAA GCT
102  D   H   K   A   E   A   Y   A   Q   T   V   P   Y   G   I   P   Q   I   K   A
364 CCA GCC GTA CAC GCT CAA GGT TAT AAA GGT GCT AAC GTC AAA GTA GCT GTC CTT GAT ACT
122 P   A   V   H   A   Q   G   Y   K   G   A   N   V   K   V   A   V   L   D   T
424 GGA ATC CAC GCC GCA CAC CCT GAC TTA AAT GTT GCA GGC GGT GCT AGC TTC GTC CCT TCA
142  G   I   H   A   A   H   P   D   L   N   V   A   G   G   A   S   F   V   P   S
484 GAG CCA AAT GCC ACA CAA GAC TTT CAA TCA CAC GGA ACT CAC GTA GCC GGA ACC ATT GCT
162  E   P   N   A   T   Q   D   F   Q   S   H   G   T   H   V   A   G   T   I   A
544 GCC CTT GAT AAC ACG ATT GGT GTT CTT GGG GTT GCG CCA AGC GCC TCC TTG TAT GCC GTT
182  A   L   D   N   T   I   G   V   L   G   V   A   P   S   A   S   L   Y   A   V
604 AAA GTA TTA GAT CGT TAC GGC GAT GGA CAA TAC AGC TGG ATT ATC AGC GGT ATT GAA TGG
202  K   V   L   D   R   Y   G   D   G   Q   Y   S   W   I   I   S   G   I   E   W
664 GCT GTT GCC AAT AAC ATG GAT GTC ATC AAT ATG AGC TTA GGC GGA CCA AAC GGC TCC ACA
222  A   V   A   N   N   M   D   V   I   N   M   S   L   G   G   P   N   G   S   T
724 GCG CTT AAA AAT GCT GTA GAC ACA GCG AAT AAC CGC GGA GTA GTT GTC GTT GCC GCA GCA
242  A   L   K   N   A   V   D   T   A   N   N   R   G   V   V   V   V   A   A   A
784 GGG AAT TCA GGT TCA ACT GGC TCT ACT AGC ACC GTT GGC TAT CCT GCA AAA TAC GAC TCT
262  G   N   S   G   S   T   G   S   T   S   T   V   G   Y   P   A   K   Y   D   S
844 ACA ATT GCT GTT GCC AAC GTG AAC AGC AAC AAT GTC AGA AAC TCG TCT TCA AGC GCA GGT
282  T   I   A   V   A   N   V   N   S   N   N   V   R   N   S   S   S   S   A   G
904 CCT GAA TTA GAT GTT TCT GCA CCT GGT ACA TCT ATT TTA AGT ACA GTA CCA AGC AGT GGA
302  P   E   L   D   V   S   A   P   G   T   S   I   L   S   T   V   P   S   S   G
964 TAT ACA TCT TAT ACG GGA ACA TCG ATG GCA TCT CCT CAT GTA GCA GGA GCA GCA GCG CTT
322  Y   T   S   Y   T   G   T   S   M   A   S   P   H   V   A   G   A   A   A   L
1024 ATC CTT TCT AAG TAT CCG AAT CTA TCA ACT TCT CAG GTT CGT CAG CGC TTA GAA AAT ACA
342  I   L   S   K   Y   P   N   L   S   Q   V   R   Q   R   L   E   N   T   N   T
1084 GCA ACA CCG CTT GGT AAT TCG TTC TAT TAC GGA AAA GGA TTA ATT AAC GTT CAA GCG GCT TCT AAC TAA
362  A   T   P   L   G   N   S   F   Y   Y   G   K   G   L   I   N   V   Q   A   A S   N  Stop
```

Figure 11: The full length for the isolated keratinase gene (ker gene) and its educed amino acid sequence. Amino acid sequence starts with methionine (Met-) residue, and is represented by the single-letter code.

Table 8: Sequence analysis of ker gene isolated from strain C_4 using NCBI's BLASTX2.6.0+ program.

Description	Organism	Sequence	Accession
Alkaline serine proteinase	*B. pumilus*	100	ACO94164.1
Lehensis serine protease	*B. lehensis*	100	AFP23380.1
Alkaline serine protease	*B. pumilus*	99	BAE79641.1
Dehairing protease precursor	*B. pumilus*	99	AAR19220.1
Peptidase S8	*B. pumilus*	99	WP_026050071.1
Alkaline serin proteinase	*B. pumilus*	99	BAA93474.1
Protease	*B. pumilus*	99	ADK63096.1
Peptidase S8	MULTISPECIES: *Bacillus*	99	WP_008348814.1
Serine alkaline protease (subtilisin E)	*B. stratosphericus LAMA585*	99	EMI14709.1
Subtilisin Carlsberg	*Bacillus pumilus* ATCC 7061	99	EDW22774.1
Peptidase S8	*B. pumilus*	99	WP_034620013.1
Peptidase S8	*B. pumilus*	99	WP_041093123.1
Peptidase S8	*B. stratosphericus*	99	WP_039962807.1
Peptidase S8	*B. invictae*	99	WP_045034875.1
Peptidase S8	*B. pumilus*	99	WP_044140726.1
Peptidase S8	*B. aerophilus*	99	WP_041507592.1
Peptidase S8	*B. altudinis*	99	WP_039167642.1
Peptidase S8	*B. pumilus*	98	WP_012009474.1
Serine alkaline protease	*B. circulans*	98	ADN04910.1
Serine alkaline protease (subtilisin E)	*B. pumilus*	98	KIL22204.1
Serine alkaline protease,	*B. pumilus*	98	CAO03040.1
Peptidase S8	*B. sp.* DW5-4	98	WP_034323660.1
Peptidase S8	*B. pumilus*	98	WP_041117216.1
Peptidase S8	*B. safensis*	98	WP_034282323.1
Peptidase S8	*B. pumilus*	98	WP_034663897.1
Serine alkaline protease (subtilisin E)	*B. pumilus*	97	KIL10386.1
Peptidase S8	*B. pumilus*	97	WP_041110188.1
Peptidase S8	*B. sp.* WP8	96	WP_039183048.1
Keratinase precursor	*B. pumilus*	96	ACM47735.1
Serine alkaline protease (subtilisin E)	*Bacillus sp.* HYC-10	96	KIL09959.1
Peptidase S8	*B. xiamenensis*	96	WP_008359041.1
Organic solvent tolerant protease	*B. pumilus*	96	AAU88064.1
Keratinase	*B. pumilus*	96	ADK11996.1
Peptidase S8	*B.safensis*	96	WP_029706931.1
Peptidase S8	*B. pumilus*	96	WP_041089929.1
Peptidase S8	*B. safensis*	96	WP_044335827.1
MULTISPECIES: peptidase S8	*Bacillus*	96	WP_025093353.1
Peptidase S8	*B. pumilus*	95	WP_024426548.1
Serine alkaline keratinase	*Brevibacillus brevis*	95	AGO58466.1
Alkaline serine protease precursor	*B. pumilus*	95	ACM07731.1

Serine alkaline keratinase	*B. circulans*	94	AGN91700.1

5.3. *Bacillus* sp. C_4 genome sequencing and *de novo* assembly

To search for further putative keratinase genes and to provide the genomic information involved in keratin degradation, the whole genome sequencing of *Bacillus* sp. C_4 strain was performed by next generation sequencing using an Illumina Miseq system, according to the manufacturer's protocol for 2x150 bp chemistry by Eurofins Genomics, Germany. Both genomic shotgun (SG) and 8Kbp mate paired libraries sequencing were performed.

After trimming in shotgun sequencing, 7,416,279 paired-end reads (2,200,746,668 bp) with an average length of 148 bp were obtained (Table 9). However, in mate-paired (8 KB) library sequencing, 5 865 628 paired-end reads and 5 957 589 reads, were obtained (Table 10).

Table 9: Trimmomatic report for SG reads.

Surviving Pairs

Total Entries: 14 832 558

Total Length: 2 200 746 668

Min Length: 70

Max Length: 150

Mean Length: 148

Modal Length: 150

N50: 150

N90: 150

GC Content: 41,9

#A: 639 762 482

#T: 639 837 822

#G: 460 371 913

#C: 460 774 135

#N: 316

#Others 0

Table 10: Processed read data for the cleaned LJD read sequences (i.e. adaptor sequences and bad quality bases were removed).

	LJD forward	LJD reverse	singleton
Suffix			
Total Entries	5 865 628	5 865 628	5 957 589
Total Length	650 167 636	624 280 354	612 849 897
Mean Length	110	106	102

Expected Insert Size (based on lab measurements):

Library Type	Size
8kb	7865 bp

De novo sequence assembly and scaffolding was done by Eurofins Genomics in Germany using the Newbler assembler software v2.9. The number of resulting contigs was in total 117 with lengths ranging from 100 bp to 282,994 bp and an N_{50} value of 72817 bp. The number of scaffolds was two with 5,386 bp and 3,659,360 bp respectively and an N_{50} value of 3,659,360bp, (Table 11). The scaffold 2 was the phiX library added before sequencing to estimate the sequencing quality. Various filtering steps are applied to the illumina sequences in order to remove phiX related reads. Nevertheless a small fraction of phiX reads may remain. They have been omitted with some doublette sequences when subjected to NCBI for genome annotation.

The results indicate that *Bacillus sp.*C_4 genome consist of one circular chromosome containing 3,659,360 bp with an average G+C content of 41.4%. No plasmids were found during the genome analysis, and none were found by agarose gel electrophoresis (data not shown). The G+C content was in good agreement with that of the closely related species *Bacillus pumilus* Ku bf1 (41, 6%) and *Bacillus pumilus* W3 (41.39%), respectively. (**Balsingh *et al,* 2016; Guan *et al,* 2015**)

Table 11: Assembly statistics.

Total Entries (Contigs)	117
Min Length	100
Max Length	282 994
Mean Length	31 122
N50	72 817
N90	22 028
G+C Content	41,3%

Scaffold 1

Length(bp)	3,659,360
G+C Content	41.4%
N50	3,659,360

Scaffold 2

Length (bp)	5 386(is phiX)
G+C Content	44.7%

5.4. Genome Annotation

Genome assembly is followed by gene prediction/annotation, a computational process in which regions of the DNA containing coding genes are identified. The whole genome sequence of *Bacillus* sp. $C_4$was annotated by submitting it to the web based automated Bacterial Annotaion System (BASys) and the prokaryotic annotation pipeline at the National Center for Biotechnology Information, Bethesda, U.S.A. (NCBI).

The annotation results indicated that *Bacillus sp* C_4 belongs to the *B. pumilus* group of organisms. NCBI Prokaryotic genome annotation pipeline predicted a total of 3,698 genes, 3,596 coding sequences (CDS), 71 RNA genes and 31 pseudogenes. The RNA coding genes predicted include 60 tRNAs, 10 rRNAs, and 1 non-coding RNA (ncRNA), (Table 12 and Figure 12). The genome was found to be 99.6 % complete when analyzed.

This whole-genome *Bacillus pumilus* C_4 sequence has been deposited at NCBI GenBank under the accession number CP011109. The bio-project number was PRJNA278012.

Table 12: General features of *Bacillus pumilus* C_4genome.

Features	**Chromosome**
Molecular shape	Circular
Genome size (bp)	3,659,360
C+G content (%)	41.4
Total number of genes	3,698
Protein coding genes (CDSs)	3,596
Pseudo genes	31
rRNAs (5S, 16S, 23S)	10
tRNAs	60
n.c RNA	1
Frameshifted genes	15

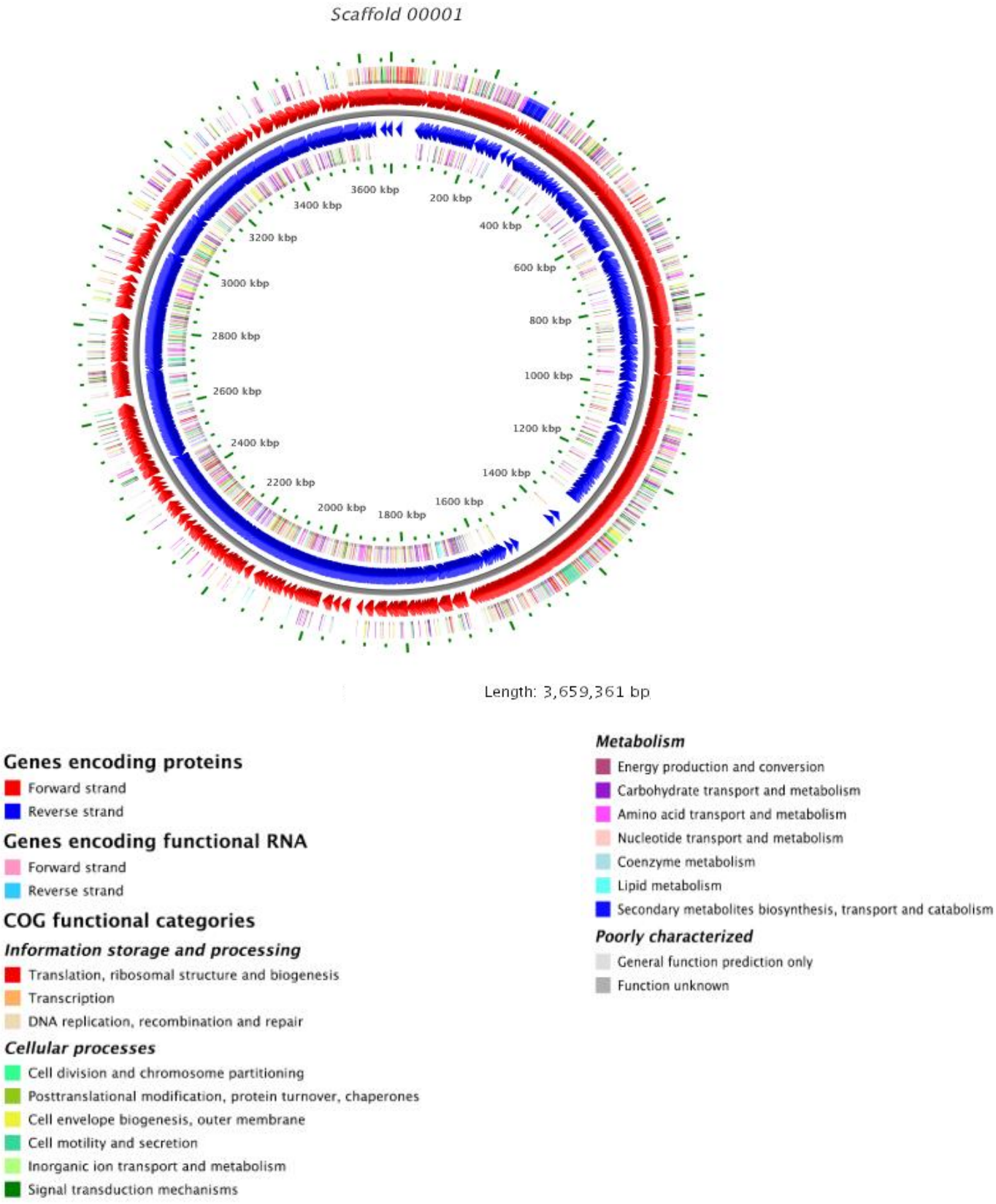


Figure 12: Circular representation of *Bacillus pumilus* C_4 genome.

The outer and inner circles indicate the chromosomal location in kilobases from base 1, the start of the replication origin(Each tick is 40Kb). Circles in red and bleu colours are coding sequencestranscribed in the

Clockwise direction and in the counterclockwise direction, respectively. Genes displayed in the second and

fifth circles are color-coded according to different functional categories:

BASys and NCBI annotations of *Bacillus pumilus* C_4 genome revealed the presence of a number of additional genes that are known to be involved in protein degradation. Eighty one candidate genes for peptidase were found overall in the draft genome. Six (06) genes were annotated as putative peptidase S8, five (05) serine protease related genes, two (02) genes as putative zinc-metallo-proteases and one (01) as putative cysteine protease. Among the six peptidase S8 related genes, two genes were predicted as subtilisins based on BASys annotation. The first one is the Ker gene (accession number: KX184831) previously isolated in this study. It was designated Ker1 gene. The second one is a homolog to Ker1 and was named Ker2. This gene was subsequently submitted to NCBI GenBank (Accession Number: KX184832), Figure 13.

```
1      ATGAAAGGGAGAGTGGGTATCGTGAAGGTCAAATCATTTGGAGCAGGGCT
51     TTGTATGGCAAGTATGCTACTGGCATCGATGACATTCGGCGCATCAGATG
101    TATCTGCAAAGGATCAAGCAAAAAAAGAGTATATGATCGGTTTTTCTTCT
151    TCCGTTCAAGACAAAACACAAAAACAGCTTGTTGAAAAGGCCGGCGGTCA
201    TGTGAAAGAATCAATAGAGAAAATAGATATGATGAAAGTTTCGTTGAACG
251    AAGCATCAAAAGAAAAACTACATCAAGCAAAAGAAGTTACTTTTATTGAG
301    GAAGACCAAAAAGCAAAAACAAGCGGTCAAAGCGTTCCTTATGGCATCAA
351    AAGCATCAAAGCACAAAAGGTACATAAACGAGGATATGCCGGACAGAATG

401    TCAAAGTAGCTGTTCTTGACAGTGGAATCGACGGCAAGCATGAAGATTTA

451    TGACCCGCATGAACACGGAACTCACGTTGCAGGCACGATCGCAGCATTAG

501    CATGTAACTGGTGGTGTCAGCTTTGTTCCAACAGAGTCCGATCCGCTCGT

551    ATAATAAAGTAGGCGTCGTTGGTGTGGCACCAAAAGCTTCGATCTATGCG

601    GTGAAAGTGGCAGATGAAAATGGTGACGGCTACTATAGCTGGATCATTAA

651    AGGCATTGAATGGGCGATTGAGAATGAGATGGATGTCATCAATATTAGTA

701    TGGGGGGAGCAAGTGAATCGGAGGCGCTGAAAGAAGCGGTAGATCGAGCA
```

Figure 13: The full length for the ker 2 gene.

Using *de novo* sequencing and genome annotation we could identify a second gene for putative keratinase, named Ker2 gene. The ORF of this gene has 1152 bp, an equal amount of base pairs as Ker1gene, with an ATG start codon and a TAA stop codon (Figure 13). The two genes had a gene sequence similarity of 67% using blastn at NCBI.

The deduced amino acid sequence of ker2 gene of *B. pumilus* C_4contained 383 amino acid residues and the three residues at the active site that form the catalytic triad in all subtilisins are Asp-140, His-172 and Ser-329. Comparison of this amino acid sequence of the ker2 gene using the online BLASTX showed 100% identity with serine alkaline protease (subtilisin E) of *B. stratosphericus* LAMA 585 (Accession number EMI12150.1) and peptidase S8 of multispecies *Bacillus* (Accession number WP_035390997.1), 99% identity with subtilisin

Carlsberg of *Bacillus* sp. M 2-6 (Accession numberEIL84986.1), peptidase S8 of *Bacillus pumilus* (Accession number WP_029575389.1) and subtilisin Carlsberg of *Bacillus pumilus* (Accession number KIL26870.1), Table 13.

Table 13: Sequence comparison of the *Ker 2* gene of *Bacillus pumilus*. C_4 using BLASTX 2.6.0+ program.

Description	Organism	sequence	Accession number
Serine alkaline protease (subtilisin E)	*B. stratosphericus* LAMA 585	100	EMI12150.1
Peptidase S8	MULTISPECIES: *Bacillus*	100	WP_035390997.1
Subtilisin Carlsberg	*B. sp.* M 2-6	99	EIL84986.1
Peptidase S8	MULTISPECIES: *Bacillus*	99	WP_034647494.1
Peptidase S8	*B. pumilus*	99	WP_029575389.1
Subtilisin Carlsberg	*B. pumilus*	99	KIL26870.1
Peptidase S8	*B. altitudinis*	99	WP_035702958.1
Peptidase S8	*B. pumilus*	99	WP_026050107.1
Pubtilisin	*B. altitudinis* 41KF2b	99	KDE30915.1
Peptidase S8	*B. sp.* DW5-4	95	WP_034319985.1
Peptidase S8	*B. pumilus*	95	WP_044141213.1
Peptidase S8	*B. pumilus*	94	WP_034620505.1
Peptidase S8	*B. pumilus*	94	WP_034661158.1
Subtilisin Carlsberg	*B. pumilus*	94	KIL17080.1
Subtilisin Carlsberg	*B. pumilus* ATCC7061	94	EDW21217.1
Peptidase S8	*B. pumilus*	93	KDE52880.1
Peptidase S8	*B. safensis*	93	WP_034622393.1
Peptidase S8	multispecies: *Bacillus*	93	WP_029708034.1
Peptidase S8	*B. pumilus*	93	WP_041117873.1
Peptidase S8	*B. safensis*	93	KEP30825.1
Peptidase S8	*B. safensis*	93	WP_029706051.1
Peptidase S8	*B. pumilus*	93	WP_041109684.1
Subtilisin Carlsberg	*B. pumilus*	93	KIL21504.1
Subtilisin Carlsberg	*B. pumilus*	92	KIL11523.1
Peptidase S8	*B. sp.* HYC-10	92	WP_008361817.1
Peptidase S8	*B. safensis*	92	WP_034280781.1
Peptidase S8	*B. sp.* WP8	92	WP_039183179.1
Subtilisin	*B. safensis* FO-36b	92	KDE29455.1
Peptidase S8	*B. safensis*	91	WP_046312283.1
Peptidase S8	*B.pumilus*	91	WP_041086842.1
Subtilisin Carlsberg	*B. pumilus*	90	KIL15277.1

The results indicated that the protease producer *Bacillus pumilus* C_4 has at least two genes encoding subtilisin-like serine proteases belonging to the peptidase S8 Family (Clan SB; Sub family S8A). Most keratinases isolated from the *Bacillus* sp. belong to the subtilisins which are typical members of the serine protease family (**Lange *et al,* 2016**). Keratinases are a special kind of proteolytic enzymes that display the capability of degrading keratins (**Liu *et al,* 2013**).

Serine proteases are the largest family of proteolytic enzymes (**McKerrow *et al,* 1993**) and are grouped into 14 clans with 27 families, in the MEROPS database (the peptidase database at http://merops.sanger.ac.uk/) (**Rawlings *et al,* 2006**). According to MEROPS, subtilases are the second largest family of serine proteases in terms of number of sequences and characterized peptidases. They are grouped under the S8 family of SB clan and are characterized by a catalytic triad of three amino acids, namely aspartate, histidine, and serine. They are further classified into two sub families namely prokaryotic subtilisins or subtilisin-like serine proteases (S8A) and eukaryotic kexins (S8B) (**Rawlings *et al,* 2008)**.

These proteases show broad substrate specificity, have usually a molecular mass in the range 18–90 kDa. They are generally active at neutral and alkaline pH, with optima at pH 7–11 (**Rao *et al,* 1998**) and irreversibly inhibited by PMSF (Phenyl-methane-sulfonyl-fluoride) (**Powers *et al,* 2002**). This is in agreement with our earlier findings (**Fellahi 2009**). Our previous study (**Fellahi, 2009**) revealed that the partial purified protease fraction from *Bacillus pumilus* C_4 is optimally active at pH of 8.5 and 10. It was inhibited by the serine peptidase inhibitor, PMSF.

A single kind of keratinolytic protease is not sufficient for efficient keratin degradation, and this suggests that degradation of this recalcitrant material requires the cooperative action of multiple enzymes (**Yamamura *et al,* 2002**). *Bacillus pumilus* C_4 can utilize and decompose compact hard keratinaceous materials such as chicken feathers and sheep wool as sole carbon and nitrogen source (**Fellahi *et al,* 2014**). The presence of all these protease-related genes may explain the broad substrate specificity of this protein degrader strain and especially the remarkable keratinolytic activity observed from its culture broth toward both α- and β-keratin in less than three days. So far, the vast majority of the identified keratinase producing organisms appear to be able to hydrolyze only β-keratin in chicken feather and few are known to hydrolyze both α- and β-keratin. (**Gupta *et al,* 2013**).

5.5. Identification of the enzyme fraction related to protein utilization by *Bacillus pumilus* strain C_4

The crude protease fraction produced by growing *Bacillus pumilus* C_4 on the optimized medium showed after partial purification followed by SDS-PAGE and staining with Coomassie brilliant blue R-250 several bands, (Figure 14). One distinct and one weaker band with molecular weights of about 36 and 28 kDa respectively,

when comparing with the molecular weight marker, were subjected to protein identification by nano HPLC-ESI-MS/MS and Mascot database search.

The Mascot search result using the *B. pumilus* entries from the Ref Seq protein data base and the amino sequences retrieved from the Ker1and Ker2genes can be seen in table 14.

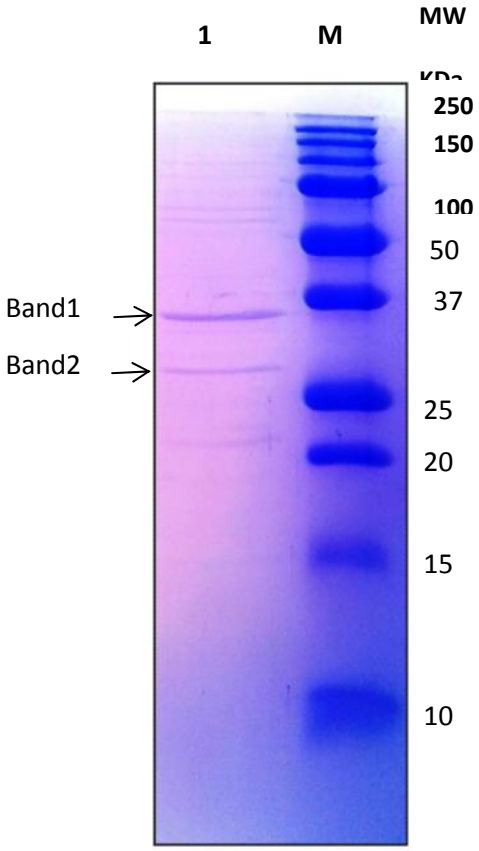


Figure 14: The dialyzed proteolytic enzyme investigated on SDS-PAGE using a 12 % gel. Two bands at 36 and 28 kDa, respectively, were excised for protein identification (Lane 1). Lane M corresponds to molecular weight protein marker.

Table 14: Protein identification by MASCOT Search on the purified enzyme from *Bacillus pumilus* strain C_4 using Ref Seq protein data base and the two amino acid sequences retrieved from Ker1 and Ker2 genes.

SDS-PAGE band	Identified protein	MASCOT peptides identified	MASCOT ion score[a]
1 and 2	Keratinase 1	GVVVVAAAGNSGSTGSTSTVGYPAK	146
		YDSTIAVANVNSNNVR	122
		YPNLSTSQVR	58
		QRLENTATPLGNSFYYGK	81
		LENTATPLGNSFYYGK	102
		GLINVQAASN	85
1	Keratinase 2	VGVVGVAPK	61
		VADENGDGYYSWIIK	80
		SGTSMASPHVAGAAAVILSK	107
		HPNLTNDELR	54
		HPNLTNDELRER	34
1	Peptidase S8 (gi|648268958)	VGVVGVAPK	61
		VADENGDGYYSWIIK	80
		SGTSMASPHVAGAAAVILSK	107
		LGEPFYYGAGLVNVQK	106

[a] : Individual ion scores >26 indicates identity or extensive homology at 95 % level of confidence

Mascot Search results indicated that the partial purified protease fraction in fact consists of two enzymes, corresponding to the previously identified genes Ker1and Ker2 of the *B. pumilus* strainC_4. When using only the *B. pumilus* entries from the RefSeq protein data the peptides from MS/MS data were identified as peptidase S8.

Figure 15 illustrates the peptides from Ms/Ms identified in keratinase 1, keratinase 2 as well as peptidase S8 (gi|648268958) and their distribution in the respective enzymes. As can be seen keratinase 2 shares three peptides with peptidase S8 (gi|648268958) while keratinase 1 does not have any peptide in common with the identified Peptidase S8.

Keratinase 1

1 MCVKKKNVMT SVLLAVPLLF SAGFGGSIAN AETASKSESE KSYIVGFKAS
51 ATTNSSKKQA VTQNGGKLEK QYRLINAAQV KMSEQAAKKL EHDPSIAYVE
101 EDHKAEAYAQ TVPYGIPQIK APAVHAQGYK GANVKVAVLD TGIHAAHPDL
151 NVAGGASFVP SEPNATQDFQ SHGTHVAGTI AALDNTIGVL GVAPSASLYA
201 VKVLDRYGDG QYSWIISGIE WAVANNMDVI NMSLGGPNGS TALKNAVDTA
251 NNR**GVVVVAA AGNSGSTGST STVGYPAKYD STIAVANVNS NNVR**NSSSSA
301 GPELDVSAPG TSILSTVPSS GYTSYTGTSM ASPHVAGAAA LILSK**YPNLS**
351 **TSQVRQRLEN TATPLGNSFY YGKGLINVQA ASN**

Keratinase 2

1 MKGRVGIVKV KSFGAGLCMA SMLLASMTFG ASDVSAKDQA KKEYMIGFSS
51 SVQDKTQKQL VEKAGGHVKE SIEKIDMMKV SLNEASKEKL HQAKEVTFIE
101 EDQKAKTSGQ SVPYGIKSIK AQKVHKRGYA GQNVKVAVLD SGIDGKHEDL
151 HVTGGVSFVP TESDPLVDPH EHGTHVAGTI AALDNK**VGVV GVAPK**ASIYA
201 VK**VADENGDG YYSWIIK**GIE WAIENEMDVI NISMGGASES EALKEAVDRA
251 YDNGILIVAS AGNAGSYGSL NTIDYPAKYS SVMAVASVDQ RKQRAFDSSV
301 GEEVEVSAPG VSTLSTIPHN EYGYK**SGTSM ASPHVAGAAA VILSKHPNLT**
351 **NDELRER**LTK

Peptidase S8 (gi|648268958)

1 MASMLLASMT FGASDVSAKD QAKKEYMIGF SSSVQDKTQK QLVEKAGGHV
51 KESIEKIDMM KVSLNEASKE KLHQAKEVTF IEEDQKAKTS GQSVPYGIKS
101 IKAQKVHKRG YTGQNVKVAV LDSGIDGKHE DLHVTGGVSF VPTESDPLVD
151 PHEHGTHVAG TIAALDNK**VG VVGVAPK**ASI YAVK**VADENG DGYYSWIIK**G
201 IEWAIENEMD VINISMGGAS ESEALKEAVD RAYDNGILIV ASAGNAGSYG
251 SLNTIDYPAK YSSVMAVASV DQRKQRAFDS SVGEEVEVSA PGVSTLSTIP
301 HNEYGYK**SGT SMASPHVAGA AAVILSK**HPN LTNDEVRERL TKTATK**LGEP**
351 **FYYGAGLVNV QK**AAR

Figure 15: The distribution of peptides from MS/MS in keratinase 1, keratinase 2 and peptidase S8 (gi|648268958). Matched peptides are shown in bold red.

Analysis of the protein fraction related to protease activity produced by *Bacillus pumilus* C_4 through nLC-ESI-MS/MS followed by Mascot database search confirmed that the two genes ker1 and ker2 encode keratinolytic proteases and this bacterial strain produces at least two different keratinolytic proteases belonging to the Protease S8 family.

Recent findings suggest that several proteases may have keratinolytic activity but that such activity only leads to full keratin decomposition if several different keratinolytic enzymes act together (**Huang *et al,* 2015; Lange *et al,* 2014**).Huang found that the combination of three novel proteases from *Onygena corvine* belonging to three protease families (S8, M28, and M3) can efficiently degrade keratin. Also this finding is in agreement with earlier conclusions (**Yamamura *et al,* 2002**). Yamamura and his group found that for the bacterium *Stenotrophomonas* sp D-1 it was not sufficient with one protease for effective keratin degradation but two different keratinases were needed. This suggests that keratin degradation requires the cooperative action of multiple enzymes. With such complexity, it is relevant to compare microbial keratin decomposition with the microbial decomposition of well-studied polymers such as cellulose and chitin.

The above results suggest that the broad substrate specificity and the high keratin-degrading activity towards both alpha and beta keratins observed from culture broth of *B. pumilus* C_4 originates from the production of more than one keratinolytic proteases. This characteristic makes *Bacillus pumilus* C_4 as a potent candidate in a cost effective pretreatment step of keratin rich wastes in waste refinery as the two waste fractions, avian feather and sheep wool, do not have to be separated before hydrolyzation of the protein into valuable feedstuff and for the biogas production.

Conclusion & Perspectives

Conclusion and Perspectives

The current study was carried out with the objective of improving the production of the extracellular protease of *Bacillus* sp. C_4, screening for some proteases related to the keratin-degrading ability of this strain and identifying the secreted proteases in the optimized medium and their relationship to the previously identified proteases for this bacterium.

Based on a two-step strategy, One-Variable-at-a-Time approach followed by Central Composite Design (CCD), an overall of 1.8-fold enhancement in extracellular protease production was achieved in the optimized medium as compared to the un-optimized medium. Under the optimal conditions the extracellular protease production started earlier than that produced in the non-optimized conditions.

By genomics approaches; PCR amplification, *de novo* sequencing and genome annotation, for the protease producer and keratin degrading *Bacillus pumilus* C_4 at least two genes encoding subtilisin-like serine proteases (*ker1*, KX184831 and *Ker2*, KX184832 genes) were identified. The annotation results also indicated that *Bacillus sp*. C_4 belongs to the *B. pumilus* group of organisms and suggested eighty-one candidate genes for protease and peptidase overall in the draft genome.

nLC-ESI-MS/MS followed by Mascot database search of the partial purified protease fraction produced in the optimized medium confirmed that the two genes *ker1* and *ker2* encode proteases and this bacterial strain produces at least two different enzymes belonging to the Protease S8 family. This finding could explain the broad substrates specificity and the remarkable keratinilytic activity of *Bacillus pumilus* C_4 towards both types of keratins. The literature supported that keratin degradation requires the cooperative action of multiple enzymes.

Taken together, these findings makes *Bacillus pumilus* C_4 as a potent candidate in a cost effective pretreatment step of keratin-rich wastes in waste refinery as the two waste fractions, avian feather and sheep wool, do not have to be separated before hydrolyzation of the protein into valuable feedstuff for the biogas production. As well as these enzymes could possibly

find a use for degrading keratinaceous wastes and provide proteins, peptides, and amino acids as valuable ingredients for animal feed.

Further confirmation of the keratinolytic activity of these two identified proteases, could be achieved via keratin zymography. Cloning of the two genes, *ker1* and *ker2*, into a suitable efficient expression vector for gene expression purposes will be needed in an attempt to enzyme purification and biochemical characterization. The results more greatly encourage the isolation and characterization of all the predicted protease genes in *Bacillus pumilus* C_4genome to identify the ones involved in keratin utilization and that may help understanding the keratin decomposition mechanism which is till now still unclear.

References

References

1. **Adessi C., Matton G., Ayala G., Turcatti G., Mermod J.J., Mayer P., and Kawashima E. (2000)** Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. *Nucleic Acids Research*. **28(20)**: e87.
2. **Adigüzel A.C., Bitlisli B.O., Yaşa I. and Eriksen N.T. (2009)** Sequential secretion of collagenolytic, elastolytic and keratinolytic proteases in peptide-limited cultures of two *Bacillus cereus* strains isolated from wool. *Journal of Applied Microbiology*. **107**: 226–234.
3. **Ahuja S.K., Ferreira G.M. and Morreira A.R. (2004)** Application of Plackett-Burman design and response surface methodology to achieve exponential growth for aggregated shipworm bacterium. *Biotechnology and Bioengineering*. **85**: 666-675.
4. **Allpress J.D., Mountain G. and Gowland P.C. (2002)**Production, purification and characterization of an extracellular keratinase from *Lysobacter* NCIMB 9497. *Letters in Applied Microbiology*. **34(5)**: 337–342.
5. **Anisworth S.J. (1994)** Soap and Detergent. *Chemical Engineering News*. **72**: 34-59.
6. **PowersArgos P. (1987)** A sensitive procedure to compare amino acid sequences. *Journal of Molecular Biology*. **193:** 385–396.
7. **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. and Zagnitko O. (2008)** The RAST server: rapid annotations using subsystems technology. *BMC Genomics*. **9**:75.
8. **Bakhtiar S., Estiveira R.J. and Hatti-Kaul R. (2005)** Substrate specificity of alkaline protease from alkaliphilic feather-degrading *Nesterenkonia* sp. AL20. *Enzyme and Microbial Technology*. **37**: 534–540.
9. **Balsingh J., Radhakrishna S. and Ulaganathan K. (2016)** Draft genome sequence of *Bacillus pumilus* ku-bf1 isolated from the gut contents of wood boring *Mesomorphus* sp. *Frontiers in Microbiology*.**7**: 1037.
10. **Banerjee R. and Bhattacharyya B.C. (1993)** Kinetic properties of extracellular alkaline proteases of *Rhizopus oryzae*. *Journal of Fermentation and Bioengineering*. **75**:380–382.
11. **Barba C., Mendez S., Roddick-Lanzilotta A., Kelly R., Parraand J.L. and Coderch L. (2008)** Cosmetic effectiveness of topically applied hydrolysed keratin peptides and lipids derived from wool. *Skin Research Technology*. **14**: 243–248.
12. **Barone J.R., Schmidt W.F. and Liebner C.F.E. (2005)** Thermally processed keratin films. *Journal of Applied Polymer Science*. **97**: 1644–1651.
13. **Bas D. and Boyaci I. H. (2007)** Modeling and optimization I: Usability of response surface methodology. *Journal of Food Engineering***78**: 836–845.
14. **Bayoudh A., Gharsallah N., Chamkha M., Dhouib A., Ammar S. and Nasri M. (2000)** Purification and characterization of an alkaline protease from *Pseudomonas aeruginosa* MN1.*Journal of Industrial Microbiology and Biotechnology*. **24**:291–295.
15. **Beckloff N., Starkenburg S., Freitas T. and Chain P. (2012)** Bacterial Genome Annotation. *Methods in molecular biology*. **881**: 471-503.
16. **Bedouet L., Arnold F., Batina P. and Robreau G. (1998)** Potential application of the *Clostridium tyrobutyricum* flagellin as substrate for proteinase detection. *Biotechnology Letters*. **20**: 1011–1016.

17. **Beg Q.K., Saxena R.K. and Gupta R. (2002)** De-repression and subsequent induction of protease synthesis by *Bacillus mojavensis* under fed-batch operations. *Process Biochemistry*. **37**:1103–1109.
18. **Beg Q. K., Sahai V. and Gupta R. (2003)** Statistical media optimization and alkaline protease production from *Bacillus mojavensis* in a bioreactor. *Process Biochemistry*. **39**: 203-209.
19. **Behjati S. and Tarpey P. S. (2013)** What is next generation sequencing? *Archives of disease in childhood. education and practice edition*. **98**: 236–238.
20. **Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N. and Bourne P.E. (2000)** The Protein Data Bank. *Nucleic Acids Research*. **28**: 235–242.
21. **Bernal C., Cairó J. and Coello N. (2006)** Purification and characterization of a novel exocellular keratinase from *Kocuria rosea*. *Enzyme and Microbial Technology*. **38**: 49–54.
22. **Bernal C., Vidal L., Valdivieso E. and Coello N. (2003)**Keratinolytic activity of *Kocuria rosea*. *World Journal of Microbiology and Biotechnology*. **19(3)**: 255–261.
23. **Bernhardt K., Schrempf H. and Goebel W. (1978)** Bacteriocin and antibiotic resistance plasmids in *Bacillus cereus* and *Bacillus subtilis. Journal of Bacteriology*. **133**: 897-903.
24. **Besemer J., Lomsadze A. and Borodovsky M. (2001)** GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Research.***26**: 1107–1115.
25. **Bhosale S.H., Rao M.B., Deshpande V.V. and Srinivasan M.C. (1995)** Thermostability of high-activity alkaline protease from *Conidiobolus coronatus* (NCL 86.8.20). *Enzyme and Microbial Technolo*gy **17**: 136–139.
26. **Binnewies T.T., Motro Y., Hallin P.F., Lund O., Dunn D., La T., Hampson D.J., Bellgard M., Wassenaar T.M. and Ussery D.W. (2006)** Ten years of bacterial genome sequencing: comparative-genomics-based discoveries. *Functional and Integrative Genomics***6**:165–185.
27. **Brandelli A., Daroit D.J. and Riffel A. (2010)** Biochemical features of microbial keratinases and their production and applications. *Applied Microbiology and Biothnology*. **85(6)**: 1735–1750.
28. **Branton D., Deamer D.W., Marziali A., Bayley H., Benner S.A., Butler T., Di Ventra M., Garaj S., Hibbs A., Huang X., Jovanovich S.B., Krstic P.S., Lindsay S., Ling X.S., Mastrangelo C.H., Meller A., Oliver J.S., Pershin Y.V., Ramsey J.M., Riehn R., Soni G.V., Tabard-Cossa V., Wanunu M., Wiggin M. and Schloss J.A.(2008)** The potential and challenges of nanopore sequencing. *Nature Biotechnology*. **26**: 1146–1153.
29. **Braslavsky I., Hebert B., Kartalov E. and Quake S.R. (2003)** Sequence information can be obtained from single DNA molecules. *Proceedings of the National Academy of Sciencesof the United States of America*. **100**: 3960–3964.
30. **Breithaupt H. (2001)** The hunt for living gold. *EMBO Reports.***2**: 968–971.
31. **Bressollier P., Letourneau F., Urdaci M. and**Verneuil **B. (1999)** Purification and characterization of a keratinolytic serine proteinase from *Streptomyces albidoflavus. Applied and Environmental Microbiology*. **65**: 2570–2576.
32. **Brutt E.H. and Ichida J.M. (1999)** Occurrence of feather-degrading bacilli in the plumage of birds. *The Auk***116**: 364–372.
33. **Cai C.G. and Zheng X.D. (2009)**Medium optimization for keratinase production in hair substrate by a new *Bacillus subtilis* KD-N2 using response surface methodology.*Journal of Industrial Microbiology and Biotechnology*.**36(7)**: 875–883.

34. **Cao Z.J., Lu D., Luo L.S., Deng Y.X., Bian Y.G., Zhang X.Q. and Zhou M.H. (2012)** Composition analysis and application of degradation products of whole feathers through a large scale of fermentation. *Environmental Science and Pollution Research*. **19**: 2690-2696.
35. **Cao Z.J., Zhang Q., Wei D.K., Chen L., Wang J., Zhang X.Q. and Zhou M.H. (2009)**Characterization of a novel *Stenotrophomonas* isolate with high keratinase activity and purification of the enzyme. *Journal of Industrial Microbiology and Biotechnology*. **36**: 181–188.
36. **Chakrabarti S.K., Matsumura N. and Ranu R.S. (2000)** Purification and characterization of an extracellular alkaline serine protease from *Aspergillus terreus* (IJIRA 6.2).*Current Microbiology*. **40**: 239–244.
37. **Chandra S. J., Pallavi P., Anuradha B. S. and Reddy S. R. (2015)** Optimization of bioprocess for enhanced production of alkaline protease by *Bacillus subtilis* SHmIIIa through Plackett-Burman design. *African journal of Microbiology Research*. **9(28)**: 1738-1744.
38. **Chauhan B. and Gupta R. (2004)** Application of statistical experimental design for optimization of alkaline protease production from *Bacillus* sp. RGR-14. *Process Biochemistry*. **39**: 2115-2122.
39. **Cheng S.W., Hu M.N., Shen S.W., Takagi H., Asano M. and Tsai Y.C. (1995)** Production and characterization of keratinase of a feather-degrading *Bacillus licheniformis* PWD-1. *Bioscience, Biotechnology and Biochemistry*. **59**: 2239–2243.
40. **Cheng S.-W., Wang Y.-F. and Wanga M.-L. (2012)** Statistical Optimization of medium compositions for alkaline protease production by newly isolated *Bacillus amyloliquefaciens*. *Chemical and Biochemical Engineering Quarterly*.**26(3)**: 225–231.
41. **Chenna R., Sugawara H., Koike T., Lopez R., Gibson T.-J., Higgins D.-G. and Thompson J.-D. (2003)**Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research*. **31(13)**:3497–500.
42. **Cheung A.L., Ying P., Fischetti V.A. (1991)** A method to detect proteinase activity using unprocessed X-ray films. *Analytical Biochemistry*. **193**: 20–23.
43. **Chitte R.R., Nalawade V.K. and Dey S. (1999)** Keratinolytic activity from the broth of a feather degrading thermophilic *Streptomyces thermoviolaceus* strain SD8. *Letters in Applied Microbiology*. **28**: 131-136.
44. **Chi Z., Ma C., Wang P. and Li H. F. (2007)** Optimization of medium cultivation conditions for alkaline production by the marine yeast *Aureobasidium pullulans*. *Bioresource Technology*. **98**: 534- 538.
45. **Cliffe A.J. and Law B.A.(1982)** A new method for the detection of microbial proteolytic enzyme in milk. *Journal of Dairy Research*. **49**: 209 – 219.
46. **Collins F. S., Morgan M. and Patrinos A. (2003)** The human genome project: lessons from large-scale biology. *Science*. **300 (5617)**: 286–290.
47. **Corrêa A.P.F., Daroit D.J. and Brandelli A. (2010)**Characterization of a keratinase produced by *Bacillus* sp. P7 isolated from an Amazonian environment. *International Biodeterioration and Biodegradation.* **64**(1): 1-6.
48. **Cortezi M., Cilli E. and Contiero J. (2008)***Bacillus amyloliquefaciens*: a new keratinolytic feather degrading bacteria. *Current Trends in Biotechnology and Pharmacy*. **2:** 170–177.
49. **Coward-Kelly G., Chang V.S., Agbogbo F.K. and Holtzapple M.T. (2006)** Lime treatment of keratinous materials for the generation of highly digestible animal feed: 1. Chicken feathers. *Bioresource Technology.***97**: 1337-1343.
50. **Dalev P. G. (1990)** An enzyme-alkaline hydrolysis of feather keratin for obtaining a protein concentrate for fodder. *Biotechnological Letters*. **12**: 71–72.

51. **Dalev P.G. (1994)** Utilization of waste feathers from poultry slaughter for production of a protein concentrate. *Bioresource Technology*. **48**: 265–267.
52. **Darah I. and Ibrahim C.O. (1996)** Effect of agitation on lignin-degrading enzyme production by Phanerochaete chrysosporium grown in shake-flask cultures. *Asia Pacific Journal of Molecular Biology and Biotechnology*. **4(3):** 174-182.
53. **Daroit D.J. and Brandelli A. (2014)** A current assessment on the production of bacterial keratinases. *Critical Review of Biotechnology*. **34**: 372–384.
54. **Daroit D.J., Corrêa A.P.F. and Brandelli A. (2009)**Keratinolytic potential of a novel *Bacillus* sp. P45 isolated from the Amazon basin fish *Piaractus mesopotamicus*.*International Biodeterioration and Biodegradation.* **63**: 358–363.
55. **Dayarian A., Michael T.P. and Sengupta A.M. (2010)** SOPRA: scaffolding algorithm for paired reads via statistical optimization. *BMC Bioinformatics.* **11**: 345.
56. **Debette J. (1991)** Isolation and characterization of an extracellular proteinase produced by a soil strain of *Xanthomonas maltophila*. C*urrent Microbiology*, **22**, 85-90.
57. **Delcher A.L., Harmon D., Kasif S.,** White O**. and** Salzberg S.L**. (1999)** Improved microbial gene identification with GLIMMER. *Nucleic Acids Research*. **27(23)**: 4636–4641.
58. **Demain A.L. and Adrio J.L. (2008)** Contributions of microorganisms to industrial biology. *Molecular Biotechnology*. **38**: 41-55.
59. **Desai S.S., Hegde S., Inamdar P., Sake N. and Aravind M.S. (2010)** Isolation of keratinase from bacterial isolates of poultry soil for waste degradation.*Engineering in Life Science*. **10(4)**: 361–367.
60. **De Souza P. M., De Assis Bittencourt M. L., Caprara C. C., De Freitas M., De Almeida R. P. C., Silveira D., Fonseca Y. M., Ferreira Filho E. X., Junior A. P. and Magalhães P. O. (2015)** A biotechnology perspective of fungal proteases. *Brazilian Journal of Microbiology*. **46(2)**: 337-346.
61. **DiGuistini S., Liao N.Y., Platt D., Robertson G., Seidel M.,** Chan S.K**.,** Docking T.R**.,** Birol I**.,** Holt R.A**.,** Hirst M**.,** Mardis E**.,** Marra M.A**.,** Hamelin R.C**.,** Bohlmann J**.,** Breuil C**. and** Jones S.J**. (2009)***De novo* genome sequence assembly of a filamentous fungus using Sanger, 454 and Illumina sequence data. *Genome Biology*. **10 (9)**: R94.
62. **Ding S. and Sun H. (2009)** Preparation method of cutin dispelling cosmetics and use method. Patent: CN101396328.
63. **Dressman D., Yan H., Traverso G., Kinzler K.W. and Vogelstein B. (2003)** Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proceedings of the National Academy of Sciences of the United States of America*. **100(15)**: 8817-8822.
64. **Edwards A. and Caskey C.T. (1991)** Closure strategies for random DNA sequencing. *Methods*. **3(1)**: 41-47.
65. **Ekblom R. and Wolf J.B.W. (2014)** A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*. **7(9)**: 1026–1042.
66. **El-Dadaly H. and Zaid K. A**. **(1999)** Microbial degradation of native keratin in batch fermentation. *Pakistan Journal of Biological Sciences*. **2**: 627-634.
67. **El-Metwally S., Hamza T., Zakaria M. and Helmy M. (2013)** Next-generation sequence assembly: four stages of data processing and computational challenges.*PLoS Computational Biology*.**9(12): e1003345.**
68. **Encarna P. and Elena F.K. (2011)** Tape, in particular, adhesive tape, for the treatment of skin disorders comprising at least one hyperkeratosis inhibitor and/or at least one keratinolytic agent. Patent: WO2011050947.

69. **Esawy M.A. (2007)** Isolation and partial characterization of extracellular keratinase from a Novel Mesophilic *Streptomyces albus* AZA. *Research Journal of Agriculture and Biological Science*. **3**: 808- 817.
70. **Evans K.L., Crowder J. and Miller E.S. (2000)** Subtilisins of *Bacillus* spp. Hydrolyze keratin and allow growth on feathers.*Canadian Journal of Microbiology*. **46(11)**: 1004–1011.
71. **Fakh-fakh N., Haddar A., Hmidet N., Farikha F. and Moncef N. (2010)** Application of statistical experimental design for optimization of keratinases production by *Bacillus pumilus* A1 grown on chicken feather and some biochemical properties. *Process Biochemistry*. **45**: 617–626.
72. **Fakhfakh N., Kanoun S., Manni L. and Nasri M. (2009)** Production and biochemical and molecular characterization of a keratinolytic serine protease from chicken feather-degrading *Bacillus licheniformis* RPk.*Canadian Journal of Microbiology*. **55(4)**: 427–436.
73. **Fedurco M., Romieu A., Williams S., Lawrence I. and Turcatti G. (2006)** BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Research*. **34(3)**: e22.
74. **Fellahi S. (2009)** Isolation, identification, and proteolytic activity of some protease producing bacteria. M.Sc. Dissertation. Alexandria: Alexandria University.
75. **Fellahi S., Zaghloul T., Feuk-Lagerstedt E. and Taherzadeh M. (2014)** A *Bacillus* strain able to Hydrolyze alpha- and beta-keratin. *Journal of Bioprocess and Biotechnology*. **4**: 1000181-1000187.
76. **Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J-F., Dougherty B.A. and Merrick J.M. (1995)** Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*. **269**: 496–512.
77. **Forgács G., Alinezhad S., Mirabdollah A., Feuk-Lagerstedt E. and Horváth I.S. (2011)** Biological treatment of chicken feather waste for improved biogas production. *Journal of Environmental Science*. **23**: 1747-1753.
78. **Forgács G, Lundin M, Taherzadeh M, Horvath I (2013)** Pretreatment of chicken feather waste for improved biogas production. *Applied Biochemistry and Biotechnol*ogy. **169**: 2016-2028.
79. **Frankena J., Koningstein GM., Van Verseveld H.W. and Stouthamer A.H. (1986)** Effect of different limitations in chemostat cultures on growth and production of exocellular protease by *Bacillus licheniformis. Applied Microbiology and Biotechnology*. **24**: 106–112.
80. **Frankena J., Van Verseveld H.W and Stouthamer A.H. (1985)** A continuous culture study of he bioenergetic aspects of growth and production of exocellular protease in *Bacillus licheniformis. Applied Microbiology and Biotechnology*. **22**:169–176.
81. **Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A., Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G. and Kelley J.M. (1995)** The minimal gene complement of *Mycoplasma genitalium. Science***270**: 397–404.
82. **Fraser R.D. and Parry D.A. (2008)** Molecular packing in the feather keratin filament. *Journal of Structural Biology*.**162**:1–13.
83. **Freddi G., Mossotti R. and Innocenti R.(2003)** Degumming of silk fabric with several proteases. *Journal of Biotechnology*. **106**: 101-112.
84. **Friedrich J., Gradisar H., Mandin D. and Chaumont J.P. (1999)** Screening fungi for synthesis of keratinolytic enzymes. *Letters in Applied Microbiology*. **28**: 127–130.
85. **Friedrich J., Gradisar H., Vrecl M. and Pogacnik A. (2005)***In vitro* degradation of porcine skin epidermis by a fungal keratinase of *Doratomyces microsporus*. *Enzyme and Microbial Technology*. 36:455–460.

86. **Gallegos N.G., Jun J. and Hageman J.H. (1996)** Preparation of general proteinase substrates using 3,5-dinitrosalicylaldehyde. *Journal of Biochemical and Biophysical Methods***33**: 31–41.
87. **Ghaemi Oskouie S.F., Tabandeh F., Yakchali B. and Eftekhar F. (2008).** Response surface optimization of medium composition for alkaline protease production by *Bacillus clausii*. *Biochemical Engineering Journal.* **39(1)**: 37-42.
88. **Giongo J.L., Lucas F.S., Casarin F., Heeb P. and Brandelli A. (2007)** Keratinolytic proteases of *Bacillus* species isolated from the Amazon basin showing remarkable de-hairing activity. *World Journal of Microbiology and Biotechnology*.**23**: 375–382.
89. **Global Proteases Market – Grpwth, Trends and Forecasts (2017 – 2022) (2017)** https://www.researchandmarkets.com/research/dm5c8j/global_proteases
90. **Gobinath S.C.B., Anbu P., Lakshmipriya T., Tang T-H., Chen Y., Hashim U., Ruslinda A.R. and Arshad M.K. (2015)** Biotechnological aspects and perspective of microbial keratinase production. *Biomedical Research International*. **2015(2)**:1–10.
91. **Gonnella G. and Kurtz S. (2012)** Readjoiner: a fast and memory efficient string graph-based sequence assembler. *BMC Bioinformatics*. **13**: 82.
92. **Goodwin S., Gurtowski J., Sayers S. E., Deshpande P., Schatz M. and McCombie W.R. (2015**) Oxford Nanopore Sequencing and de novo Assembly of a Eukaryotic Genome. *Genome Research*. **25**: 1750–1756.
93. **Goodwin S., McPherson J. D. and McCombie W. R. (2016)** Coming of age: ten years of next-generation sequencing technologies. *Nature Review Genetics*. **17(6)**: 333-351.
94. **Gousterova A., Braikova D., Goshev I., Christov P., Tishinov K., Vasileva-Tonkova E., Haertle T. and Nedkov P. (2005)** Degradation of keratin and collagen containing wastes by newly isolated *Thermoactinomycetes* or by alkaline hydrolysis. *Letters in Applied Microbiology*.**40**: 335-340.
95. **Gousterova A., Nustorova M., Goshev I., Christov P., Braikova D., TishinovK., Haertlé T. and Nedkov P. (2003)** Alkaline hydrolysate of waste sheep wool aimed as fertilizer. *Biotechnology & Biotechnological Equipment*. **17(2)**: 140-145.
96. **Gousterova A., Nustorova M., Paskaleva D., Naydenov M., Neshev G. and Vasileva-Tonkova E. (2012)** Assessment of feather hydrolysate from thermophilic actinomyces for soil amendment and biological control application. *International Journal of Environmental Research*. **6**: 467-474.
97. **Gradisar H., Friedrich J., Krizaj I. and Jerala R. (2005)** Similarities and specificities of fungal keratinolytic proteases: comparison of keratinases of *Paecilomyces marquandii* and *Doratomyces microsporus* to some known proteases.*Applied and Environmental Microbiology*. **71 (7)**: 3420–3426.
98. **Grazziotin A., Pimentel F.A., de Jong E.V. and Brandelli A. (2006)** Nutritional improvement of feather protein by treatment with microbial keratinase. *Animal Feed and Science Technology*. **126**: 135–144.
99. **Grazziotin A., Pimentel F.A., de Jong E.V. and Brandelli A. (2008)** Poultry feather hydrolysate as a protein source for growing rats. *Brazilian Journal of Veterinary Research and Animal Science*. **45**: 61–67.
100. Guan Z.B**.,** Cai Y.J**.,** Zhang Y.Z**.,** Zhao H**. and** Liao X.R**. (2015)** Complete genome sequence of *Bacillus pumilus* W3: A strain exhibiting high laccase activity. *Journal Biotechnology*.**207**: 8-9.
101. **Gupta R., Beg Q. K., Khan S. and Chauhan B. (2002c)** An overview on fermentation, downstream processing and properties of microbial alkaline proteases. *Applied Microbiology and Biotechnology*. **60**: 381–395.
102. **Gupta R., Beg Q. K. and Lorenz P.(2002b)** Bacterial alkaline proteases: molecular approaches and industrial applications. *Applied Microbiology and Biotechnology*. **59**:15-32.

103. Gupta R., Beg Q. K. and Puri S. (2002a) Optimization of alkaline protease production from *Bacillus* sp. by response surface methoology. *Current Microbiology*. **44**: 286-290.

104. Gupta R., Rajput R., Sharma R. and Gupta N. (2013b) Biotechnological applications and prospective market of microbial keratinases. *Applied Microbiology and Biotechnology*. **97**: 9931–9940.

105. Gupta R., Ramnani P. (2006) Microbial keratinases and their prospective applications: an overview. *Applied Microbiology and Biotechnology*. **70(1)**: 21–33.

106. Gupta R., Sharma R. and Beg Q. K. (2013a) Revisiting microbial keratinases: next generation proteases for sustainable biotechnology. *Critical Reviews in Biotechnology*. **33(2)**: 216-228.

107. Gupta R., Tiwary E., Sharma R., Rajput R. and Nair N. (2013c) Microbial Keratinases: Diversity and applications. In*Thermophilic Microbes in Environmental and Industrial Biotechnology*. **33** pp 881-904.

108. Gupta S. and Singh R. (2013) Statistical modeling and optimization of keratinase production from newly isolated *Bacillus subtilis* RSE163. *International Journal of Advanced Biotechnology and Research*. **4**:167–174.

109. Gupta S. and Singh R. (2014) Hydrolyzing proficiency of keratinases in feather degradation. *Indian Journal of Microbiology*. **54(4)**: 466-470.

110. Gushterova A., Vasileva-Tonkova E., Dimova E., Nedkov P. and Haertlé T. (2005) Keratinase production by newly isolated Antarctic actinomycete strains. *World Journal of Microbiology and Biotechnology*. **21 (6)**: 831–834.

111. Hagen H.E., Klager S.L., McKerrow J.H. and Ham P.J. (1997) A simple assay for the detection of native and recombinant protease activity. *Analytical Biochemistry*. **251**: 121–122.

112. Hameed A., Keshavarz T. and Evans C.S. (1999) Effect of dissolved oxygen tension and pH on the production of extracellular protease from a new isolate of *Bacillus subtilis* K2, for use in leather processing. *Journal of Chemical Technology and Biotechnology*. **74**: 5–8.

113. Hanlon G.W. and Hodges N.A. (1981) Bacitracin and protease production in relation to sporulation during exponential growth of *Bacillus licheniformis* on poorly utilized carbon and nitrogen sources. *Journal of Bacteriology*. **147**: 427–431.

114. Han M., Luo W., Gu Q. and Yu X. (2012) Isolation and characterization of a keratinolytic protease from a feather-degrading bacterium *Pseudomonas aeruginosa* C11. *African Journal of Microbiological Research.* **6(9)**: 2211–2221.

115. Hartley B.S. (1960) Proteolytic enzymes.*Annual Review of Biochemestry*. ***29*** 45–72.

116. Hassan M.A., Haroun B.M., Amara A.A. and Serour **E. A. (2013)** Production and characterization of keratinolytic protease from new wool-degrading *Bacillus* species isolated from Egyptian ecosystem. *Biomedical Research International*. **2013**:175012.

117. Hernandez D., Francois P., Farinelli L., Osteras M. and Schrenzel J. (2008)*De novo* bacterial genome sequencing: millions of very short reads assembled on a desktop computer. *Genome Research*. **18**: 802–809.

118. Hmidet N., El-Hadj Ali N., Haddar A., Kanoun **S.,** Sellami-Kamoun A. **and** Nasri **M. (2009)** Alkaline proteases and thermostable α-amylase co-produced by *Bacillus licheniformis* NH1: Characterization and potential application as detergent additive. *Biochemical Engineering Journal*. **47**:71-79.

119. Hong G.F. (1981) A method for sequencing single-stranded cloned DNA in both directions. *Bioscience reports*, vol. 1(3): 243-252.

120. Hossain M.S., Azad A.K., Abu Sayem S.M., Mostafa G. and Hoq M.M. (2007) Production and partial characterization of feather-degrading keratinolytic serine protease from *Bacillus licheniformis* MZK-3.*Journal of Biological Science*. **7(4)**: 599–606.

121. **Hossain M.S., Azimi N. and Skiena S. (2009)** Crystallizing short-read assemblies around seeds. *BMC Bioinformatics*. **10**: S16.

122. **Huang Y., Busk P. K., Herbst F-A. and Lange L. (2015)** Genome and secretome analyses provide insights into keratin decomposition by novel proteases from the non-pathogenic fungus Onygena corvina. *Applied Microbiology and Biotechnology*. **99**: 9635–9649.

123. **Hubner U., Bock U. and Schugerl K. (1993)** Production of alkaline serine protease subtilisin Carlsberg by *Bacillus licheniformis* on complex medium in a stirred tank reactor. *Applied Microbiology and Biotechnology*. **40**: 182–188.

124. **Hui Z., Doi H., Kanouchi H., Matsuura Y., Mohri S., Nonomura Y. and Oka T.(2004)** Alkaline serine protease produced by *Streptomyces* sp. degrades PrP^{Sc}. *Biochemical and Biophysical Research Communication*. **321**:45-50.

125. **Huse S. M., Huber J. A., Morrison H. G., Sogin M. L. and Welch D. M. (2007)** Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology*. **8(7)**: R143.

126. **Illumina I. (2010)** Genomic sequencing, [Homepage of Illumina Inc.], [Online]. Available:https://www.illumina.com/documents/products/datasheets/datasheet_genomic_sequence.pdf.

127. **Illumina I. (2016) An introduction to Next generation sequencing technology** [Homepage of Illumina Inc.], [Online]. https://www.illumina.com/content/dam/illumina marketing/documents/products/illumina_sequencing_introduction.pdf.

128. **Imelfort M. and Edwards D. (2009)***De novo* sequencing of plant genomes using second-generation technologies. *Briefings in Bioinformatics*. **10(6)**: 609-618.

129. **Industrial Enzymes Market by type (carbohydrases, proteases, non-starch polysaccharides & others), application (food & beverage, cleaning agents, animal feed & others), brands & by region— global trends and forecasts to 2020. (2015)** www.bccresearch.com. http:// www.marketsandmarkets.com/Market-Reports/industrial-enzymesmarket-237327836.html

130. **Infante I., Morel M.A., Ubalde M.C., Martinez-Rosales C., Belvisi S. and Castro-Sowninski S. (2010)** Wool-degrading *Bacillus* isolates: Extracellular protease production for microbial processing of fabrics. *World Journal of Microbiology and Biotechnology*. **26**: 1047–1052.

131. **Inhs D.A., Schimdt W. and Ritcher F.R. (1999)** Proteolytic enzyme cleaner.US patent number 5961366.

132. Ip CL., Loose M., Tyson J.R., de Cesare M., Brown B.L., Jain M., Leggett R.M., Eccles D.A., Zalunin V., Urban J.M., Piazza P., Bowden RJ., Paten B., Mwaigwisya S., Batty E.M., Simpson J.T., Snutch TP., Birney E., Buck D., Goodwin S., Jansen H.J., O'Grady J. **and** Olsen H.E.**(2015)**MinION analysis and reference consortium: Phase 1 data release and analysis. *F1000 Research*.**4**: 1075.

133. **Iqbal S.M., Akin D. and Bashir R. (2007)** Solid-state nanopore channels with DNA selectivity. *Nature Nanotechnology*. **2**: 243–248.

134. **Jani S.A., Chudasama C.J., Patel D.B., Bhatt P.S. and Patel H.N. (2012)** Optimization of Extracellular Protease Production from Alkali Thermo Tolerant Actinomycetes: *Saccharomonospora viridis SJ-21*. *Bulletin of Environment, Pharmacology and Life Sciences*. **1(6)**: 84 – 92.

135. **Jaouadi B, Abdelmalek B, Fodil D, Ferradj FZ, Rekiz H, Zarai N and Bejar S (2010)** Purification and characterization of a thermostable keratinolytic serine alkaline proteinase from *Streptomyces* sp. strain AB1 with high stability in organic solvents. *Bioresource Technology*. **101**: 8361-8369.

136. **Jaouadi B., Abdelmalek B., Jaouadi N. Z. and Bejar S. (2011)** The Bioengineering and Industrial Applications of Bacterial Alkaline Proteases: the Case of SAPB and KERAB. Progress in Molecular and

Environmental Bioengineering - From Analysis and Modeling to Technology Applications, Prof. Angelo Carpi (Ed.), InTech, DOI: 10.5772/23850.

137. **Jasmin C., Chellappan S., Sukumaran R.K., Elyas K.K., Bhat S.G. and Chandrasekaran M. (2010)** Molecular cloning and homology modelling of a subtilisin-like serine protease from the marine fungus, *Engyodontium album* BTMFS10. *World Journal of Microbiology and Biotechnology.* **26:**1269–1279.
138. **Jeong J.H., Lee O.M., Jeon Y.D., Kim J.D. and Lee N.R.(2010)** Production of keratinolytic enzyme by a newly isolated feather-degrading *Stenotrophomonas maltophilia* that produces plant growth-promoting activity. *Process Biochemistry*. **45**: 1738–1745.
139. **Jisha V. N., Smitha R. B., Pradeep S., Sreedevi S., Unni K. N., Sajith S., Priji P., Josh M. S., Benjamin S. (2013)** Versatility of microbial proteases. *Advances in Enzyme Research*.**1 (3)**: 39-51.
140. **Kalisz H.M. (1988)** Microbial proteinases. *Advanced Biochemical Engineering Biotechnology*. **36:** 1–65.
141. **Kannahi M. and Ancy R.J. (2012)** Keratin degradation and enzyme producing ability of *Aspergillusflavus* and *Fusarium solani* from soil. *Journal of Chemical and Pharmaeutical Research*. **4**: 3245–3248.
142. **Kanekar P.P., Nilegaonkar S.S., Sarnaik S.S. and Kelkar A.S. (2002)** Optomization of protease activity of alkaliphilic bacteria isolated from an alkaline lake in India. *Bioresource Technology*. **85**: 87-93.
143. **Karthikeyan R., Balaji S. and Sehgal P.K. (2007)** Industrial applications of keratins—a review. *Journal of Science and Industrial Research*. **66(9)**: 710–715.
144. **Khardenavis A.A., Kapley A. and Purohit H.J. (2009)** Processing of poultry feathers by alkaline keratin hydrolyzing enzyme from *Serratia* sp. HPC 1383. *Waste Management*. **29**: 1409–1415.
145. **Khusro A. (2015)** Statistical Approach for Optimization of Independent Variables on Alkali-Thermo Stable Protease Production from *Bacillus Licheniformis* Strain BIHPUR 0104. *Electronic Journal of Biology*.**11(3)**: 93-97.
146. **Kihm D.J., Hutton M.T., Hanlin J.H. and Johnson E.A. (1988)** Zinc stimulates sporulation in *Clostridium botulinum* 113B. *Current Microbiology*. **17**: 193-198.
147. **Kluskens L.D., Voorhorst W.G.B., Siezen R.J., Schwerdtfeger R.M., Antranikian G., Van der Oost J. and De Vos W.M. (2002)** Molecular characterization of fervidolysin, a subtilisin-like serine protease from the thermophilic bacterium *Fervidobacterium pennivorans*. *Extremophiles*.**6**: 185–194.
148. **Kojima M., Kanai M., Tominaga M., Kitazume S., Inoue A. and Horikoshi K. (2006)** Isolation and characterization of a feather-degrading enzyme from *Bacillus pseudofirmus* FA30-01.*Extremophiles*. **10(3)**: 229–235.
149. **Kopan R and Fuchs E (1989)** A new look into an old problem: keratins as tools to investigate determination, morphogenesis, and differentiation in skin.*Genes & Development*.**3 (1)**: 1-15
150. **Kornillowicz-Kowalska T. and Bohacz J. (2011)** Biodegradation of keratin waste: theory and practical aspects. *Waste Management*. **31(8)**: 1689–701.
151. **Krishnan S., Prapulla S. G., Rajalakshmi D., Misra M. C. and Karanth N. G. (1998)**Screening and selection of media components for lactic acid production using Plackett-Burman design. *Bioprocess Engineering*. **19:** 61-65.
152. **Krieger Y., Bogdanov-Berezovsky A., Gurfinkel R., Silberstein E., Sagi A. and Rosenberg L. (2012)** Efficacy of enzymatic debridement of deeply burned hands. *Burns*. **38**: 108-112.
153. **Kshetri P., Ningombam O. and Ningthoujam D. S. (2016)** Optimization of alkaline protease production by alkaliphilic *Bacillus* sp. KW2 in low cost medium using statistical approaches. *Applied Microbiology.* **2(3):**DOI: 10.4172/2471-9315.1000117.

154. **Kublanov I.V., Tsiroulnikov K.B., Kaliberda E.N.,Rumsh L. D., Haertlé T. and Bonch-Osmolovskaya E. A. (2009)** Keratinase of an anaerobic thermophilic bacterium *Thermoanaerobacter* sp. Strain 1004-09 isolated from a hot spring in the Baikal rift zone. *Microbiology*. **78(1)**: 67–75.

155. **Kumar A.G., Swarnalatha S., Gayathri S., Nagesh N. and Sekaran G. (2008a)** Characterization of an alkaline active thiol forming extracellular serine keratinase by the newly isolated *Bacillus pumilus*. *Journal of Applied Microbiology*. **104**: 411–419.

156. Kumar C.G**. and** Takagi H**. (1999)** Microbial alkaline proteases: from a bioindustrial viewpoint. *Biotechnology Advances.***17(7)**: 561-94.

157. **Kumar D.**, **Kumar A.**, **George N. and Gupta N. (2016)** Production optimization and characterization of an alkaline thermostable protease and its application as laundry additive. *International Journal of Scientific Engineering and Applied Science*.**2(6)**: 9-43.

158. **Kumari M., Sharma A. and Jagannadham M.V. (2012)** Religiosin B, a milk-clotting serine protease from *Ficus religiosa*. *Food Chemistry*. **131**: 1295-1303.

159. **Kumar R., Balaji S., Uma T.S., Mandal A.B. and Sehgal P.K. (2010)** Optimization of influential parameters for extracellular keratinase production by *Bacillus subtilis* (MTCC9102) in solid state fermentation using horn meal - a biowaste management. *Applied Biochemistry and Biotechnology*. **160**: 30-39.

160. **Kumar D., Savitri, Thakur N., Verma R. and Bhalla T. C. (2008b)** Microbial proteases and application as laundry detergent additive. *Research Journal of Microbiology*. **3**: 661-672.

161. **Lakshmi B.K.M. and Hemalatha K.P.J. (2016)** Production of alkaline protease from *Bacillus licheniformis* through statistical optimization of growth media by response surface methodology. *Fermentation Technology*. **5(130),** DOI:10.4172/2167-7972.1000130.

162. **Lakshmi B.K.M., Ratna P.V., Ambika Devi K. and Hemalatha KP.J. (2014)**Screening, optimization of production and partial characterization of alkaline protease from haloalkalophilic *Bacillus* sp. *International Journal of Research in Engineering and Technology*. **3(2)**: 435-443.

163. **Laemmli U. K. (1970)** Cleavage of the structural proteins during the assembly of the head of the bacteriophage T4. *Nature*. **227:** 680 – 685.

164. **Land M., Hauser L., Jun S.-R., Nookaew I., Leuze M.R., Ahn T.-H., Karpinets T., Lund O., Kora G., Wassenaar T., Poudel S. and Ussery D. W. (2015)** Insights from 20 years of bacterial genome sequencing. *Functional and Integrative Genomics*. **15**: 141–161.

165. **Lange L., Busk P.K. and Huang Y. (2014)** Use of a microbial composition for the degradation of keratinaceous materials. Denmark Patent WO 2014/169920 A2.

166. **Lange L., Huang Y. and Busk P. K. (2016)** Microbial decomposition of keratin in nature—a new hypothesis of industrial relevance. *Applied Microbiology and Biotechnology*. **100**: 2083–2096.

167. **Lateef A., Oloke J.K., Gueguim E.B., Sobowale B.O., Ajao S.O. and Bello B.Y. (2010)** Keratinolytic activities of a new feather-degrading isolate of *Bacillus cereus* LAU 08 isolated from Nigerian soil. *International Biodeterioration and Biodegradation*. **64**: 162–165.

168. **Luria S.E., Adams J.N. and Ting R.C. (1960)** Transduction of lactose-utilizing ability among strains of *Escherichia coli* and *Shigella dysenteriae* and the properties of the transducing phage particles. *Virology*. **12**: 348-390.

169. **Lee H., Suh D. B., Hwang J. H. and Suh H. J. (2002)** Characterization of a keratinolytic metalloprotease from *Bacillus* sp. SCB-3. *Applied Biochemistry and Biotechnology***. 97**: 123-133.

170. Leighton T. J. and Doi R. H. (1971) The stability of messenger ribonucleic acid during sporulation in *Bacillus subtilis*. *Journal of Biological Chemistry*. **264**: 3189-3195.

171. Liang X., Bian Y., Tang X.F., Xiao G. and Tang B. (2010) Enhancement of keratinolytic activity of a thermophilic subtilase by improving its autolysis resistance and thermostability under reducing conditions. *Applied Microbiology and Biotechnology*. **87(3)**: 999–1006.

172. Li H. (2012) Exploring single-sample SNP and INDEL calling with whole genome *de novo* assembly. *Bioinformatics*. **28**: 1838–1844.

173. Li J., Shi P.J., Han X.Y., Meng K., Yang P.L., Wang Y.R., Luo H.Y., Wu N.F., Yao B. and Fan Y.L. (2007) Functional expression of the keratinolytic serine protease gene sfp2 from *Streptomyces fradiae* var. k11 in *Pichia pastoris.Protein Expression Purification*. **54**: 79–86.

174. Lin H.H. and Yin L.J. (2010) Feather meal and rice husk enhanced keratinase production by *Bacillus licheniformis* YJ4 and characteristics of produced keratinases. *Journal of Marine Science and Technology*. **18(3)**: 458-465.

175. Lin H.H., Yin L.J. and Jiang S.T. (2009) Cloning, expression, and purification of *Pseudomonas aeruginosa* keratinase in *Escherichia coli* AD494(DE3) pLysS expression system. *Journal of Agricultural and Food Chemistry*. **57**: 3506–3511.

176. Liu B**.,** Zhang J**.,** Li B**.,** Liao X**.,** Du G**. and** Chen J**. (2013)**Expression and characterization of extreme alkaline, oxidation-resistant keratinase from *Bacillus licheniformis* in recombinant *Bacillus subtilis* WB600 expression system and its application in wool fiber processing. *World Journal of Microbiology and Biotechnology*.**29(5)**: 825-832.

177. Liu J., Xing J., Chang T., Ma Z. and Liu H. (2005) Optimization of nutritional conditions for nattokinase production by *Bacillus natto* NLSSE using statistical experimental methods. *Process Biochemistry*. **40**: 2757–2762.

178. Liu L., Li Y., Li S., Hu N., He Y., Pong R., Lin D., Lu L., and Law M. (2012) Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology*. **2012**: 1-11.

179. Loman N. J., Constantinidou C., Chan, J. Z., Halachev M., Sergeant M., Penn C. W., Robinson E. R. and Pallen M. J. (2012) High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nature Reviews Microbiology*.**10(9)**: 599-606.

180. Lowe T.M. and Eddy S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research.***25**: 955–964.

181. Lowry O. H., Rosebrough N. J., Farr A. L. and Randell R. J. (1951) Protein measurement with the Folin phenol reaction. *Journal of Biological Chemistry*. **193**: 265–275.

182. Lynch M.H., O'Guin W.M., Hardy C., Mark L. and Sun T.T. (1986) Acidic and basic hair/nail 'hard' keratins: their co-localisation in upper cortical and cuticle cells of the human hair follicle and their relationship to 'soft' keratins. *Journal of Cell Biology*. **103**:2593-2606.

183. Macedo A.J., da Silva W.O.B., Gava R., Driemeier D., Henriques J.A.P. and Termignoni C. (2005) Novel keratinase from *Bacillus subtilis* S14 exhibiting remarkable dehairing capabilities. *Applied Environmental Microbiology*. **71**: 594–596.

184. Manafi A., Hashemlou A., Momeni P. and Moghimi H.R. (2008) Enhancing drugs absortion through third-degree burn wound eschar. *Burns*. **34**: 698–702.

185. Mao W., Pan R. and Freedman D. (1992) High production of alkaline protease by *Bacillus licheniformis* in fed-batch fermentation using a synthetic medium. *Journal of Industrial Microbiology*. **11**: 1–6.

186. Mardina V., Yusof F. and Alam MD. Z. (2015) Statistical optimization of physichochemical factors for protease production by *Bacillus licheniformis* on skim latex serum effluent fortified media. *Journal of Engineering Science and Technology*. **6(1)**: 42–52.

187. Mardis E. R. (2013) Next-generation sequencing platforms. *Annual Review and Analytical Chemistry*. **6**: 287–303.

188. Margulies M., Egholm M., Altman W.E., Attiya S., Bader J.S., Bemben L.A., Berka J.**,** Braverman M.S.**,** Chen Y.J.**,** Chen Z.**,** Dewell S.B.**,** Du L.**,** Fierro J.M.**,** Gomes X.V.**,** Godwin B.C.**,** He W.**,** Helgesen S.**,** Ho C.H.**,** Irzyk G.P.**,** Jando S.C.**,** Alenquer M.L.**,** Jarvie T.P.**,** Jirage K.B.**,** Kim J.B.**,** Knight J.R.**,** Lanza J.R.**,** Leamon J.H.**,** Lefkowitz S.M.**,** Lei M.**,** Li J.**,** Lohman K.L.**,** Lu H.**,** Makhijani V.B.**,** McDade K.E.**,** McKenna M.P.**,** Myers E.W.**,** Nickerson E.**,** Nobile J.R.**,** Plant R.**,** Puc B.P.**,** Ronan M.T.**,** Roth G.T.**,** Sarkis G.J.**,** Simons J.F.**,** Simpson J.W.**,** Srinivasan M.**,** Tartaro K.R.**,** Tomasz A.**,** Vogt K.A.**,** Volkmer G.A.**,** Wang S.H.**,** Wang Y.**,** Weiner M.P.**,** Yu P.**,** Begley R.F**. and** Rothberg J.M**. (2005)** Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. **437**: 376–380.

189. Markowitz V.M., Chen I.M.A., Palaniappan K., Chu K., Szeto E., Grechkin Y., Ratner A., Anderson I., Lykidis A., Mavromatis K., Ivanova N.N. and Kyrpides N.C. (2009) The integrated microbial genomes system: an expanding comparative analysis resource. *Nucleic Acids Research*. **38**: 382–390.

190. Martin J.A. and Wang Z. (2011) Next-generation transcriptome assembly. Nat Rev Genet **12**: 671–682.

191. Matikeviciene V., Masiliuniene D. and Grigiskis S. (2009) Degradation of keratin containing wastes by bacteria with keratinolytic activity. *Environment,Technology, Resources*. **1**: 284-289.

192. Mazotto A.M., Cedrola S.M.L., Lins U., Rosado A.S., Silva K.T., Chaves J.Q., Rabinovitch L., Zingali R.B. and Vermelho A.B. (2010) Keratinolytic activity of *Bacillus subtilis* AMR using human hair. *Letters in Applied Microbiology*. **50(1)**: 89–96.

193. McKerrow J.H., Sun E., Rosenthal P.J., Bouvier J. (1993) The proteases and pathogenicity of parasitic protozoa. *Annual Review of Microbiology*. **47**: 821–853.

194. McKittrick J., Chen P.Y., Bodde S.G., Yang W., Novitskaya E.E. and Meyers M.A. (2012) The structure, functions, and mechanical properties of keratin. *JOM*. **64 (4)**:449–468.

195. Medvedev P. and Brudno M. (2009) Maximum likelihood genome assembly. *Journal of Computetional Biology*. **16**: 1101–1116.

196. Medvedev P., Georgiou K., Myers G. and Brudno M. (2007) Computability of models for sequence assembly. In: Giancarlo R. and Hannenhalli S., editors. Algorithms in bioinformatics. Springer Berlin Heidelberg. pp. 289–301.

197. Menon A. S. and Goldberg A. L. (1987) Protein substrates activate the ATP-dependent protease La by promoting nucleotide binding and release of bound ADP. *Journal of Biological Chemistry*.**262**:14929–14934.

198. Meyers M.A., Chen P., Lin A.Y. and Seki Y. (2008) Biological materials: structure and mechanical properties. *Progress in Materials Science*. **53**: 1–206.

199. Miller J.R., Delcher A.L., Koren S., Venter E., Walenz B.P., Brownley A.**,** Johnson J.**,** Li K.**,** Mobarry C**. and** Sutton G**.(2008)** Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics*. **24**: 2818–2824.

200. Miller J. R., Koren S. and Sutton G. (2010) Assembly algorithms for next-generation sequencing data. Genomics. **95**: 315–327.

201. Mitsuiki S., Ichikawa M., Oka T., Sakai M., Moriyama Y., Sameshima Y., Goto M. and Furukawa K. (2004) Molecular characterization of a keratinolytic enzyme from an alkaliphilic *Nocardiopsis* sp. TOA-1.*Enzyme and Microbial Technology*. **34(5)**: 482–489.

202. Mohamedin A.H. (1999) Isolation identification and some cultural conditions of a protease producing thermophilic *Streptomyces* strain grown on chicken feather as a substrate. *International Biodeterioration and Biodegradation*. **43**:13-21.

203. Montgomery D. C. (Ed) (2005) Design and Analysis of Experiments: Response surface method and designs,8th ed. New Jersey: John Wiley and Sons, Inc.

204. More S.V., Khandelwal H.B., Joseph M.A. and Laxman R.S. (2013) Enzymatic degumming of silk with microbial proteases. *Journal of Natural Fibers***10(2)**: 98–111.

205. Muller-Hellwiga S., Groschupb M. H., Pichnerc R., Gareisc M., Martlbauerd E., Scherera S. and Loessner M. J. (2006) Biochemical evidence for the proteolytic degradation of infectious prion protein PrP^{sc} in hamster brain homogenates by foodborne bacteria. *Systematic and Applied Microbiology*. **29:** 165-171.

206. Murdan S. (2002) Drug delivery to the nail following topical application. *International Journal of Pharmacy*. **236**: 1-26.

207. Myers E.W.**,** Sutton G.G.**,** Delcher A.L.**,** Dew I.M.**,** Fasulo D.P.**,** Flanigan M.J.**,** Kravitz S.A.**,** Mobarry C.M.**,** Reinert K.H.**,** Remington K.A.**,** Anson E.L.**,** Bolanos R.A.**,** Chou H.H.**,** Jordan C.M.**,** Halpern A.L.**,** Lonardi S.**,** Beasley E.M.**,** Brandon R.C.**,** Chen L.**,** Dunn P.J.**,** Lai Z.**,** Liang Y.**,** Nusskern D.R.**,** Zhan M.**,** Zhang Q.**,** Zheng X.**,** Rubin G.M.**,** Adams M.D**. and** Venter J.C.**(2000)** A whole-genome assembly of *Drosophila*, *Science*. **287** (5461): 2196–2204.

208. Nagal S. and Jain P.C. (2010) Feather degradation by strains of *Bacillus* isolated from decomposing feathers.*Brazilian Journal of Microbiology*. **41**: 196–200.

209. Nakazato **T.,** Ohta **T., and** Bono **H. (2013)** Experimental design-based functional mining and characterization of high-throughput sequencing data in the sequence read archive. *PLOS One*.**8(10)**: e77910.

210.Naleway S-E., Taylor J.R.A., Porter M-M., Meyers M-A. and McKittrick J. (2016) Structure and mechanical properties of selected protective systems in marine organisms.*Materials Science and Engineering C*. **59**:1143-1167.

211. Nascimento W.C.A. and Martin M.L.L.(2006) Studies on the stability of protease from *Bacillus* sp. and its compatibility with commercial detergent. *Brazilian Journal of Microbiology*. **37**: 307–311.

212. Nassar F.R., Abdelhafez A.A., El-Tayeb T.S. and Abu-Hussein S.H. (2015) Response surface methodology for the optimization of proteases production by a novel egyptian isolate *Bacillus amyloliquefaciens* 35s.*British Microbiology Research Journal*. **6(5)**: 255-269.

213. Nawrocki E.P. (2014) Annotating functional RNAs in genomes using Infernal. *Methods in Molecular Biology*. **1097**: 163–197.

214. Nawrocki E.P., Burge S.W., Bateman A., Daub J., Eberhardt R.Y., Eddy S.R., Floden E.W., Gardner P.P., Jones T.A., Tate J. and Finn **R. D. (2015)** Rfam 12.0: updates to the RNA families database. *Nucleic Acids Research.***43**: D130–D138.

215. Nawrocki E.P. and Eddy S.R. (2013) Infernal 1.1: 100-fold faster RNA Homology searches. *Bioinformatics*. **29**: 2933–2935.

216. Ng C.S., Wu P., Fan W.L., Yan J., Chen C.K., Lai Y.T., Wu S.M., Mao C.T., Chen J.J., Lu M.Y., Ho M.R., Widelitz R.B., Chen C.F., Chuong C.M. and Li W.H. (2014) Genomic organization, transcriptomic

analysis, and functional characterization of avian α- and β-keratins in diverse feather forms. *Genome Biology Evolution*. **6**: 2258–2273.

217. Ng M. and Auld D.S. (1989) A fluorescent oligopeptide energy transfer assay with broad applications for neutral proteases. *Analytical Biochemistry*. **183**: 50–56.

218. Niedringhaus T. P., Milanova D., Kerby M. B., Snyder M. P. and Barron A. E. (2011) Landscape of Next-generation sequencing technologies. *Analytical chemistry*. **83(12)**: 4327–4341.

219. Nilegaonkar S.S., Zambare V.P., Kanekar P.P., Dhakephalkar P.K. and Sarnaik S.S. (2007) *Production and partial characterization of dehairing protease from Bacillus cereus* MCM B-326.*Bioresource Technology*. **98(6):** 1238–1245.

220. Ogino H., Watanabe F., Yamada M., Nakagawa S., Hirose T., Noguchi A., Yasuda M. and Ishikawa H. (1999) Purification and characterization of organic solvent-stable protease from organic solvent-tolerant *Pseudomonas aeruginosa* PST-01. *Journal of Bioscience and Bioengineering*. **87**:61–68.

221. O'Hara M.B. and Hageman J.H. (1990) Energy and calcium ion dependence of proteolysis during sporulation of *Bacillus subtilis* cells. *Journal of Bacteriology*. **172**: 4161–4170.

222. Oh S., Rheem, S., Sim, J., Kim, S. and Baek, Y. (1995) Optimizing conditions for the growth of Lactobacillus casei YIT 9018 in tryptone-yeast extract-glucose medium by using response surface methodology. *Applied and Environmental Microbiology*. **61**: 3809–3814.

223. Okoroma E., Garelick H., Abiola O. and Purchase D. (2012) Identification and characterisation of a*Bacillus licheniformis* strain with profound keratinase activity for degradation of melanised feather. *International Biodeterioration and Biodegradation*. **74**: 54-60.

224. Olajuyigbe F. M. and Falade A. M. (2014) Purification and partial characterization of serine alkaline metalloprotease from *Bacillus brevis* MWB-01. *Bioresources and Bioprocessing*. **1**: 8.

225. Onifade A., Al-Sane N., Al-Musallam A. and Al-Zarban S. (1998) A review: potentials for biotechnological applications of keratin-degrading microorganisms and their enzymes for nutritional improvement of feathers and other keratins as livestock feed resources. *Bioresource Technology*. **66(1)**: 1–11.

226. Orvis J., Crabtree J., Galens K., Gussman A., Inman J.M., Lee E., Nampally S., Riley D., Sundaram J.P., Felix V., Whitty B., Mahurkar A., Wortman J., White O. and Angiuoli S.V. (2010) Ergatis: a web interface and scalable software system for bioinformatics workflows. *Bioinformatics***26**: 1488–1492.

227. Oskouie S.F.G., Tabandeh F., Yakhchali B. and Eftekhar F. (2008) Response surface optimization of medium composition for alkaline protease production by *Bacillus clausii. Biochemical Engineering Journal.* **39**: 37–42.

228. Outtrup H., Dambmann C., Christiansen M. and Aaslyng D.A. (1995)*Bacillus* sp. JP 395 Method of making and detergent composition. US patent number 5466594.

229. Ouyang S., Thibaud-Nissen F., Childs K.L., Zhu W. and Buell C.R. (2009) Plant genome annotation methods. *Methods in Molecular Biology*. **513**: 263–282.

230. Ozsolak F., Kapranov P., Foissac S., Kim S.W., Fishilevich E., Monaghan A.P., John B. and Milos P.M. (2010) Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell*. **143**: 1018–1029.

231. Pant**G.,** Prakash**A., Pavani J.V.P., Bera S., Devirama G.V.N.S., Kumara A., Panchpurib M., Prasuna R. G. (2015)** Production, optimization and partial purification of protease from *Bacillus subtilis. Journal of Taibah University for Science*. **9**: 50–55.

232. Papadopoulos M. C., El Boushy A. R., Roobdeen, A. E. and Ketelaars E. H. (1986) Effects of processing time and moisture content on amino acid composition and nitrogen characteristics of feather meal. *Animal Feed and Science. Technology*. **14**: 279-290.

233. Pareek C. S., Smoczynski R. and Tretyn A. (2011) Sequencing technologies and genome sequencing. *Journal of Applied Genetics*. **52**: 413–435.

234. Park G.T. and Son H.J. (2009) Keratinolytic activity of *Bacillus megaterium* F7-1, a feather-degrading mesophilic bacterium.*Microbiolgy Research*. **164(4)**: 478–485.

235. Paszkiewicz K. and Studholme D. J. (2010) *De novo* assembly of short sequence reads. *Briefings in Bioinformatics*. **11(5)**: 457-472.

236. Paul T., Das A., Mandal A., Jana A., Maity C., Adak A., Halder S.K., DasMohapatra P.K., Pati B.R. and Mondal K.C. (2014) Effective dehairing properties of keratinase from *Paenibacillus woosongensis* TKB2 obtained under solid state fermentation. *Waste and Biomass Valorisation*. 5(1): 97–107.

237. Peltola H., Soderlund H. and Ukkonen E. (1984) SEQAID: a DNA sequence assembling program based on a mathematical model. *Nucleic Acids Research*. **12**: 307–321.

238. Petinate S.D.G., Branquinha M.H., Coelho R.R.R., Vermelho A.B., DeSimone S.G. (1999) Purification and partial characterizationof an extracellular serine-proteinase of *Streptomyces cyaneus* isolated from Brazilian cerrado soil. *Journal of Applied Microbiology*.**87**: 557–563.

239. Pettersson E**.,** Lundeberg J**. and** Ahmadian A**. (2009)** Generations of sequencing technologies. *Genomics*.**93(2)**:105-111.

240. Pillai P. and Archana G. (2008) Hide depilation and feather disintegration studies with keratinolytic serine protease from a novel *Bacillus subtilis* isolate. *Applied Microbiology and Biotechnology*. **78:** 643–650.

241. Poly F., Read T., Tribble D.R., Baqar **S.,** Lorenzo**M. and** Guerry **P. (2007)** Genome sequence of a clinical isolate of *Campylobacter jejuni* from Thailand. *Infection Immunity***75**: 3425–3433.

242. Pop M. (2009) Genome assembly reborn: recent computational challenges. *Briefings in Bioinformatics*. **10(4)**: 354-366.

243. Pop M., Phillippy A., Delcher A.L. and Salzberg S.L. (2004) Comparative genome assembly. *Brief Bioinformatics*. **5**: 237–248.

244. Poza M., de Miguel T., Sieiro C. and Villa T.G. (2001) Characterization of a broad pH range protease of *Candida caseinolytica*. *Journal of Applied Microbiology*. **91**: 916–921.

245. Powers J.C., Asgian J.L., Ekici Ö.D. and James K.E.(2002) Irreversible inhibitors of serine, cysteine, and threonine proteases. *Chemical Reviews*. **102(12)**: 4639–750.

246. Prakash P., Jayalakshmi S.K. and Sreeramulu K. (2009) Production of keratinase by free and immobilized cells of *Bacillus halodurans* strain PPKS-2: partial characterization and its application in feather degradation and dehairing of the goat skin. *Applied Biochemistry and Biotechnology*. **160(7)**: 1909-1920.

247. Prasad H.V., Kumar G., Karthik L. and Rao K.V.B. (2010) Screening of extracellular Keratinase producing bacteria from feather processing areas in Vellore, Tamil Nadu, India. *Journal of Scientific Research*. **2(3)**: 559–565.

248. Puhl A.A., Selinger L.B. and Mc Allister T.A., (2009)*Actinomadura keratinilytica* sp. nov., a keratin degrading actinobacterium isolated from bovine manure compost. *International Journal of Systematic Evolution and Microbiology*.**59**: 828–834.

249. Purchase, D. **(2016)** Microbial keratinases: characteristics, biotechnological applications and potential. In: The Handbook of Microbial Bioresources. Gupta, Vijai K. and Sharma, Gauri D. and Tuohy, Maria G. and Gaur, Rajeeva, eds. CAB International Publishing, Wallingford, UK, pp. 634-674. ISBN 9781780645216.

250. Puri S., Beg Q.K. and Gupta R. (2002) Optimization of alkaline protease production from *Bacillus* sp. using response surface methodology. *Current Microbiology*. **44**: 286–290.

251. Quail M.A., Smith M., Coupland P., Otto T.D., Harris S.R., Connor T.R., Bertoni A., Swerdlow H. P. and Gu Y. (2012)A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers.*BMC Genomics*. **13**: 341.

252. Rai S.K., Konwarh R. and Mukherjee A.K. (2009) Purification, characterization and biotechnological application of an alkaline b-keratinase produced by *Bacillus subtilis* RM-01 in solidstate fermentation using chicken-feather as substrate. *Biochemical Engineering Journal*. **45**: 218-225.

253. Rajamani S., Hilda A. (1987) Plate assay to screen fungi for proteolytic activity. *Current Science*. **56**: 1179–1181.

254. Ramamurthy V, Upadhyay CM, Kothari RM (1991) An optimized protocol for the preparation and application of acid protease. *Journal of Biotechnology*. **21**: 187-196.

255. Rao M.B., Tanksale A.M., Ghatge M.S. and Deshpande V.V. (1998) Molecular and biotechnological aspects of microbial proteases.*Microbiology and Molecular Biology Review*. **62(3)**:597–635.

256. Rathakrishnan P. and Nagarajan P. (2013) Optimization of the production of protease by *Bacillus cereus* with response surface methodology using groundnut shell. *International Journal of Pharmaceutical, Chemical and Biological Sciences*. **3(2)**: 200-209.

257. Rawlings N. D. and Barrett A. J.(1993) Evolutionary families of peptidases. *Biochemical Journal*. **290**: 205–218.

258. Rawlings N., Morton F. R. and Barrett A. J. (2006) MEROPS: the peptidase database. *Nucleic Acids Research*. **34**: D270–D272.

259. Rawlings N., Morton F., Kok C., Cong J. and Barrett A. (2008) MEROPS: the peptidase database. *Nucleic Acids Research*. **36**:D320–D325.

260. Reddy L.V.A., Wee Y-J., Yun J-S. and Ryu H-W.(2008) Optimization of alkaline protease production by butch culture of *Bacillus* sp. RKY3 through Plackett-Burman and response surface methodological approaches. *Bioresource Technology*. **99**: 2242-2249.

261. Reddy P. R. M., Reddy G. and Seenayya G.(1999) Production of thermostable β-amylase and pullulanase by *Clostridium thermosulfurogenes* SV2 in solid state fermentation: screening of nutrients using Plackett-Burman design. *Bioprocess Engineering*. **21:** 175-179.

262. Reinhardt J.A.**,** Baltrus D.A.**,** Nishimura M.T.**,** Jeck W.R.**,** Jones C.D**. and** Dangl J.L**.** (2009) *De novo* assembly using low-coverage short read sequence data from the rice pathogen *Pseudomonas syringae pv. Oryzae*. *Genome Research*. **19(2)**: 294–305.

263. Reuter J.A.**,** Spacek D.V**. and** Snyder M.P**. (2015)** High-Throughput Sequencing Technologies. *Molecular Cell*. **58(4)**: 586-597.

264. Richardson E. J. and Watson M. (2013) The automatic annotation of bacterial genomes. *Briefings in Bioinformatics*. **14(1)**: 1-12.

265. Riffel A., Daroit D.J. and Brandelli A. (2011) Nutritional regulation of protease production by the feather-degrading *bacterium Chryseobacterium* sp. kr6. *New Biotechnology*. **28**:153–157.

266. Riffel A., Lucas F., Heeb P. and Brandelli A. (2003) Characterization of a new keratinolytic bacterium that completely degrades native feather keratin.*Archive Microbiology*. **179**:258–265.

267. Riessen, S. and Antranikian, G. (2001) Isolation of *Thermoanaerobacter keratinophilus* sp. nov., a novel thermophilic, anaerobic bacterium with keratinolytic activity. *Extremophiles*. **5**: 399-408.

268. Rothberg, J.M. and Leamon, J.H. (2008) The development and impact of 454 sequencing. *Nature Biotechnology*. **26:** 1117–1124.

269. Sabotic J. and Kos J. (2012) Microbial and fungal protease inhibitors—current and potential applications. *Applied Microbiology and Biotechnology*. **93**:1351–1375.

270. Safarik I. (1987a) A modified procedure for the preparation of insoluble chromogenic substrates for the determination of proteolytic activity. *Journal of Biochemical and Biophysical Methods*. **14**: 355–357.

271. Safarik I. (1987b) A simple method for determination of proteolytic activity. *Biotechnol Techniques*.**1**: 135–136.

272. Safarik I. (1987c) Thermally modified azocasein – a new insoluble substrate for the determination of proteolytic activity. *Biotechnology and Applied Biochemistry*. **9**: 323–324.

273. Safarik I. (1988a) An inexpensive insoluble chromogenic substrate for the determination of proteolytic activity. *Journal of Industrial Microbiology*. **3**: 259–261.

274. Safarik I. (1988b) An insoluble chromolytic substrate for the determination of proteolytic activity. *Journal of Biochemical and Biophysical Methods*. **17**:193–198.

275. Safarik I. (1989) Spectrophotometric determination of proteolytic activity in coloured solutions. *Journal of Biochemical and Biophysical Methods*. **19**: 201–206.

276. Saha S., Dhanasekaran D., Shanmugapriya S. and Latha S**.(2013)***Nocardiopsis* sp. SD5: a potent feather degrading rare actinobacterium isolated from feather waste in Tamil Nadu, India. *Journal of Basic Microbiology*. **53(7)**: 608–616.

277. Sambrook J., Frisch E. F. and Maniatis T. (eds). **(1989)** Molecular Cloning: A Laboratory Manual, 2nd ed. Cold spring Harbor laboratory press, New York.

278. Sangali S. and Brandelli, A. (2000) Feather keratin hydrolysis by a *Vibrio* sp. strain kr2.*Journal of Applied Microbiology*. **89**:735–743.

279. Saxena R. and Singh R. (2010) Statistical optimization of conditions for protease production from *Bacillus* sp. *Acta Biologica Szegediensis*. **54(2):**135-141.

280. Schechler I. and Berger A.(1967) On the size of the active site in proteases I papain. *Biochemical and Biophysical Research Communications*. **27**: 157-162.

281. Schrooyen P.M.M., Dijkstra P.J., Oberthur R.C., Bantjes A**. and** Feijen J**. (2001)** Partially carboxymethylated feather keratins. 2. Thermal and mechanical properties of films. *Journal of Agricultural Food and Chemistry*. **49(1)**: 221–230.

282. Schrooyen P.M.M. and Radulf B. (2004) Keratin-based products and methods for their productions. USPatent 0210039.

283. Selvam K. and Vishnupriya B. (2012) Biochemical and molecular characterization of microbial keratinase and its remarkable applications. *International Journal of Pharmaceutical and Biological Archive*. **3(2)**: 267–275.

284. Sharma A. and Satyanarayana T. (2013) Comparative Genomics of *Bacillus* species and its Relevance in Industrial Microbiology. *Genomics Insights*.**6**: 25–36.

285. Sharma K.- M., Kumar R., Panwar S. and Kumar A. (2017) Microbial alkaline proteases: Optimization of production parameters and their properties. *Journal of Genetic Engineering and Biotechnology*. **15(1)**: 115-121.**Shendure J., Porreca G.J., Reppas N.B., Lin X., McCutcheon J.P., Rosenbaum A.M., Wang M.D., Zhang K., Mitra R.D. and Church G.M. (2005)** Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*.**309(5741)**: 1728-1732.

286. Shimogaki H., Takeuchi K., Nishino T., Odhera M., Kudo T., Ohba K., Iwama M. and Irie M. (1991) Purification and properties of a novel surface-active agent- and alkaline-resistant protease from *Bacillus* sp. Y. *Agricultural and Biological Chemistry*. **55**: 2251–2258.

287. Shrinivas D. and Nayak G.R. (2011) Characterization of alkaline thermostable keratinolytic protease from thermoalkalophilic *Bacillus halodurans* JB 99 exhibiting dehairing activity. *International Biodeterioration and Biodegradation*. **65:** 29–35.

288. Silveira S.T., Casari F., Gemelli S. and Brandelli A. (2010) Thermodynamics and kinetics of heat inactivation of a novel keratinase from *Chryseobacterium* sp. strain kr6. *Applied Biochemistry and Biotechnology*.**162(2)**: 548–560.

289. Simpson J.T. and Durbin R. (2012) Efficient *de novo* assembly of large genomes using compressed data structures. *Genome Research*. **22**: 549–556.

290. Singh R., Kumar M., Mittal A. and Mehta P.K. (2016) Microbial enzymes: Industrial progress in 21st century.*3 Biotech*. **6**:174.

291. Sivakumar T., Shankar T., Vijayabaskar P. and Ramasubramanian V. (2013) Optimization of cultural condition for keratinase production using *Bacillus cereus* TS1. *Insight Microbiology*. **3**: 1–8.

292. Son H.J., Park H.C., Kim H.S. and LeeC.-Y. (2008) Nutritional regulation of keratinolytic activity in *Bacillus pumilis*. *Biotechnology Letters*. **30**: 461–465.

293. Stothard P**. and** Wishart D.S**. (2006)** Automated bacterial genome analysis and annotation. *Current Opinion in Microbiology*. **9**: 505–510.

294. Spyros T. (2003) Use of dual compartment mixing container for enzyme mixtures useful to treat acne. Patent: US6627192.

295. Suganthi C., Mageswari A., Karthikeyan S., Anbalagan M., Sivakumar A. and Gothandam K.M. (2013) Screening and optimization of protease production from a halotolerant *Bacillus licheniformis* isolated from saltern sediments. *Journal of Genetic Engineering and Biotechnology*. **11**: 47–52.

296. Suh H.J. and Lee H.K. (2001) Characterization of a keratinolytic serine protease from *Bacillus subtilis* KS-1. *Journal of Protein Chemistry*.**20**: 165-169.

297. Sumana D., Sudarshan M., Thakur A.R. and Ray Chaudhuri S. (2013) Degumming of raw silk fabric with help of marine extracellular protease. *American Journal of Biochemistry and Biotechnology*. **9(1)**: 12–18.

298. Sumantha A., Larroche C. and Pandey A. (2006) Microbiology and industrial biotechnology of food-grade proteases: A perspective. *Food Technology and Biotechnology*.**44(2)**: 211–220.

299. Sundus H., Mukhtar H. and Nawaz A. (2016) Industrial applications and production sources of serine alkaline proteases: A Review. *Journal of Bacteriology and Mycology*.**3(1)**: 00051.

300. Suntornsuk W., Tongjun J., Onnim P., Oyama H., Ratanakanokchai K., Kusamran T. and Oda K. (2005) Purification and characterization of keratinase from a thermotolerant feather-degrading bacterium. *World Journal of Microbiology and Biotechnology*. **21**: 1111-1117.

301. Suzuki Y., Tsujimoto Y., Matsui H. and Watanabe K. (2006) Decomposition of extremely hard-to-degrade animal proteins by thermophilic bacteria. *Journal of Bioscience and bioengineering*. **102**: 73-81.

302. Syed D.G., Lee J.C., Li W.J., Kim C.J. and Agasar D. (2009) Production, characterization and application of keratinase from *Streptomyces gulbargensis*. *Bioresource Technology*. **100**: 1868-1871.

303. Szabo I., Benedek A., Szabo I.M. and Barabas G. (2000) Feather degradation with a thermotolerant *Streptomyces graminofaciens* strain. *World Journal of Microbiology and Biotechnology*. **16**: 253-255.

304. Takami H., Nakamura S., Aono R. and Horikoshi K. (1992) Degradation of Human Hair by a Thermostable Alkaline Protease from Alkaliphilic *Bacillus* sp. No. AH-101.*Bioscience, Biotechnology and Biochemistry***.56 (10)**: 1667–1669.

305. Tapia D.M.T. and Contiero J. (2008) Production and partial characterization of keratinase produced by a microorganism isolated from poultry processing plant wastewater. *African Journal of Biotechnology*. **7(3)**: 296–300.

306. Tari C., Genckal H. and Tokatli F.(2006) Optimization of a growth medium using a statistical approach for the production of an alkaline protease from a newly isolated *Bacillus* sp. L21. *Process Biochemistry*. **41**: 659-665.

307. Tatusov R.L., Koonin E.V. and Lipman D.J. (1997) A genomic perspective on protein families.*Science* **278**: 631–637.

308. Tatusova T**.,** DiCuccio M**.,** Badretdin A**.,** Chetvernin V**.,** Nawrocki EP**.,** Zaslavsky L**.,** Lomsadze A**.,** Pruitt KD**.,** Borodovsky M**. and** Ostell J**. (2016)** NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research.***44(14)**: 6614-6624.

309. Thys R.C.S., Lucas F.S., Riffel A., Heeb P. and Brandelli A. (2004) Characterization of a protease of a feather-degrading *Microbacterium* species. *Letters in Applied Microbiology*. **39(2)**: 181–186.

310. Tiwary E. and Gupta R. (2010) Medium optimization for a novel 58 kDa dimeric keratinase from *Bacillus licheniformis* ER-15: biochemical characterization and application in feather degradation and dehairing of hides. *Bioresource Technology*. **101**: 6103–6110.

311. Tomar R., Kumar R. and Jagannadham M.V. (2008) A stable serine protease, wrightin, from the latex of the plant *Wrightia tinctoria* (Roxb.) R. Br.: purification and biochemical properties. *Journal of Agricultural Food and Chemistry*. **56**:1479-1487.

312. Tork S., Aly M.M. and Nawar L. (2010) Biochemical and molecular characterization of a new local keratinase producing *Pseudomomanas*sp., MS21. *Asian Journal of Biotechnology*. **2(1)**: 1–13.

313. Ulfig K. (2003) Studies of keratinolytic and keratinophilic fungi in sewage sludge by means of a multi-temperature hair baiting method. *Polish Journal of Environmental Studies*. **12**: 461–466.

314. Van Dijk E.L**.,** Auger H**.,** Jaszczyszyn Y**. and** Thermes C**. (2014)** Ten years of next-generation sequencing technology. *Trends in Genetics*. **30(9)**: 418-426.

315. Van Domselaar G.H., Stothard P., Shrivastava S., Cruz J.A., Guo A., Dong X., Lu P., Szafron D., Greiner R. and Wishart D.S. (2005) BASys: a web server for automated bacterial genome annotation. *Nucleic Acids Research*. 33: 455–459.

316. Van Putten A.B., Spitzenberger F., Kretzmer G., Hitzmann B., Dors M., Simutis R. and Schugerl K. (1996) Improvement of the production of subtilisin Carlsberg alkaline protease by *Bacillus licheniformis* by on-line monitoring and control in a stirred tank reactor. *Journal of Biotechnology*. **49**: 83–93.

317. Varela H., Ferrari M.D., Belobradjic L., Weyrauch R. and Loperena M.L. (1996) Effect of medium composition on the production by a new *Bacillus subtilis* isolate of protease with promising unhairing activity. *World Journal of Microbiology and Biotechnology*. **12**: 643–645.

318. Vasileva-Tonkova E., Gousterova A. and Neshev G. (2009) Ecologically safe method for improved feather wastes biodegradation. *International Biodeterioration and Biodegradation*. **63**: 1008-1012.

319. Venil C.K**. and** Lakshmanaperumalsamy P**. (2009)** Application of response surface methodology in medium optimization for protease production by the new strain of *Serratia marcescens* SB08. *Polish Journal of Microbiology*. **58(2)**: 117-124.

320. Verma A**.,** Singh H**.,** Anwar S**.,** Chattopadhyay A**.,** Tiwari K.K**.,** Kaur S**. and** Dhilon G.S**. (2016)** Microbial keratinases: industrial enzymes with waste management potential. *Critical Reviews in Biotechnology*. **13**: 1-16.

321. Vermelho A.B., Meirelles M.N.L., Lopes A., Petinate S.D.G., Chaia A.A. and Branquinha M.H. (1996) Detection of extracellular proteases from microorganisms on agar plates. *Memórias do Instituto Oswaldo Cruz*.**91**: 755–760.

322. Vidal L., Christen P. and Coello M.N. (2000) Feather degradation by *Kocuria rosea* in submerged culture. *World Journal of Microbiology and Biotechnology*. **16(6)**:551–554.

323. Villa A.L.V., Aragao M.R.S., Santos E.P., Mazotto A.M., Zingali R.B., Souza E.P. and Vermelho A.B. (2013) Feather keratin hydrolysates obtained from microbial keratinases: effect on hair fiber. *BMC Biotechnology*. **13**:15.

324. Voelkerding K. V., Dames S. A. and Durtschi1 J. D. (2009) Next-generation sequencing: from basic research to diagnostics. *Clinical Chemistry*. **55(4)**: 641–658.

325.Wang, B., Yang, W., Mc Kittrick, J. and Meyers, MA. (2016) Keratin: Structure, mechanical properties,occurrence in biological organisms, and efforts at bioinspiration. *Progress in Materials Science*. **76**: 229–318.

326. Wang H.Y., Liu D.M., Liu Y., Cheng CF., Ma Q.Y., Huang Q. and Zhang Y.Z. (2007) Screening and mutagenesis of a novel *Bacillus pumilus* strain producing alkaline protease for dehairing. *Letters in Applied Microbiology*. **44**: 1–6.

327.WangJ., HaoS., LuoT., YangQ.and Wang B. (2016b) Development of feather keratin nanoparticles and investigation of their hemostatic efficacy. *Materials Science and Engineering C*.**68**: 768–773.

328. Wang J-J., Borwornpinyo R., Odetallah N. and Shih J. C. H. (2005) Enzymatic degradation of a prion-like protein, Sup35NM-His6. *Enzyme and Microbial Technology*. **36**: 758-765.

329. Wang J.J., Rojanatavorn K. and Shih J.C.H. (2004). Increased production of *Bacillus* keratinase by chromosomal integration of multiple copies of the kerA gene. *Biotechnology and Bioengineering*. **87**: 459–464.

330. Wang L., Cheng G., Ren Y., Dai Z**.,** Zhao Z.S**.,** Liu F**.,** Li S**.,** Wei Y**.,** Xiong J**.,** Tang X.F**. and** Tang B**. (2015)** Degradation of intact chicken feathers by *Thermoactinomyces* sp. CDF and characterization of its keratinolytic protease. *Applied Microbiology and Biotechnology*. **99(9)**: 3949–3959.

331. Wang S.L., Hsu W.T., Liang T.W., Yen Y.H., Wang C.L. (2008) Purification and characterization of three novel keratinolytic metalloproteases produced by *Chryseobacterium indologenes* TKU014 in a shrimp shell powder medium. *Bioresource Technology*. **99**:5679–5686.

332. Wang X and Parsons C.M. (1997) Effect of processing systems on protein quality of feathermeals and hog hair meals. *Poultry Science*. **76**: 491-496.

333. Ward O.P. (2011) Proteases. In: Moo-Young, M. (Ed.), Comprehensive Biotechnology, 2nd ed, Waterloo, Canada, pp. 571-582.

334. Watanabe K. (2004) Collagenolytic proteases from bacteria. *Applied Microbiology and Biotechnology*. **63**: 520-526.

335. Wetterstrand, K.A. (2011)02/04-last update, DNA sequencing costs: Data from the NHGRI large-scale genome sequencing program [Homepage of National Human Genome Research Institute, NIH], [Online]. Available: http://www.genome.gov/sequencingcosts [2011, 07/26].

336. Wheeler D.A., Srinivasan M., Egholm M., Shen Y., Chen L., McGuire A., He W., ChenY.-J., MakhijaniV., RothG. T., GomesX., TartaroK., NiaziF., TurcotteC. L., IrzykG. P., LupskiJ. R., ChinaultC., SongX-z., LiuY., YuanY., NazarethL., QinX., MuznyD. M., MarguliesM., WeinstockG. M., GibbsR. A. and Rothberg J. M. (2008) The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452:**872– 876.

337. Wikstrom M.B. (1983) Detection of microbial proteolytic activity by a cultivation plate assay in which different proteins adsorbed to a hydrophobic surface are used as substrates. *Applied and Environmental Microbiology*. **45**: 393–400.

338. Wikstrom M.B., Elwing H. and Linde A. (1981) Determination of proteolytic activity: a sensitive and simple assay utilizing substrates adsorbed to a plastic surface and radial diffusion in gel. *Analytical Biochemistry*. **118**: 240–246.

339. Yamada S., Wirtz D. and Coulombe P.A. (2002) Pairwise assembly determines the intrinsic potential for self-organization and mechanical properties of keratin filaments. *Molecular Biology of the Cell*. **13**:382–391.

340. Yamamura S., Morita Y., Hasan Q., Yokoyama K. and Tamiya E (2002) Keratin degradation: a cooperative action of two enzymes from *Stenotrophomonas* sp. *Biochemical and Biophysical Research Communication*. **294**: 1138–1143.

341. Yang Y. (2012) Skin-whitening and freckle-dispelling essence and preparation method thereof. Patent: Cn102612104.

342. YangY., Zhang **S.,** Howe**K.,** Wilson **D. B.,** Moser **F.,** Irwin **D. and** Thannhauser **T. W. (2007)** A comparison of nLC-ESI-MS/MS and nLC-MALDI-MS/MS for GeLC-based protein identification and iTRAQ-based shotgun quantitative proteomics.*Journal of Biomolecular Techniques*.**18(4)**: 226-237.

343. Yoshioka M., Miwa T., Horii H., Takata M., Yokoyama T., Nishizawa K., Watanabe M., Shinagawa M. and Murayama Y. (2007) Characterization of a proteolytic enzyme derived from a *Bacillus* strain that effectively degrades prion protein. *Journal of Applied Microbiology*. **102(2)**: 509–515.

344. Yue X.Y., Zhang B., Jiang D.D., Liu Y.J. and Niu T.G. (2011) Separation and purification of a keratinase as pesticide against root-knot nematodes. *World Journal of Microbiology and Biotechnology*. **27**: 2147-2153.

345. Zaghloul T.I**. (1998)** Cloned *Bacillus subtilis* alkaline protease (aprA) gene showing high level of keratinolytic activity.*Applied Biochemistry and Biotechnology*.**72**: 199-205.

346. Zambare V., Nilegaonkar S. and Kanekar P. (2011) A novel extracellular protease from *Pseudomonas aeruginosa* MCM B-327: enzyme production and its partial characterization. *New Biotechnology*. **28**: 173-181.

347. Zarai Jaouadi N, Rekik H, Ben Elhoul M, Zohra Rahem F**.,** Hila C.G**.,** Ben Aicha HS**.,** Badis A**.,** Toumi A**.,** Bejar S**. and** Jaouadi B**. (2015)** A novel keratinase from *Bacillus tequilensis* strain Q7 with promising potential for the leather bating process. *International Journal of Biological Macromolecules*. **79**: 952–964.

348. Zhang B., Jiang D.D., Zhou W.W., Hao H.K. and Niu T.G. (2009) Isolation and characterization of a new *Bacillus* sp. 50-3 with highly alkaline keratinase activity from *Calotes versicolor* faeces. *World Journal of Microbiology and Biotechnology*. **25**: 583–590.

349. Zhang J., **ChiodiniR.**, **BadrA.andZhang G. (2011)** The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics*. **38(3)**: 95–109.

350. Zhou X**.,** Ren L**.,** Meng Q**.,** Li Y**.,** Yu Y**. and** Yu J**. (2010a)** The next-generation sequencing technology and application. *Protein Cell*. **1(6)**:520-536.

351. Zhou X. G., Ren L. F., Li Y. T., ZHANG M., YU Y.D. and YU J. (2010b) The next-generation sequencing technology: A technology review and future perspective. *Science China Life Science*. **53**: 44–57.

Fellahi *et al. AMB Expr (2016) 6:42*
DOI 10.1186/s13568-016-0213-0

AMB Express

ORIGINAL ARTICLE Open Access

Identification of two new keratinolytic proteases from a *Bacillus pumilus* strain using protein analysis and gene sequencing

Soltana Fellahi[1,2], Abdelwaheb Chibani[1], Elisabeth Feuk-Lagerstedt[2*] and Mohammad J. Taherzadeh[2]

Abstract

The *Bacillus* strain (CCUG 66887) has a high capacity to excrete keratinase with the ability to degrade *both* alpha- and beta keratin. In this study we aimed to show the characteristics of the keratinolytic protease and to identify its gene by using liquid chromatography–electrospray ionization tandem mass spectrometry methods (nanoHPLC–ESI–MS/MS) followed by Mascot data base search. The results showed that the enzyme in fact consists of two different keratinases, both with a molecular mass of 38 kDa. Further, DNA sequencing generated the open reading frame (ORF) of one of the genes (*Ker1*), and de novo genome sequencing identified the ORF of the second gene (*Ker2*). The two keratinase genes contain 1153 base pairs each and have a gene similarity of 67 %. In addition, the *Bacillus* strain was classified as *Bacillus pumilus* and its genes were annotated in the GeneBank at NCBI (accession: CP011109.1). Amino acid sequences alignment with known *B. pumilus* proteases indicated that the two keratinases of *B. pumilus* strain C_4 are subtilisin-like serine proteases belonging to the Protease S8 family. Taken together, these result suggest the two keratinases as promising candidates for enzymatic processing of keratinous wastes in waste refinery.

Keywords: *Bacillus pumilus*, Keratinase, α-Keratin, β-Keratin, NanoHPLC–ESI–MS/MS, DNA sequencing

Introduction

Annually, just the global feather waste from the poultry processing industry reaches 8.5 million tons. At present, the poultry feathers are dumped, buried, used for land filling, or incinerated, resulting in environmental challenges in terms of storage, handling, emission control, and ash disposal (Agrahari and Wadhwa 2010). Poultry feathers are also turned into feather meal used as animal feed because of the high protein content. However, the use of waste for animal feed is becoming tighter (Commission of the European Communities 2000). Additionally, the high treatment costs make the process economically unfeasible. An environmentally and economically promising process to recover the feather waste is to produce renewable energy by e.g. anaerobic digestion. In this process, not only does the valuable methane result as a byproduct, but also digested residues are formed. The latter can safely be used as a fertilizer, since pathogens presented in the feather waste have been eradicated in the process (Salminen and Rintala 2002a, b).

The recalcitrant keratin is the major compound in several biological materials. It is also the waste product in poultry, slaughterhouse, leather- and fur processing industries and consists of feather, hair, horn, hoof, nails, claws, wool, and bristles (Kornillowicz-Kowalska and Bohacs 2011). While some of the materials like hair or wool, to a great extent, are composed of the helix form of α-keratin, other materials such as feather are largely composed of the flat form of β-keratin. Among the two different types of keratin structures, the content of sulfur varies giving the keratin a softer or harder structure and affects the degradation of the keratinous material to a greater extent (Brandelli et al. 2010). A large number of microorganisms have been reported to produce keratinases (Brandelli et al. 2010; Gupta and Ramnani 2006; Onifade et al. 1998), and among bacteria, the best studied are organisms from the genus *Bacillus* (Gobinath et al.

*Correspondence: elisabeth.feuk@hb.se
[2] Swedish Centre for Resource Recovery, University of Borås, Borås, Sweden
Full list of author information is available at the end of the article

2014). Keratinases (EC 3.4.99.11) are serine- or metalloproteases (Gupta and Ramnani 2006), and many bacterial keratinases have been sequenced, cloned, and characterized indicating a sequence similarity with the subtilisin family, Family S8, of serine proteases (Rawlings and Barret 1993).

The isolation and characterization of a keratin-degrading bacterium, *Bacillus* sp C_4, has been reported by this lab. The proteolytic activity was broadly specific, and the bacterium could grow and produced a significant level of keratinase when using wool or chicken feather as substrates. A total hydrolysis of the keratinous waste was obtained in less than 3 days (Fellahi et al. 2014). Also this proteolytic enzyme has shown activity and stability over a broad pH range with two distinct optima, one at pH 8.5 and the other at pH 11, indicating that it might be not one but two enzymes. Its activity was completely inhibited by phenylmethylsulfonyl fluoride (PMSF) pointing out that the enzyme is a serine protease (Fellahi 2009).

In an attempt to increase our understanding of the *Bacillus* strain's ability to simultaneously hydrolyze *both* α- and β-keratin, we in this study aimed to show the characteristics of the keratinolytic protease and to identify its gene. So far, the vast majority of the identified keratinase-producing organisms appears to be able to hydrolyze only the β-keratin in the chicken feather (Gupta et al. 2013), which gives the keratinolytic protease from this strain a potential for simultaneous degradation of both types of keratin in waste refinery.

Materials and methods

Bacterial strain and medium

The microorganism used in this study was *Bacillus* sp. C_4; CCUG 66887. It has earlier been isolated from the compost and identified using biochemical tests and 16S rDNA technique (GenBank accession: FJ214667) (Fellahi et al. 2014). Before the strain was used for protease production it was grown at 37 °C for 24 h on peptone yeast extract medium containing Bactopeptone, 10 g/l; Yeast extract, 5 g/l and NaCl, 5 g/l.

Keratinase gene sequence determination

Multiple sequence alignment with CLUSTALW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) (Chenna et al. 2003) was used to align keratinase genes from different *Bacillus pumilus* strains to investigate the resemblance among the strains to be able to choose one strain for designing the first sequencing primer set (F: TTAGAA GCCGCTTGAACGTTA, R: ATGTGCGTGAAAAA GAAAAATGTG). Genomic DNA was isolated from strain C_4 using MasterPure™ Gram Positive DNA Purification Kit (Epicentre), and the DNA was sent together with the sequence for the first primer set to Eurofins Genomics, Germany where the primers were synthesized and both DNA strands sequenced by Sanger method. From the retrieved two DNA sequences, a new primer set (F: AAGTATTAGATCGTTACGGCGATG GAC, R: CCAAGAACACCAATCGTGTTATCAAGG) was designed and once again sent to Eurofins Genomics together with genomic DNA. This procedure was repeated a third time with primer (F: TTGCCAACGT GAACAGCAAC) to determine the open reading frame (ORF) of the gene.

De novo sequencing and genome annotation

To be able to search the genome of strain C_4 for additional putative keratinase genes, de novo sequencing of the whole genome using the instrument MiSeq and the MiSeq Control Software 2.3.0.3 was performed by Eurofins Genomics, Germany. The sequence assembly and scaffolding was done using the Newbler assembler software v2.9. The genome sequence was annotated using the prokaryotic annotation pipeline at the National Center for Biotechnology Information, Bethesda, USA (NCBI).

Partial purification of proteases for nanoHPLC–ESI–MS/MS

The protease production from *Bacillus* C_4 strain was done according to Fellahi and coworkers. (Fellahi et al. 2014). In short: the C_4 strain was grown in 50 ml of modified Schaeffer's medium (Leighton and Doi 1971) containing Beef extract, 3 g/l; Bactopeptone, 5 g/l; KCl, 2 g/l; Yeast extract, 2 g/l; pH 7, and supplemented with 2 mM $MgSO_4.7H_2O$; 1 mM $CaCl_2$; 0.1 mM MnCl2; 1 mM $FeSO_4$; and 0.1 % (w/v) glucose. The production was done using a 250-ml E-flask with a 2 % inoculum size in a shaker incubator (Excella 24, New Brunswick Scientific) at 37 °C, 160 rpm. After 24 h, the cell-free supernatant was received by centrifugation at 8000 rpm for 15 min at 4 °C (Optima Max-XP, Beckman Coulter).

For the identification of the enzyme by nanoHPLC–ESI–MS/MS the cell-free supernatant was precipitated by adding NH_3SO_4 to 65 % saturation at 4 °C and slowly mixing in a shaker incubator (Excella 24, New Brunswick Scientific) for 1 h. The precipitate was collected by centrifugation at 12,000 rpm for 30 min at 4 °C (Optima Max-XP, Beckman Coulter). The pellet was re-suspended in 500 µl of 20 mM Tris–HCl buffer, pH 8, followed by dialysis overnight against the same buffer. Proteolytic activity was measured as described by Cliffe and Law (1982), using Hide Powder Azure (HPA, Sigma) as substrate. Approximately 20 µg of the purified enzyme was run on a 12 % SDS-PAGE-gel, according to Laemmli (1970) along with a molecular weight protein marker (All Blue Protein Precision Standard, Bio-Rad).

Protein identification by nanoHPLC–ESI–MS/MS and data base search

Protein identification of the crude enzyme was performed by Proteome Factory AG, Germany. Two protein spots from a 12 % SDS-PAGE-gel were cut out and digested in-gel by trypsin (Promega, Mannheim, Germany) and analyzed by nanoHPLC–ESI–MS/MS. The LCMS system consisted of an Agilent 1100 nanoHPLC system (Agilent, Waldbronn, Germany), PicoTip electrospray emitter (New Objective, Woburn, MA, USA), and an Orbitrap XL or LTQFT. The retrieved peptides were analyzed using an ultra-mass spectrometer (ThermoFisher Scientific, Bremen, Germany). Peptides were first trapped and desalted on the enrichment column Zorbax 300SB-C18, 0.3 mm × 5 mm (Agilent) for 5 min (solvent: 2.5 % acetonitrile/0.5 % formic acid), then separated on a Zorbax 300SB-C18, 75 µm × 150 mm column (Agilent) using a linear gradient from 10 to 32 % B (Solvent A: 5 % acetonitrile in water, Solvent B: acetonitrile. Both solvents contained 0.1 % formic acid). Ions of interest were data-dependently subjected to MS/MS according to the expected charge state distribution of the peptide ions. MS/MS ion search of the Mascot search engine (Matrix Science, London, England) was performed, and only peptide matches with a score of 20 or above were accepted. Proteins were identified against the *B. pumilus* entries from the RefSeq protein database available at NCBI, which was appended to an existing bacterial database. The search results were also run against the amino acid sequences retrieved from the *Ker1* and *Ker2* genes.

Amino acid sequence alignment

The amino acid sequences of the retrieved enzymes from the ultra-mass spectrometer analysis were compared to other proteases produced by the *B. pumilus* strains using the Basic Local Alignment Search Tool Blastp 2.2.3.1 available at NCBI.

Results

Keratinase gene sequence determination

The multiple sequence alignment followed by synthesis of primers and Sanger sequencing resulted in an ORF of a keratinase gene, named *Ker1*, which contains 1153 base pairs. The gene was subsequently submitted to the NCBI GenBank (Accession Number: KX184831).

De novo sequencing and genome annotation

The de novo sequencing of the whole genome indicated that the organism is a *B. pumilus* and that it possesses a genome of 3.6 million base pairs. The organism was annotated to NCBI (Accession Number: CP011109.1) using the prokaryotic annotation pipeline, and the result indicated that this *B. pumilus* strain has around 4000 genes. One of these genes is a homolog to *Ker1* and was named *Ker2.* The gene was subsequently submitted to NCBI GenBank (Accession Number: KX184832).

Partial purification of proteases from strain C_4 for nano-HPLC-ESI-MS/MS

The crude protease fraction showed after SDS-PAGE and staining with Coomassie brilliant blue R-250 one distinct band and one weaker band with molecular weights of about 28 and 36 kDa, respectively (Fig. 1).

Protein identification by nanoHPLC–ESI–MS/MS and data base search

The Mascot Search Result using the *B. pumilus* entries from the RefSeq protein data base and the amino sequences retrieved from the *Ker1* and *Ker2* genes can be seen in Table 1. The band corresponding to 28 kDa only contained keratinase 1, while the band corresponding to 38 kDa contained keratinase 1 and keratinase 2. These results indicated that the partial purified keratinolytic protease in fact consists of two enzymes, both with a molecular weight of 38 kDa, and corresponding to the identified genes *Ker1* and *Ker2* of the *B. pumilus* strain

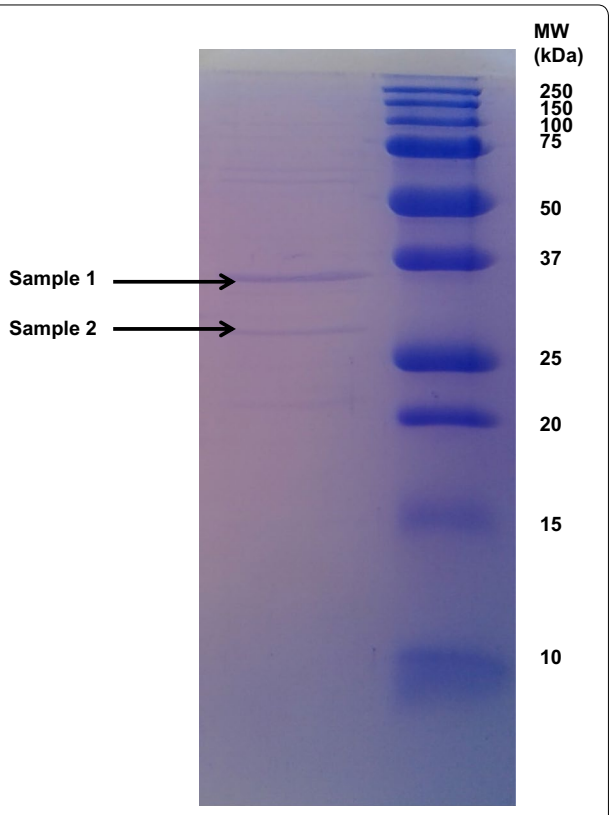


Fig. 1 The dialyzed proteolytic enzyme investigated on SDS-PAGE using a 12 % gel. Two bands at 36 and 28 kDa, respectively, were excised for protein identification

Table 1 Protein identification by MASCOT Search on the purified enzyme from *Bacillus* strain C_4 using RefSeq protein data base and the two amino acid sequences retrieved from *Ker1* and *Ker2*

SDS-Page lane	Identified protein	Sequence identity	MW (kDa)	MASCOT peptides identified	MASCOT ion score[a]
1 and 2	Keratinase 1	WP_008348814.1	38.8	QRLENTATPLGNSFYYGK	81
				GVVVVAAAGNSGSTGSTSTVGYPAK	146
				YDSTIAVANVNSNNVR	122
				LENTATPLGNSFYYGK	102
				GLINVQAASN	85
1	Keratinase 2	WP_017357922.1	38.3	VGVVGVAPK	61
				VADENGDGYYSWIIK	80
				SGTSMASPHVAGAAAVILSK	107
				HPNLTNDELR	53
				HPNLTNDELRER	34
1	Peptidase S8	gi|648268958	38.8	VGVVGVAPK	61
				VADENGDGYYSWIIK	80
				LGEPFYYGAGLVNVQK	106
				SGTSMASPHVAGAAAVILSK	107

[a] Individual ion scores >26 indicates identity or extensive homology at 95 % level of confidence

C_4. When using the *B. pumilus* entries from the RefSeq protein data the peptides from MS/MS identified Peptidase S8. The result also showed a resemblance between keratinase 2 and peptidase S8.

Figure 2 illustrates the peptides from Ms/Ms identified in keratinase 1, keratinase 2 as well as peptidase S8 and their distribution and coverage in the respective enzymes. As can be seen keratinase 2 shares three peptides with peptidase S8 but also has unique one while keratinase 1 does not have any peptides in common with the identified Peptidase S8.

Amino acid sequence alignment

By comparing the amino acid sequences of the two retrieved keratinase enzymes with known proteases of other *B. pumilus* strains the results indicated that these enzymes are subtilisin-like serine proteases belonging to the Protease S8 family (Tables 2, 3).

Discussion

The fast growth rate of the microorganisms, the accessibility for genetic engineering and the short time for the production and purification steps make them the ideal source for production of proteases (Rao et al. 1998). The far most popular source of commercial alkaline proteases is from the *Bacillus* species. The main reason for this is their ability to produce large amounts of alkaline proteases having significant proteolytic activity and stability at high pH as well as high temperature (Jacobs 1995; Yang et al. 2000). So far, the vast majority of the identified keratinase producing organisms appear to be able to hydrolyze only β-keratin in chicken feather and few are known to hydrolyze both α- and β-keratin. (Gupta et al. 2013).

In this study two keratinases from the *Bacillus* sp C_4 strain were identified. The project started by trying to find a keratinase gene in the genome of the strain by attaching primers designed for the *Ker A* gene of *Bacillus licheniformis* PWD-1 identified by Lin and coworkers (1995). After conventional PCR followed by agarose gel electrophoresis we found many different gene products due to unspecific binding of the primers but no gene product comparable in size with a keratinase gen. Still we decided to DNA sequence three fragments by the Sanger method which resulted in three DNA sequences with the lengths of 236–396 base pairs. By using NCBI blast network service (http://www.blast.ncbi.nlm.nih.gov/Blast.cgi), we found that these gene sequences might belong to the genome of a *B. pumilus* strain. With this information at hand we aligned genes from twelve different *B. pumilus* strains (Table 4) using CLUSTAL W2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/). The outcome of the process indicated high resemblance between many sequences, mostly in the beginning and the end of the different aligned genes. One of the strains, *B. pumilus* strain A1, produces a keratinase (Fakhfakh-Zouari et al. 2010) which has comparable qualities with the keratinolytic enzyme produced by *Bacillus* sp C_4 (Fellahi 2009). Both enzymes are active and stabile over a broad pH range and also their activity is completely inhibited by PMSF suggesting that they both are serine proteases. The keratinase precursor gene from the *B. pumilus* strain A1 was for this reason chosen for primer design of the first pair of sequencing primers for the keratinase

Keratinase 1

```
  1 MCVKKKNVMT SVLLAVPLLF SAGFGGSIAN AETASKSESE KSYIVGFKAS
 51 ATTNSSKKQA VTQNGGKLEK QYRLINAAQV KMSEQAAKKL EHDPSIAYVE
101 EDHKAEAYAQ TVPYGIPQIK APAVHAQGYK GANVKVAVLD TGIHAAHPDL
151 NVAGGASFVP SEPNATQDFQ SHGTHVAGTI AALDNTIGVL GVAPSASLYA
201 VKVLDRYGDG QYSWIISGIE WAVANNMDVI NMSLGGPNGS TALKNAVDTA
251 NNRGVVVVAA AGNSGSTGST STVGYPAKYD STIAVANVNS NNVRNSSSSA
301 GPELDVSAPG TSILSTVPSS GYTSYTGTSM ASPHVAGAAA LILSKYPNLS
351 TSQVRQRLEN TATPLGNSFY YGKGLINVQA ASN
```

Keratinase 2

```
  1 MKGRVGIVKV KSFGAGLCMA SMLLASMTFG ASDVSAKDQA KKEYMIGFSS
 51 SVQDKTQKQL VEKAGGHVKE SIEKIDMMKV SLNEASKEKL HQAKEVTFIE
101 EDQKAKTSGQ SVPYGIKSIK AQKVHKRGYA GQNVKVAVLD SGIDGKHEDL
151 HVTGGVSFVP TESDPLVDPH EHGTHVAGTI AALDNKVGVV GVAPKASIYA
201 VKVADENGDG YYSWIIKGIE WAIENEMDVI NISMGGASES EALKEAVDRA
251 YDNGILIVAS AGNAGSYGSL NTIDYPAKYS SVMAVASVDQ RKQRAFDSSV
301 GEEVEVSAPG VSTLSTIPHN EYGYKSGTSM ASPHVAGAAA VILSKHPNLT
351 NDELRERLTK
```

Peptidase S8

```
  1 MASMLLASMT FGASDVSAKD QAKKEYMIGF SSSVQDKTQK QLVEKAGGHV
 51 KESIEKIDMM KVSLNEASKE KLHQAKEVTF IEEDQKAKTS GQSVPYGIKS
101 IKAQKVHKRG YTGQNVKVAV LDSGIDGKHE DLHVTGGVSF VPTESDPLVD
151 PHEHGTHVAG TIAALDNKVG VVGVAPKASI YAVKVADENG DGYYSWIIKG
201 IEWAIENEMD VINISMGGAS ESEALKEAVD RAYDNGILIV ASAGNAGSYG
251 SLNTIDYPAK YSSVMAVASV DQRKQRAFDS SVGEEVEVSA PGVSTLSTIP
301 HNEYGYKSGT SMASPHVAGA AAVILSKHPN LTNDEVRERL TKTATKLGEP
351 FYYGAGLVNV QKAAR
```

Fig. 2 The distribution of peptides from MS/MS in keratinase 1, keratinase 2 and peptidase S8. Matched peptides are shown in *bold black*

gene. After Sanger sequencing the complete ORF of the keratinase gene (*Ker1*) was identified. To find resemblance with other genes from *B. pumilus* strains, nucleotide BLAST was used. The gene showed a high similarity to several serine protease genes from different *B. pumilus* strains. As much as 99 % similarity was found with the peptidase S8 gene of the *B. pumilus* strain W3, GenBank accession: CP011150.1 (Zheng-Bing et al. 2015). This result confirmed that the isolated fragment of the genome encodes one of the serine protease from the *Bacillus* sp. C_4 strain. By de novo sequencing and annotation of the genome, we could identify a homolog to the peptidase S8 gene, the *Ker2* gene with equal amount of base pairs as *Ker1.* The two genes had a gene sequence similarity of 67 %.

The crude protease showed after partial purification followed by SDS-PAGE-gel and staining with Coomassie brilliant blue R-250 one distinct band and one weaker band. Their molecular weights were estimated to about 28 and 36 kDa, respectively, when comparing with the molecular weight marker. When the bands on the gel were analyzed by a nanoHPLC–ESI–MS/MS system the result indicated two proteins with a molecular weight of 38. 8 and 38.3 kDa, respectively (Table 1). The molecular size difference between the excised gel bands and the molecular size of the identified enzymes may be due to a degradation prior to the SDS-PAGE. When the proteins were identified by Mascot Search the result indicated that the two proteins in fact are corresponding to the two genes *Ker1* and *Ker2.* That the crude enzyme in fact contains two different keratinases is in agreement with earlier conclusions (Yamamura et al. 2002). Yamamura and his group found that for the bacterium *Stenotrophomonas* sp D-1 it was not sufficient with one protease for effective keratin degradation but two different keratinases were needed. This suggests that degradation requires the cooperative action of multiple enzymes. We are well aware that *Bacillus* sp C_4 may need more than two keratinases for optimal degradation although we have not found any additional.

Table 2 Amino acid sequence alignment of keratinase 1 with proteases of *Bacillus* strains using Blastp 2.2.3.1

Description	Organism	Amino acid sequence identity (%)	Accession number
Alkaline serine proteinase	*B. pumilus*	100	ACO94164.1
Lehensis serine protease	*B. lehensis*	100	AFP23380.1
Alkaline serine protease	*B. pumilus*	99	BAE79641.1
Dehairing protease precursor	*B. pumilus*	99	AAR19220.1
Peptidase S8	*B. pumilus*	99	WP_026050071.1
Alkaline serin proteinase	*B. pumilus*	99	BAA93474.1
Protease	*B. pumilus*	99	ADK63096.1
Peptidase S8	MULTISPECIES: *Bacillus*	99	WP_008348814.1
Serine alkaline protease (subtilisin)	*B. stratosphericus LAMA585*	99	EMI14709.1
Subtilisin Carlsberg	*Bacillus pumilus* ATCC 7061	99	EDW22774.1
Peptidase S8	*B. pumilus*	99	WP_034620013.1
Peptidase S8	*B. pumilus*	99	WP_041093123.1
Peptidase S8	*B. stratosphericus*	99	WP_039962807.1
Peptidase S8	*B. invictae*	99	WP_045034875.1
Peptidase S8	*B. pumilus*	99	WP_044140726.1
Peptidase S8	*B. aerophilus*	99	WP_041507592.1
Peptidase S8	*B. altudinis*	99	WP_039167642.1
Peptidase S8	*B. pumilus*	98	WP_012009474.1
Serine alkaline protease	*B. circulans*	98	ADN04910.1
Serine alkaline protease (subtilisin E)	*B. pumilus*	98	KIL22204.1
Serine alkaline protease, preproprotein	*B. pumilus*	98	CAO03040.1
Peptidase S8	*B.* sp. DW5-4	98	WP_034323660.1
Peptidase S8	*B. pumilus*	98	WP_041117216.1
Peptidase S8	*B. safensis*	98	WP_034282323.1
Peptidase S8	*B. pumilus*	98	WP_034663897.1
Serine alkaline protease (subtilisin E)	*B. pumilus*	97	KIL10386.1
Peptidase S8	*B. pumilus*	97	WP_041110188.1
Peptidase S8	*B.* sp. WP8	96	WP_039183048.1
Keratinase precursor	*B. pumilus*	96	ACM47735.1
Serine alkaline protease (subtilisin E)	*Bacillus* sp. HYC-10	96	KIL09959.1
Peptidase S8	*B. xiamenensis*	96	WP_008359041.1
Organic solvent tolerant protease	*B. pumilus*	96	AAU88064.1
Keratinase	*B. pumilus*	96	ADK11996.1
Peptidase S8	*B. safensis*	96	WP_029706931.1
Peptidase S8	*B. pumilus*	96	WP_041089929.1
Peptidase S8	*B. safensis*	96	WP_044335827.1
MULTISPECIES: peptidase S8	*Bacillus*	96	WP_025093353.1
Peptidase S8	*B. pumilus*	95	WP_024426548.1
Serine alkaline keratinase	*Brevibacillus brevis*	95	AGO58466.1
Alkaline serine protease precursor	*B. pumilus*	95	ACM07731.1
Serine alkaline keratinase	*B. circulans*	94	AGN91700.1

When comparing the amino acid sequences of the two retrieved keratinases with known proteases of other *B. pumilus* strains using Blastp we found that keratinase 1 and 2 are subtilisin-like serine proteases belonging to the Protease S8 family (Tables 2, 3). These proteases show broad substrate specificity, have usually a molecular mass in the range 18–90 kDa. They are generally active at neutral and alkaline pH, with optima at pH 7–11 (Rao et al. 1998) and irreversibly inhibited by PMSF (Powers et al. 2002). This is in agreement with our earlier findings (Fellahi 2009).

As a conclusion, the present study confirmed that the keratinolytic protease produced by the non-genetically

Table 3 Amino acid sequence alignment of keratinase 2 with proteases of *Bacillus* strains using Blastp 2.2.3.1

Description	Organism	Amino acid sequence identity (%)	Accession number
Serine alkaline protease (subtilisin E)	*B. stratosphericus* LAMA 585	100	EMI12150.1
Peptidase S8	MULTISPECIES: *Bacillus*	100	WP_035390997.1
Subtilisin Carlsberg	*B. sp.* M 2-6	99	EIL84986.1
Peptidase S8	MULTISPECIES: *Bacillus*	99	WP_034647494.1
Peptidase S8	*B. pumilus*	99	WP_029575389.1
Subtilisin Carlsberg	*B. pumilus*	99	KIL26870.1
Peptidase S8	*B. altitudinis*	99	WP_035702958.1
Peptidase S8	*B. pumilus*	99	WP_026050107.1
Pubtilisin	*B. altitudinis* 41KF2b	99	KDE30915.1
Peptidase S8	*B. sp.* DW5-4	95	WP_034319985.1
Peptidase S8	*B. pumilus*	95	WP_044141213.1
Peptidase S8	*B. pumilus*	94	WP_034620505.1
Peptidase S8	*B. pumilus*	94	WP_034661158.1
Subtilisin Carlsberg	*B. pumilus*	94	KIL17080.1
Subtilisin Carlsberg	*B. pumilus* ATCC7061	94	EDW21217.1
Peptidase S8	*B. pumilus*	93	KDE52880.1
Peptidase S8	*B. safensis*	93	WP_034622393.1
Peptidase S8	MULTISPECIES: *Bacillus*	93	WP_029708034.1
Peptidase S8	*B. pumilus*	93	WP_041117873.1
Peptidase S8	*B. safensis*	93	KEP30825.1
Peptidase S8	*B. safensis*	93	WP_029706051.1
Peptidase S8	*B. pumilus*	93	WP_041109684.1
Pubtilisin Carlsberg	*B. pumilus*	93	KIL21504.1
Pubtilisin Carlsberg	*B. pumilus*	92	KIL11523.1
Peptidase S8	*B. sp.* HYC-10	92	WP_008361817.1
Peptidase S8	*B. safensis*	92	WP_034280781.1
Peptidase S8	*B. sp.* WP8	92	WP_039183179.1
Subtilisin	*B. safensis* FO-36b	92	KDE29455.1
Peptidase S8	*B. safensis*	91	WP_046312283.1
Peptidase S8	*B.pumilus*	91	WP_041086842.1
Subtilisin Carlsberg	*B. pumilus*	90	KIL15277.1

Table 4 *Bacillus pumilus* strains used for gene alignment in CLUSTAL W2

B. pumilus strain	GI number and accession version	Gene
NJM4	gi|226938414|gb|FJ869878.1|	Alkaline serine proteinase
N/A	gi|38373993|gb|AY458140.1|	Dehairing protease precursor
N/A	gi|7415641|dbj|AB029082.1|	Alkaline serine proteinase
SG2	gi|301131525|gb|GQ398415.1|	Protease gene
SGMM8	gi|290472378|gb|GU143024.1|	Protease (Alp) gene
bppA	gi|87886606|dbj|AB211527.1|	Alkaline serine
A1	gi|222353759|gb|FJ619651.1|	Keratinase precursor
115b	gi|52843271|gb|AY743586.1	Organic solvent protease gene
KS12	gi|300429855|gb|HM219183.1|	Keratinase gene
TMS55	gi|221193393|gb|FJ584420.1|	Alkaline serine protease precursor
3–19	gi|297342830|gb|AY754946.2|	Subtilisin like serine
sapB	gi|186928863|emb|AM748727.1|	Serine alkaline protease preprotein

modified *B. pumilus* strain C_4 consists of two different enzymes belonging to the Protease S8 family. This may explain why the strain is able to simultaneously hydrolyze both α- and β-keratin in less than three days. It also makes the bacterium a potent candidate in a cost effective pre-treatment step of keratinase rich waste in waste refinery as the two waste fractions, avian feather and sheep wool, do not have to be separated before hydrolyzation of the protein into valuable feedstuff for the biogas production.

Authors' contributions

SF has contributed with planning and implementation of the research work as well as interpretation of data and article preparation. AC has contributed with planning and design of research work as well as interpretation of data. EFL has contributed with planning and implementation of the research work as well as interpretation of data and article preparation. MJT has contributed with planning and implementation of research work as well as revising the article critically. All authors read and approved the final manuscript.

Author details

[1] Department of Biology, Faculty of Sciences of Nature and Life, Mostaganem University, Mostaganem, Algeria. [2] Swedish Centre for Resource Recovery, University of Borås, Borås, Sweden.

Acknowledgements

The authors thank Dr. Johan O. Westman for his advice relating to the molecular biology investigations.

Competing interests

The authors declare that they have no competing interests.

Fundings

This study was supported by the University of Borås and Sparbanksstiftelsen Sjuhärad in Sweden, as well as Mostaganem University in Algeria.

Received: 10 May 2016 Accepted: 21 June 2016
Published online: 30 June 2016

References

Agrahari S, Wadhwa N. Degradation of chicken feather a poultry waste product by keratinolytic bacteria isolated from dumping site at Ghazipur poultry processing plant. Int J Poult Sci. 2010;9(5):482–9.

Brandelli A, Daroit DJ, Riffel A. Biochemical features of microbial keratinases and their production and applications. Appl Microbiol Biot. 2010;85(6):1735–50.

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. 2003;31(13):3497–500.

Cliffe AJ, Law BA. A new method for the detection of microbial proteolytic enzymes in milk. J Dairy Res. 1982;49(02):209–19.

Commission of the European Communities. White paper on food safety. Off J. 2000;C076:1–4.

Fakhfakh-Zouari N, Hmidet N, Haddar A, Kanoun S, Nasri M. A novel serine metallokeratinase from a newly ilsolated *Bacillus pumilus* A1 grown on chicken feather meal: biochemical and molecular characterization. Appl Ciochem Biotechnol. 2010;162:329–44.

Fellahi S. Isolation, identification, and proteolytic activity of some protease-producing bacteria. M.Sc. Dissertation. Alexandria: Alexandria University; 2009.

Fellahi S, Zaghloul T, Feuk-Lagerstedt E, Taherzadeh M. A *Bacillus* strain able to Hydrolyze alpha- and beta-keratin. J Bioprocess Biotechnol. 2014;4:1000181–7.

Gobinath SCB, Anbu P, Lakshmipriya T, Tang T-H, Chen Y, Hashim U, Ruslinda AR, Arshad MK. Biotechnological aspects and perspective of microbial keratinase production. Biomed Res Int. 2014;2015:1–10.

Gupta R, Ramnani P. Microbial keratinases and their prospective applications: an overview. Appl Microbiol Biotechnol. 2006;70(1):21–33.

Gupta R, Sharma R, Qasim KB. Revisiting microbial keratinases: next generation proteases for sustainable biotechnology. Crit Rev Biotechnol. 2013;33(2):216–28.

Jacobs MF. Expression of the subtilis in Carlsberg-encoding gene in *Bacillus licheniformis* and *Bacillus subtilis*. Gene. 1995;152:67–74.

Kornillowicz-Kowalska T, Bohacs J. Biodegradation of keratin waste: theory and practical aspects. Waste Manag. 2011;31(8):1689–701.

Laemmli UK. Cleavage of the structural proteins during the assembly of the head of the bacteriophage T4. Nature. 1970;227:680–5.

Leighton TJ, Doi RH. The stability of messenger ribonucleic acid during sporulation in *Bacillus subtilis*. J Biol Chem. 1971;246:3189–95.

Lin X, Kelemen DW, Miller ES, Shih JC. Nucleotide sequence and expression of kerA, the gene encoding a keratinolytic protease of *Bacillus licheniformis* PWD-1. Appl Environ Microbiol. 1995;61(4):1469–74.

Onifade A, Al-Sane N, Al-Musallam A, Al-Zarban S. A review: potentials for biotechnological applications of keratin-degrading microorganisms and their enzymes for nutritional improvement of feathers and other keratins as livestock feed resources. Bioresour Technol. 1998;66(1):1–11.

Powers JC, Asgian JL, Ekici ÖD, James KE. Irreversible inhibitors of serine, cysteine, and threonine proteases. Chem Rev. 2002;102(12):4639–750.

Rao MB, Tanksale AM, Ghatge MS, Deshpande VV. Molecular and biotechnological aspects of microbial proteases. Microbiol Mol Biol Rev. 1998;62(3):597–635.

Rawlings ND, Barret AJ. Evolutionary families of peptidases. Biochem J. 1993;290:205–18.

Salminen EA, Rintala JA. Anaerobic digestion of organic solid poultry waste-a review. Bioresour Technol. 2002a;83(1):13–26.

Salminen EA, Rintala JA. Semi-continuous anaerobic digestion of solid poultry slaughterhouse waste: effect of hydraulic retention time and loading. Water Res. 2002b;36(13):3175–82.

Yamamura S, Morita Y, Hasan Q, Yokoyama K, Tamiya E. Keratin degradation: a cooperative action of two enzymes from *Stenotrophomonas* sp. Biochem Biophys Res Commun. 2002;294:1138–43.

Yang JK, Shih IL, Tzeng YM, Wang SL. Production and purification of protease from a *Bacillus subtilis* that can deproteinize crustacean wastes. Enzyme Microb Technol. 2000;26:406–13.

Zheng-Bing G, Yu-Jie C, Yan-Zhou Z, Zxiang-Ru L. Complete genome sequence of *Bacillus pumilus* W3: a strain exhibiting high laccase activity. J Biotechnol. 2015;207:8–9.