National and Kapodistrian University of Athens
School of Science, Faculty of Biology,
Sector of Botany, Group of Microbiology

Doctorate thesis

"Bacterial hydrolases of thermophilic origin and their application in plant biomass valorisation"

Anastasia P. Galanopoulou

Athens,
July 2018

Εθνικό και Καποδιστριακό Πανεπιστημίο Αθηνών,
Σχολή Θετικών Επιστημών, Τμήμα Βιολογίας,
Τομέας Βοτανικής, Ομάδα Μικροβιολογίας

Διδακτορική Διατριβή

"Βακτηριακές υδρολάσες θερμόφιλης προέλευσης και αξιοποίηση τους στην αποικοδόμηση της φυτικής βιομάζας"

Αναστασία Π. Γαλανοπούλου

Αθήνα,
Ιούλιος 2018

Advising Committee:

- Dimitris Hatzinikolaou, Associate Professor, Biology Department, National and Kapodistrian University of Athens (supervisor)
- George Diallinas, Professor, Biology Department, National and Kapodistrian University of Athens
- Constantinos Vorgias, Professor, Biology Department, National and Kapodistrian University of Athens

Defence Committee:

- Dimitris Hatzinikolaou, Associate Professor, Biology Department, National and Kapodistrian University of Athens (supervisor)
- George Diallinas, Professor, Biology Department, National and Kapodistrian University of Athens
- Constantinos Vorgias, Professor, Biology Department, National and Kapodistrian University of Athens
- Dimitris Kekos, Professor, School of Chemical Engineering, National Technical University of Athens
- Katherine - Maria Pappas, Associate Professor, Biology Department, National and Kapodistrian University of Athens
- Konstantinos Kormas, Professor, Department of Ichthyology & Aquatic Environment, University of Thessaly
- George Skretas, Research Associate Professor, National Hellenic Research Foundation

TABLE OF CONTENTS

AKNOWLEDGEMENTS

The present Thesis could not have been accomplished without the significant contribution of many people, to whom I am very grateful.

Trying to name the majority of them, I should start by expressing my sincere gratitude to my PhD advisor, Associate Professor Dimitris Hatzinikolaou, for our long time collaboration; It was he, who introduced me in the field of Microbial Biotechnology and guided me in my first experimental attempts. More importantly, he encouraged independency in my work, trusted my opinion and offered all the necessary scientific and economical support. I am more than grateful for all of these.

I would also like to acknowledge the support of the two additional members of my PhD committee; Professor George Diallinas and I, had numerous conversations regarding my work. I will always appreciate his thoughtful and very honest comments; I truly enjoyed these moments which were more than helpful for me. Professor Konstantinos Vorgias' contribution was also of great importance, as he gave me limitless access to his lab, something absolutely necessary for several parts of my PhD.

Further, a big thank, goes to Professor Ed Bayer from Weizmann Institute of Science, who hosted me in his lab for almost a year in total. Besides his warm hospitality, I have to acknowledge that working in a lab like his, gave me the perspective of the research environment in a top Institute, like Weizmann. It was an amazing experience that I will always cherish.

Additionally, I would like to express my appreciation to the three Research Associates of the Microbiology group; Dr Sotiris Amilis, Dr Stathis Katsifas and Dr Alexandros Savvides for all their organizational and scientific support. I do believe that without Sotiri's help I would have never managed to perform even the simplest *E. coli* transformation.

Of great importance was also the contribution of all my friends; the PhD students, post-docs, master and undergraduate students from the lab of Dr Hatzinikolaou, and overall the group of Microbiology, the lab of Dr Vorgias and the lab of Dr Bayer in Weizmann Institute. I would like to thank each and every one individually for all the nice moments and our discussions that I will never forget. I really hope that even our ways seem to be parting, we will still stay in touch.

I would also like to acknowledge the contribution of Professor Dimitris Kekos from the school of Chemical Engineering - NTUA, Assistant Professor Katherine Pappas from the Biology Department of UOA, Professor Konstantinos Kormas from the Department of Ichthyology & Aquatic Environment, University of Thessaly and Research Associate Professor George Skretas from the National Hellenic Institute for participating in my examining committee and their thoughtful comments on my Thesis. A special thank goes to Dr Skretas and his post doc student Dr Dimitra Zarafeta for our smooth collaboration in the framework of several projects throughout these years. Additionally, I would like to thank Dr Kormas for all his support, every time I asked him for; including his support for my application for a FEMS research fellowship and also our conversations regarding the *Geobacillus icigianus* SP50 genome project.

For the accomplishment of several parts of my PhD I had the pleasure to collaborate with several scientists from different fields. With a chronological order, I have to acknowledge the contribution of the Assistant Professor Nikolaos Psaroudakis from the Chemistry Department of UOA for his help on the synthesis of the substrates used for the characterization of the β-P-glucosidase, Research Associate Professor Evangelia Chrysina from the National Hellenic Institute for taking over the structural characterization of the same enzyme; I am waiting for these results with excitement. Further, I have to acknowledge the contribution of all the co-authors of my scientific papers and also Stavroula Delenta for providing me the pretreated substrates used for the hydrolysis reactions. '*Geobacillus icigianus* SP50' project, couldn't have been accomplished without the contribution of the PhD candidate Stelios Varassas who isolated the DNA of the strain, Dr Alexandra Meziti who spent several hours helping me with the analysis of the genome and the Research Associates Dr Zoi Litou and Dr Nikos Papandreou, both from the Biology Department of UOA, who helped me with the BLAST and HMMER searches. Additionally, the suggestions of Dr Ioannis Theologidis regarding the horizontal transfer were valuable for my analysis. Lastly, I am grateful to Professor Yoval Shoham, the senior researcher Smadar Shulami and the PhD candidadate Tal Zelter, all from Technion Institute of Technology - Israel, for sharing all their secrets regarding the transformation of *Geobacillus stearothermophilus* T-1 with me.

Most important even though last, I would like to thank my family and my friends for all their precious support.

Abstract

The present thesis lies within the general framework of second generation biorefineries. Its subject specifically addresses the exploitation of thermophilic bacteria and their enzymes towards the bioconversion of lignocellulose to biofuels and other chemicals. In particular, it is focused on the study of the hydrolytic potential against cellulose and xylan of 103 thermophilic bacteria isolated from the volcanic habitat of Santorini (Meintanis, Chalkou et al. 2006). The majority of these strains belong to the genus *Geobacillus* and was able to degrade xylan, while only a few also presented cellulolytic activity (Stathopoulou, Galanopoulou et al. 2012).

Within this context, the PhD thesis can be divided into four parts; the first three are interlinked since they are all related to the degradation of xylan and cellulose by the Santorini *Geobacillus* strains. In the last one - also related to cellulose degradation by *Geobacilli* - a different *Geobacillus* isolate is employed, *G. stearothermophilus* T-1, a strain for which the genome is known, the xylanolytic system is well studied and there are available genetic manipulation techniques for its transformation (Shulami, Zaide et al. 2007).

Specifically, in the first part of the thesis, selected strains isolated from the volcanic habitat of Santorini were screened for genes encoding xylanolytic and cellulolytic enzymes. The screening of the desired genes was performed using polymerase chain reaction (PCR) and appropriately designed primers. Primer design was based in the conserved aminoacid regions of all the biochemically characterized xylanolytic and cellulolytic enzymes from *Geobacilli* and related strains, deposited in CAZy database (www.cazy.org). The outcome of this effort was the detection of two xylanases of the glycoside family GH10, three *β*-xylosidases of the glycoside families GH39, GH43 and GH52 respectively, one *β*-P-glucosidase of the glycoside family GH1 and one endoglucanase of the glycoside family GH5. Biochemical characterization of the enzymes validated their optimum xylanolytic or cellulolytic activities at neutral or slightly acid pH values. In addition, all enzymes presented significant thermostability at 60 and 65 ˚C while the maximum activity for all of them was observed at temperatures between 60 ˚C and. 70 ˚C.

In the second part of the thesis, one xylanase of the glycoside family GH10, the *β*-xylosidase of the glycoside family GH52 and the endoglucanase GH5, all cloned and characterized in the previous part, were used to construct a tetravalent designer cellulosome able to act at elevated

temperatures. A fourth complementary activity, that from the *β*-glucosidase of the bacterium *Caldicellulosiruptor saccharolyticus* (DSMZ 8903), was additionally employed. All four genes were combined with a dockerin domain, in order for the chimaeric products to interact with the corresponding cohesins of the cellulosome's scaffoldin. The resulted cellulosomal complex remained stable for at least 6 hours of incubation at 60 ˚C, while at the same time it hydrolyzed corn stover up to 66 % more efficiently compared to the free enzymes. The produced designer cellulosome, constitutes the first thermocellulosome characterised so far, confirming for the first time the feasibility of employing the designer cellulosome technology in biorefinery processes requiring elevated temperatures (Galanopoulou, Moraïs et al. 2016).

The third part of the thesis concerns the genome analysis for one of the cellulolytic strains isolated from the volcanic habitat of Santorini, namely, '*Geobacillus icigianus* SP50'. We decided to further investigate this particular strain for two reasons; first, the cellulolytic phenotype is a rare trait, not well documented among *Geobacilli*. Second, '*Geobacillus icigianus* SP50' was the only strain among the *Geobacillus* strains of Santorini, bearing the gene encoding for the endoglucanase of the glycoside hydrolase family GH5, already characterized in the first part of the thesis. Genome sequencing of '*Geobacillus icigianus* SP50' allowed us to identify that the GH5 endoglucanase is a part of a genetic locus, unique in Geobacilli, that contains also four additional genes; a transcriptional regulator of the LacI family and three genes constituting together an ATP-dependent oligosaccharide transporter. We propose that all these, together with the GH5 endoglucanase, form a genetic cluster, which is responsible for the degradation and utilization of a broad range of *β*-D-glucans. Intriguingly, genes of this *β*-D-glucan utilization cluster, were not detected in any other *Geobacillus* genome while the neighboring regions were significantly conserved. In contrast though, some of the genes of the cluster were found conserved in representatives of other bacterial genera (*Thermicanus aegyptius*) and even families (*Paenibacillus elgii),* indicating that the cluster has possibly been acquired through horizontal gene transfer from phylogenetically distant bacteria.

Finally, in line of exploiting the cellulolytic ability of *Geobacillus* strains, we attempted to transform the non-cellulolytic strain *G. stearothermophilus T-1* to a cellulolytic one through heterologous expression of a cellulase from the thermophilic bacterium *Thermobifida fusca* YX.

The procedure we chose was based on well-established transformation protocols of the strain using a derivative of the pNW33N plasmid (Zeigler 2001, Zeigler 2002, Shulami, Zaide et al. 2007). The resulted bifunctional phenotype could be of exceptional biotechnological interest since it could be incorporated in one-step biorefinery processes requiring elevated temperatures.

The outcome of the thesis, raises questions that could be further examined:

Regarding their enzymatic ability, it is already known that *Geobacilli,* besides xylanolytic hydrolases, produce a great variety of enzymes with auxiliary activities acting on hemicellulose. These enzymes, could be cloned and biochemically characterized in order to evaluate their contribution in lignocellulose degradation processes.

The creation of functional designer thermocellulosomes, sets the basis for their incorporation in various biotechnological applications requiring elevated-temperatures. Among them, their incorporation in one-step biorefineries for ethanol production is of great interest and an ongoing project for our lab.

Regarding the glucanolytic strain *"Geobacillus icigianus* SP50", one of the immediate goals of our lab is to investigate the regulation of the *β*-glucan utilization cluster.

In the same line, the genetically modified cellulolytic strain *G. stearothermophilus* T-1, could be incorporated in the hydrolysis of lignocellulosic substrates at elevated temperatures. To further improve the cellulolytic phenotype of the transformed *G. stearothermophilus* T-1 strain, we would like to attempt the introduction of the corresponding DNA cassette into the genome of the microorganism.

Finally, it would be of great interest to develop the appropriate methodology for the genetic manipulation of the thermophilic strains isolated from the volcano of Santorini. In this way, the observed phenotype could be further investigated or appropriately modified for biotechnological purposes.

Περιληψη

Η παρούσα διδακτορική διατριβή εντάσσεται στο γενικότερο πλαίσιο των βιοδιυληστηρίων δεύτερης γενιάς και ειδικότερα στην αξιοποίησης των θερμόφιλων μικροοργανισμών και των προϊόντων τους στη βιομετατροπή της λιγνοκυτταρινούχου βιομάζας σε βιοκαύσιμα και άλλα ενδιάμεσα χημικά. Πιο συγκεκριμένα, η εργασία εστιάστηκε στην υδρόλυση της κυτταρίνης και της ξυλάνης από 103 θερμόφιλα βακτηριακά στελέχη τα οποία συλλέχθηκαν από το ηφαίστειο της Σαντορίνης (Meintanis, Chalkou et al. 2006). Τα στελέχη αυτά στην πλειοψηφία τους ανήκουν στο γένος *Geobacillus* και εμφανίζουν ξυλανολυτική ικανότητα ενώ ορισμένα διαθέτουν επιπλέον την ικανότητα υδρόλυσης της κυτταρίνης (Stathopoulou, Galanopoulou et al. 2012).

Υπό αυτό το πλαίσιο, η διδακτορική εργασία μπορεί να υποδιαιρεθεί σε τέσσερα επί μέρους τμήματα. Τα πρώτα τρία είναι αλληλένδετα μεταξύ τους και σχετίζονται με την αποικοδόμηση της κυτταρίνης και της ξυλάνης από τα στελέχη *Geobacillus* που απομονώθηκαν από το ηφαίστειο της Σαντορίνης. Στο τελευταίο τμήμα,-το οποίο επίσης σχετίζεται με την αποικοδόμηση της κυτταρίνης από τους γεωβάκιλλους,- επιλέξαμε ένα διαφορετικό στέλεχος, το στέλεχος *G. stearothermophilus T-1*, για το οποίο το γονιδίωμα είναι γνωστό, το ξυλανολυτικό του σύστημα είναι εκτενώς μελετημένο και επιπλέον έχουν ήδη αναπτυχθεί τα κατάλληλα εργαλεία μετασχηματισμού του (Shulami, Zaide et al. 2007).

Ως σημείο αφετηρίας, επιλεγμένα στελέχη από το ηφαιστειακό ενδιαίτημα της Σαντορίνης ελέγχθηκαν για γονίδια που κωδικοποιούν κυτταρινολυτικά και ξυλανολυτικά ένζυμα. Ο εντοπισμός των επιθυμητών γονιδίων πραγματοποιήθηκε με την αλυσιδωτή αντίδραση πολυμεράσης (PCR) και χρήση κατάλληλων μορίων εκκινητών (primers). Ο σχεδιασμός των primers, βασίστηκε στις συντηρημένες περιοχές της αμινοξικής αλληλουχίας όλων των βιοχημικά χαρακτηρισμένων ξυλανολυτικών και κυτταρινολυτικών ενζύμων του γένους Geobacillus αλλά και συγγενικών ειδών τα οποία βρίσκονται κατατεθειμένα στη βάση δεδομένων CAZy (www.cazy.org). Η παραπάνω πορεία, οδήγησε στον εντοπισμό δύο ξυλανασών της οικογένειας γλυκοσιδικών υδρολασών (GH) 10, τριών β-ξυλοσιδασών των οικογενειών GH39, GH43 και GH52, μίας β-P-γλυκοζιδάσης της οικογένειας GH1, και μίας ενδογλυκανάσης της οικογένειας GH5. Ο βιοχημικός χαρακτηρισμός των ενζύμων αυτών, επιβεβαίωσε την αντίστοιχη ξυλανολυτική ή κυτταρινολυτική τους δράση. Επιπλέον, φάνηκε

πως τα ένζυμα αυτά δρουν σε ουδέτερο έως ελαφρώς όξινο περιβάλλον, παρουσιάζουν ιδιαίτερη θερμοσταθέροτητα στους 60 και 65 ˚C ενώ εμφανίζουν μέγιστη ενεργότητα μεταξύ 60 ˚C και 70 ˚C.

Στο δεύτερο τμήμα της διδακτορικής διατριβής, η μία ξυλανάση GH10, η β-ξυλοζιδάση GH52 και η ενδογλυκανάση GH5 που κλωνοποιήθηκαν στο προηγούμενο στάδιο της διδακτορικής διατριβής και επιπλέον μία β-γλυκοζιδάση από το βακτήριο *Caldicellulosiruptor saccharolyticus* (DSMZ 8903) χρησιμοποιήθηκαν για την κατασκευή ενός κυτταρινοσώματος ικανού να δρά σε υψηλές θερμοκρασίες. Στα ένζυμα αυτά εισήχθηκε μια εξειδικευμένη επικράτεια αγκίστρωσης (dockerin) έτσι ώστε το προκύπτον χιμαιρικό ένζυμο να είναι ικανό να προσδεθεί στην αντίστοιχη επικράτεια σύνδεσης (cohesin) της πρωτεΐνης στήριξης (scafoldin) του κυτταρινοσώματος. Το σύμπλοκο που προέκυψε, αποδείχτηκε ότι παραμένει σταθερό στους 60 ˚C για τουλάχιστον έξι ώρες επώασης και οτι στο συγκεκριμένο χρονικό διάστημα, υδρολύει το άχυρο καλαμποκιού έως και 66% πιο αποδοτικά σε σύγκριση με τα αντίστοιχα ελεύθερα ένζυμα.(Galanopoulou, Moraïs et al. 2016). Η παρούσα εργασία παρουσιάζει ιδιαίτερο ενδιαφέρον καθώς περιγράφει για πρώτη φορά την κατασκευή θερμοκυτταρινοσωμάτων και αποδεικνύει την λειτουργικότητα τους (Galanopoulou, Moraïs et al. 2016).Τα παραπάνω ανοίγουν το δρόμο για την αξιοποίηση των δομών αυτών σε βιοτεχνολογικές εφαρμογές που απαιτούν υψηλές θερμοκρασίες.

Το τρίτο μέρος της παρούσας εργασίας αφορά στην ανάλυση του γονιδιώματος ενός από τα απομονωθέντα στελέχη του ηφαιστείου της Σαντορίνης, του στελέχους "*Geobacillus icigianus* SP50". Η μελέτη του συγκεκριμένου στελέχους, παρουσιάζει ενδιαφέρον για δύο λόγους: Αρχικά, ο κυτταρινολυτικός φαινότυπος είναι σπάνιος και όχι καλά μελετημένος για το γένος των γεωβακίλλων. Δεύτερον, το στέλεχος "*Geobacillus icigianus* SP50" ήταν το μοναδικό στη συλλογή, στο οποίο εντοπίστηκε το γονίδιο της ενδογλυκανάσης GH5 που χαρακτηρίστηκε στο πρώτο μέρος της εργασίας. Από την ανάλυση του γονιδιώματος του εν λόγω στελέχους, εκτός από το γονίδιο της GH5 κυτταρινάσης, εντοπίστηκαν τέσσερα επιπλέον γονίδια: το γονίδιο ενός μεταγραφικού ρυθμιστή της οικογένειας LacI καθώς και τρία γονίδια τα οποία κωδικοποιούν για έναν ATP-εξαρτώμενο μεταφορέα ολιγοσακχαριτών. Προτείνουμε πως αυτά τα πέντε γονίδια μαζί αποτελούν μία ομάδα γονιδίων υπεύθυνη για την αποικοδόμηση και πρόσληψη ενός

μεγάλου εύρους β-D-γλυκανών από τα κύτταρα. Ενδιαφέρον παρουσιάζει το γεγονός ότι τα επιμέρους γονίδια της συγκεκριμένης ομάδας, δεν εντοπίστηκαν σε κανένα άλλο γονιδίωμα του γένους *Geobacillus* ενώ αντίθετα, σημαντικά συντηρημένα εμφανίστηκαν στο γονιδίωμα αντιπροσώπων διαφορετικών βακτηριακών γενών (*Thermicanus aegyptius*) αλλά και οικογενειών (*Paenibacillus elgii*). Το γεγονός αυτό υποδεικνύει ότι η συγκεκριμένη ομάδα γονιδίων ενσωματώθηκε, πιθανά, στο γονιδίωμα του '*Geobacillus icigianus* SP50' μέσω οριζόντιας μεταφοράς.

Τέλος και το τελευταίο κομμάτι της διδακτορικής διατριβής, επικεντρώθηκε στη διάσπαση της κυτταρίνης από το γένος *Geobacillus*. Πιο συγκεκριμένα, επιχειρήθηκε η μετατροπή του μη κυτταρινολυττικού στελέχους *G. stearothermophilus* T-1 σε κυτταρινολυτικό μέσω ετερόλογης έκφρασης μίας κυτταρινάσης του βακτηρίου *Thermobifida fusca* YX. Η μεθοδολογία που ακολουθήθηκε βασίστηκε σε πρωτόκολλα μετασχηματισμού του συγκεκριμένου στελέχους με ένα παράγωγο του πλασμιδιακού φορέα pNW33N (Zeigler 2001, Zeigler 2002, Shulami, Shenker et al. 2014). Ο προκύπτον μικτός φαινότυπος είναι ιδιαίτερης βιοτεχνολογικής σημασίας καθώς μπορεί να χρησιμοποιηθεί σε υψηλής θερμοκρασίας βιομετατροπές ενός σταδίου.

Από το σύνολο των επιμέρους τμημάτων της διατριβής που αναλύθηκαν παραπάνω, θεωρούμε ότι προκύπτουν ερωτήματα που χρήζουν περαιτέρω μελέτης:

Σε ενζυμικό επίπεδο, είναι γνωστό ότι οι γεωβάκιλοι διαθέτουν μία πληθώρα ενζύμων τα οποία δρουν βοηθητικά στην αποικοδόμηση της ημικυτταρίνης. Τα ένζυμα αυτά θα μπορούσαν να κλωνοποιηθούν και να χαρακτηριστούν βιοχημικά έτσι ώστε να εκτιμηθεί η συνεργιστική δράση τους στην αποτελεσματικότερη αποικοδόμηση της λιγνοκυτταρίνης.

Η δημιουργία των θερμοκυτταρινοσωμάτων και η αποτελεσματική ικανότητα υδρόλυσης αυτών σε υψηλές θερμοκρασίες, επιτρέπει την εξαγωγή μίας σειράς υποθέσεων οι οποίες θα μπορούσαν να εξεταστούν. Μεταξύ αυτών, ιδιαίτερο ενδιαφέρον για το εργαστήριο μας, έχει η ένταξη τους σε βιοδιυλιστήρια ενός σταδίου για την παραγωγή αιθανόλης.

Όσον αφορά το κυτταρινολυτικό στέλεχος '*Geobacillus icigianus* SP50', ένας από τους άμεσους στόχους μας είναι η διερεύνηση της ρύθμισης έκφρασης της β-D-γλυκανολυτικής ομάδας. Στην ίδια κατεύθυνση, θα μπορούσε να ενταχθεί το γενετικά τροποποιημένο στέλεχος *G. stearothermophilus* T-1 για τη χρήση του σε διεργασίες υδρόλυσης της λιγνοκυτταρίνης ενός σταδίου. Ως προς την περαιτέρω βελτιστοποίηση του στελέχους, άμεσος στόχος είναι η ενσωμάτωση του υπεύθυνου τμήματος DNA για τον διάσπαση της κυτταρίνης, στο γονιδίωμα του μικροοργανισμού. Τέλος, ιδιαίτερο ενδιαφέρον θα είχε η ανάπτυξη της κατάλληλης μεθοδολογίας για το γενετικό χειρισμό των βακτηριακών στελεχών τα οποία έχουν απομονωθεί από το ηφαίστειο της Σαντορίνης έτσι ώστε να μελετηθεί εκτενέστερα ο υπάρχον φαινότυπός τους ή να τροποποιηθεί επιπλέον έτσι ώστε να ενταχθούν σε διάφορες βιοτεχνολογικές διεργασίες.

1. Introduction

1.1 BIOREFINERY; AN ALTERNATIVE TO THE PETROLEUM BASED ECONOMY

Fossil fuels constitute a finite resource of energy and carbon-based chemicals that is being rapidly consumed. The inevitable peak in crude oil extraction volumes is estimated that has already happened in 2006, while for the alternative forms of hydrocarbon deposits and natural gas it is foreseen within the next couple of decades (Nehring 2009, Miller and Sorrell 2014). Consequently, there is an immediate demand for alternative energy and reduced carbon sources which, in contrast to the conventional ones, should be renewable and environmentally friendly. Towards this direction, different sources are currently examined such as nuclear fission, water electrolysis, wind and wave power, solar radiation, geothermal heat and biomass (Figure 1.1.1).

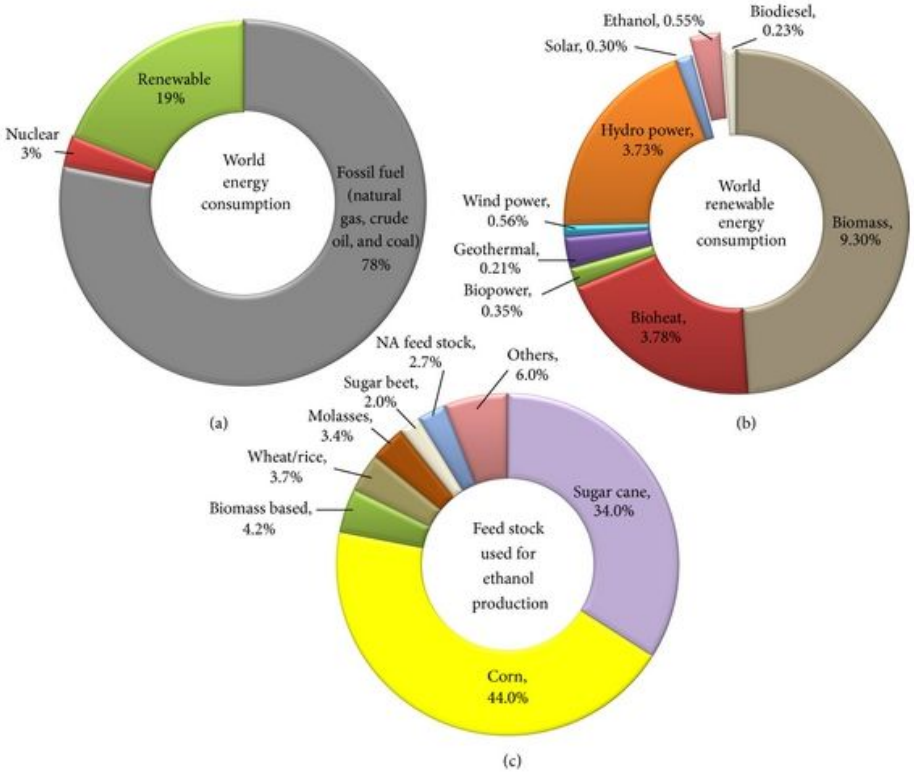


Figure 1.1.1: Contribution of renewable energy and biofuels in the total energy consumed in the world in 2011. (a) distribution of world energy consumption; (b) distribution of world renewable energy consumption, and (c) distribution o different feedstock currently used for ethanol production (Balan 2014).

The term biorefinery defines the cascade of chemical and biotechnological processes that integrate towards the production of fuels, power, and value-added chemicals from biomass.

Biorefinery is analogous to today's petroleum refinery, which produces multiple fuels and products from petroleum. By producing several products, a biorefinery takes advantage of the various components in biomass and their intermediates, therefore maximizing the value derived from the biomass feedstock. As an energy production process it has several advantages over the above mentioned alternatives, such as the small contribution to the green house phenomenon - since it is based on renewables - and the well-established technology applied in order to deliver safe and reliable energy. Thus, biomass derived fuels currently cover the 10% of the global energy demands and many countries, including members of the European Union, have set even higher goals for the near future (European-Biomass-Association 2016, World-Energy-Council 2016).

In general, four biomass types are already used or considered favorable for biorefinery applications. Each one is of different origin and has unique structural properties that require the implementation of different technologies for its exploitation:

1. ***Woodland biomass***, releases energy upon combustion and currently represents the largest portion of biomass used for the production of energy, especially in the developing countries. Even though it represents a very simple approach for energy production, significant concerns are raised regarding the harmful pollutants, like carbon monoxide and particulate matter, released with wood smoke. Additionally, demanding energy needs in combination with poor tree-harvesting control, lead to deforestation (World-Energy-Council 2016).
2. ***Biomass from agricultural plants*** dedicated for biofuels production is already widely used in many countries for the production of the so-called 1st generation biofuels. In this case, easily hydrolyzed compounds, like starch from corn and sucrose from sugarcane or oil from soybean and palm trees can be converted to ethanol or biodiesel, respectively. However 1st generation biofuels raise a lot of ethical and environmental concerns, since their production has resulted in a change-of-use for fields traditionally exploited for food production (developed countries) or in an aggressive deforestation for the creation of arable land (developing word). In addition, high volume growth of the corresponding

plants requires the use of significant amounts of water, pesticides and fertilizers (Ho, Ngo et al. 2014, World-Energy-Council 2016).

3. ***Algae*** are an alternative source for the production of energy. These photosynthetic microorganisms use sunlight, carbon dioxide, inorganic nutrients and water to grow and accumulate lipids that can be further converted to biofuels (mostly biodiesel). However, the technology is not yet cost competitive and production is energy intensive (Chew, Yap et al. 2017, Eppink, Olivieri et al. 2017).
4. ***Waste-based biorefineries*** could contribute to energy and high-value products generation, while at the same time they could alleviate the processing costs for the constantly increasing waste volumes of our society (Hoornweg D. 2012). Present technologies, focus on the conversion of the organic part of the wastes that constitutes approximately 50% of solid waste mass and can be relatively easily converted to fuels and other commodities. (Hoornweg D. 2012, Scoma, Rebecchi et al. 2016). The significant amounts of municipal wastes that are currently being produced, render them as a favourable raw material towards this direction. However, in order for them to be efficiently exploited, many parameters such as the regional and seasonal variation of their composition and volume have to be unraveled. On the other hand, agricultural wastes, and particularly lignocellulose, offer a more homogenous organic mix that is already being used as an alternative source for ethanol and other biofuels production - 2nd generation biofuels (World-Energy-Council 2016). However significant research remains to be done in order to make biorefinery from any type of wastes a cost competitive approach (Menon and Rao 2012, Balan 2014, Saini, Saini et al. 2015, Scoma, Rebecchi et al. 2016).

1.2 LIGNOCELLULOSE AS RAW MATERIAL FOR BIOREFINERIES

1.2.1 Composition and structure of lignocellulose

Lignocellulose represents the most abundant and renewable source of reduced carbon on Earth and its utilization as a raw material does not have any impact on land use. The three major components of lignocellulose, cellulose, hemicellulose and lignin (Figure 1.2.1) are briefly described below:

1. **Cellulose** is the main structural component of plant biomass representing 30 to 50% of the total dry weight of the plant. It is found in the plant cell walls and its main role is to confer rigidity to the plant. At the molecular level, cellulose consists of cellobiose units linked together via β-(1,4)-glycosidic bonds forming linear fibers. The later, are being further assembled via hydrogen bonds and weak van der Waals forces to longer linear units which can be further polymerized to crystalline structures (Lynd, Weimer et al. 2002).

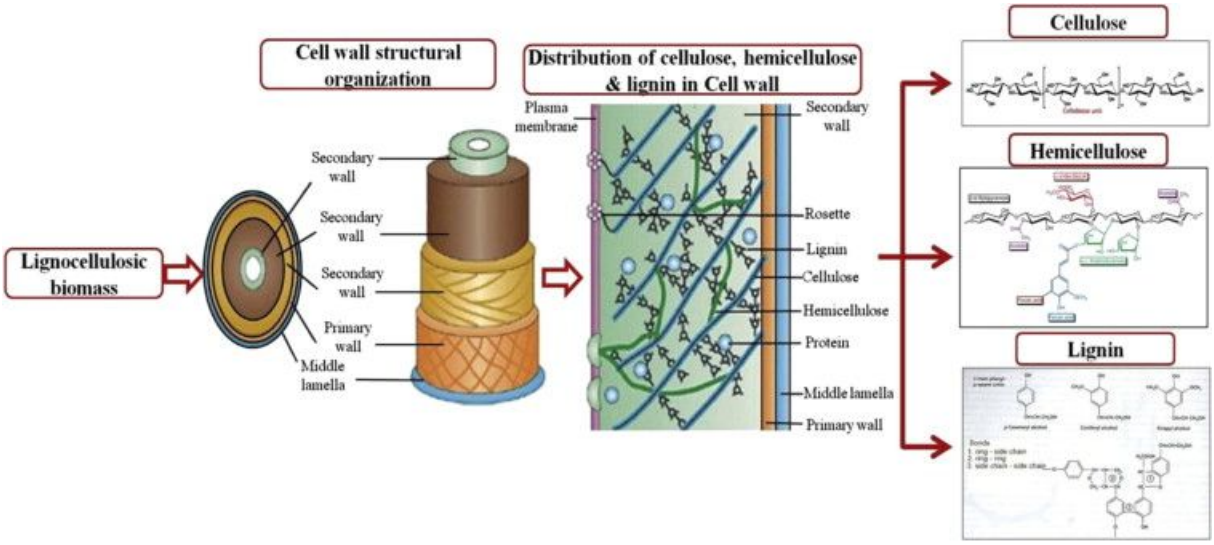


Figure.1.2.1: Diagrammatic illustration of the framework of lignocellulose. The chemical structure of the three main components-cellulose, hemicellulose and lignin- are also given (Menon and Rao 2012).

2. **Hemicellulose** represents the 25-35% of the total plant biomass. It contributes in the structural stability of the plant and also to the protection of cellulose fibers against enzymatic degradation. Hemicellulose is a heterogenous mix of branched heteropolymers consisting of D-xylose, D-glucose, D-mannanose, D-galactose, D-glucuronic acid and L-arabinose. The polymers are named after the principal monomer component as xylan, mannan, galactan or arabinan, respectively. The ratio of the various hemicelluloses varies among different plants, however in most cases xylan is the principal hemicellulosic polysaccharide; especially in the cell walls of land plants, it represents up to 30% – 35% of the total dry weight and in the hardwood from angiosperms it accounts for approximately 15% – 30% of the total dry weight. The hemicellulose layer is located

between cellulose fibers and lignin, forming non-covalent bonds with the first and covalent ones with the second (Beg, Kapoor et al. 2001, Menon and Rao 2012).

3. **Lignin** is a hydrophobic, cross-linked aromatic polymer which represents the 15 - 20% of plant biomass dry weight. It interferes with the microbial degradation of the plant while also offers rigidity and flexibility. Lignin occurs from the polymerization of the three major units, p-cumaryl alcohol, coniferyl alcohol and sinapyl alcohol. The ratio of the monomers varies between different plants, wood tissues and cell wall layers (Leisola, Pastinen et al. 2012, Menon and Rao 2012).

1.2.2 Typical approaches to convert lignocellulose to biofuels

As a result of the complex nature of lignocellulose, the production of 2nd generation biofuels and chemicals requires more demanding conditions compared to 1st generation biofuels (Balan 2014) (Figure 1.2.2). There are three different general approaches that can be followed for the exploitation of this complex biopolymer; (i) the thermochemical (Elliott, Biller et al. 2015), (ii) the biochemical, and (iii) the hybrid one which involves both thermochemical and biochemical steps. The incorporation of microbial or enzymatically mediated steps in the biochemical and the hybrid approaches, is considered favorable in terms of sustainability and production costs.

Generally, at the biochemical route, the following steps have to be implemented upon harvesting of the lignocellulosic byproduct (Figure 1.2.2):

1. **Biomass pretreatment:** In an attempt to defend itself against invading microorganisms, plant biomass has been evolved to be significantly recalcitrant. Thus, the first step for the biochemical utilization of lignocellulose, is to render the carbohydrate polymers more accessible to hydrolytic enzymes by breaking the bonds between lignin and cellulose or hemicellulose (Figure 1.2.3). Towards this direction, various methods have been employed, with efficiencies that significantly depend on the inherent properties of the individual materials. The most prominent methods examined so far include (i) physical pretreatment (e.g. size reduction by grinding or milling combined with use of water at high temperatures and pressures), (ii) chemical pretreatment under alkaline conditions (e.g. NaOH, alkaline hydrogen peroxide, lime), acid conditions (e.g. dilute sulfuric acid,

use of organic solvents under acidic conditions - organosolv) or neutral conditions (e.g. liquid hot water, ozonolysis), (iii) biological pretreatment incorporating fungi or ligninolytic enzymes. Among them, the physical and the biological pretreatments are more advantageous, since they do not involve or produce as many enzyme inhibitors as the chemical pretreatment, a fact that affects the subsequent enzymatic and fermentation steps (Capolupo and Faraco 2016, Arevalo-Gallegos, Ahmad et al. 2017).

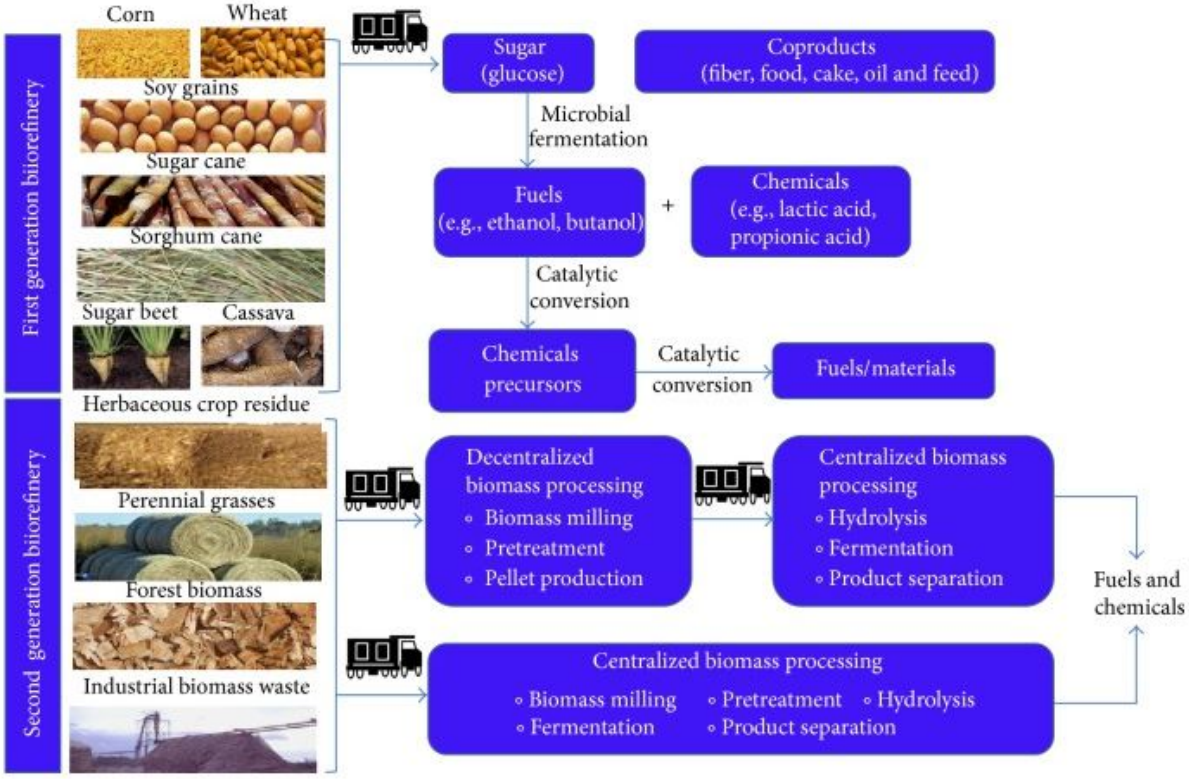


Figure 1.2.2: Different feedstocks used in the first and second generation biorefinery and required procedures for producing biofuels, biochemicals, food and feed (Balan, 2014).

2. **Enzymatic hydrolysis:** The aim of this process is to convert the polymers of biomass to simpler organic molecules, mainly hexoses, pentoses, and low molecular weight phenolics (Figure 1.2.4). For this purpose, a variety of enzymes is required (Paragraph 1.4) that, in the simplest hydrolytic scheme, have to be separately produced and then allowed to interact with the substrate. This extra step significantly increases the cost of the overall process, leading to a quest for one-vessel alternatives referred in literature as SSF (Simultaneous Saccharification and Fermentation) or CBP (Consolidated

BioProcessing). In the first one-vessel approach, the enzymes are added directly into the bio-conversion vessel and hydrolysis proceeds in parallel with the fermentative production. In the second approach, enzyme production and product formation are performed by the same microorganism (or consortium of microorganisms). (Balat 2011, Blumer-Schuette, Brown et al. 2014).

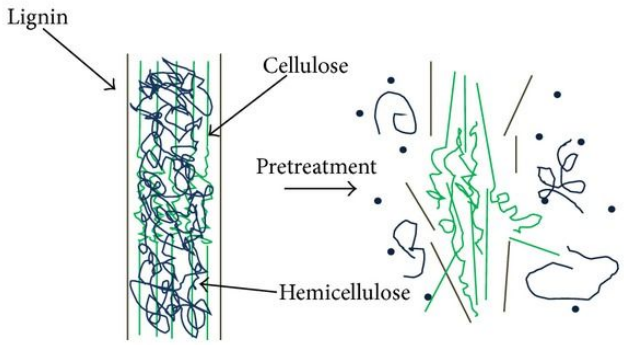


Figure 1.2.3: Deconstruction of lignocellulose into cellulose, hemicellulose and lignin (Lee, Hamid et al. 2014)

Another cost-reducing factor is the employment of thermophilic or even extremophilic enzymes and microorganisms. Performing hydrolysis at elevated temperatures, offers many advantages like, reduced microbial contamination risks, increased transport rates, higher catalytic efficiencies and reduction of incubation times. Elevated temperatures may additionally increase productivity by alleviating the product inhibition effect through evaporation, as in the cases where a volatile final product, such as ethanol, is produced through SSF or CBP processes (Yeoman, Han et al. 2010, Bhalla, Bansal et al. 2013). Finally, the incorporation of extremophiles and their enzymes could also be combined and partially contribute to a pretreatment step that involves elevated temperatures (Blumer-Schuette, Kataeva et al. 2008, Blumer-Schuette, Brown et al. 2014).

3. **Microbial fermentation and co-products generation**: The bioconversion of the simple sugars to ethanol and other value-added products can be conducted by various microorganisms like the natural glucose-fermenting yeast *Saccharomyces cerevisiae* and bacterium *Zymomonas mobilis*, and the natural xylose fermenting yeasts of the *Pichia*, *Candida* and *Kluyveromyces* genera. Bacteria, such as *Zymomonas mobilis*, *Escherichia coli* and *Klebsiella oxytoca*, are also of particular interest, since they can reduce the fermentation time to a few hours only. Since the majority of organisms in nature are able to ferment either hexoses or pentoses, several attempts are also being made for the

construction of genetically modified strains able to simultaneously ferment both types of sugars (Stephanopoulos 2007, Balat 2011, Peralta-Yahya, Zhang et al. 2012).

4. **Product separation:** as biomass – based biofuel production is being optimized and reaches commercial viability, the last process step required for optimization is product recovery. For ethanol and other volatile substances, the method of choice is usually distillation. For other fermentation derived products the selected recovery method depends on the chemical nature of the particular compound. As a result, several methods have been employed including, precipitation, membrane separation and ultrafiltration (Ahmetović, Martín et al. 2010).

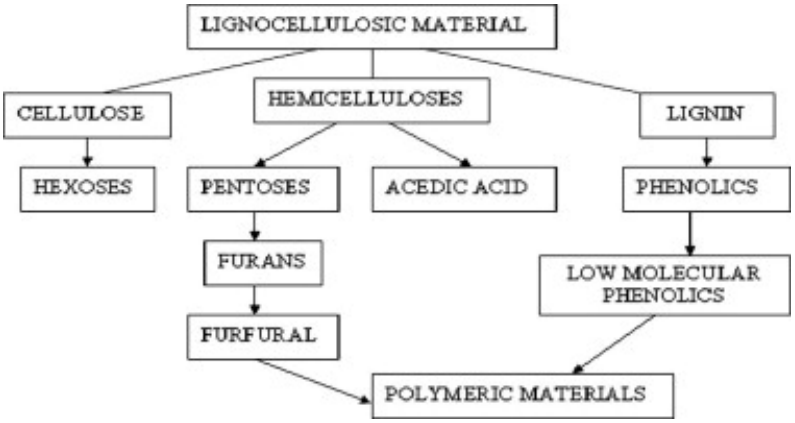


Figure 1.2.4: Main degradation products occurring during hydrolysis of lignocellulosic material (Balat 2011).

1.3 LIGNOCELLULOSE DEGRADERS

Plant biomass degradation ability is a trait encountered in relatively few organisms, scattered over the tree of life. Until very recently, this trait was attributed only to some fungal and bacterial representatives. However, extensive research on the field, has revealed that some archaea, single cell eukaryotes and even some invertebrates - like plant-parasitic nematodes, cockroaches and termites - also possess genes coding for lignocellulolytic enzymes (Figure 1.3.1) (Cragg, Beckham et al. 2015). Interestingly though, there is not any single organism known, that contains all the required genetic information for the complete lignocellulose deconstruction (Koeck, Pechtl et al. 2014). As a result, in nature this process requires the synergistic action of two or more degraders.

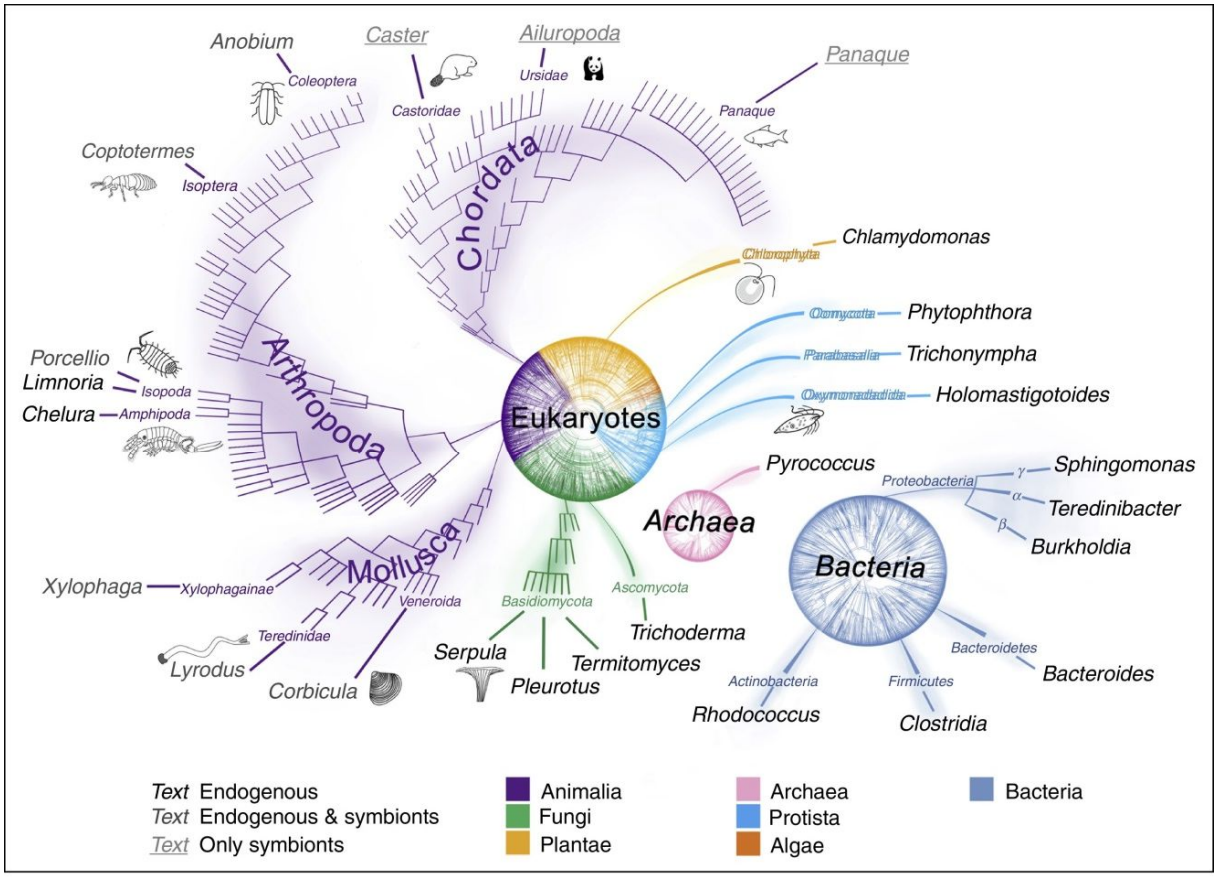
1.3.1 Lignocellulose degrading bacteria

Among bacteria, free living lignocellulolytic representatives have been reported in both soil and marine environments. Common soil degraders belong to the phyla *Firmicutes*, *Actinobacteria*, *Proteobacteria* and *Bacteroidetes* (Koeck, Pechtl et al. 2014). Representatives of wood degrading marine bacteria are some gram negative rods and some species of the phylum *Bacteroidetes* belonging to the *Cytophaga - Flavobacteria* group (Björdal 2012). In addition to the free living bacteria, lignocellulolytic bacteria also live symbiotically in the gut of some invertebrates and also of some vertebrates such as fish and ruminants. Based on recent metagenomic studies, the majority of the symbiotic bacteria fall into the same taxonomic groups as their soil counterparts (Cragg, Beckham et al. 2015).

Bacterial degraders harbour genes for cellulose, hemicellulose and pectin depolymerisation. Recently a number of Actinobacteria, α-Proteobacteria and γ-Proteobacteria have been identified bearing ligninase genes (Bugg, Ahmad et al. 2011). It is noteworthy that, despite the great variety of enzymes in the Domain of Bacteria, an individual *species* able to completely degrade lignocellulose by itself has not been isolated so far. In contrast, there are several bacterial groups that are able to efficiently hydrolyze the specific components of lignocellulose. For example representatives of the genus *Geobacillus* are able to efficiently degrade and utilise xylan while, in contrast, members of the genus *Clostridium*, like *Clostridium thermocellum,* are able to efficiently degrade cellulose.

Environmental parameters like oxygen availability, temperature and salt tolerance can also affect the cellulolytic strategy followed by the microorganisms (Lynd, Weimer et al. 2002). For example, aerobic bacteria do not require physical contact with their substrate and produce substantial amounts of extracellular or cell associated enzymes for its degradation (Figure 1.3.2 A). On the other hand, in order for the majority of anaerobic bacteria to hydrolyse lignocellulose, attachment of the cells on the substrate is a prerequisite. Upon attachment on plant-biomass, degradation occurs via poly-enzymatic complexes, the cellulosomes. These lego-like protein superstructures consist of a non-catalytic subunit called scaffoldin, on which enzymatic subunits are attached (Figure 1.3.2 B). This specific, intra-subunit assembly occurs through the cohesin modules of the scaffoldin and the dockerin modules of the enzymes (Bayer, Belaich et al. 2004).

Scaffoldins can additionally possess carbohydrate-binding modules (CBMs) which specifically bind to the polysaccharide components of plant biomass (Shoham, Lamed et al. 1999). The advantages of employing cellulosomes in plant biomass degradation are thought to be; (i) the concerted enzyme activity in close proximity to the bacterial cell, enabling optimum synergism between the cellulosomal enzymes, and (ii) the minimization of the distance over which cellulose hydrolysis products must diffuse, allowing efficient uptake of these oligosaccharides by the host cell.



Current Opinion in Chemical Biology

Figure 1.3.1: The sparse and localised distribution of selected organisms capable of lignocellulose or cellulose degradation mapped onto the Tree of Life, with highest taxonomic ranks colour-coded as shown in key. Genus names of organisms degrading lignocellulose using endogenous enzymes shown in bold, those with endogenous plus symbiont-derived enzymes shown printed pale and those with only symbiont-derived enzymes shown underlined (Cragg, Beckham et al. 2015).

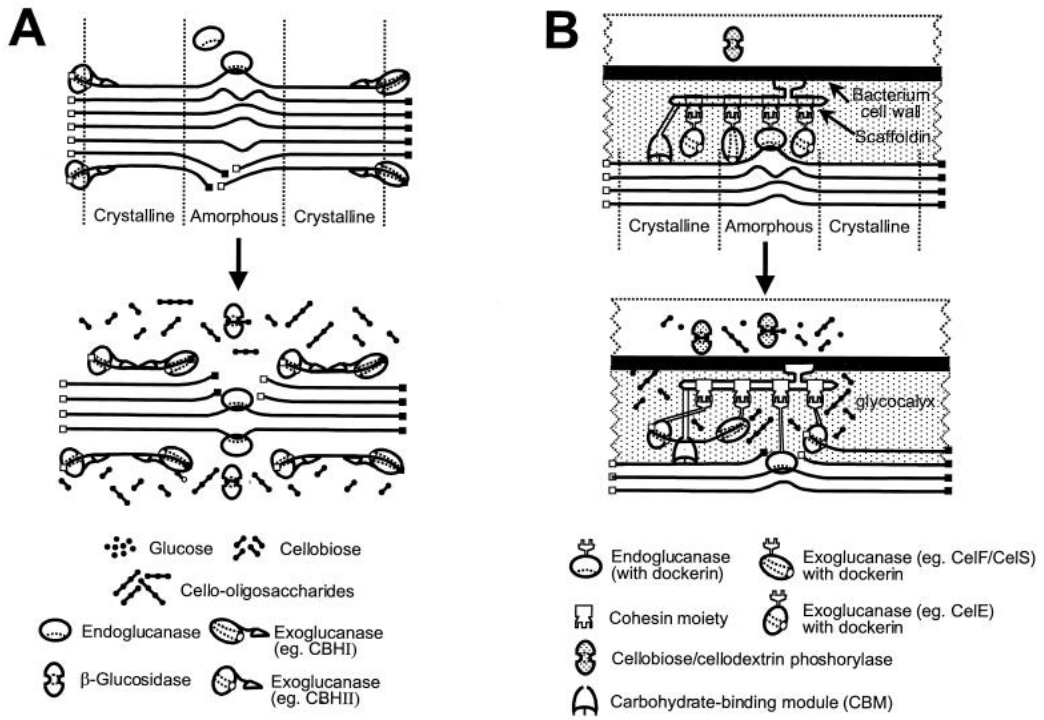


Figure 1.3.2: Schematic representation of the hydrolysis of amorphous and microcrystalline cellulose by noncomplexed (A) and complexed (B) cellulase systems. The solid squares represent reducing ends, and the open squares represent nonreducing ends. Amorphous and crystalline regions are indicated. Cellulose, enzymes, and hydrolytic products are not shown to scale (Lynd, Weimer et al. 2002).

1.3.2 Lignocellulose degradation by fungi

Lignocellulolytic traits are encountered within all fungal taxa; from the protist-like *Chytridomycetes* to the advanced *Basidiomycetes*. The two cellulolytic ascomycetes, *Hypocrea jecorina* (*Trichoderma reesei*) and the xylanolytic *Aspergillus niger* have been extensively studied for over 50 years and until today are considered the most important organisms for cellulase/hemicellulase production (De Souza 2013). In addition to the above, Agaricomycotina, representatives are also well known for their lignocellulolytic potential. Based on the type of wood decay, Agaricomycotina are traditionally distinguished in white rot and brown rot (Grigoriev, Cullen et al. 2011, Cragg, Beckham et al. 2015). White-rot fungi, like *Phanerochaete chrysosporium* (Kersten and Cullen 2007) bear a significant amount of ligninolytic, cellulolytic and/ or hemicellulolytic enzymes causing a white or yellow, soft and spongy result to the rotted wood. On the other hand, brown rot fungi like *Serpula lacrymans*, as implied by their name, cause a brown discoloration to the rotten wood and also wood cracking and shrinking. Brown rot

fungi have evolved multiple times from the predecessors of current white rot fungi.Their main difference is the lack of ligninolytic enzymes and key cellulases and their replacement by a non-enzymatic mechanism, the chelator-mediated Fenton (CMF) system. In this case, the microorganism produces extracellular hydrogen peroxide, which interacts with the iron found in the environment, causing oxidative degradation of the wood (Floudas, Binder et al. 2012, Arantes, Milagres et al. 2014).

1.4 Enzymes and associated modules involved in lignocellulose degradation

The decomposition of plant biomass requires the contribution of various enzymes, categorized to the following families:

1. **Glycoside Hydrolases** (GHs) which are responsible for the hydrolysis of the glycosidic bond between two or more carbohydrates and another chemical moiety.
2. **Carbohydrate Esterases** (CEs) catalyzing the de-O or de-N-acylation of substituted saccharides.
3. **Polysaccharide Lyases** (PLs) that cleave uronic acid-containing polysaccharide chains, generating an unsaturated hexenuronic acid residue and a new reducing end
4. **Auxilary-activity redox enzymes** (AAs), incuding ligninolytic enzymes and lytic polysaccharide mono-oxygenases, the action of which help the original GH, CE and PL enzymes to gain access to their substrate.

Finally, particular domains of the above enzymes or scaffoldins, possess a carbohydrate-binding ability (the Carbohydrate-Binding Modules - CBMs), which further enhance the affinity of the enzymes with their substrate (Lombard, Golaconda Ramulu et al. 2014).

The contribution of each different enzyme and CBM to the efficient biomass degradation depends on the specific characteristics of the substrate and still novel enzymes are reported to enhance the overall degradation process (Bugg, Ahmad et al. 2011, Hiras, Wu et al. 2016).

In the paragraph that follows, are described the two classification approaches of these enzymes and CBMs based on their amino-acid sequence similarity (CAZy classification) or their

biochemical profile (IUBMB Enzyme nomenclature). In addition, they key enzymatic activities, according to the IUBMB enzyme nomenclature, for cellulose and xylan degradation are discussed (Paragraphs 1.4.1 and 1.4.2).

1.4.1 Classification of enzymes and associated modules based on amino-acid sequence similarities

According the IUBMB Enzyme nomenclature, all enzymes are classified according to their substrate specificity and occasionally on their molecular mechanism; such a classification though, is not intended to and, consequently, does not always reflect the structural features of these molecules. In the 1990s, an alternative classification method, based on amino-acid sequence similarities was proposed by Bernard Henrissat (Henrissat 1991, Henrissat and Davies 1997). Since there is a direct relationship between sequence and folding similarities, such a classification;

1. Reflects the structural features of these enzymes
2. Helps to reveal the evolutionary relationship of the enzymes
3. Provides a convenient tool to derive mechanistic information
4. Illustrates the difficulty of deriving relationships between family membership and substrate specificity (Lombard, Golaconda Ramulu et al. 2014).

According to this sequence-based classification, all carbohydrate-related enzymes can be classified in different families, available in CAZy database (www.cazy.org) (Lombard, Golaconda Ramulu et al. 2014). Currently (2 July 2018) there are 153 GH families, 28 PL families, 16 CE families and 15 AA families described in CAZy with these numbers to be constantly increasing as new sequences and biochemical data are accumulated.

1.4.2 Hydrolases acting on cellulose

Biochemical analysis of various bacterial and fungal cellulolytic systems, has separated the key enzymes of cellulose hydrolysis in three distinct categories; endocellulases, exoglucanases and cellobiosidases.

1. **Endoglucanases** or 1,4-*β*-D-glucan-4-glucanohydrolases (E.C 3.2.1.4), act at random internal amorphous sites within the cellulose polysaccharide chain, generating oligosaccharides of various lengths and, consequently, new chain ends (Lynd, Weimer et al. 2002).
2. **Exoglucanases** act in a processive manner on either the reducing or nonreducing ends of cellulose polysaccharide chains, liberating either glucose or cellobiose as major products. Exoglucanases can also act on microcrystalline cellulose, by peeling cellulose polysaccharide chains from the microcrystalline structure. Based on the cellulose- chain end (reducing or not) that these cellulases act, and also the corresponding released products, exocellulases can be classified as; (i) 1,4-*β*-D-glucan glucanohydrolases or cellodextrinases (E.C. 3.2.1.74), (ii) non reducing end 1,4-*β*-D-glucan cellobiohydrolases (E.C. 3.2.1.91) and (iii) reducing end 1,4-*β*-D-glucan cellobiohydrolases (E.C. 3.2.1.176) (Barr, Hsieh et al. 1996, Lynd, Weimer et al. 2002).
3. ***β*-*G*lucosidases** also known as *β*-glucoside-glucohydrolases (E.C.3.2.1.21), act on cellobiose or other oligosaccharides released from the action of endocellulases and exocellulases releasing glucose monomers. Some bacteria though, possess a phosphoenolpyruvate-dependent phosphotransferase system (PTS) for the uptake and further hydrolysis of cellobiose. In this case, instead of *β*-glucosidases, bacteria produce ***β*-P-glucosidases** (E.C 3.2.1.86) acting on P-cellobiose (Lai and Ingram 1993, Lynd, Laser et al. 2008).

According to the amino-acid based classification, all the above enzymes are distributed in sixteen GH families as summarized in Table 1.4.1.

1.4.3 Hydrolases acting on xylan

In a way similar with cellulose, the complete decomposition of xylan requires the interaction of a number of main-chain and side-chain-cleaving enzyme activities (Figure 1.4.1) (Subramaniyan and Prema 2002, Kumar, Singh et al. 2008). Among them, prominent role have:

1. **Endo-*β*-1,4-xylanases** or *β*-1,4-D-xylan xylanohydrolases (E.C. 3.2.1.8) which hydrolyse the (1→4)-*β*-D-xylosidic linkages in xylans at internal sites of the xylan backbone.
2. **Oligosaccharide reducing-end xylanases** (E.C. 3.2.1.156) releasing xylose units from the non-reducing end of xylooligosaccharides.

Table 1.4.1: Distribution of the various enzymes acting on cellulose in different GH families.

E.C	Enzymatic activity	GH families
3.2.1.4	Endoglucanases	5, 6, 7, 8, 9, 12, 44, 45, 48, 51, 74
3.2.1.74	1,4-*β*-D-Glucan glucanohydrolases	1, 3, 5, 9
3.2.1.91	Non reducing end 1,4-*β*-D-glucan cellobiohydrolases	7, 48
3.2.1.176	Reducing end 1,4-*β*-D-glucan cellobiohydrolases	1, 3, 5, 9, 30
3.2.1.21	*β*-*G*lucosidases	1, 3, 5, 9, 30, 116
3.2.1.86	*β*-P-glucosidases	1, 4

3. **endo-1,3-*β*-D-xylanases** (E.C. 3.2.1.32), responsible for the hydrolysis of (1→3)-*β*-D-xylosidic linkages in xylans at internal sites of the xylan backbone.
4. **xylan 1,-4-*β*-xylosidases** also known as 1,4-*β*-D-*β*-xylosidases (E.C. 3.2.1.37) and **xylan 1,-3-*β*-xylosidases** also known as 1,3-*β*-D-xylosidases (E.C. 3.2.1.72), are responsible for the cleavage of xylose residues from the non-reducing termini of 1,4-*β*-D-xylans and 1,3-*β*-D-xylans, respectively, and are also able to act on xylobiose.

According to the amino-acid based classification, all the above enzymes are distributed in seventeen GH families as summarized in Table 1.4.2

1.5 THERMOPHILIC REPRESENTATIVES OF BACILLUS SENSU LATO; CLASSIFICATION AND BIOTECHNOLOGICAL INTEREST

1.5.1 Classification of *Bacillus sensu lato*; a brief overview

It was Nathan R. Smith, Francis E. Clark, and Ruth E. Gordon who first used the term *Bacillus* in the 1930's for all "rod-shaped bacteria capable of aerobically forming refractile endospores that are more resistant than vegetative cells to heat, drying, and other destructive agencies" (Gordon, Haynes et al. 1973). Some years later, in 1949, the same scientists proposed for all thermophilic representatives to be assigned to two species; *B. stearothermophilus* and *B. coagulans*. However even since these days, it was predicted that inevitably, more thermophilic species will be found as more strains are being discovered (Gordon and Smith 1949). Indeed, in 1993, based on phenotypic characteristics, White et al. re-examined *Bacillus* classification uncovering an "enormous diversity" within the genus (White, Sharp et al. 1993). As a result, thermophiles fell into 18 phenetic clusters, some of them corresponding to already known species.

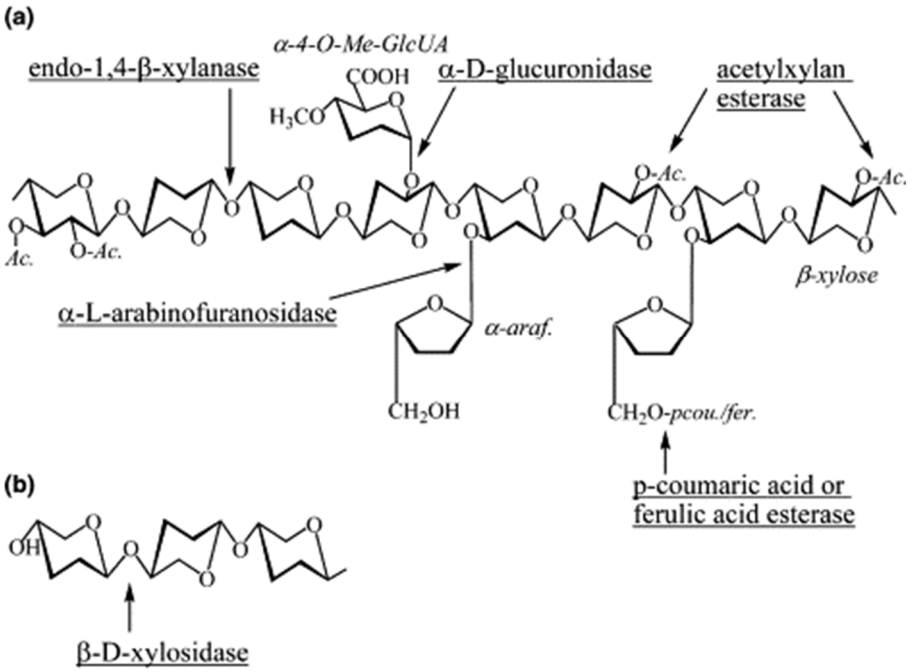


Figure 1.4.1 (a) Structure of xylan and the sites of its attack by xylanolytic enzymes. The backbone of the substrate is composed of 1,4- *β*-linked xylose residues. Ac., Acetyl group; α-araf., α-arabinofuranose; α-4-O-Me-GlcUA, α-4-O-methylglucuronic acid; pcou., P-coumaric acid; fer., ferulic acid. (b) Hydrolysis of xylo-oligosaccharide by *β*-xylosidase (Collins, Gerday et al. 2005).

However, the revolution in the taxonomy of the genus occurred when Woese introduced the small subunit of rRNA as a taxonomical tool (Fox, Stackebrandt et al. 1980). Since then, a re-organization of *Bacillus sensu lato* started (Ash, Farrow et al. 1991) that continues with fast pace until today (Aliyu, Lebre et al. 2016). During the two previous decades, the taxonomical groups reported in the present work - the genera *Aneurinibacillus* (Shida, Takagi et al. 1996) and *Geobacillus* (Nazina, Tourova et al. 2001) - have been established based on their 16s rRNA sequence similarities. Later on though, further re-classifications occurred and nowadays the genus *Aneurinibacillus* is classified within the Paenibacillaceae family (Vos, Ludwig et al. 2015). *Geobacillus,* recently divided to *Geobacillus*, *Aeribacillus* (Minana-Galbis, Pinzon et al. 2010) and *Parageobacillus* (Aliyu, Lebre et al. 2016), all of them belonging to the populous Bacillacaeae family. In addition to the above, additional genera with thermophilic representatives are the three genera of Paenibacillaceae family *Brevibacillus*, *Paenibacillus* and *Ureibacillus* and the three genera of Bacillaceae family *Alicyclobacillus*, *Anoxybacillus* and *Thermobacillus* (Zeigler 2001, Zeigler 2013).

Table 1.4.2:Distribution of the various enzymes acting on cellulose in different GH families.

E.C	Enzymatic activity	GH family
3.2.1.8	endo-*β*-1,4-xylanases or *β*-1,4-D-xylan xylanohydrolases	5, 8, 10, 11, 30, 43, 51, 98, 141
3.2.1.156	oligosaccharide reducing-end xylanases	8
3.2.1.32	endo-1,3-*β*-D-xylanases	10, 11, 26
3.2.1.37	xylan 1,-4-*β*-xylosidases	1, 3, 30, 39, 43, 51, 52, 54, 116, 120
3.2.1.72	1,3-*β*-D-xylosidases	Not assigned in any GH family

Besides the differences in their 16s rRNA sequences, G+C contents, fatty acid and polar lipid profiles and other morphological/physiological characteristics used for their differentiation, the members of all these genera present some common characteristics; they are all sporulative, aerobic or facultative anaerobic, have an optimum growth at elevated temperatures (around 60 ˚C) and they are all encountered in very divergent environments that range from hot springs and compost piles to ice cores in Antarctica and deap sea sediments (Zeigler 2001, Zeigler 2014).

1.5.2 Thermophilic *Bacilli* as a source of enzymes and cell products with a biotechnological interest

A significant number of enzymes from thermophilic *Bacilli* are already employed (Table 1.5.1) or are currently being evaluated for their incorporation into industrial or research processes. The great interest on these enzymes is mainly attributed to their thermophile nature, which - as already discussed in Paragraph 1.2.2 - offers compatibility with all processes requiring elevated temperatures. Some of these enzymes, already employed in the industrial sector are the following (McMullan, Christie et al. 2004, Hussein, Lisowska et al. 2015, Chakdar, Kumar et al. 2016);

1. **Proteases** deployed in leather processing, peptide synthesis and food industry.
2. **Carboxylesterases** and **lipases** used in the cosmetic industry for the synthesis of emollient esters, cheese production and synthesis of sugar esters. The fine-chemical industry also examines the incorporation of these enzymes in the production of optically active building blocks for applications in synthetic chemistry.
3. **L-arabinose isomerases** exploited in the production of D-galactose-based food sweeteners.
4. **Pyrimidine nucleoside phosphorylases**, employed in the production of modified nucleosides, widely used in the treatment of diverse human tumors and viral infections.
5. **Xylanases** and **arabinoxylanases** employed in the pulp and paper industry, deinking during paper recycling processes, improvement of animal feed, improvement of dough's quality at bakery industry, biofuels production and production of high-value-added prebiotics for the pharmaceutical industry.

In addition to the enzymes produced by thermophilic Bacilli, the structural glycoproteins of their cell-surface layers (S-layers) are also considered of great interest as affinity membranes or nanoparticles for carrying enzymes, antibodies or ligands (Hussein, Lisowska et al. 2015).

1.5.3 Thermophilic *Bacilli* for whole-cell applications and genetic engineering

Besides the incorporation of their discrete enzymes, *Geobacilli* and related bacteria, are also used as whole-cell catalysts for various applications such as biorefinement of linen fibers, bioremediation of environmental pollutants and production of fuel and chemicals (Hussein, Lisowska et al. 2015). Especially for the latter case, emphasis has been given towards the production of genetically engineered strains with increased yields for ethanol and other organic compounds or enhanced hydrolytic potential.

Table 1.5.1: A selection of U.S. Patents for *Geobacillus*[*] products or processes (Zeigler 2001)

Product or Process	Patent No.
α-arabinofuranosidase	US05434071
acetate kinase	US05610045
alpha-amylase	US05824532, US05849549
arabino furanoside	US05491087
biological indicator for sterilization	US05073488, US05223401, US05252484, US05418167
BsrFI restriction endonuclease	US06066487
Catalase	US06022721
cellobiose fermentation	US06102690
DNA polymerase	US05747298, US05830714, US05834253,

	US05874282, US06013451, US06066483, US06100078, US06238905
ethanol production	US05182199
glucose-6-phosphate dehydrogenase	US04331762
liquefying starch	US05756714
maleate dehydrogenase	US04331762
neutral proteases	US06103512
perillyl compounds	US05487988
polynucleotide phosphorylase	US04331762
prenyl diphosphate synthase	US06225096
pyruvate kinase	US04331762
riboflavin glucoside	US06190888
superoxide dismutase	US05772996
Xylanase	US05434071
Xylosidase	US05489526

*The information of this Table was collected in 2001, consequently many of the strains considered as *Geobacillus* these days, today might have been reclassified to another related thermophilic genus.

The genetic modifications, include the introduction of autonomously replicating vectors carrying enzymes like a pyruvate decarboxylase (Van Zyl, Taylor et al. 2014), a cellulase and an α-amylase (Suzuki, Yoshida et al. 2013). Additionally, for the optimization of the expression levels, different regulatory elements have been examined, including both native and synthetic promoters (Suzuki, Yoshida et al. 2013, Lin, Rabe et al. 2014, Pogrebnyakov, Jendresen et al. 2017) and also ribosome binding sites (Pogrebnyakov, Jendresen et al. 2017).

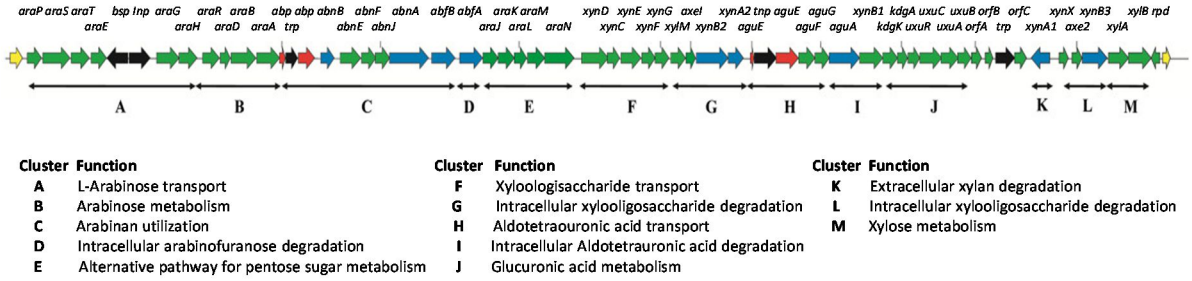
In addition to the above, a few attempts of interfering with the chromosomal DNA have also been accomplished; Cripps et al. (2009) were the first to construct a double - knockout strain of *Geobacillus* for enhanced ethanol production (Cripps, Eley et al. 2009). Through homologous recombination, they inactivated the two enzymes responsible for the production of lactate and acetate (two co-products produced along with ethanol during fermentation) delivering a strain with ethanol as the only fermentation product. Some years later, Suzuki and his colleagues were the first to develop a counter-selection system for *Geobacillus kaustophilus* HTA426 to introduce marker-free genes into the chromosomal DNA (Suzuki, Murakami et al. 2012).This technique utilizes the toxicity against 5-fluoroorotic acid conveyed by the pyrF gene product to select for integration events. However, the creation of a double-knockout parent strain, which finally results in auxotrophic mutants, make this approach impractical for commercial strain production. Recently, Bacon and his colleagues proposed an alternative counter-selection system based on kanamycin resistance and the fact that increased concentration of X-Glu leads to a considerable reduction in the size of colonies (Bacon, Hamley-Bennett et al. 2017). The final mutant is not dependent on any auxotrophy or antibiotic thus making this approach more favorable for industrial purposes.

Research in the field is still ongoing and intensive with many limitations yet to be overcome. For example, the existing transformation protocols are established for a few, not commercial or non-sequenced strains. Furthermore, the available transformation protocols are considered very laborious and with low transformation efficiencies. Finally, even though there are available plasmids and *E. coli* - *Geobacillus* shuttle vectors, more suitable ones for a broader range of strains and applications must be developed (Kananaviciute and Citavicius 2015).

1.5.4 Genomic data and hydrocarbon utilization by thermophilic *Bacilli*

G. stearothermophilus T-6, was the first and until to date the best characterized hemicellulolytic *Geobacillus* representative. For this particular strain, the molecular determinants for hemicelluloluse degradation and utilization are clustered together in a 39.7 kb-long genetic locus (Figure 1.5.1). This, can be further subdivided into thirteen gene clusters on the basis of their predicted function; Clusters A to E are responsible for arabinan degradation and utilization while clusters F to M are responsible for the same processes on xylan (De Maayer, Brumm et al. 2014).

Y. Shoham and his group systematically investigated the regulation of expression of the majority of these clusters in *G stearothermophilus* T-6. In particular, they characterized the clusters A to D, containing the L-arabinan degradation and utilization systems (Shulami, Raz-Pasteur et al. 2011), cluster F with the xylo-oligosaccharide transporters (Shulami, Zaide et al. 2007), cluster K comprised of the extracellular xylanase XynA1 and cluster J that includes the glucuronic acid utilization system of the strain (Shulami, Gat et al. 1999).Through the above studies, the tight regulation of the hemicellulose system was revealed, including several mechanisms like induction-repression, catabolite repression, transition phase regulation and quorum sensing at low cell densities (Gat, Lapidot et al. 1994, Shulami, Gat et al. 1999, Shulami, Zaide et al. 2007).



Cluster	Function
A	L-Arabinose transport
B	Arabinose metabolism
C	Arabinan utilization
D	Intracellular arabinofuranose degradation
E	Alternative pathway for pentose sugar metabolism

Cluster	Function
F	Xyloologisaccharide transport
G	Intracellular xylooligosaccharide degradation
H	Aldotetraouronic acid transport
I	Intracellular Aldotetrauronic acid degradation
J	Glucuronic acid metabolism

Cluster	Function
K	Extracellular xylan degradation
L	Intracellular xylooligosaccharide degradation
M	Xylose metabolism

Figure 1.5.1. Schematic diagram of the *G. stearothermophilus* T-6 hemicellulose utilization locus. Each arrow represents a gene in the locus, with genes encoding predicted transposons colored in black, while open reading frames interrupted by transposons are colored in red. Genes encoding glycosyl hydrolases are colored in blue. The *G. stearothermophilus* T-6 hemicellulose utilization locus can be subdivided into thirteen gene clusters on the basis of their predicted function (De Maayer, Brumm et al. 2014).

For the regulation of xylan's degradation and utilization in particular, Shoham's group proposed that *Geobacillus* cells, initially sense the presence of xylan in their environment detecting minute amounts of xylose, which presumably exist as a by-product of xylan hydrolysis by neighboring microorganisms. Xylose is most likely detected by a two-component system, comprised of a

class I histidine kinase xylose-sensor protein, XynD, and the response regulator XynC (Figure 1.5.1). In the presence of extracellular xylose, XynD phosphorylates XynC, and as a result, the phosphorylated form activates the expression of a dedicated xylo-oligosaccharide ABC transporter XynEFG (Shulami, Zaide et al. 2007). The facilitated entrance of xylosaccharides into the cell and their subsequent hydrolysis to xylose, allows the induction of the system by xylose via inactivation of the XylR repressor. In the absence of xylose, XylR, which is not a part of the hemicellulolytic gene cluster, specifically binds to the *xynDC* promoter region and to five additional promoters; the promoter regions of the extracellular xylanase XynA the uncharacterized xylose epimerase (XynB), the acetyl xylan esterase (Axe1), the intracellular xylanase (XynA2), the *xyn*X-*axe*2-*xynB* operon, the xylose kinase, and the xylose isomerase operon (*xylAB*) as well as its own gene (*xylR*) (Shulami, Shenker et al. 2014).

Opposed to xylose induction, the xylanolytic system of *G. stearothermophilus* T-6 is repressed by the presence of glucose, cassamino acids and generally nutrients' excess (Shulami, Shenker et al. 2014). Additional layers of regulation are quorum sensing, since it has been observed that the expression of XynA is induced in the exponential phase of growth and also the uncharacterised protein XynX which seems to play a repressing role on XynA (Shulami, Shenker et al. 2014).

Comparative analysis of the hemicellulolytic utilization cluster of *G. stearothermophilus* T-6 with the corresponding sections of the available genomes at 2014, revealed that this system is significantly conserved among the genus members. De Maayer and his colleages showed that the hemicellulose utilization locus described for *G. stearothermophilus* T-6, is also present in the complete or partial genomes of 17 out of 24 *Geobacillus* strains examined (De Maayer, Brumm et al. 2014). However, among the 17 positive strains, only 7 out of 41-68 proteins were conserved among all strains. In the same line, another study showed that the majority of the hemicellulolytic genes of Geobacilli are rather part of the flexible than the core gene pool (Bezuidt, Pierneef et al. 2016). All the above, translate into extensive differences in the hydrolytic and utilization strategies employed by *Geobacillus* spp. for hemicellulose polymers (De Maayer, Brumm et al. 2014) and might be linked with the presence of specific *Geobacillus* spp. to specific environmental niches (Bezuidt, Pierneef et al. 2016).

The sequenced genomes of *Geobacillus* spp. and representatives of related genera are exponentially increased (Figure 1.5.2) but still the genomic information especially for the newly characterized genera of *Aeribacillus* and *Aneurinibacillus* is quite limited. Thus, the information derived from newly sequenced genomes reveals the existence of novel enzymes like glycosyl hydrolases, carbohydrate binding domains and auxiliary activities giving improved insights into the ecological adaptations and also the biotechnological potential of these organisms (Bezuidt, Pierneef et al. 2016).

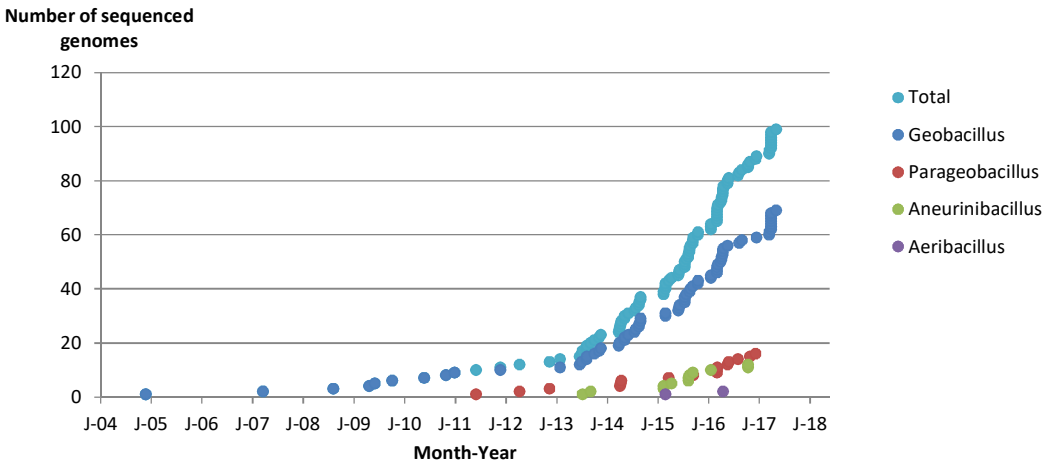


Figure 1.5.2: Sequenced genomes of *Geobacillus spp.*, *Parageobacillus spp., Aneurinibacillus spp.* and *Aeribacillus spp.*deposided in the genomic database of NCBI until January of 2017.

1.6 Aim of the doctorate thesis

The present thesis, aimed in the exploitation of selected thermophilic bacterial strains isolated from the volcanic habitat of Santorini. Our goal was the incorporation of these strains and their hydrolases to plant biomass degradation applications.

In particular, the present work can be summarized in the following four research objectives;

1. The first aim of the present thesis was the detection, production and characterization of the enzymes responsible for plant biomass degradation.
2. The second aim, was the incorporation of representative *Geobacillus* enzymes from the first part of the thesis to designer cellulosomes, leading to multiprotein complexes suitable for use in high-temperature biorefinery applications.

3. In parallel, we managed to isolate a few strains presenting cellulolytic activity. Since this trait is not well documented in the genus, but at the same time it would be of great biotechnological interest, in the framework of the present thesis, the genome of the cellulolytic strain '*G. icigianus* SP50' was sequenced and analysed in order to investigate the genomic information responsible for this phenotype in *Geobacilli.*.
4. Finally, since cellulolytic activity is of great biotechnological interest, the incorporation and expression of a cellulase by the strain *G. kaustophilus* T-1 was also investigated.

2. Materials and Methods

2.1 GENETIC SCREENING OF 10 THERMOPHILIC STRAINS FOR CELLULOLYTIC AND XYLANOLYTIC GENES

2.1.1 Strains, media and DNA extraction

10 bacterial thermophilic strains were selected according to their relatively high biomass degrading activity or their phylogenetic identity (Stathopoulou, Galanopoulou et al. 2012, Stathopoulou, Savvides et al. 2013). All strains are deposited in ATHUB culture collections (WDCM reg.# 1094). For DNA extractions, all bacteria were grown aerobically overnight in Nutrient Broth (PanReac-AppliChem, Darmstadt, Germany) at 60 °C. DNA was extracted with a modified CTAB protocol (http://jgi.doe.gov/wp-content/uploads/2014/02/JGI-Bacterial-DNA-isolation-CTAB-Protocol-2012.pdf).

2.1.2 Phylogenetic analysis of 16s RNA

The 16S ribosomal sequences of the 10 selected strains were amplified with the universal primers pA and p1492 (Lane 1991) and PCR conditions described elsewhere (Stathopoulou, Savvides et al. 2013). In order to classify our strains to alreadty described otus, addiotional 16S rRNA sequences were obtained from the 16s ribosomal sequences database, NCBI (|NR_026515.1 |*A.pallidus*_DSM3670, |NR_115287.1| *G. thermodenitrificans*_BGSC94A1, |NR_025109.1| *G. subterraneus*_34, |NR_115286.1| *G. thermoleovorans*_BGSC96A1. |NR_115285.1| *G. kaustophilus*_BGSC90A1, |NR_074989.1| *G. kaustophilus*_HTA426, |NR_074931.1| *G. thermoleovorans*_CCB_US3_UF5, NR_102795.1| *G.thermoglucosidasius*_C56-YS93, |NR_115284.1| *G.stearothermophilus*_BGSC9A20, |NR_028708.1 | *G.caldoxylosilyticus*_S1812, |NR_112216.1| *A.thermoaerophilus*_DSM10154).

Sequences were aligned using the default settings for non-coding sequences of MUSCLE implemented in MEGA6 software (Tamura, Stecher et al. 2013). Neighbor-joining phylogenetic trees based on the 16S rRNA gene sequences, were calculated by maximum –likelihood method and 100 bootstrap replicates. The parameters used were optimized by MEGA6 based on our dataset (Tamura, Stecher et al. 2013). Tree visualization was conducted with iTOL (Letunic and Bork 2016). The sequences used were 1376 bp long (positions 93 to 1460 of the sequence GK_RS14960 16S ribosomal RNA from *G. kaustophilus* HTA426). Only clades with bootstrap values equal to, or greater of 60% are shown.

2.1.3 Degenerate primers and PCR conditions for the detection of the cellulolytic and xylanolytic genes

For the detection of the xylanolytic and cellulolytic genes, degenerate primers were designed (Table 2.1.1) using the following general procedure:

1. Amino acid sequences of characterized enzymes from *Geobacillus* strains, or Bacillaceae representatives in the cases that not enough information from *Geobacillus* were available, were collected from CAZy database (Lombard, Golaconda Ramulu et al. 2014) and SWISS-PROT database (Bairoch and Apweiler 2000). These sequences belong to the same GH family and share the same enzymatic activity (EC). In some cases, additional sequences of putative genes were also added (Appendix I, Tables 10 to 13).
2. Using SignalP 4.1 server (Petersen, Brunak et al. 2011), the signal peptides of the amino acid sequences were identified and ,if present, they were removed.
3. All the sequences, without the signal peptides, were aligned by MultAlin software (Corpet 1988) in order for the conserved regions to be revealed.
4. Based on the conserved domains, degenerate primers were designed. Whenever needed, inosine was added to increase degeneracy.

The PCR conditions were optimized for every strain individually.

2.1.4 Cloning

After running the PCR product of the degenerate primers in agarose gels, the bands of the expected lengths were purified and cloned into pCR 2.1- topo vectors (Thermo Fischer Scientific) in order to be sequenced. The nucleotide sequences of the inserts were compared with these deposited in the nucleotide collection of NCBI (nr/nt) using BLAST (Altschul, Gish et al. 1990). Based on the sequences with the highest similarity, new primers were designed in order to clone the complete open reading frames replacing the signal peptides with His-Tags (Table 2.1.2).

Table 2.1.1: Degenerate primers for the identification of enzymes of various GH families

GH family	**Enzymatic activity**	**Forward Primer (5'-3')**	Reverse Primer(5'-3')
GH1	6-phospho-*β*-glucosidase (3.2.1.86)	CAARTCGAAGGRGCGKSRAACGAA	AACGWCCAAASAYRGTARCCKWKCACAT
GH5(i)	endoglucanase	TACGTNATHATHGAYTGGCA	GTTCCCCAYTCNSWNACRAA
GH5(ii)	(3.2.1.4)	GCYAAYGARCCRAAYGG	TCYGARTYCCCCAYTC
GH9	endo-1,4-*β*-xylanase (3.2.1.8)	GGIGGITGGYAYGAYGCIGA	SYIGKICGRTGRTGIGGRTG
GH10	endo-1,4-*β*-xylanase (3.2.1.8)	CARCCIGARGARGGIMRITT	GTRTGRTYRTCIGCDATICC
GH11	endo-1,4-*β*-xylanase (3.2.1.8)	ACCGGNAAYTTYGINGTNGTNGG	CCTTCNGTNGGNADNACYTG
GH43	*β*-xylosidase (3.2.1.37)	AWGGRGMRTGGTCKGAICCGATTT	CGYCKSSCRAYAAACGYYKGIGWA
GH48	reducing-end acting cellobiohydrolase (3.2.1.176)	GARGCICCIGAYYAYGGICAY	CCICKYTGIWAIGTRTTIAK
GH52	*β*-xylosidase (3.2.1.37)	TTCTTYAAGCNCAYCAYWSNCC	CCICKYTGIWAIGTRTTIAK

A: adenine, C: cytosine, G: guanine, T: thymine, R: A or G, Y: C or T, S: G or C, W: A or T, K: G or T, M: A or C, B: C or G or T, D: A or G or T, H: A or C or T, V: A or C or G, I: inosine

Table 2.1.2: Primers for the complete open reading frames of hydrolases. With upper case letters are indicated the nucleotides of the gene sequence. With bold lower case letters are indicated the restriction sites

GH family	Enzymatic activity	Forward Primer (5'-3')	Reverse Primer (5'-3')	Vector	Comments
GH1	6-phospho-*β*-glucosidase (3.2.1.86	gatc**catatg**GAGATCGCCATCTGAAACC	ccaat**ggatcc**TACAGCTCGGCACCGTTCG	pET15b	pET15b. Cloned with NdeI, BamHI
GH5	endoglucanase (3.2.1.4)	ttc**catatg**ATGGAGCGTACACCAGTGGAA	cta**gctcagc**CTACTACTCTTTGAACAAACGTTT	pET15b	pEt15b. Cloned with NdeI, NheI
GH10 (extracellular)	endo-1,4-*β*-xylanase (3.2.1.8)	tat**ggatcc**TAAAACTGAACAATCATACTCT	ata**ggatcc**TCATCACTTATGATCGATAATAG	pET15b	His-Tag from pET15b.Cloned with BamHI
GH10 (cell associated)	endo-1,4-*β*-xylanase (3.2.1.8)	tat**ggatcc**ATGAACAGCTCCCTCCCCTCC	tat**ggatcc**TTATCAGACACTCACTGCCCTCCA	pET15b	His-Tag from pET15b.Cloned with BamHI
GH39	*β*-xylosidase (3.2.1.37)	tat**ggatcc**GATGAAGGTYGTAAAYGTGCC	tat**ggatcc**TCATGAAGARTACGATGYATTTC	pET15b	His-Tag from pET15b.Cloned with BamHI
GH43	*β*-xylosidase (3.2.1.37)	cccc**ctcgag**SYSAAAATYAAAAAYCCYATYTTAACYGG	cccc**ctcgag**CTACTAYAAYTCTTTGTACAAAAARTAGTCRAARTC	pET15b	His-Tag from pET15b.Cloned with XhoI
GH52	*β*-xylosidase (3.2.1.37)	cga**gatctG**ATGSCAACCAATSTATTTTT	tat**aagctt**TYATTCYCCCTCCCTMTRC	pET32a	His-Tag from pEt32a.Cloned with BglI and HindIII.

A: adenine, C: cytosine, G: guanine, T: thymine, R: A or G, Y: C or T, S: G or C

2.2 CHARACTERIZATION OF THE ISOLATED ENZYMES

2.2.1 Construction of the mutant GH1-*β*-P- glucosidase$_{378}$

For the replacement of the catalytic nucleophile glutamic acid 358 of the polypeptide sequence of the GH1-*β*-P-glucosidase to a neutral glycine, the site directed mutagenesis protocol of Stratagene was used (Stratagene, USA). According to the instructions of the company, the above set of primers and PCR program were used accordingly:

$F_{SDM:}$ 5'- CTG CCG ATT TTG ATT ACA GGC AAC GGC TTA GGT GAG TTC -3'

$R_{SDM:}$ 5'- GAA CTC ACC TAA GCC GTT GCC TGT AAT CAA AAT CGG CAG -3'

(Underlined are indicated the nucleic acids of the site directed mutagenesis)

PCR program:

In. Denaturation:	95 ˚C	30 sec	
Denaturation:	95 ˚C	30 sec	18 cycles
Annealing:	55 ˚C	60 sec	
Extension	68 ˚C	7 min 30 sec	

For the PCR reaction Kapa HiFI polymerase (Kapa Biosystems,USA) was used.

As DNA template the pET15b vector harboring the wild type *b*-P-glucosidase of *Geobacillus sp.* SP24 was used (Table 2.1.2)

The success of site directed mutagenesis was validated by sequencing the plasmid of selected colonies.

2.2.2 Expression and purification of the cloned enzymes

Plasmids harboring the isolated genes (Table 2.1.2) were expressed in *Escherichia coli* BL21 (DE3) (New England Biolabs Ipswitch MA), and the corresponding His-tagged proteins were

purified on a Ni-nitriloacetic acid (NTA) column (Macherey—Nagel, Düren, Germany) according to the manufacturer's protocol. The progress of the purification procedure was validated through SDS-PAGE. Purified protein samples were dialyzed against phosphate buffer and stored at 4 °C filter sterilized or at −20 °C in 50 % (v/v) glycerol.

2.2.3 Production and purification of phosphorylated substrates for the biochemical characterization of the GH1 6-phospho-*β*-glucosidase

6-P-*β*-Cellobiose, 6-P-*β*-gentiobiose, 6-P-*β*-lactose and 6-P-*β*-pnpG were enzymatically synthesized by cellobiose, gentiobise, lactose and pnpG respectively and partially purified with barium as reported previously (Thompson, Lichtenthaler et al. 2002). For the phosphorylation of the disaccharides, the gene of *β*-glucoside kinase (BglK) from *Klesbiella pneumoniae* was cloned into pET15b vector and the enzyme was produced from *E.coli* BL21 cells as described elsewere (Suzuki, Okazaki et al. 2012).

The enrichment and separation of the phosphorylated substrates from the remaining not phosphorylated, were further achieved through anion exchange chromatography with a home-made Q-sepharose column. In the presence of 100mM piperizine pH10, the phosphorylated substrates were selectively bound on the column and consequently elucidated with an 100 mM NaCl solution.

In order to detect the product of the phosphorylation reactions, 1 µl of the appropriate dilution of the reactions were spotted onto silica gel 60 TLC F_{254} plates (Merck) and developed with *n*-butanol/acetic acid/water 2:1:1 (v:v:v). 1 µl of the not phosphorylated substrates as well as glucose and P-glucose in standard concentration (10g/L) were run in parallel as standards. Glycosides were visualized by spaying with orcinol 0.2% (w/v) in H_2SO_4 20% (v/v) followed by heating at 100 ˚C for 10-15 min.

The yield of the reactions as well as the purity of the phosphorylated substrate, in terms of the ratio of the phosphorylated to the not phosphorylated dissacharides, were estimated indirectly through enzymatic reactions against the not phosphorylated di-saccharides: 6-P-*β*-Cellobiose, 6-P-*β*-gentiobiose and 6-P-*β*-pnpG reaction mixtures were incubated overnight with *β*-glucosidase from *C. saccharolyticum* (Paragraphs 2.3.2 and 2.3.3) while 6-P-*β*-lactose with *β*-galactosidase

from *E.coli* (Sigma-Aldrich). Thus, the remaining not phosphorylated disaccharides were completely hydrolysed to glucose. For the calculations of the yield and purity, the total amounts of sugars realized from the samples of the phosphorylated sugars, were compared with the initial glucose equivalents of the disaccharide samples before the reaction of phosphorylation.

2.2.4 Enzyme activity and protein concentration assays

β-Xylosidase was assayed using *p*-nitro-phenyl-*β*-D-xylopyranoside (pNP-X) as a substrate in a final concentration of 2 mM in 100 mM MES buffer adjusted to pH 6.5. Properly diluted enzyme samples were added to a final reaction volume of 150 μL. Following 5 min of incubation at 60 °C, the reaction was terminated by immediately placing the samples into an ice bath and adding 150 μL of 1 M Na_2CO_3. Samples (200 μL) of the above reaction mixture were placed in a microplate and the extent of hydrolysis was estimated by measuring the absorbance of the pNP released at 410 nm.

β-6, P-glucosidase was assayed using *p*-nitro-phenyl-*β*-D-glycopyranoside (P-pNP-G) as well as P-cellobiose as substrates. P-pNP-G was used in a final concentration of 0.65 mM in 100 mM MES, pH 6.5. 25 μl of properly diluted enzyme were added to the reaction mix and the final volume was 350 μl. Following 10 min of incubation at 60 °C, the reaction was terminated by immediately placing the samples into an ice bath and adding equal volume of 1 M Na_2CO_3. The extent of hydrolysis was estimated by measuring the absorbance of the pNP released at 410 nm. P-cellobiose was used in a final concentration of 1.05 mM in 100 mM MES, pH 6.5. 10 μl of properly diluted enzyme was added to the reaction mix and the final volume was 200 μl. After 10 min at 60 °C, the enzymatic activity was estimated indirectly by measuring the realised glucose with the GOD-PAP test (BIOSIS, Greece). Glucose production was estimated using a glucose standard curve.

Xylanase activity was determined using 1% w/v (final conc.) beechwood xylan as substrate in 100 mM MES buffer pH 6.5 at 200 μL total reaction volume (Bailey et al. 1992). Samples were incubated for 10 min at 60 °C in a thermoshaker at 1000 rpm. Following incubation, the reaction was terminated by immediately placing samples into an ice bath and adding 400 μL DNS reagent (Miller 1959). Residual xylan solids were removed by centrifugation (5 min at 13000xg) and the

supernatant was transferred to a new tube and boiled for 5 min. Aliquots (200 μL) were transferred to a 96-well microplate and the absorbance at 540 nm was measured. Xylooligosaccharide production was estimated using a xylose standard curve.

Endo-glucanase activity was estimated in the same way with xylanase but with carboxy methyl cellulose (CMC) at 1.5% w/v (final conc.) as substrate, omission of the centrifugation step (CMC is soluble), and using glucose for the DNS reference curve.

All reactions were conducted at least in triplicate. For all enzyme assays, special attention was paid for proper sample dilutions in order not to exceed 10% of substrate conversion and to ensure linearity between reaction rate and working enzyme concentration. Enzyme activities were expressed in Units (U). I Units is defined as the amount of enzyme necessary to catalyse the production of 1 μmole of product per minute under the above-described pH and temperature conditions.

Protein concentration was estimated both by sample absorbance at 280 nm, based on the known amino acid composition of the proteins (Gasteiger, Hoogland et al. 2005, Moraïs, Barak et al. 2011) and by the Bradford (1976) method.

2.2.5 Biochemical characterization

The effect of temperature on enzyme activity was determined by performing, the appropriate for each enzyme, standard assay described above at the appropriate temperature range. Thermostability was evaluated by incubation of properly diluted enzyme samples at the indicated temperature at pH 6.5 (100 mM MES buffer). Aliquots were withdrawn at specific time intervals and assayed for the corresponding enzyme activity.

pH optima were determined in a similar manner at 60 °C and a buffer range from 4 to 9 (100 mM citrate-phosphate or phosphate buffers). Temperature stability was studied at pH 6 (100 mM phosphate buffer) by incubation of a properly diluted purified enzyme sample at the indicated temperatures. Aliquots were withdrawn at specific time intervals and assayed for enzymatic activity at the regular assay conditions.

Michaelis-Menten kinetics were studied at 60 °C and phosphate buffer 100 mM pH 6 for *β*-xylosidase and MES 100 mM pH 6.5 for the rest enzymes. K_Ms were determined by measuring the initial reaction rates for initial substrates concentrations ranging from 0.1 to 20 mM. For *β*-xylosidase, the competitive inhibition constants, K_Is, were determined at 5, 10, 40 and 80 mM initial concentrations of either xylose or arabinose in the reaction mixture and determination of the corresponding apparent K_M's with pNPX as substrate, as described above. Inhibition by various metal cations was performed at the regular assay mixture using either 1 or 10 mM of the corresponding salt. For *β*-phospho-glucosidase, the inhibitory effect of cellobiose, lactose, glucose, galactose, fructose P-fructose and P-glucose where tested by adding 50 mM of each one in the reaction mixture and using P-pnpG as a substrate. The linear and non-linear regression routines of SigmaPlot software (version 12.0, Systat Software Inc.) were used for data analysis.

Reduced MW determination and purity assessment was performed through standard SDS-PAGE followed by regular Coomasie staining (Blum, Beier et al. 1987).

2.2.6 Preliminary crystallographic analysis of GH1 6-*β*-P-glucosidase and its inactivated mutant 6-*β*- P-glucosidase$_{378}$

The crystallization of the free enzymes was optimized using the vapor diffusion method and commercially available crystallization screens at the robotic facility (crystallization robot: robotOryx Nano, Douglas Instruments, UK) of National Research Foundation, Greece. For the optimization the temperature was kept constant at 19 ˚C, while different pH values from 6.5 to 7.5 at the presence of various concentration of PEG 3350 and PEG 4000 were tested.

2.3 Construction and studies on temperature effect of a designer thermocellulosome

2.3.1 Strains, media and DNA extraction

Geobacillus sp. isolates SP24 and SP50 (Stathopoulou et al. 2012) were used as sources for endo-glucanase Cel5 (SP50), endo-xylanase Xyn10 (SP50) and *β*-xylosidase Xyd52 (SP24). Strains are deposited in ATHUB culture collections (WDCM reg.No 1094) under the above designation. *β*-glucosidase Bgl1 was obtained from *Caldicellulosiruptor saccharolyticus* (DSMZ 8903). For DNA extraction, Geobacilli were grown aerobically overnight in Nutrient Broth

(PanReac-AppliChem, Darmstadt, Germany) at 60 °C while *C. saccharolyticus* was grown for 48 h in DSMZ medium 640 at 70 °C in sealed anaerobic flasks following helium sparging. DNA was extracted with a modified CTAB protocol (http://jgi.doe.gov/wp-content/uploads/2014/02/JGI-Bacterial-DNA-isolation-CTAB-Protocol-2012.pdf). An additional *β*-glucosidase, Bgl1 from *Clostridium thermocellumin* in pET28a vector, was provided by the lab of Professor Ed Bayer.

2.3.2 Cloning

Wild-type Cel5, Xyn10, and Xyd52 genes were amplified and cloned in the pET plasmid vectors as previously described (Table 2.1.2). Bgl1 from *Calidicellulosiruptor saccharolyticum* was cloned into pET15b vector with BamHI and the following primers (With upper case letters are indicated the nucleotides of the gene sequence. With bold lower case letters are indicated the restriction sites):

F_{primer}: tat**ggatcc**gatgAGTTTCCCAAAAGGATT

R_{primer}: cgc**ggatcc**TTATTACGAATTTTCCTTTATATA

The chimaeric enzyme-dockerin constructs were assembled by replacing the hydrolase gene from the corresponding plasmids constructed in previous studies (Moraïs et al. 2010) with the above-mentioned genes (Table 2.3.1). The construction of the recombinant scaffoldin used in the present study (Scaf•BTFA) has been described previously (Moraïs et al. 2010).

2.3.3 Expression and purification

Plasmids harboring the genes for Cel5, Bgl1 *C.th*,Bgl1 *C.s,* Xyn10sp24, Xyn10sp50, Xyd39, Xyd43 Xyd52 as well as the genes for the chimaeric enzyme constructs were expressed in Escherichia coli BL21 (DE3) (New England Biolabs, Ipswitch MA), and the corresponding His-tagged proteins were purified on a Ni-nitriloacetic acid (NTA) column (Macherey – Nagel, Düren, Germany) according to the manufacturer's protocol with the difference of using Tris Buffered Saline (TBS) instead of phosphate buffer. For the scaffoldin plasmid, Scaf•BTFA, an extra purification step, based on the affinity of the CBM3 to phosphoric acid-swollen cellulose (PASC) (Haimovitz et al. 2008), was additionally employed. The progress of the purification

procedure was validated through SDS-PAGE. Purified protein samples were dialyzed against TBS buffer and stored in 50% (v/v) glycerol at -20 °C.

Protein concentration was estimated both by sample absorbance at 280 nm, based on the known amino acid composition of the proteins (Gasteiger et al. 2005; Moraïs et al. 2011), and by the Bradford (1976) method.

2.3.4 Enzyme assays and characterization

β-Glucosidase was estimated using p-nitro-phenyl-*β*-D-glucopyranoside (pNP-G) as substrate at a final concentration of 2 mM in 100 mM MES buffer adjusted to pH 6.5. Properly diluted enzyme samples were added to a final reaction volume of 150 μL. Following 15 min of incubation at 60 °C, the reaction was terminated by immediately placing the samples into an ice bath and adding 150 μL of 1 M Na2CO3. Samples (200 μL) of the above reaction mixture were placed in a microplate and the extent of hydrolysis was estimated by measuring the absorbance of the pNP released at 410 nm. *β*-Xylosidase, xylanase and endoglucanase activitites were assayed as previously described in Paragraph 2.2.5.

All reactions were conducted at least in triplicate. For all enzyme assays, special attention was paid for proper sample dilutions in order not to exceed 10% of substrate conversion and to ensure linearity between reaction rate and working enzyme concentration. Enzyme activities were expressed in Units (U), defined as the amount of enzyme necessary to catalyze the production of 1 μmole of product per minute under the above-described pH and temperature conditions.

Table 2.3.1: Primers for the enzyme-dockerin chimaeras. With upper case letters are indicated the nucleotides of the gene sequence. With bold lower case letters are indicated the restriction sites.

Construct acronym	**Cloning product**	**Forward Primer (5'-3')**	**Reverse Primer (5'-3')**	**Dockerin bearing cloning vector**	**Restriction enzymes used**
Cel5_sp50-docb	Endo glucanase GH5-dockerin b	gt**cat*atg***caccatcaccatcaccat CAGCGTACACCAGTGGAAG	gtg**actagt**CTCTTTGAACAAACGT TTCC	p43-b[(2)] (pET21a)	*Nde*I, *Spe*I
Cel5_sp50-doct	Endoglucanase GH5-dockerin t	tct**gctagc***atg*caccatcaccatcaccat CAGCGTACACCAGTGGAAG	gtg**actagt**CTCTTTGAACAAACGT TTCC	p10B-t[(1)] (pET21a)	*Nhe*I, *Spe*I
Bgl1_cs-doct	*β*-Glycosidase GH1-dockerin t	tct**gctagc***atg*caccatcaccatcaccat AGTTTCCCAAAAGGATTTTTG	at**actagt**CGAATTTTCCTTTATAT ACTGCTG	p10B-t[(1)] (pET21a)	*Nhe*I, *Spe*I
Bgl1_*cth*-doct	*β*-Glycosidase GH1-dockerin t	tct**gctagc**ATGCACCATCACCATCACCA TTCAAAGATAACTTTCCCAAAAG	ata**actagt**AAAACCGTTGTTTTTGA TTACTTC	p10B-t[(1)] (pET21a)	*Nhe*I, *Spe*I
Xyn10_sp24-doct	Xylanase GH10-dockerin t	tct**gctagc**TAAAACTGAACA ATCATACTCT	ata**actagt**CTTATGATCGATAATAG	p10B-t[(1)] (pET21a)	*Nhe*I, *Spe*I
Cel5_sp50-docf	Endoglucanase GH5-dockerin f	tct**gctagc***atg*caccatcaccatcaccat CAGCGTACACCAGTGGAAG	gtg**actagt**CTCTTTGAACAAACGT TTCC	p10A-f[(2)] (pET21a)	*Nhe*I, *Spe*I
Xyn10_sp50-docf	Xylanase GH10-dockerin f	cta**gctagc**CATCATCATCATCA TCACAGC	cc**actagt**cTTATGATCGATAATA GCCCAAT	p10A-f[(2)] (pET21a)	*Nhe*I, *Spe*I

Xyd39_sp24-doca	*β*-Xylosidase dockerin a	GH39-	ctt**gctagc**GATGAAGGTYGTAAAYGTGCC	ctg**gagctc**TGAAGARTACGATGYATTTC		*Nhe*I, *Sac*I
Xyd43_sp24-doca	*β*-Xylosidase dockerin a	GH43-	ctt**gctagc**ATGCACCATCACCATCACCATGCGAAAATTAAAAATCCTAT	ctg**gagctc**TAATTCTTTGTACAAAAAGTAG		*Nhe*I, *Sac*I
Xyd52_sp24-doca	*β*-Xylosidase dockerin a	GH52-	ctt**gctagc***atg*caccatcaccatcaccatCCAACCAATCTATTTTTCAAC	ctg**gagctc**YTCYCCCTCCTCMARCCAYAAAATRC	pXyn11XBM-a[(1)] (pET21a)	*Nhe*I, *Sac*I

A: adenine, C: cytosine, G: guanine, T: thymine, R: A or G, Y: C or T, S: G or C

(1) obtained from (Moraïs, Barak et al. 2010), **(2)** obtained from (Moraïs, Barak et al. 2011)

The effect of temperature on wild-type and chimaeric enzyme activity was determined by performing the standard assay described above at various temperatures between 40 and 75°C. Thermostability was evaluated by incubation of properly diluted enzyme samples at the indicated temperature at pH 6.5 (100 mM MES buffer). Aliquots were withdrawn at specific time intervals and assayed for the corresponding enzyme activity.

2.3.5 Cellulosome assembly

Equimolar amounts of the different proteins constitutive of the designer cellulosome were prepared and incubated for 2 h at 37 °C in TBS buffer pH 7.4 supplemented with 5 mM CaCl2 and 0.0025 % (w/v) Tween 20 (final conc.). The electrophoretic mobility of the proteins was analyzed by SDS and non-denaturing PAGE (Vazana et al. 2012) with the difference that in the latter, both upper and lower gels as well as running buffer were set to pH 8.5.

2.3.6 Thermostability of scaffoldin complexed with individual enzymes

The structural stability of the complex at different temperatures was assessed for the complex of samples were incubated at various temperatures (100 mM MES pH 6.5) and equal volume aliquots were withdrawn at different time intervals. The structural integrity of the complex was assessed in all sample aliquots by SDS and non-denaturing PAGE as described above. Band density in PAGE gels were quantified using ImageJ freeware(Schneider, Rasband et al. 2012).

2.3.7 Determination of the stability and functionality of CBM

The effect of temperature on the stability and functionality of the scaffoldin's CBM was examined using the complex of Scaf•BTFA with the chimaeric enzyme Xyn10sp50-f as a model. An affinity pull-down assay using microcrystaline cellulose (Avicel) combined with the determination of xylanase activity was employed as follows: Following complex formation, the Scaf•BTFA-Xyn10sp50-f cellulosome samples were incubated at different temperatures. For each temperature, equal-volume aliquots were removed at various time intervals and placed in 1.5 mL Eppendorf tubes containing TBS buffer pH 7.4 supplemented with 5 mM CaCl2, 0.0025 % w/v Tween 20 and Avicel (Sigma-Aldrich, Taufkirchen, Germany) at a final concentration of 10% w/v. The resulting suspensions were gently stirred with mini magnetic stirrers for 2 h at room temperature to allow the CBM of the cellulosome to bind to Avicel (Vazana et al. 2012).

Then, the supernatant fluids containing all unbound protein molecules, were collected following a 10-min centrifugation step at 12000 rpm. Precipitated Avicel was additionally washed with 0.5 mL of 100 mM TBS, in order to remove the remaining free and weakly bound components. Xylanase activity determination was subsequently performed, in order to determine the percentages of bound cellulosome (stable and functional CBM) and unbound cellulosomal components (denatured CBM) at each condition. The xylanase activity of the supernatant fluids washings and Avicel was estimated with the standard assay procedure. For Avicel-associated activity, the assay was performed by adding the substrate solution (xylan) directly on the Avicel pellet after washing. The entire procedure was repeated with the same concentrations of the free chimaeric enzyme Xyn10.sp50-f as control.

2.3.8 Thermostability of the tetravalent complex

The structural stability of the tetravalent complex was examined at 60 °C using gel permeation chromatography (GPC). For this purpose a custom made column of high length-to-internal diameter ratio was employed packed with Sephacryl S-200 resin (GE Healthcare, Piscataway, NJ). The resin was packed in a 9.5-cm internal diameter glass column between two Econo-Flow adaptors (BIORAD, München, Germany Cat# 738-0015) in order to form a bed of 95 cm length. TBS at a flowrate of 0.5 mL/min was used for elution using a benchtop peristaltic pump (ISMATEC, Wertheim, Germany). The void volume was determined using Blue Dextran (MW>2000 kDa), and the column was calibrated with the following standard proteins; *β*-amylase (200 kDa), alcohol dehydrogenase (150 kDa), albumin (66 kDa), carbonic anhydrase (29 kDa) and cytochrome c (12.4 kDa) (Sigma-Aldrich). Following formation of the tetravalent complex, the sample (500 μL) was incubated at 60 °C. At 2, 6 and 24 h, aliquots of 100 μl were withdrawn and introduced into the column. The A280 of the eluate was monitored with a Pharmacia LKB Uvicord-SII UV Detector. The signal from the detector was recorded through an A/D converter (LinkLab Ltd, Athens, Greece) coupled to a data acquisition platform based on DAZYLab software (MCC, Norton, MA, USA).

2.3.9 Hydrolytic performance of the cellulosome

The hydrolytic performance of the tetravalent cellulosome was evaluated on corn stover, a major agricultural lignocellulose byproduct. The material was subjected to size reduction (< 0.8 mm) in

a laboratory mill followed by mild alkali pretreatment (0.1 g NaOH per g dry biomass at 90 °C for 3 h). The cellulose - hemicellulose - lignin content (TAPPI analysis) of the neutralized and dried pretreated biomass samples was 50 - 37 - 11 w/w, respectively (max SD ± 17%). Equimolar amounts (1 nmole) of scaffoldin and each of the four dockerin-bearing enzymes were used for complex formation, and the resulting designer cellulosome was incubated at 60 °C, up to 24 h in 1 mL of 100 mM MES buffer pH 6.5, containing 20 mg of dry pretreated biomass (thermoshaker at 800 rpm). In parallel, the hydrolysis was conducted with the same amounts of the free chimaeric and wild-type enzymes under the same conditions. The progress of the hydrolysis reaction was followed by monitoring the release of reducing sugars using the DNS method.

2.4 Growth tests and genome analysis of '*Geobacillus icigianus* SP50'

2.4.1 Growth of '*Geobacillus icigianus* SP50' on different carbon sources

Modified LB medium (mLB) was used for *'Geobacillus icigianus* SP50*'* standard growth (Chen, Wojcik et al. 1986). The growth ability of *G. stearothermophilus* T-1 in different carbon sources, was tested in 0.6 mL batch cultures in 1.5 mL Eppendorf tubes at 60 ˚C with orbital shaking at 900 rpm. The medium used was based on MG medium (Zeigler 2001) by replacing glucose with the indicated carbon source. An overnight *'Geobacillus icigianus* SP50*'* culture grown in mLB agar was used as inoculum and suspended in the appropriate volume of MG medium.Three independent cultures were conducted for each carbon source with an initial OD of 0.1.

2.4.2 DNA extraction and draft genome assembly

DNA extraction of '*Geobacillus icigianus* SP50', was conducted according to the standard CTAB protocol of the DOE Joint Genome Institute (William, Feil et al. 2012). Genome sequencing was outsourced to MR DNA (Texas, USA) as follows: The initial concentration of DNA was evaluated using the Qubit® dsDNA HS Assay Kit (Life Technologies). The sample was appropriately diluted to achieve the recommended DNA input of 50 ng at a concentration of 2.5 ng/μL. The library was prepared using Nextera DNA Sample preparation kit (Illumina) following manufacturer's recommendations. The sample underwent the simultaneous fragmentation and addition of adapter sequences during a limited-cycle (5 cycles) PCR in which

unique indices were added to the samples. Following library preparation, the final concentration of the DNA was measured using the Qubit® dsDNA HS Assay Kit (Life Technologies), and the average library size was determined using the Agilent 2100 Bioanalyzer (Agilent Technologies). The library was diluted (to 10.0 pM) and sequenced paired end for 300 cycles using the HiSeq 2500 system (Illumina). The quality control of the contigs, and the putative transcriptome and proteome were predicted through MiGA website (Rodriguez-R, Gunturu et al. 2018).

2.4.3 Functional classification of the *'Geobacillus icigianus* SP50' proteome and detection of the carbohydrate-active enzymes (CAZymes)

Functional annotation of the predicted protein-coding genes was conducted through the Rapid Annotation using Subsystem Technology (RAST) server (http://rast.nmpdr.org) (Aziz, Bartels et al. 2008). Additionally, the CAZome of *'Geobacillus icigianus* SP50*'* was identified by running local HMMER v3.1b2 (http://hmmer.org/) against dbCAN database (release 6, last update 9 Dec 2017) (Yin, Mao et al. 2012) and analyzed as follows: All alignments with a signature domain coverage greater than 0.35 and (1) an E-value lower than $1x10^{-5}$ for sequences greater than 80 aa or (2) alignments of smaller length with E-value < $1x10^{-3}$ were collected. From these, all queries with an E-value lower than $1x10^{-18}$ and signature domain coverage greater than 0.35 were accepted as true positives and were assigned to the CAZy family of the subject sequence. The remaining sequences were further inspected by searching against the entire Uniprot/SwissProt (Consortium 2017) and NCBI's conserved domain database (CDD) (Marchler-Bauer, Derbyshire et al. 2015).

2.4.4 Taxonomic classification based on 16s ribosomal RNA sequences

The classification of *'Geobacillus icigianus* SP50*'* up to the genus level, was held by assigning all rRNA sequences found in the genome of '*Geobacillus icigianus* SP50' to the RDP naive Bayesian Classifier (Wang, Garrity et al. 2007).

2.4.5 Phylogenetic analysis based on recN, rpoB and spo0A DNA sequences

The DNA sequences of *recN*, *rpoB* and *spo0A* were retrieved by local BLASTn from the genome of '*Geobacillus icigianus* SP50' and all 69 *Geobacillus* genomes retrieved from the GenBank assembly database (Kitts, Church et al. 2016) (Appendix II, Table 1). The DNA sequences of the

Uniprot KB entries Q75TB2 (*spo0A*), Q5L405 (*rpoB*) and Q75TB4 (*recN*) were used as queries for the blast search. Five out of the 70 genomes (*Geobacillus* sp. WCH70, *Geobacillus* sp. Y4.1MC1, *Geobacillus* sp. 44B, *Geobacillus* sp. 44C and *Geobacillus galactosidasius* strain DSM 187511) did not have all three selected markers and thus excluded from the analysis. Additionally, the corresponding sequences of *Parageobacillus thermoglucosidasius* DSM2542 were used as outgroup. Mafft v.7(auto settings) (Katoh and Standley 2013) was used to align the individual DNA sequences of the selected gene-markers from each genome and then poorly aligned regions were removed using the default settings in Gblocks v0.91b (Talavera and Castresana 2007). All alignments were concatenated to produce a final alignment of 66 taxa of 5718 nt each. MEGA 6 (Hall 2013, Tamura, Stecher et al. 2013) was used to produce a Neighbor-Joining phylogenetic tree with the maximum likelihood statistic method (substitution model GTR+G=I) and 100 bootstrap replicates.

2.4.6 Phylogenetic metric calculations

Average Nucleotide Identity (ANI) and Average Amino acid Identity (AAI) values of pairwise comparisons were computed with ANI/AAI matrix tool through enveomics collection (Rodriguez-R and Konstantinidis 2016) for all 71 genomes described in the previous paragraph.

2.4.7 Screening for glycoside hydrolases with putative *β-D*-glucanase activity in *Geobacillus* pangenome

All putative protein sequences of the 70 *Geobacillus* genomes, were compared with all the GH families known to include the activities below; cellulase (E.C. 3.2.1.4), cellulose-1,4-*β*-cellobiosidase (non-reducing end) (3.2.1.91), cellulose-1,4-*β*-cellobiosidase (reducing end) (3.2.1.76), licheninase (3.2.1.73) and glucan endo-1,3-*β*-D-glucosidase (3.2.1.39). These families, 20 in total, were identified by performing local HMMER v3.1b2 (http://hmmer.org/) against dbCAN database (release 6, last update 9 Dec 2017) (Yin, Mao et al. 2012) and analyzed as described in paragraph 2.5.3. For all positive hits, the closest characterized enzymes were identified by running local blastp BLAST+2.6.0 against dbCAN database (release 6, last update 9 Dec 2017). Especially for the family GH5, blastp was performed against the sequences of DBCAN enriched with the sequence of a novel GH5 hydrolase from another *Geobacillus* sp.

characterized by our lab as an arabinoxylan-specific endo *β*-1,4-xylanase (3.2.1.-) with no detectable activity on CMC (unpublished data).

2.4.8 Screening for the *β-D*-Glucan utilization locus of '*Geobacillus icigianus* SP50'

The DNA sequences of the five genes of the *β*-glucan utilization locus of '*Geobacillus icigianus* SP50' were used as queries to conduct blastn megablast of the BLAST+ v.2.6.0 package against (1) the RefSeq chromosome records for non-human organisms blast database (other_genomic), screened May 2018 (Coordinators 2016) and also (2) the local database consisting of all 70 *Geobacillus* genomes described in paragraph 2.5.5. All matches with an e-value ≤ 0.01 and sequence coverage, equal with, or greater than the half of the length of the queries, were considered as positive hits.

2.4.9 Genome analysis of the *β*-D-Glucan utilization locus

Potential promoter regions and transcription terminators were predicted using the Softberry programs BProm (Solovyev V. 2011).

Alignments of the *β*-glucan utilization locus and the surrounding region of '*G. icigianus* SP50' were conducted with (1) the phylogenetically closest *Geobacillus* strains as well as with (2) representative genomes containing the closest homologues of the *β-Glucan* utilization locus of '*G.icigianus* SP50' as found in Paragraph 2.5.8. In order to obtain higher quality genomes of the selected strains, the draft genomes were assembled to scaffoldins by alignment against the closely related complete or higher quality draft genomes, using the Multi-Draft based Scaffolder (MeDuSa) (Bosi, Donati et al. 2015). The assembled scaffoldins were subsequently annotated through the RAST annotation server (Aziz, Bartels et al. 2008, Overbeek, Olson et al. 2014) and finally aligned using Mauve (Darling, Mau et al. 2004). Easyfig was used to visualize the linear comparisons among the genomes (Sullivan, Petty et al. 2011).

2.5 Heterologous expression of the cellulase Cel6A-*T.f* in *G. stearothermophilus* T-1

2.5.1 Cellulase Cel6A-*T.f*; Production and biochemical characterisation

The Cel6A cellulase gene of *Thermobifida fusca* strain YX, harboured in the pET28a vector, was kindly provided by the lab of Professor Edward Bayer, from Weizmann Institute of Sciences, Rehovot, Israel. The expression, purification, activity and protein estimation as well as the assays followed for the biochemical characterisation of Cel6*A* were conducted according to Paragraphs 2.2.2, 2.2.4 and 2.2.5, respectively.

2.5.2 *G. stearothermophilus* T-1 growth conditions

Modified LB medium (mLB) was used for *Geobacillus* standard growth (Chen, Wojcik et al. 1986). In order to examine the growth ability of *G. stearothermophilus* T-1 in different carbon sources, 0.6 mL batch cultures in 1.5 mL Eppendorf tubes at 60 ˚C, with orbital shaking at 900 rpm, were conducted. The medium used was based on MG medium (Zeigler 2001) by replacing glucose with the indicated carbon source. An overnight *G. stearothermophilus* T-1 culture grown in mLB agar, suspended to the appropriate volume of MG medium, was used as inoculum. Three independent cultures were conducted for each carbon source at an initial OD_{600} of 0.1.

2.5.3 Plasmid construction

All the constructs used in the present study are derivatives of the *Geobacillus – E. coli* shuttle vector pNW35N (Shulami, Shenker et al. 2014) (Figure 2.5.1). pNW35Ni was synthesized to contain two 500 bp regions of *G. stearothermophilus* T-1 and the operator and terminator of *xynA of G. stearothermophilus T-1* (*xynA*-T-1) flanking the cellulase *cel6*A* gene sequence (Genescript®, China). Cel6*A**, was the *cel6A* of *Thermobifida fusca* with codon optimization for *Geobacillus stearothermophilus*, designed to contain the signal peptide of *xynA*-T-1 instead of its native one. For pNW35Nii construction, the sequence of *cel6A** with the corresponding regulatory elements, was amplified using as DNA template the plasmid PNW35Ni and subsequently was cloned into the pNW35N vector between *BamHI* and *Eco*RI sites. pNW35Niii was designed to be the same like pNW35Ni but with two homologous regions of 1100 bp instead of two regions of 500 bp. For the construction of pNW35Niii, an intermediate construct pNW35N(iii), consisting of the two homologous regions, was initially designed. Then, using a restriction free method, the Cel6A was incorporated into the desired position. Briefly for the restriction free method, two complement pairs of primers were designed, both containing parts of the vector (pNW35N(iii)) as well as the insert (*cel6A**) (Table 2.5.1). After amplification, the

PCR products were digested for 5 h with *Dpn*I (Thermo Fischer Scientific, USA). Aliquots of these two were mixed together with ligase in order for the desired plasmid product to occur. For all the transformations *E.coli* DH5a cells (Novagen) were used grown in LB supplemented with 1% of glucose and the appropriate antibiotic. All PCRs were performed by using High-Fidelity DNA Polymerase (New England Biolabs, USA) with the exception of the PCRs for the restriction free method which were performed using Kapa Hi-Fi Hot Start polymerase (Kapa Biosystems, USA). All primers used, are given in Table 2.5.1.

2.5.4 Protoplast transformation of *G. stearothermophilus* T-1

Protoplast transformation of *G. stearothermophilus* T-1 with plasmid DNA, was based on the protocol described by Wu and Welker (Wu and Welker 1989) with all modifications described at (Shulami, Shenker et al. 2014). Chloramphenicol 6 μg/mL was used as the selective antibiotic.

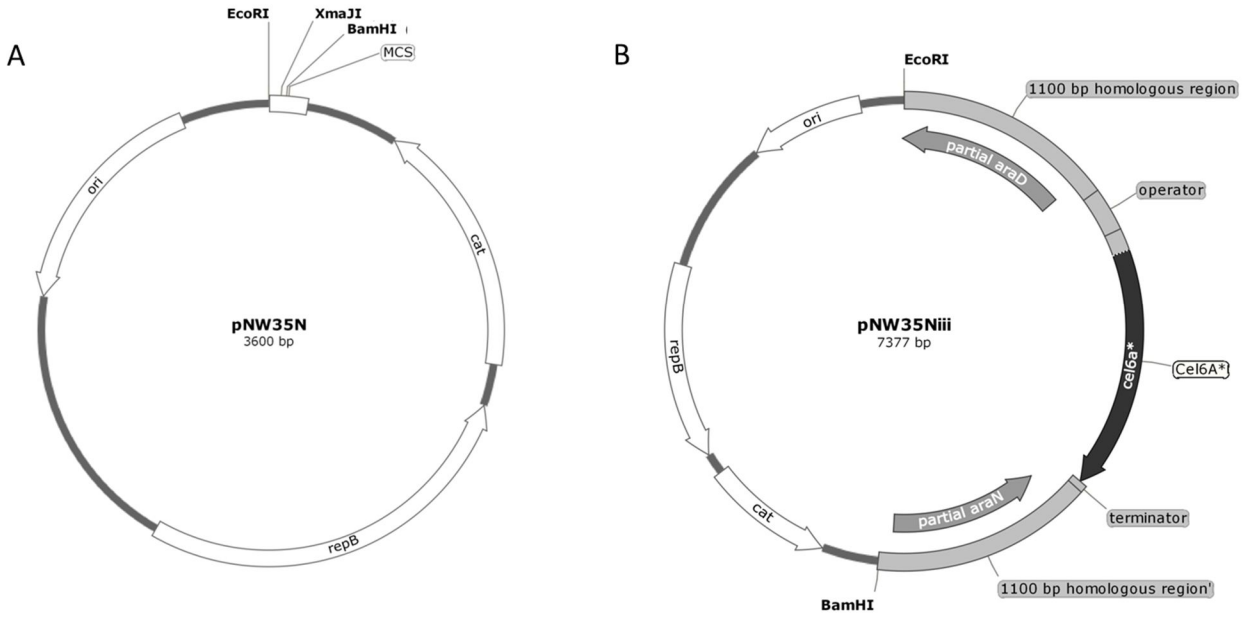


Figure 2.5.1: *Geobacillus – E. coli* shuttle vectors pNW35N (A) and pNW35Niii (B). White arrows represent the genes of pNW35N vector; the gene responsible for chloramphenicol resistance, chloramphenicol acetyltransferase, from *Staphylococcus aureus* expressed in both *E. coli* and *Geobacillus* strains (*cat*), the gene for the replication-initiation protein which enables replication in *Geobacillus* strains (*repB)* and the origin of replication of *E. coli* (*ori*). On pNW35Niii, the grey arrows indicate the DNA sequences derived from *G. stearothermophilus* T-1 strain; the two homologous 1100 bp regions located next to the desired site of insertion, the operator and terminator of XynA-T-1 and the

signal peptide of XynA-T-1 which is a part of Cel6a* gene. The dark grey arrow represents the gene of cellulase Cel6A upon codon optimization for *G. stearothermophilus* T-1 (Cel6a*).

2.5.5 DNA isolation and manipulation

Genomic DNA from *G. stearothermophilus* T-1 was isolated using the TIANamp Bacteria DNA kit (TIANGEN, China). Plasmid DNA from the transformed *G. stearothermophilus* T-1 strains was purified using the QIAprep® Spin Miniprep Kit (Qiagen, Germany) according to the protocol described for plasmid DNA from *Bacillus subtilis*.

2.5.6 Screening for cellulase activity

For the detection of the cellulolytic mutants, modified Basic Salt Medium (mBSM) agar plates containing CMC 1 %, Lactose 4 % or Lactose 4 % and Xylose 1 % (all w/v) were streaked with the transformed bacteria. After 48 h incubation at 60 ºC, the petri dishes were stained with Congo Red (Teather and Wood 1982). mBSM agar contained per liter the following: KH_2PO_4, 0.4 g; $(NH_4)_2SO_4$, 2g; MOPS-salt pH 7, 10 g; nitrilotriaetic acid, 0.2 g; NaOH, 0.125 g, Casamino acids, 3 g; $MgSO_4 \cdot 7H_2O$ 0.01 M , $CaCl_2.2H_2O$ 0.02 M , $FeSO_4.7H_2O$, $MnSO_4.4H_2O$, $ZnCl_2$ and $CuSO_4.5H_2O$ agar 1 %, Cm 7 μg/ ml.

For the detection of cellulolytic activity of the mutants in the supernatant of liquid cultures, positive colonies were used to inoculate 0.6 mL batch cultures in 1.5 mL Eppendorf tubes, incubated at 60 ˚C overnight in an orbital shaker at 900 rpm. mBSM containing CMC and xylose 1% (all w/v) was used as the growth medium. Upon incubation, cell growth was estimated photometrically through OD_{600} while the enzymatic activity of the supernatant was determined as already discussed in Paragraph 2.4.1.

Table 2.5.1: Primers for the construction of the pNW35N derivatives. With upper case letters are indicated the nucleotides of the genomic sequence. With bold lower case letters are indicated the restriction sites.

Construct acronym	Amplification target	Forward Primer (5'-3')	Reverse Primer (5'-3')	Comments
pNW35Nii	Regulatory elements of XynA-T1 from pNW35Ni.	cat**gaattc**TTATATCCCTGTTCATTGGAGTTCTC	cat**ggatcc**TCCTTTGCGGCAGCCTG	Cloned with *Eco*RI and *Bam*HI restriction enzymes
pNW35N(iii)	Homologous regions approx. 1100 bases long, among each side of the point of mutation on the genome of *G. stearothermophilus T-1.*	cat**gaattc**AGCTCCCTGATCTCATTGAC	cat**ggatcc**GATTTGCCGAATGATACGGTCTATG	Cloned with *Eco*RI and *Bam*HI restriction enzymes
pNW35Niii	Cel6A with homologous-regions of *G. stearothermophilus* T-1 ends	TTGTATAGAACACAAACAAATTATATCCCTGTTCATTGGAG *	TGTTTCGACAATCGAAATCACTCCTTTGCGGCAGCCTGCTTAAC *	Cloned by restriction free method
	Homologous regions around the point of insertion and regulatory elements of XynA-T-1 with cel6A ends.	GTTAAGCAGGCTGCCGCAAAGGAGTGATTTCGATTGTCGAAACA *	CTCCAATGAACAGGGATATAATTTGTTTGTGTTCTATACAA *	

* For the primers used for the construction of pNW35Niii by the restriction free method, underlined upper-case letters indicate complementary sequences with the vector pNW35N(iii) while upper-case letters indicate complementary sequences with the regulatory elements of the Cel6A* construct.

3. RESULTS

3.1 Criteria for selection and phylogenetic analysis of the 10 microbial strains

Under the framework of the PhD research of C. Meintanis (2005) at the Laboratory of Microbiology, Department of Biology, National and Kapodistrian University of Athens, 101 bacterial strains (Table 3.1.1) were isolated from selected regions of the volcanic environment of Santorini, Greece. All of them belong to the class of Bacilli with the majority of them being further classified into the genus *Geobacillus*.

These strains have been evaluated for possible exploitation in various biotechnological applications, namely, degradation of aliphatic hydrocarbons from crude oil (Meintanis, Chalkou et al. 2006), production of lipolytic activities (Stathopoulou, Savvides et al. 2013), and hemi-/cellulases production (Stathopoulou, Galanopoulou et al. 2012). The later, was semi-quantified for all strains through a screening procedure which incorporated solid-state cultures with xylan and cellulose as sole carbon source and subsequent visualization of the corresponding hydrolytic potentials using Congo red staining (Table 3.1.1). In accordance to the available literature, our results indicated that the majority of these strains are able to degrade plant biomass, with high preference for xylan.

Based on the two criteria given below, we chose 10 of these strains to screen for key cellulolytic and xylanolytic genes:

1. The taxonomic classification of the strains and the genome sequence availability of the corresponding taxa: We selected representatives from species with several available genome sequence data, a fact that could facilitate the isolation of the target genes. Strains SP24, SP37, SP45, SP47 and SP50 belong to the kaustophilus clade of the genus *Geobacillus* (Figure 3.1.1), a clade with many sequenced members (e.g. Takami, Takaki et al. 2004, Doi, Mori et al. 2013, Studholme 2015). We also selected representatives with limited genome information in an attempt to find relatively novel genes. Strain SP29 reveals high similarity with *Geobacillus* subterraneus (Figure 3.1.1) the genome of which, has been very recently sequenced (Poltaraus, Sokolova et al. 2016). Strains SP59,

SP87 and SP14 are members of the genus Aeribacillus (Figure 3.1.1), a genus with only two-recently sequenced genomes available (Filippidou, Jaussi et al. 2015, Poltaraus, Sokolova et al. 2016). Finally, the strain SP83, which in contrast to all other strains that belong to the Bacillaceae family, is classified within the family of Paenibacillaceae and, specifically, the genus Aneurinibacillus with six genomes available in the Genome database of NCBI (in the beginning of the PhD there were only two) (Figure 3.1.1).

2. The biochemical characteristics of the strains, have already been determined in previous studies of our group (Stathopoulou, Galanopoulou et al. 2012). Among the 101 strains tested, strains SP24, SP29, SP37, SP47, SP50 and SP83 revealed both xylanolytic and cellulolytic activities, something shown only for less than 20% of the total bacterial population (Table 3.1.1). Strains SP59 and SP87 did not produce any cellulolytic activity, but presented significantly high xylanolytic activity (Table 3.1.1). Finally, 20 strains did not exhibit any relevant activity at all. Thus we decided to further screen two respresentetatives of this group, strains SP14 and SP29, for the existence of the corresponding hydrolytic genes.

Comparison of the hydrolytic profile and the taxonomic classification among the selected strains, revealed that all *G. kaustophilus* strains present almost identical plant biomass degradation profiles (Figure3.1.1). Two out of the three *Aeribacillu*s strains present high xylanolytic activity, a characteristic found only in this taxon (Table 3.1.1). In adddition, the taxonomically distant strain SP83, presented the same xylanolytic activity levels with the representatives of the *kaustophilus* clade (Table 3.1.1). Even though our sample is limited to only 10 bacterial strains, these observations probably indicate that there are groups of hydrolytic genes conserved for each taxon.

Table 3.1.1: Comparison of endo-glucanase and xylanase activity profilesa in solid state culturesb of the 101 bacterial strains isolated from the volcanic environment of Santorini, Greece. The strains selected to be further examined in the present study are designated in bold. Adapted from (Stathopoulou, Galanopoulou et al. 2012).

Strain	Enzyme		Strain	Enzyme		Strain	Enzyme	
	Xylanase	CMCase		Xylanase	CMCase		Xylanase	CMCase
SP1	0.00	0.00	SP35	0.00	0.00	SP69	21.25	0.00
SP2	6.33	0.00	SP36	0.00	0.00	SP70	1.71	0.00
SP3	0.00	0.00	SP37	3.71	0.81	SP71	0.00	0.00
SP4	0.30	0.00	SP38	0.66	0.15	SP72	0.67	0.00
SP5	0.64	0.00	SP39	0.43	0.00	SP73	0.58	0.00
SP6	0.00	0.00	SP40	0.52	0.00	SP74	9.50	0.00
SP7	5.36	0.00	SP41	0.00	0.00	SP75	0.00	0.00
SP8	10.40	0.00	SP42	6.94	0.00	SP76	2.29	0.00
SP9	0.00	0.00	SP43	0.18	0.00	SP77	2.09	0.00
SP10	1.46	0.00	SP44	0.11	0.00	SP78	4.56	0.00
SP11	11.43	0.00	SP45	0.14	0.18	SP79	2.80	0.00
SP12	7.57	0.00	SP46	0.14	0.60	SP80	2.19	0.00
SP13	0.33	0.00	SP47	0.08	0.23	SP81	1.81	0.00
SP14	0.00	0.00	SP48	0.29	0.00	SP82	2.50	0.00
SP15	6.25	0.00	SP49	1.28	0.00	SP83	0.21	0.00
SP16	0.75	0.00	SP50	0.93	0.37	SP84	0.00	0.00
SP17	0.17	0.09	SP51	2.16	0.00	SP85	0.00	0.00
SP18	0.40	0.00	SP52	0.00	0.00	SP86	0.00	0.00
SP19	1.25	0.00	SP53	1.24	0.00	SP87	8.53	0.00
SP20	1.34	0.00	SP54	1.02	0.00	SP88	3.21	0.00
SP21	0.94	0.00	SP55	3.21	0.00	SP89	0.00	0.00
SP22	0.44	0.00	SP56	3.91	0.00	SP90	4.10	0.00
SP23	0.15	0.00	SP57	0.58	0.00	SP91	0.54	0.00
SP24	0.29	0.12	SP58	0.40	0.00	SP92	2.50	0.00
SP25	0.32	0.00	SP59	10.08	0.00	SP93	7.50	0.00
SP26	0.00	0.00	SP60	1.29	0.00	SP94	1.19	0.00
SP27	0.23	0.00	SP61	0.92	0.00	SP95	0.47	0.00
SP28	1.04	0.00	SP62	0.41	0.00	SP96	1.78	0.00
SP29	0.00	0.00	SP63	1.91	0.00	SP97	0.00	0.00
SP30	0.34	0.00	SP64	5.19	0.00	SP98	0.00	0.00
SP31	0.58	0.00	SP65	0.79	0.00	SP99	0.00	0.00
SP32	4.03	0.00	SP66	8.10	0.00	SP100	0.65	0.00
SP33	1.53	0.00	SP67	7.73	0.00	SP101	0.46	0.00
SP34	1.56	0.00	SP68	27.78	0.00			

a Activities are expressed as the ratio of the destained area diameter (mm) per 106 cfu's inoculated in each well. Data represent the mean of three plates. For all strains and enzyme activities the standard deviation (SD) ranged from 1.3 to 17.7 %

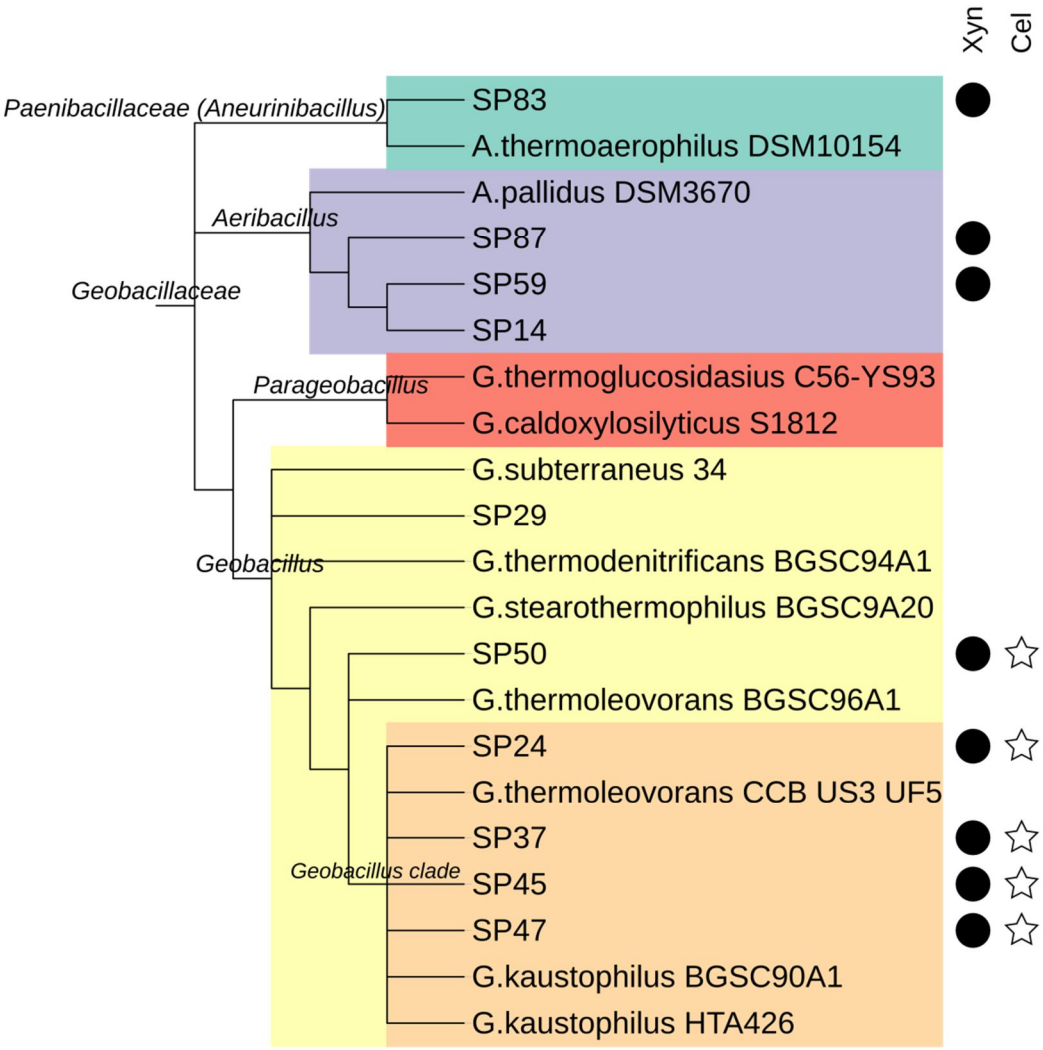


Figure 3.1.1: Phylogenetic analysis of the 10 selected strains according to their 16S rRNA partial sequences. Strains with xylanase activity are indicated with black cycles while strains with cellulase (CMCase) activity are indicated with white stars.

3.2 Genetic screening of the 10 thermophilic strains for cellulolytic and xylanolytic genes

3.2.1 Characterized cellulolytic and xylanolytic enzymes from representative Bacillaceae and Paenibacillaceae strains

In order to detect the genes of interest in the genome of the 10 selected strains, the first step was to collect all characterized cellulases and xylanases from Bacillaceae and Paenibacillaceae families,deposited in the Carbohydrate - Active enZyme database (www.cazy.org) (Lombard,

Golaconda Ramulu et al. 2014). The outcome of this work is summarized in Table 3.2 while complete lists of the related microorganisms and sequences with Uniprot numbers are set out in Appendix I, Tables 1 and 2.

As shown in Table 3.2.1, we collected 134 entries in total: 101 of which belong to Bacillaceae family and 33 to Paenibacillacae family. These sequences are classified into 15 different GH families, but since some of these families (i.e. GH5, GH8 and GH12) include more than one enzymatic activity, the total number of sequences can be further divided into 19 groups in total. All members of each single group belong to the same GH family and share the same enzymatic activity.

Concerning plant biomass degrading hydrolases from *Geobacillus, Aeribacillus* and *Aneurinibacillus* genera, at the time of the experiment, information for characterized sequences was available from the *Geobacillus* representatives only (Table 3.2.1). These 12 characterized enzymes, belong to 8 different GH families and the majority of them is related with xylan degradation. Regarding the genera of *Aeribacillus* and *Aneurinibacillus,* there were no characterized enzymes deposited in CAZY at that time and to our knowledge, neither any putative sequence nor any related literature.

Based on the above results, we decided to further screen the 10 selected strains for 8 different groups of enzymes, namely:

1. GH1 β-P-glucosidases
2. GH5 endo-1,4-β-glucanases

3,4. GH10 and GH11 endo-1,4-β-xylanases

4,5,6. GH39, GH43and GH52 xylan-1,4-β-xylosidases

7. GH43 xylan-1,4-β-xylosidases
8. GH48 cellulose-1,4-b-cellobiosidase

The rational and the outcome of our approach are discussed separately in the following paragraphs for each one of the enzymes.

Table 3.2.1: Characterized enzymes of Bacillaceae (*B*), *Geobacillus* (*G*) and Paenibacillaceae (*P*) deposited in CAZY database (August, 2016). The columns show the number of enzymes from the same taxon belonging to the same GH family and present the same enzymatic activity. The enzymatic activities are given with the corresponding enzyme commission numbers: The groups of enzymes selected for further screening are highlighted in gray background.

	Xylan degrading enzymes												Cellulose degrading enzymes								
EC	3.2.1.8			3.2.1.37			3.2.1.156			3.2.1.151			3.2.1.4			3.2.1.176			3.2.1.86		
~~GH~~	*B*	*G*	*P*	*B*	*G*	*P*	*B*	*G*	*P*	*B*	*G*	*P*	*B*	*G*	*P*	*B*	*G*	*P*	*B*	*G*	*P*
1																			4	2	
4																			2	1	
5												2	30	1	6						
8	1	-					1	-					1	-							
9													5	-	2						
10	17	6	10																		
11	29	1	9																		
12										1			1	-							
30	1	-																			
39				2	1																
43				9	3																
44															1						
48																		1			
52				3	3	1															
74												1									

** EC 3.2.1.8: endo-1,4,-b-xylanase, EC 3.2.1.37: xylan-1,4-b-xylosidase, EC 3.2.1.156: oligosaccharide reducing-end xylanase, EC 3.2.1.151: xyloglucan-specific endo-b-1,4-glucanase, EC 3.2.1.4: endo-1,4-b-glucanase, EC 3.2.1.176: cellulose 1,4-b-cellobiosidase (reducing end), EC 3.2.1.86: 6-P-b-glucosidase.
* Detailed lists of the entries of the Table, are given to Appendix I, Tables 3 to 9.

3.2.2 Detection and isolation of the GH1 *β*-P-glucosidase gene in the 10 selected strains

For the detection of GH1 *β*-P-glucosidases, we based on the DNA sequences of the characterized enzymes of *G. kaustophilus* HTA426 (Suzuki, Okazaki et al. 2012). This bacterium contains three open reading frames encoding for GH1 enzymes: GK1856, GK3214 and GK2337. All three sequences do not bear any signal peptide and the maximum amino acid identity with each other is 45%. Regarding their enzymatic activity, the first two present 6-P-*β*-glycosidase activity while GK2337 does not (Suzuki, Okazaki et al. 2012). Based on the DNA alignment of GK1856 and GK3214 (39% amino acid identity), we designed the internal primers for the GH1 family shown at Figure 3.2.1.

Four among the tested strains, gave specific PCR product (Table 3.2.2). All of the positive strains belong to the *kaustophilus* clade, just like the source strain of GK3214 and GK1856. We decided to continue further with the PCR product of strain SP24. By comparing the nucleotide sequence of the gene fragment with the DNA sequences deposited in NCBI database, it was shown that the amplified product had 99% identity with the GK3214 nucleotide sequence of *G. kaustophilus* HTA426. Based on the latter, we designed external primers for the isolation of the complete gene. After cloning in pET15b we identified that the cloned ORF was 98% identical with GK3214.

Table 3.2.2: Summary Table of the PCR screening for GH1 family genes." +" indicates presence of PCR product , " –" indicates that PCR did not work while with "N/A" are marked the strains that were not tested.

Genus	Strain	GH1 PCR	Genus	Strain	GH1 PCR	Genus	Strain	GH1 PCR
Geobacillus	SP24	+	*Aeribacillus*	SP14	N/A	*Aneurinibacillus*	SP83	-
	SP29	N/A		SP59	-			
	SP37	N/A		SP87	-			
	SP45	+						
	SP47	+						
	SP50	-						

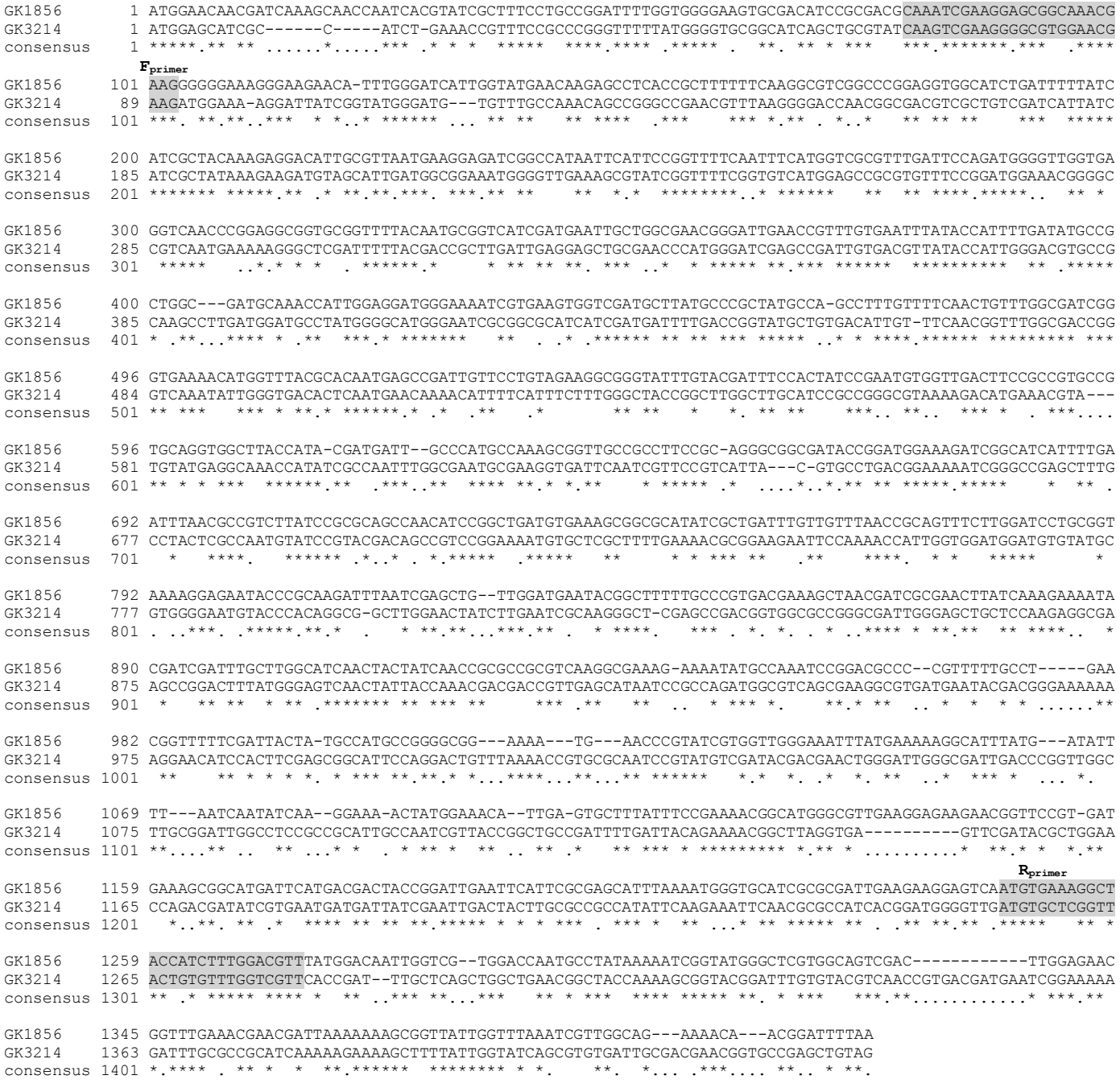


Figure 3.2.1: DNA alignment of GK3214 and GK1856. The designed inner primers are highlighted.

3.2.3 Detection and isolation of GH5 endo-1,4-β-glucanase genes from the 10 selected strains

The GH5 cellulases (E.C. 3.2.1.4) isolated from Bacillaceae and Paenibacillaceae genera are very diverge (e.g. 28% identity and 16% query coverage between A9Q742 and C5HX3) and can be classified into subfamilies, 1, 2 and 4 (Figure 3.2.2). All entries but one, are extracellular, while one bear S-layer homology domains (SLH) indicating that it possibly interacts with the bacterial cell wall. Among the entries, the most common domain architecture is a catalytic domain combined with a carbohydrate binding domain (CBM 3, CBM X2, CBM 17 or ChiC BD). However there are also GH5 cellulases, including the one isolated from the *Geobacillus* strain, containing only the cellulase domain.

Because of the great heterogeneity of GH5 cellulases, it was impossible to detect conserved regions throughout all entries long enough, to design degenerate primers. Thus, in order to overcome this obstacle, we designed two pairs of degenerate primers based on a big clade of subfamily 2 which includes the characterized, cell associated cellulase from *Geobacillus* sp 70PC53 (Uniprot No C5H6X3)(Ng, Li et al. 2009) (Figure. 3.2.2 and 3.2.3).

The first pair of primers used for PCR, gave no specific amplicons. In contrast, the second pair of primers, yielded amplicons of the anticipated molecular weight for the strains SP29 and SP50. After sequencing of these fragments, it was shown that, the fragment from the strain SP50 was identical with a fragment of C5H6X3. Consequently, we designed primers for the complete ORF and the obtained gene revealed 100% identity with the amino acid sequence of the characterized molecule.

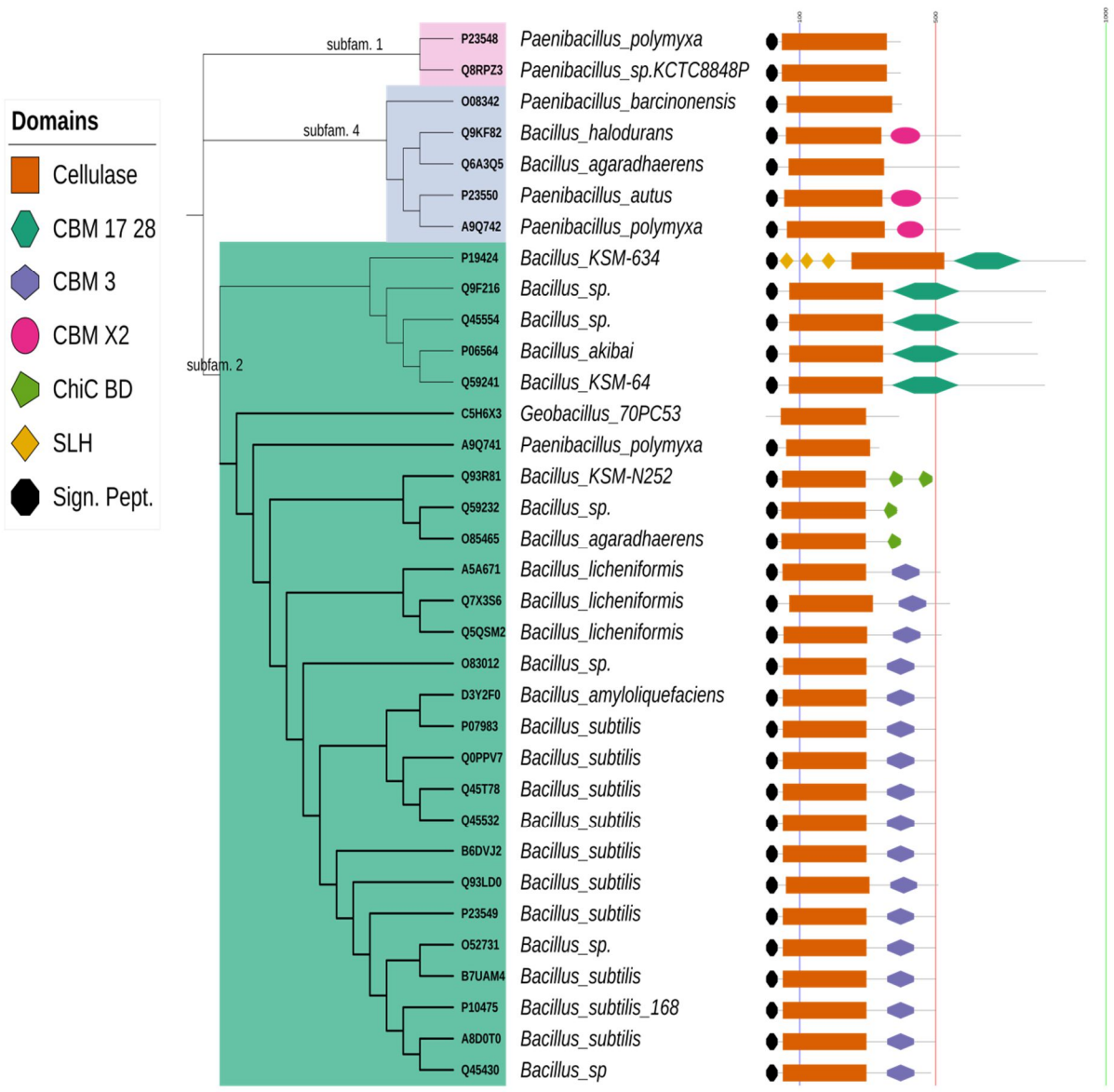


Figure 3.2.2. Phylogenetic tree of the characterized GH5 cellulases from the Bacillaceae and Paenibacillaceae families. The three subfamilies- 1,2 and 4- form independent clades and are highlighted in different colours. Uniprot numbers and the names of the microorganism sources are given next to each branch. The length and domain architecture of the entries are schematically shown at the right part of the figure while the name abbreviations of the domains are given at the Domains key. The branches of the sequences from subfamily two, used for primer design, are highlighted in bold. Cellulase: GH5 cellulase domain, CBM 3, CBM X2, CBM 17 28,ChiC BD: Carbohydrate binding domains, SLH: S-layer homology domain, Sign. Pept.: Signal peptide.

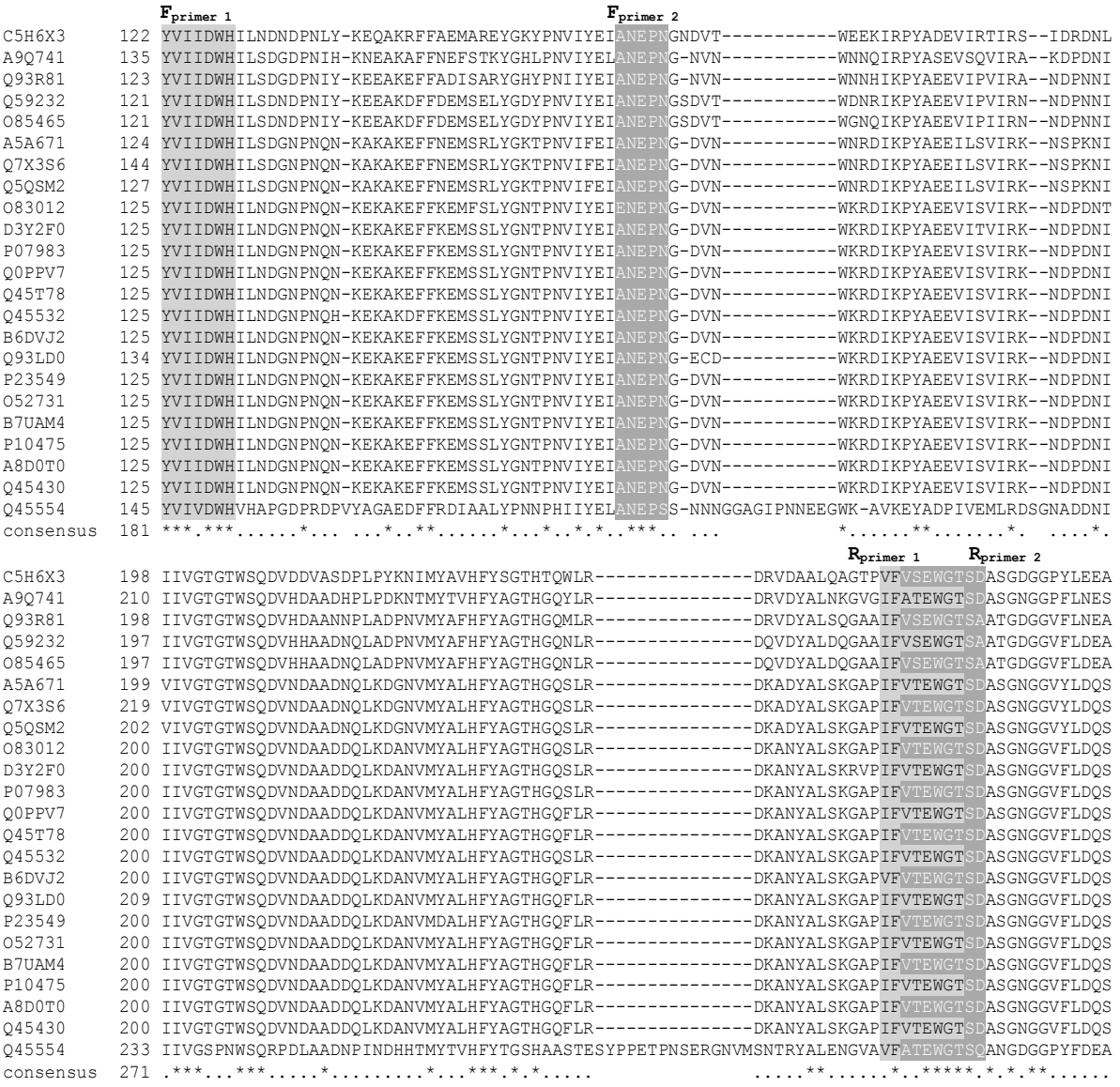


Figure 3.2.3: Part of the alignment of the amino acid sequences of the selected characterized GH5 cellulases. The residues used to design the degenerate primers are highlighted.

3.2.4 Screening for GH9 endocellulase genes from the 10 selected strains

GH9 is one of the largest families consisting mainly of endocellulases (EC 3.2.1.4). There are seven characterized representatives of this group from members of both Bacillaceae and Paenibacillaceae families (Figure 3.2.4). The six out of the seven entries are cell associated and also all but one present similar architecture bearing one catalytic domain and a cellulose binding one (CBM 3).

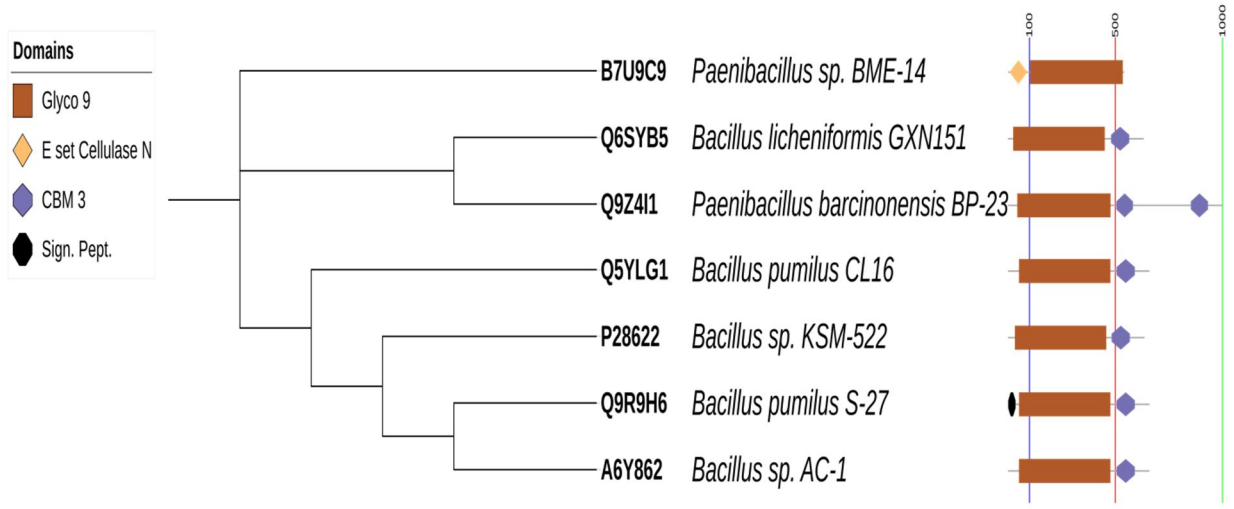


Figure 3.2.4: Phylogenetic tree of the characterized GH9 cellulases from the Bacillaceae and Paenibacillaceae families. Uniprot numbers and the names of the microorganism sources are given next to each branch. The length and domain architecture of the entries are schematically shown at the right part of the figure while the name abbreviations of the domains are given at the Domains key.Glyco 9: GH9 cellulase domain, CBM 3: Cellulose binding domain,E set Cellulase N: N-terminal Early set domain associated with the catalytic domain of cellulase, Sign. Pept.: Signal peptide.

However there are no reports for GH9 genes from any member of *Geobacillus*, *Aeribacillus* or *Aneurinbacillus*. We hypothesized that this might be due to the limited genomic information available, especially for *Aeribacillus* and *Aneurinibacillus* genera. In addition, since our strains are novel and have been isolated from an unexploited environment, they might contain genes not reported so far, including GH9 endocellulases. Consequently, we attempted to screen the ten selected strains using internal primers designed from the conserved regions of the seven characterized GH9 representative aminoacid sequences from Bacillaceae and Paenibacilaceae members (Appendix I, Table 7).

Despite the diversity of the characterised sequences, we were able to detect conserved regions and design degenerate primers (Figure 3.2.5). Following PCR, we obtained bands of various lengths, including some at the expected molecular weight. However, after sequencing, it was proven that none of these was related to the GH9 family.

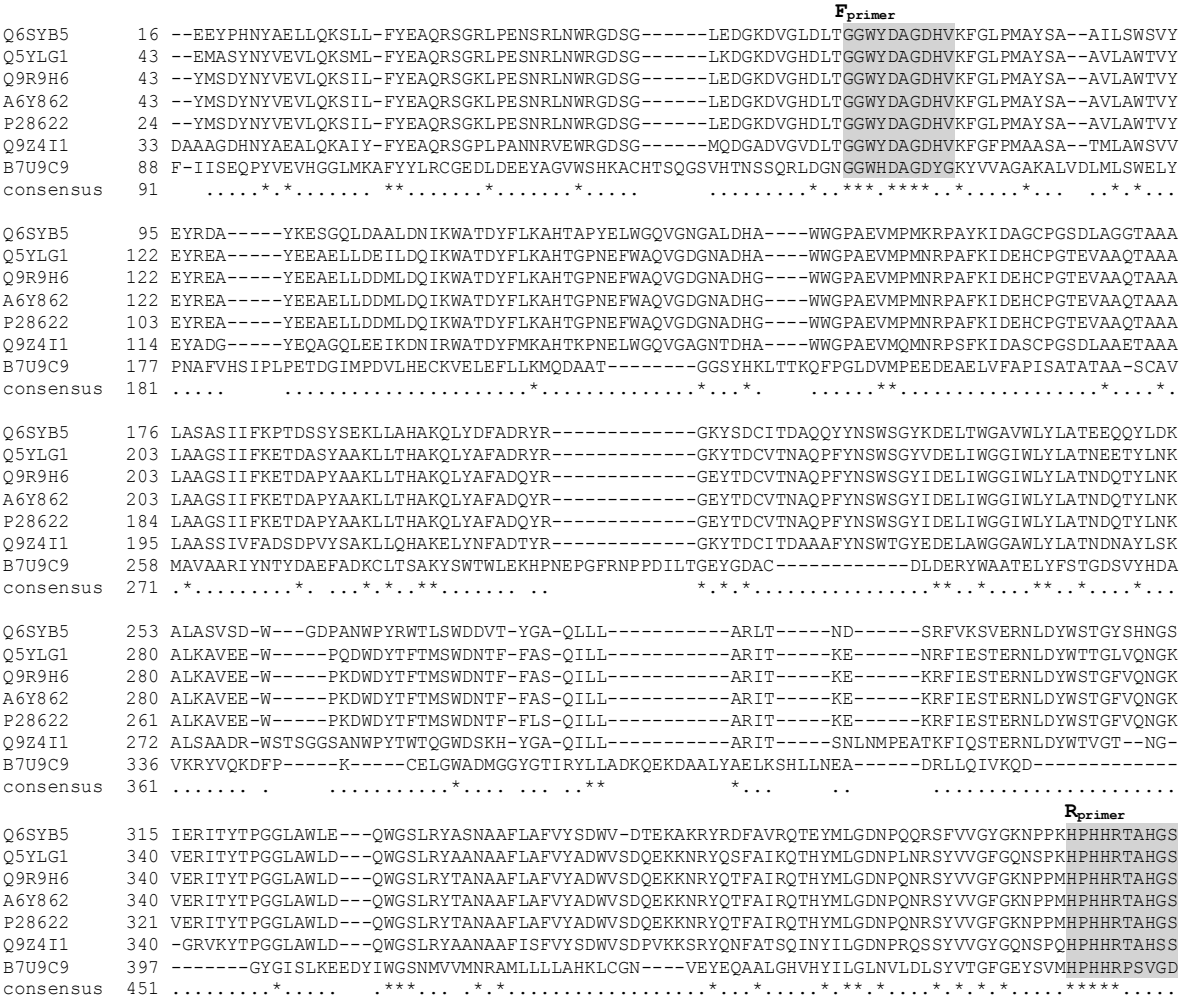


Figure 3.2.5: Part of the alignment of the amino acid sequences of selected characterized GH9 cellulases. The two sets of degenerate primers designed, are highlighted.

3.2.5 Detection and isolation of GH10 endo-β-1,4-xylanase genes from the 10 selected strains

GH10 family is mainly consisted of endo-β-1,4-xylanases (EC 3.2.1.8). Many of these enzymes have been detected in members of Bacillaceae and Paenibacillaceae and particularly six of them come from the genus *Geobacillus*. The majority of the enzymes can be divided in two distinct branches: the one with extracellular single domains xylanases and the other one with cell associated single domain xylanases (Fig 3.2.6) Exception are some *Paenibacillus* extracellular enzymes which, besides the catalytic domain, also present domains responsible for binding with substrate (CBM 4 9,CBM 9, DOMON) or bacterial cell interactions (SLH).

Based on the conserved regions shared by all *Geobacillus* entries, we were able to design degenerate primers specific for both branches (Figure 3.2.7). Specific PCR bands at the expected molecular weight were obtained for all 10 strains but SP14, SP29 and SP37 (Table 3.2.3). We randomly sequenced some representative strains resulting in two different products: the one was almost identical with an extracellular GH10 xylanase while the other with a membrane associated. Consequently, we designed specific primers for the complete ORFs of the two groups which allowed us to amplify:

1. a gene from strain SP24, that gave a protein product 99.4% identical with the membrane associated Q3YBZ9 and
2. a gene from strain SP50,with protein product 100 % identical with the extracellular D0EM78.

In addition to the genes amplified from the above strains, the membrane associated xylanase was detected in all 10 strains while the extracellular enzyme gave positive PCR only for strain SP50 (Table 3.2.3).

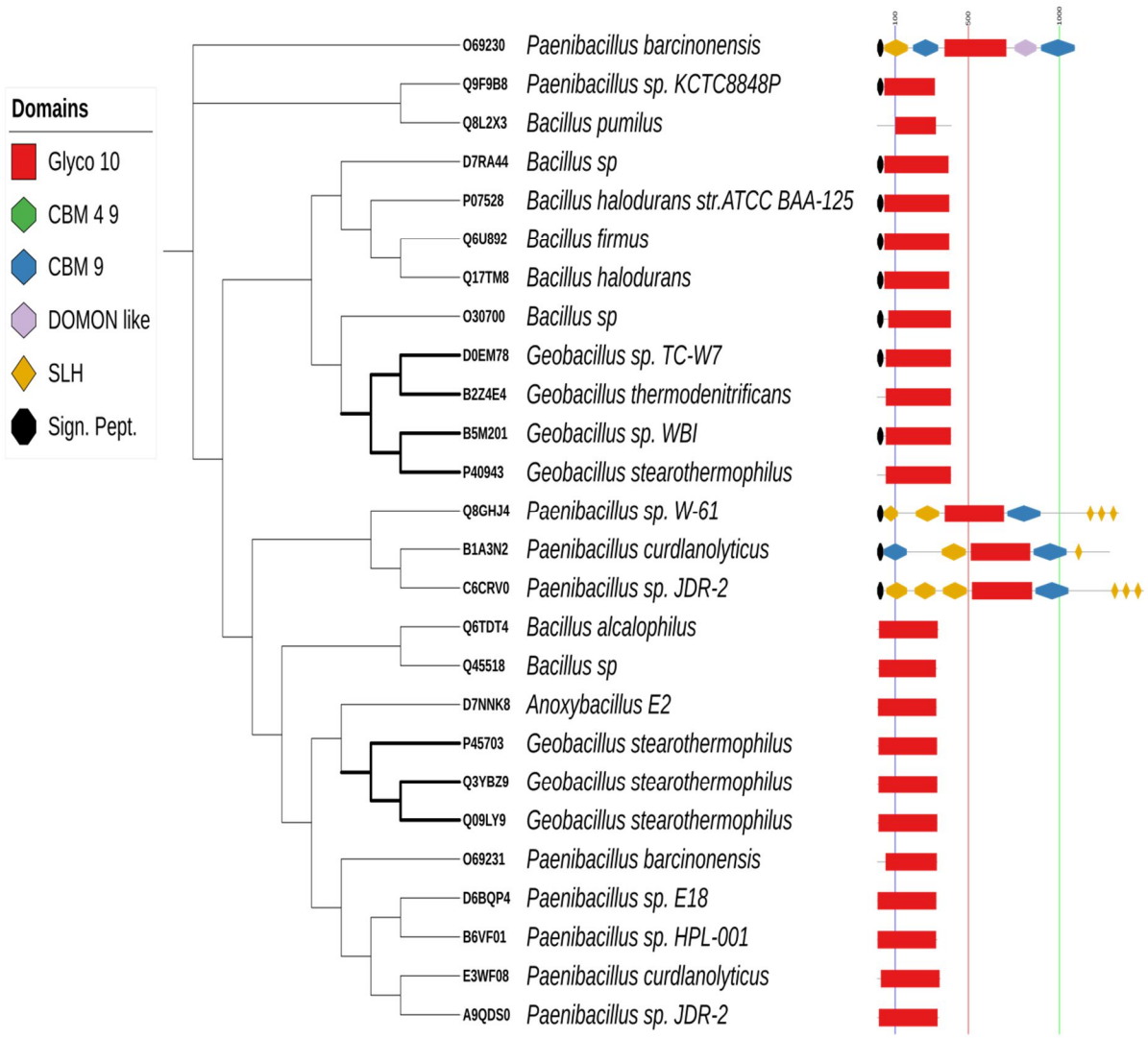


Figure 3.2.6: Phylogenetic tree of the characterized GH10 xylanases from the Bacillaceae and Paenibacillaceae families. Uniprot numbers and the names of the microorganism sources are given next to each branch. The length and domain architecture of the entries are schematically shown at the right part of the figure while the name abbreviations of the domains are given at the Domains key. The branches of the sequences, used for primer design, are highlighted in bold. Cellulase: Glyco 10: xylanase domain, CBM 4 9, CBM 9, DOMON like: Carbohydrate binding domains, SLH: S-layer homology domain, Sign. Pept.: Signal peptide.

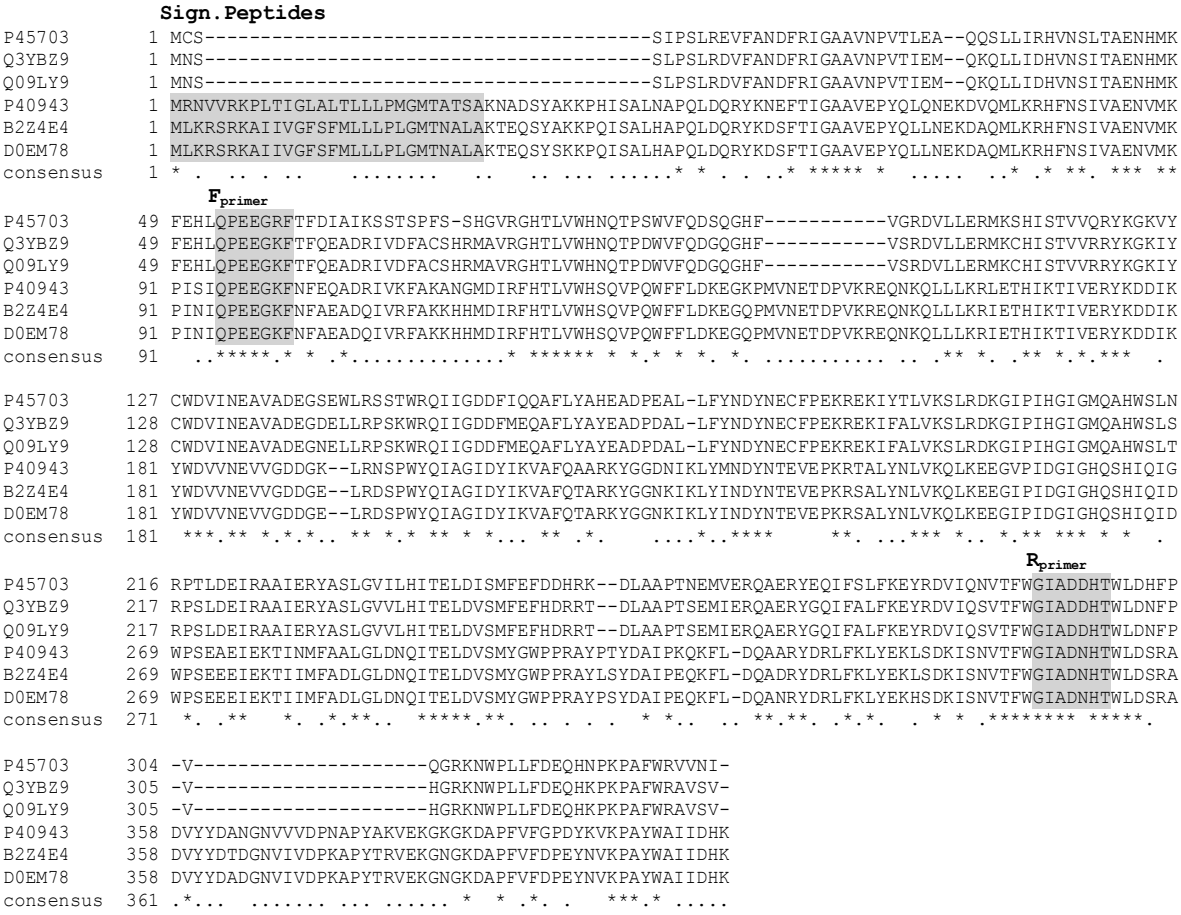


Figure 3.2.7: Amino acid sequence alignment of the six characterized GH10 endo-β-1,4-xylanases isolated from *Geobacillus* strains. The two primers, as well as the signal peptides of the extracellular enzymes are highlighted.

Table 3.2.3. Summary Table of PCR reactions with the internal (In) and external sets (Ex for the extracellular and M for the membrane associated enzymes) of primers using as template, DNA from the ten selected strains.

Genus	Strain	GH10 PCR			Genus	Strain	GH10 PCR			Genus	Strain	GH10 PCR		
		In	Ex	M			In	Ex	M			In	Ex	M
Geobacillus	SP24	+	-	+	*Aeribacillus*	SP14	-	-	+	*Aneurinibacillus*	SP83	+	-	+
	SP29	-	-	+		SP59	+	-	+					
	SP37	-	-	+		SP87	+	-	+					
	SP45	+	-	+										
	SP47	+	-	+										
	SP50	+	+	+										

3.2.6 Screening for GH11 endo-β-1,4-xylanase genes from the 10 selected strains

GH11 family contains only xylanases with representatives found in all three domains of life. Several members have been characterized, including 29 from Bacillaceae family and 9 from Paenibacillaceae. Among the three genera of interest-*Geobacillus*, *Aeribacillus* and *Aneurinibacillus*, there is only one entry reported so far, the P45705, isolated from *Geobacillus stearothermophilus* 236 (Choi and Cho 1995).

The 38 proteins from Bacillaceae and Paenibacillacae are in the majority secreted enzymes consisting of a catalytic domain and a CBM (Figure3.2.8). However, they reveal significant diversity in terms of their amino acid sequence (minimum identity 41%) and thus it was impossible to design degenerate primers from all of them. We decided to incorporate all the sequences forming a branch with P45705 from the *Geobacillus stearothermophilus* (Figure 3.2.8). This group of proteins was indeed more conserved (71% minimum identity) and upon alignment we were able to identify conserved regions and consequently design inner degenerate primers against them (Figure 3.2.9).

Following PCR, we obtained several bands for all strains, some of which were at the expected molecular weight. However, after cloning and sequencing of the target bands, we could not identify any putative GH11 sequence.

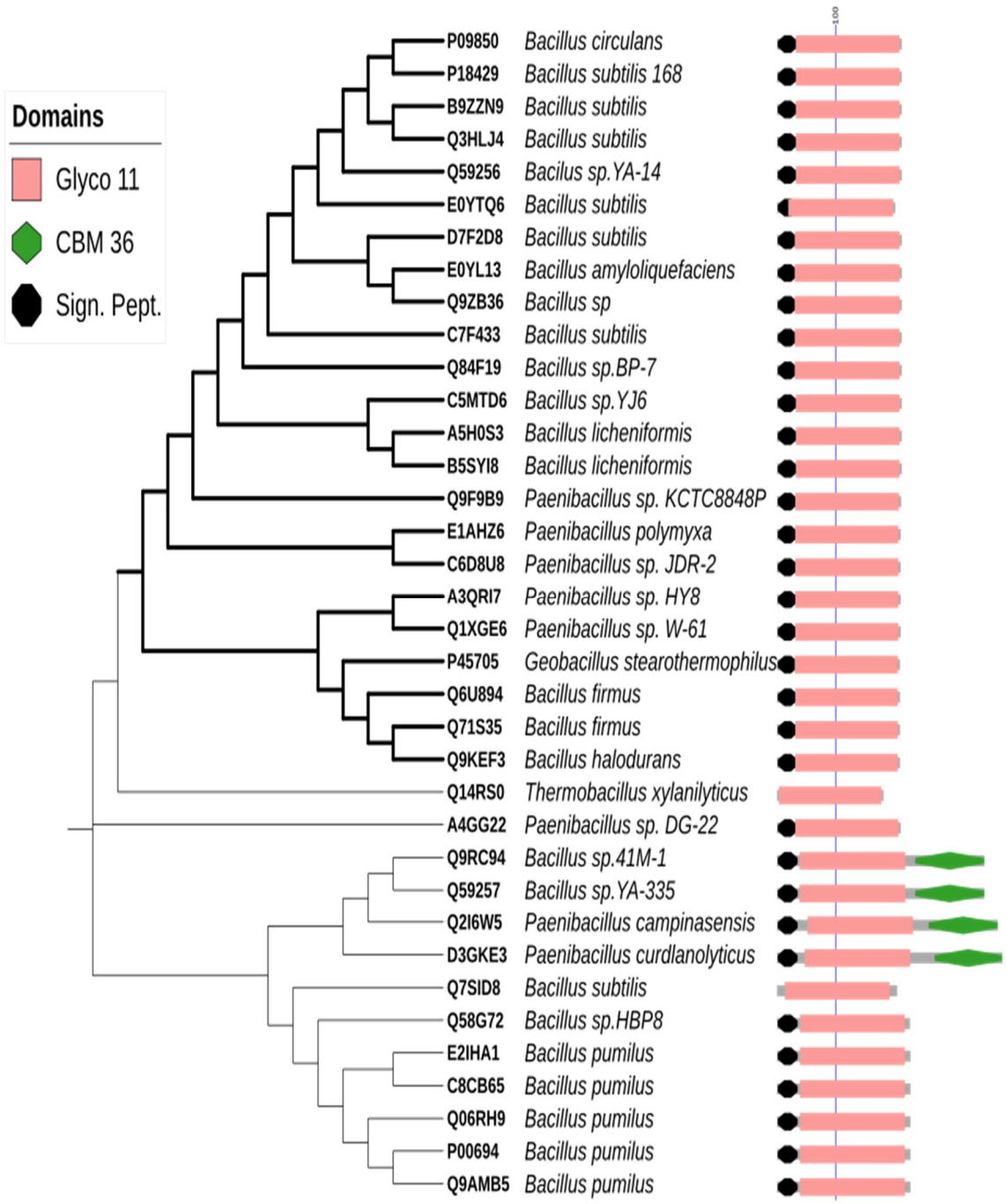


Figure 3.2.8: Phylogenetic tree of the characterized GH11 xylanases from the Bacillaceae and Paenibacillaceae families. Uniprot numbers and the names of the microorganism sources are given next to each branch. The length and domain architecture of the entries are schematically shown at the right part of the figure while the name abbreviations of the domains are given at the Domains key. The branches of the sequences, used for primer design, are highlighted in bold. Cellulase: Glyco 11: xylanase domain, CBM 36:Carbohydrate binding domain, Sign. Pept.: Signal peptide.

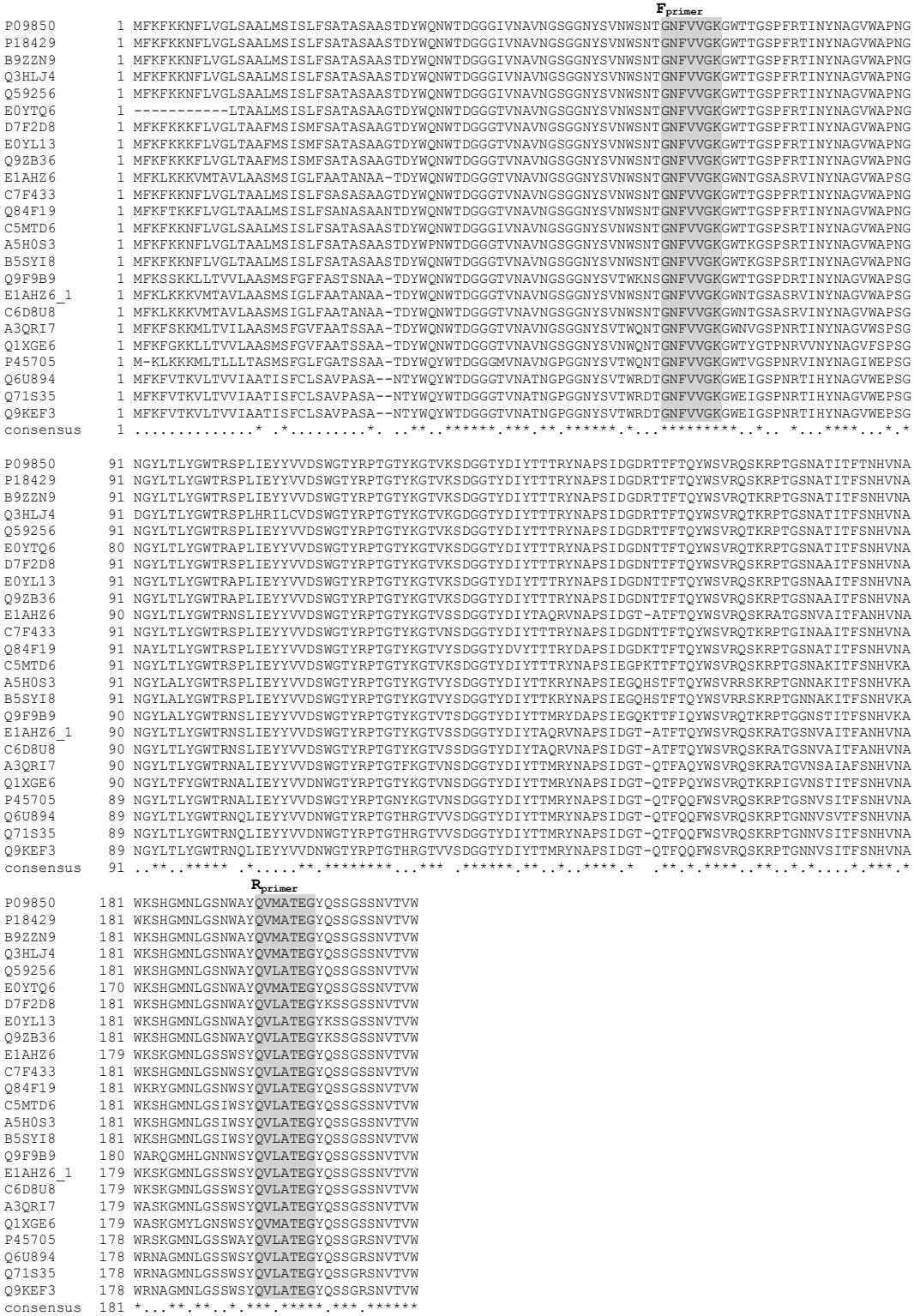


Figure 3.2.9: Amino acid sequence alignment of the 21 characterized GH11 endo-β-1,4-xylanases from Bacillaceae and Paenibacillaceae selected for designing primers. Regions used for primer design are highlighted.

3.2.7 Detection and isolation of GH39 *β*-xylosidase genes from the 10 selected strains

There are only twelve characterized bacterial GH39 enzymes deposited in CAZy database with all of them revealing *β*-xylosidase activity. Among them, two entries belong to the Bacillaceae family (Appendix I, Table 8): One is a *Bacillus halodurans* strain and the other *Geobacillus stearothermophilu*s T-6. For Paenibacillaceae family in contrast, even though there are a lot of putative GH39 genes in CAZy, none is reported as characterized. Both of the characterised enzymes from Bacillaceae members are cell-associated and are consisted of a single catalytic domain.

In order to screen for the GH39 *β*-xylosidases the 10 selected strains, we decided to give emphasis to the *Geobacillus* sequences and so we used the gene of the cell-associated *β*-xylosidase from *Geobacillus stearothermophilus* T-6 as well as the putative sequences from other *Geobacillus* strains deposited in CAZy (Appendix I, Table 10). These sequences were significantly conserved (lower sequence similarity 81% between the entries of *G. thermodenitrificans* NG80-2 and *G. thermoglucosidasius* C56-YS93), a fact that allowed us to easily design degenerate primers for inner conserved regions.

From the strains tested for the GH39 gene, only those belonging to the *Geobacillus* clade gave a positive PCR reaction (Table 3.2.4). The sequenced PCR product from strain SP50 showed significantly high identity to many entries of NCBI nr database from various *Geobacillus* strains (Table 3.2.5), allowing us to design primers for the full open reading frame. As can be seen from Table 3.2.4, similarly with the internal primers, only some *Geobacillus* strains gave a positive reaction. The resulting PCR product from strain SP50, was cloned into pET15b vector. The produced protein was 100% identical with the protein ESU73074.1 from *Geobacillus sp.* MAS1.

Table 3.2.4: Summary Table of PCR reactions for the GH39 *β*-xylosidase with the internal (In) and external sets (Ex) of primers using as template, DNA from the ten selected strains.

Genus	Strain	GH39 PCR		Genus	Strain	GH39 PCR		Genus	Strain	GH39 PCR	
		Ex	Int			Ex	Int			Ex	Int
Geobacillus	SP24	+	+	*Aeribacillus*	SP14	-	-	*Aneurinibacillus*	SP83	-	-
	SP29	NA	+		SP59	NA	-				
	SP37	+	-		SP87	NA	-				
	SP45	+	+								
	SP47	NA	-								
	SP50	+	+								

Table 3.2.5: First five results of blast search against NCBI nr database using as query the GH39 DNA fragment amplified from the genomic DNA of '*Geobacillus icigianus* SP50'.

Description	Max score	Total score	Query coverage	E value	Identity	Accesion No
Geobacillus sp. GHH01, complete genome	1042	1042	100 %	0.0	99%	CP004008.1
*Geobacillussp.*C56-T3,complete genome	1031	1031	100 %	0.0	98%	CP002050.1
Geobacillus thermoleovorans str FJAT-2391,complete genome	1030	1030	100 %	0.0	98%	QP017071.1
Geobacillus stearothermophilus str. T-6, genomic sequence	1028	1028	100 %	0.0	99%	DQ868502.2
Geobacillus sp. Y412MC52,complete genome	997	997	100 %	0.0	95%	CP002442.1

3.2.8 Detection and isolation of GH43 *β*-xylosidase genes from the 10 selected strains

There are nine characterized GH43 enzymes from Bacillaceae members revealing *β*-xylosidase activity, with the three of them, belonging to *Geobacillus* members. In contrast, there are no reports for GH43 *β*-xylosidases in Paenibacillaceae family. All characterised enzymes are consisting of a catalytic domain and they are either extracellular or cell-associated.

From previous studies in our lab, we have isolated and characterized a *β*-xylosidase from the *Geobacillus* strain SP24 (Galanopoulou 2012). The physicochemical characteristics of the purified enzyme were similar with these reported for the GH43 *β*-xylosdiases Q09LXD and Q22N4 isolated from *Geobacillus* strains (Figure 3.2.10). These proteins revealed high amino acid and DNA similarity (minimum DNA similarity 88%) thus, we designed degenerate primers based on the conserved DNA regions of these two and also the putative U2WUG3 which also revealed high DNA similarity with the two characterised.

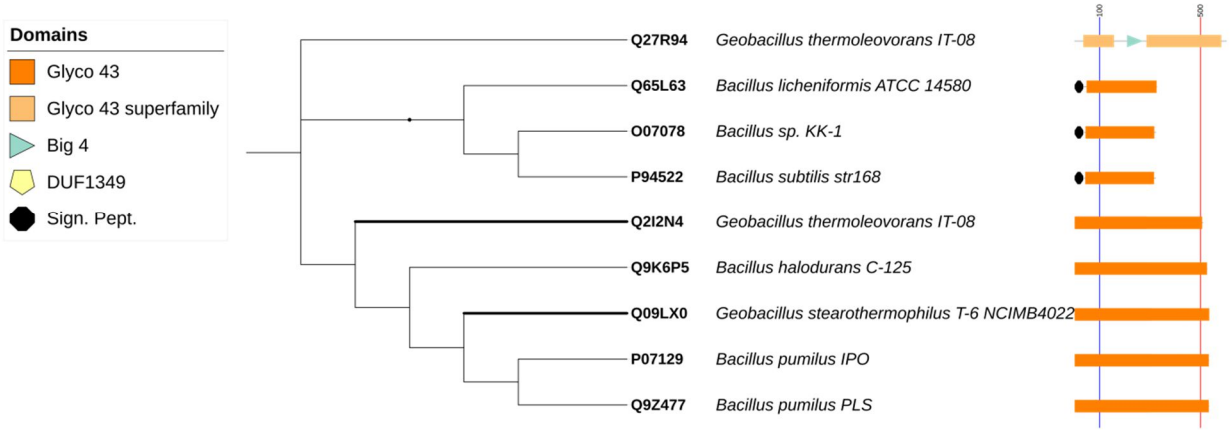


Figure 3.2.10: Phylogenetic tree of the characterized GH43 *β*- xylosidases from Bacillaceae members. Uniprot numbers and the names of the microorganism sources are given next to each branch. The length and domain architecture of the entries are schematically shown at the right part of the figure while the name abbreviations of the domains are given at the Domains key. The branches of the sequences, used for primer design, are highlighted in bold. Glyco 43 (CAZY classification subfamily 11), Glyco 43 superfamily (CAZY classification subfamily 27): xylanase domain, Big 4: bacterial Ig-like surface protein, DUF1349: conserved domain of unknown function, Sign. Pept.: Signal peptide.

PCR against DNA from all selected strains, resulted in one band at the expected molecular weight for the five out of the six *G. kaustophilus* strains and, interestingly, also from the SP83 strain (Table 3.2.6). The sequenced DNA fragment from SP24, revealed high similarity with the putative U2WUG3 GH43 *β*-xylosidase and thus we based on the aligned ends of this sequence to further design specific primers for the complete open reading frame. The complete gene revealed 96% identity with the sequence of the puatative *β*-xylosidase of *Geobacillus* GHH01 (CP002050.1).

Table 3.2.6: Summary Table of the PCR results for the detection of GH43 family genes." +" indicates presence of PCR product , " –" indicates that PCR did not work.

Genus	Strain	GH43 PCR	Genus	Strain	GH43 PCR	Genus	Strain	GH43 PCR
Geobacillus	SP24	+	*Aeribacillus*	SP14	-	*Aneurinibacillus*	SP83	+
	SP29	-		SP59	-			
	SP37	+		SP87	-			
	SP45	+						
	SP47	+						
	SP50	+						

3.2.9 Screening for GH48 Cellulose 1,4-*β*-cellobiosidase genes from the 10 selected strains

GH48 1,4-*β*-cellobiohydrolases (E.C. 3.2.1.176) are among the key enzymes involved in cellulose hydrolysis. They are among the main components of cellulosomes and other multifunctional enzymes, but they also exist as free enzymes. There are 11 characterised GH48 1,4,-*β*-cellobiohydrolases from various microorganisms deposited in CAZY with one of them belonging to a *Paenibacillus* strain (Figure 3.2.11). In contrast, even though there are some putative GH48 genes from *Bacillus*, there are not indications for any member of this family in *Geobacillus, Aeribacillus* and *Parageobacillus* genera.

The characterised enzymes, in their majority are secreted and multi-modular. They combine a GH48 catalytic domain with carbohydrate binding modules, dockerins for cellulosome formation and even catalytic domains of other specificities. Some other domains, with unclear functionality but possibly related with the localization of the proteins, are also present (FN 3 Big3 5).

Similarly with GH9 enzymes, we hypothesized that, because our strains are novel and also there are not a lot of sequenced genomes of related species available, some of the 10 selected strains - and especially those belonging to Paenibacillaceae - might contain a GH48 member. Therefore, we designed internal degenerate primers based on the catalytic domain of all available

characterized amino acid sequences available in CAZy, plus some extra putative GH48 sequences from *Paenibacilli* (Figure 3.2.12).

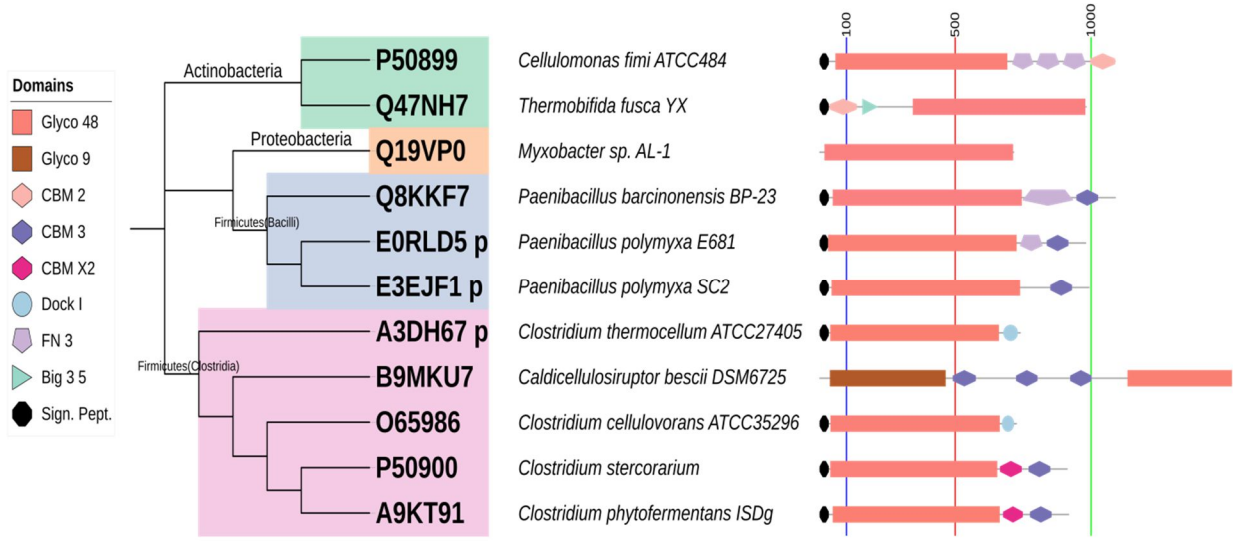


Figure 3.2.11: Phylogenetic tree of the characterized GH48 1,4-*β*-cellobiohydrolases deposited in CAZY database. Uniprot numbers and the names of the microorganism sources are given next to each branch. The length and domain architecture of the entries are schematically shown at the right part of the figure while the name abbreviations of the domains are given at the Domains key. Glyco 48: 1,4-*β*-cellobiohydrolase catalytic domain, Glyco 9: GH9 cellulase catalytic domain, CBM 2,CBM 3, CBM X2: carbohydrate binding domains, Dock I:dockerin type I,FN 3: fibronectin type 3 domain, Big 3 5: bacterial Ig-like surface protein, Sign. Pept.: Signal peptide.

Following PCR, we obtained several bands for all strains, some of which at the expected molecular weight. However, after cloning and sequencing of the target bands, it was shown that they do not correspond to any GH48 gene.

3.2.10 Detection and isolation of GH52 *β*-xylosidase genes from the 10 selected strains

In addition to GH39 and GH43 *β*-xylosidases, *Geobacilli* contain also *β*-xylosidases of the GH52 family. According to CAZY database, there are 8 characterized enzymes, all of bacterial origin. Among them, five belong to *Geobacillus* strains, one more to *Parageobacillus* and another one to *Paenibacillus* (Appendix I, Table 8). Since the four sequences presented a minimum of 52 % amino acid similarity, we designed degenerate primers based on the conserved sites shown in Figure 3.2.13.

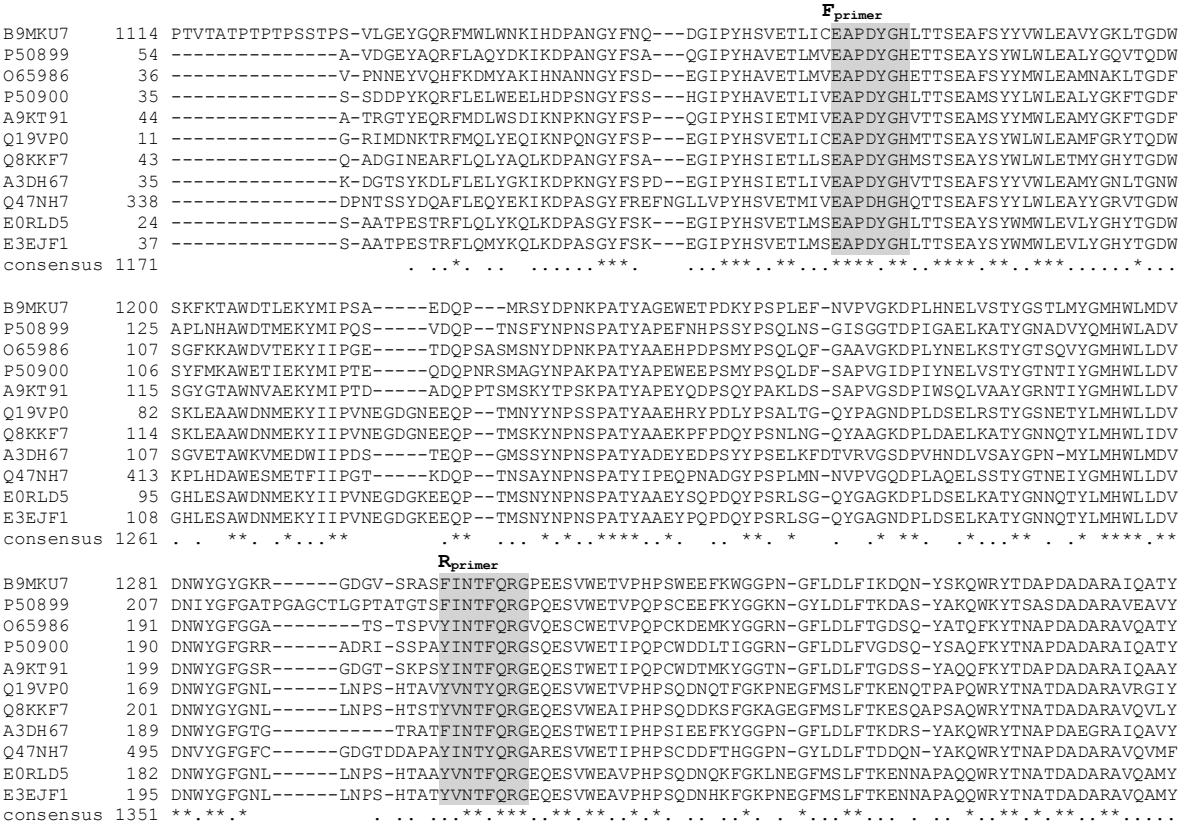


Figure 3.2.12: Alignment of parts of the eleven GH48 representatives used to design degenerate primers. Primers are highlighted.

Despite our attempts for PCR optimization, we were not able to amplify a specific PCR product. Since GH52 genes are present to all the available genomes of *Geobacillus*, we decided to reduce the degeneracy of the primers by designing a new pair based on the DNA sequences of the *Geobacillus stearothermophilus* strains. Indeed, following PCR we were able to amplify a DNA fragment of the anticipated molecular weight from the seven out of ten strains (Table 3.2.7). The open reading frame amplified from strain SP24 was further sequenced and revealed 99% amino acid identity with the putative *β*-xylosidase of *Geobacillus sp* MAS1 (Accession No. WP_023633604.1) and 97% amino acid sequence identity with the characterized *β*-xylosidase of *Geobacillus stearothermophilus* T-6 (GenBank No ABI49956.1).

Figure 3.2.13: Alignment of parts of the four GH52 *β*-xylosidases used to design degenerate primers. The two primers are highlighted in grey.

Table 3.2.7: Summary Table of the PCR results for the detection of GH52 family genes. "+" indicates presence of PCR product," –" indicates that PCR did not work

Genus	Strain	GH52 PCR	Genus	Strain	GH52 PCR	Genus	Strain	GH52 PCR
Geobacillus	SP24	+	*Aeribacillus*	SP14	-	*Aneurinibacillus*	SP83	+
	SP29	+		SP59	-			
	SP37	-		SP87	+			
	SP45	+						
	SP47	+						
	SP50	-						

3.3 Characterization of the produced thermoenzymes

In order to use the isolated enzymes in hydrolysis experiments or other applications described at the next Chapters, it was necessary to know some of their biochemical characteristics like: the activity in various temperatures and pH values, the stability under different temperatures and pH as well as their kinetic constants. Since some of the isolated enzymes were very similar with enzymes already characterized, we were able to obtain some of the needed information from the existing literature, with just a validation of some key values. For the rest, we conducted all the necessary experiments as described in the paragraphs below.

3.3.1 Biochemical and structural characterization of the GH1 β-1,4-P-glycosidase

3.3.1.1 Biochemical characterization of the GH1 β-1,4-P-glycosidase

The first step for the biochemical characterization of the GH1 β-1,4-P-glycosidase was to biochemically synthesize the appropriate phosphorylated substrates from the corresponding not phosphorylated ones. Thus, we synthesized cellobiose-6-phosphate, gentiobiose-6-phosphate, lactose-6-phosphate and the pseudo-substrate 4-nitrophenyl β-D-glucopyranoside-6-phosphate (Figure 3.3.1). The validation of the production and the level of purification were validated through thin layer chromatography as shown in Figure 3.3.2.

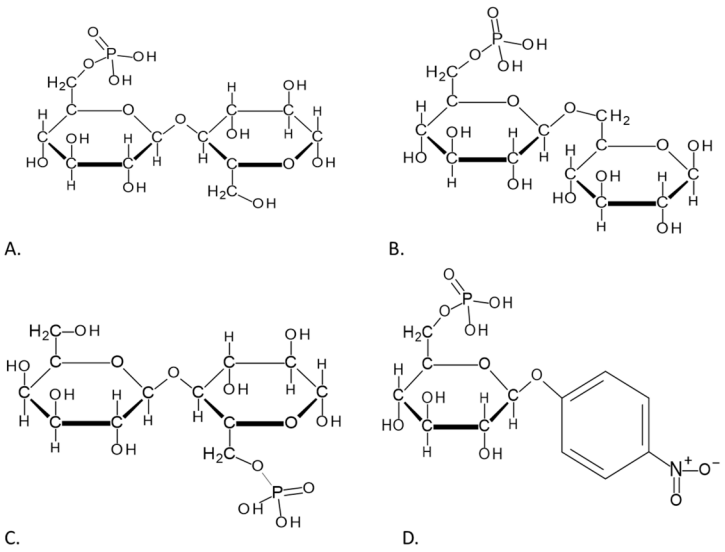


Figure 3.3.1: Chemical structure of the phosphorylated substrates used for the characterization of the GH1 β-6P-1,4-glycosidase¨: cellobiose-6-phosphate (A),gentiobisoe-6-phosphate (B),lactose-6-phosphate (C),paranitropheny-glucopyranoside-6-phosphate (D).

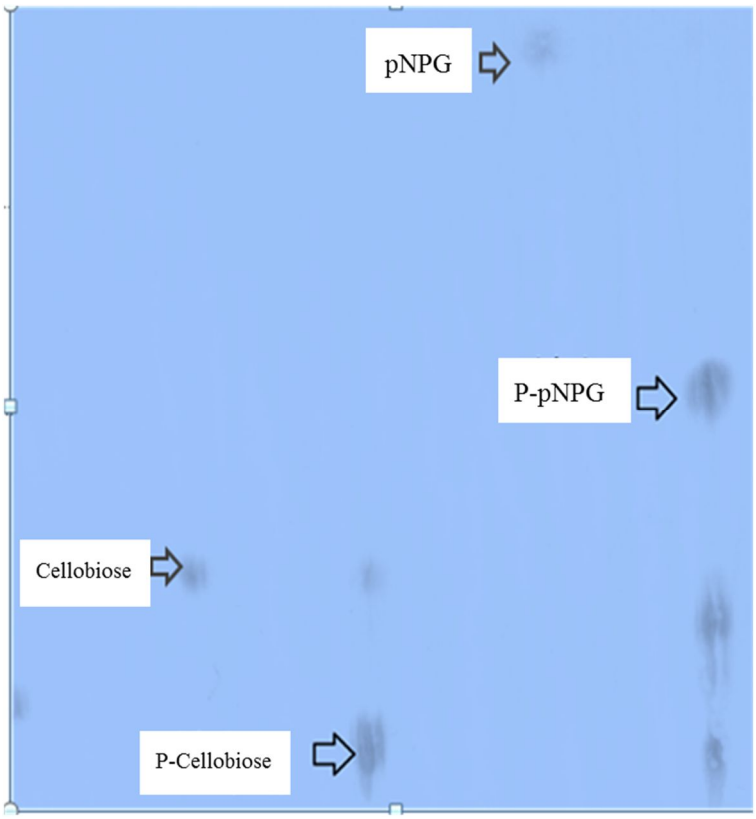


Figure 3.3.2: Thin layer chromatography for the validation of production and the estimation of purity of P-Cellobiose and P-pnpG.

Using 4-nitrophenyl β-D-glucopyranoside-6-phosphate as substrate and for 10 min incubation time, we show that GH1 β-P-glycosidase revealed the highest activity at 65 ˚C while it also retained more than 80% of the maximum activity at temperatures ranging between 60 ˚C and 70 ˚C (Figure 3.3.3).

Regarding the thermal stability, the enzyme retained almost 100 % of its initial activity upon incubation at 55 ˚C for 4 hours. In contrast, at 60 ˚C and 65 ˚C the enzymatic activity was decreased from the first hour of incubation to approximately 50 % of the initial activity (Figure 3.3.4). The effect of pH to enzymatic activity, was tested fro the pH range of 4 to 8. As can be seen in Figure 3.3.5, b-P-glycosidase revealed the optimum activity at pH 6 while it revealed aproximately 50 % of the optimum activity at pH values 5 and 7.

The enzymatic specificity of the enzyme was tested on the phosphorylated substrates described above as well as on their not phosphorylated counterparts. Upon incubation of the enzyme with the different substrates, we did not observe any enzymatic activity against the not phosphorylated oligosaccharides. In contrast, the GH1 β-P-glycosidase presents a broad specificity against 6-phospho di-hexoses, as it is able to hydrolyze all the related substrates tested (Table 3.3.1). In particular, even though the values are of the order of magnitude, the enzyme revealed the highest specificity to gentiobiose-6-phosphate and the lowest one to lactose-6-phosphate.

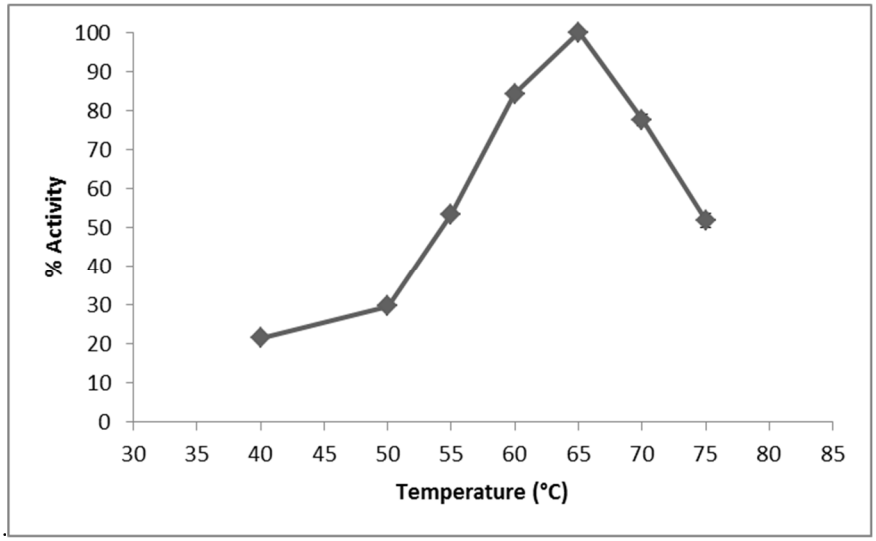


Figure 3.3.3: Activity of the GH1 β-P-glycosidase in correlation with temperature. The enzyme revealed the maximum activity at 60 ºC and more than 80 ºC of the maximum activity between 60 ºC and 70 ºC. The measurements were performed using 4-nitrophenyl β-D-glucopyranoside-6-phosphate as substrate and 10 min incubation.

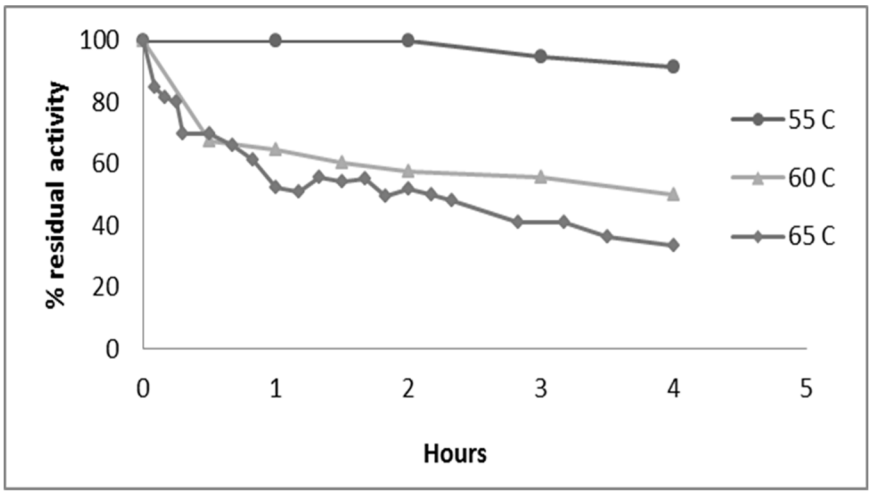


Figure 3.3.4: Thermal stability of the GH1 β-P-glycosidase at 55 ˚C, 60 ˚C and 65 ˚C for up to 4 hours incubation.

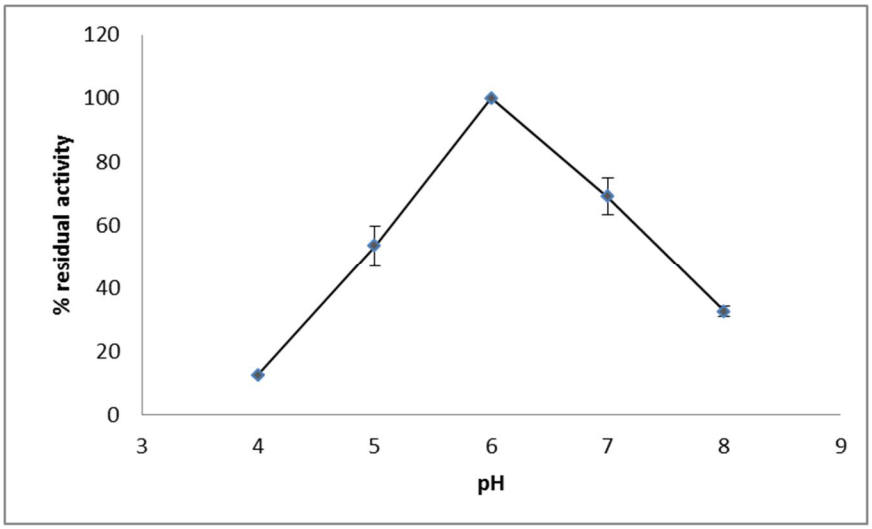


Figure 3.3.5: Activity of the GH1-*β*-P-glycosidase in correlation with pH. The optimum activity was observed at pH 6.

Lastly, we examined the impact of the presence of various mono- and di-saccharides to the reaction rate. Upon incubation of the enzyme with paranitrpopheny-glucopyranoside-6-phosphate as substrate and relatively high concentration of each putative inhibitor, the maximum decrease at the hydrolysis rate observed, was approximately 20% for P-fructose and lactose (Table 3.3.2). Thus we conclude that there is not any significant product inhibition or any type of competitive inhibition from the molecules tested.

Table 3.3.1: Michaelis-Menten constant (Km) and Maximum reaction rate (Vmax) of the hydrolyses of the phosphorylated disaccharides by the *β*-1,4-P-glycosidase.

Substrate	**K_m (mM)**	**V_{max} (µmol/min/mg)**
Cellobiose-6-phosphate	6.93±0.51	1.97±0.24
Gentiobiose-6-phosphate	4.13±2.11	1.01±0.05
Lactose-6-phosphate	9.65±0.42	1.08±0.83
P-paranitrophenyl-	0.1±0.01	1.27±0.05

Table 3.3.2: Effect of various mono- and di-saccharides on the enzymatic activity.

Sugar	Retaining activity %	Sugar	Retaining activity %
-	100 ± 2	Glucose 6-Phospahte	95 ± 1
Cellobiose	92 ± 4	Galactose	92 ± 2
Lactose	83 ± 4	Lactose-6-phosphate	84 ± 0
Fructose	98 ± 2	Fructose-6-phosphate	83 ± 4
Glucose	95 ± 3		

3.3.1.2 Structural characterization of the GH1 β-1,4-P-glycosidase

The GH1 β-1,4-P-glycosidase is considered a potential candidate for various biotechnological applications since it hydrolyzes the glycosidic bond of a broad range of β-P-di-hexoses, presents relatively high turnover rates and significant thermostability in temperatures around 60˚C. In this context, we initiated structural studies in order to elucidate the enzymatic function in the free form of the enzyme as well as in complex with its substrate analogues.

Figure 3.3.6, shows the crystals obtained using different crystallization conditions through the optimization of the process. The rhombohedral crystals obtained using 7 mg/ml protein, 0.2M Na iodide, 0.1M Bis-Tris propane pH 7.5, 20% PEG 3350 were considered adequate to continuewith the X-Ray analysis.

3.3.2 Biochemical characterization of the GH5 cellulase, GH10 β-1,4-xylanases, GH39, GH43 and GH52 β-1,4-xylosidases

Since the hydrolytic enzymes of *Geobacillus* are relatively conserved, we identified the closest representatives of the isolated ones, with available biochemical data (Table 3.3.3). However, the available information was not enough for the comparison of the isolated enzymes in order to

combine them in hydrolysis experiments. Thus we estimated, for all of them the activity and stability in relation with temperature (Table 3.3.4) and pH (Table 3.3.5).

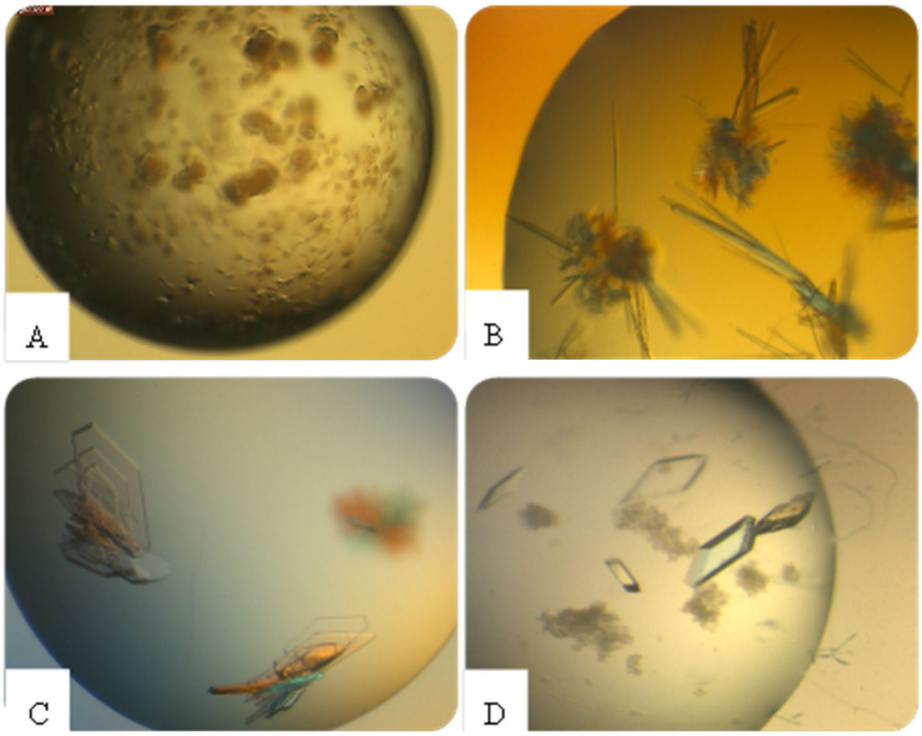


Figure 3.3.6: Crystals obtained from the free form of GH1 *β*-P-glycosidase. Crystals (urchins) of GH1 (15mg/ml) in 100 mM MES pH 6.5, 200 mM MgCl2, 25% PEG 4000 (A), Crystals (cluster of needles) of GH1 (15mg/ml) in 100 mM HEPES pH 7.5, 200 mM MgCl2, 10% PEG 4000 (B), Crystals (thin plates) of GH1 (10mg/ml), in 0.2M Na acetate, 0.1M Bis-Tris propane pH 7.5, 20% PEG 3350 (C), Single crystals (rhombohedral) of GH1 (7mg/ml), in 0.2M Na iodide, 0.1M Bis-Tris propane pH 7.5, 20% PEG 3350 (D).

All the cell associated enzymes presented a temperature optimum at around 60 ˚C to 70 ˚C while in contrast, the extracellular xylanase presented the maximum activity ten degrees higher, at 80 ˚C (Table 3.3.4). Similarly, the extracellular GH10 xylanase appeared to be significantly more stable than the rest of the enzymes even though all of them retained more than 70 % of their initial activity upon 6 h incubation at 60 ˚C (Table 3.3.4). The temperature parameters for the cell associated enzymes are in accordance with the optimum growth temperature of the majority of *Geobacillus* species which is 60 ˚C. The optimum activity at higher temperatures and the greater thermal stability are also anticipated for the excreted xylanase, since the tolerance in higher temperatures than the temperature of bacterial growth, is observed for the majority of the extracellular enzymes.

Regarding their pH profile, the cloned enzymes presented maximum activity at neutral to slightly acid pH values. Particularly, the GH5 cellulase presented optimum activity at pH 5, all xylosidases performed optimally between 5.5 and 6.5 while the two xylanases at 7.

Table 3.3.3: *Geobacillus* thermoenzymes isolated in the present study and % amino acid sequence identity with the closest relative with available biochemical data.

Enzyme	**GH**	**Bacterial source (*Geobacillus* Strain)**	**% Amino acid sequence identity with the most similar characterized enzyme**	**Uniprot Acc. No**	**Bacterial origin of the closest characterized sequence**	**Ref*.**
Cellulase	5	SP50	98.8	Q9ZFM2	*G. stearothermophilus 70PC53*	1
Xylanase (membrane associated)	10	SP24	99.4	Q3YBZ9	*G. thermoleovorans* MT-1	2
Xylanase (extracelllar)	10	SP50	95	JF792185 *	*G. stearothermophilus 71*	3
*β-Xylosidase (*cell associated)	39	SP24	98.8	Q9ZFM2	G.stearothermophilus T-6	4
β-Xylosidase (cell associated)	43	SP24	96	Q09LX0	G.stearothermophilus T-6	5
β-Xylosidase (cell associated)	52	SP24	97	KC412009 .1*	*G. stearothermophilus* 1A05583	6

There are not available Uniprot Identifiers for the entries so the GenBank Numbers are given. Ref;1: (Ng, Li et al. 2009), 2: (Wu, Liu et al. 2006), 3: (Canakci, Cevher et al. 2012), 4: (Shulami, Gat et al. 1999, Czjzek, Bravman et al. 2004, Czjzek, Ben David et al. 2005), 5: (Shallom, Leon et al. 2004, Brux, Ben-David et al. 2006), 6: (Huang, Liu et al. 2014)

Table 3.3.4:Activity and Stability of the enzymes isolated from the selected *Geobacillus* strains in correlation with temperature.

Enzyme	GH family	Bacterial source (*Geobacillus* Strain)	T_{opt} (˚C)	Temperature range for ≥ 50 % of the maximum activity (˚C)	Comments on the thermal stability of the enzyme
Cellulase	5	SP50	65	40 – 75	- Stable at 50 °C for at least one day - Stable for at least 6 h incubation at 65 °C - Upon 6 h incubation at 70 °C, retains 70% of the initial activity - Upon 6 h incubation at 75 °C retains 35% of the initial activity
Xylanase (membrane associated)	10	SP24	60	45-75	- Stable at 50 °C for at least one day - Upon 4h incubation at 60 °C retains the 80% of the maximum activity - Rapid deactivation at 70 °C
Xylanase (extracelllar)	10	SP50	80	50-90	- Stable at 50°C for several days - Almost 100% of the initial activity upon 10 h incubation at 60 °C - 50% loss of activity upon 1 h incubation at 70 °C - Rapid deactivation at 75 °C
β-Xylosidase	39	SP24	70	52-75	- Stable at 50 °C for a few days - Upon 3 d incubation at 60 °C retains the 90% of the initial activity - Upon 3 days incubation at 70 °C retains the 70 % of the initial activity
β-Xylosidase	43	SP24	65	45 – 75	- Stable at 50 °C for several days - Upon 4 h incubation at 60 °C retains 80% of the initial activity - Upon 1 h incubation at 70 °C retains 30 % of the initial activity
β-Xylosidase	52	SP24	70	50 – 75	- Stable at 50 °C for several days - Upon 4 h incubation at 60 °C retains 95% of the initial activity - Upon 4h incubation at 70 °C retains 45% of the initial activity

Table 3.3.5: Activity and Stability of the enzymes isolated from the selected *Geobacillus* strains in correlation with pH.

Enzyme	**GH family**	**Bacterial source (*Geobacillus* Strain)**	**pH_{opt}**	**pH range for ≥ 50 % of the maximum activity**	**Comments on the pH stability of the enzyme**
Cellulase	5	SP50	5	4 – 8	- Stable for at least 16 h for pH values between 5 and 9
Xylanase (membrane associated)	10	SP24	7	5.5 – 9	- Stable for at least 4 h incubation at pH values from 5 to 9
Xylanase (extracelllar)	10	SP50	7	5.4 - 9.4	- Stable for at least 20 h at pH values between 5 and 8
β-Xylosidase	39	SP24	6.5	5 – 8	- Stable for at least 1 d at pH values between 5 and 9
β-Xylosidase	43	SP24	6	4 – 7.5	- Stable for at least 5 h at pH values between 5.5 and 9
β-Xylosidase	52	SP24	5.5	4.5 - 8	- Stable for at least 5 h at pH values between 5 and 9

3.4 Construction and evaluation of a tetravalent designer thermocellulosome

3.4.1 Components of the designer thermocellulosome

After the isolation and characterization of the enzymes described in the previous paragraphs, our next aim was to incorporate the free thermophilic enzymes of *Geobacillus i*nto the multienzyme complex of designer thermocellulosome. We wanted to validate the functionality of the enzymes at the complex form as well as the performance of the cellulosome in high temperatures. Finally, we compared the hydrolytic activity of the free versus the complex enzymes on various byproducts of the agricultural industry.

In order for the cellulosome to be effective on hemicellulose we chose to incorporate enzymes with the four key (hemi)cellulolytic activites: cellulase, xylanase, *β*-xylosidase and *β*-glucosidase. Thus, we chose the GH10 xylanases, the GH39, GH43 and GH52 *β*-xylosidases and the GH5 cellulase obtained from our collection of *Geobacillus* strains. Since *Geobacilli* do not have *β*-glucosidases, we used the one from the genome of *Caldicellulosiruptor Saccharolyticus (Hong, Kim et al. 2009).*

The tetravalent scaffoldin used (Figure 3.4.1), consists of three cohesins of mesophilic origin (*Bacteroides cellulosolvens, Thermobifida fusca* and *Ruminococcus flavefaciens*) as well as a cohesin and a CBM moiety from the thermophilic bacterium *Clostridium thermocellum.* This scaffoldin has already been combined with other mesophilic enzymes for plant biomass hydrolysis studies (Moraïs, Barak et al. 2010, Moraïs, Barak et al. 2011).

For the complex formation we fused the thermoenzymes with the corresponding dockerins at the C-terminus. Since we wanted to estimate the effect of the mesophilic dockerins to the stability of the thermophilic enzymes we produced more than one chimaeric combination for each enzyme, the summary of which are given at Table 3.4.1. Based on thermal stability data described in the following paragraphs, we chose four out of the twelve chimaeric enzymes (Figure 3.4.1): GH5 cellulase-dockerin *b*, GH1 *β*-glucosidase-dockerin *a*, GH10 xylanase-dockerin *f*, GH52 xylosidase-dockerin *a*.

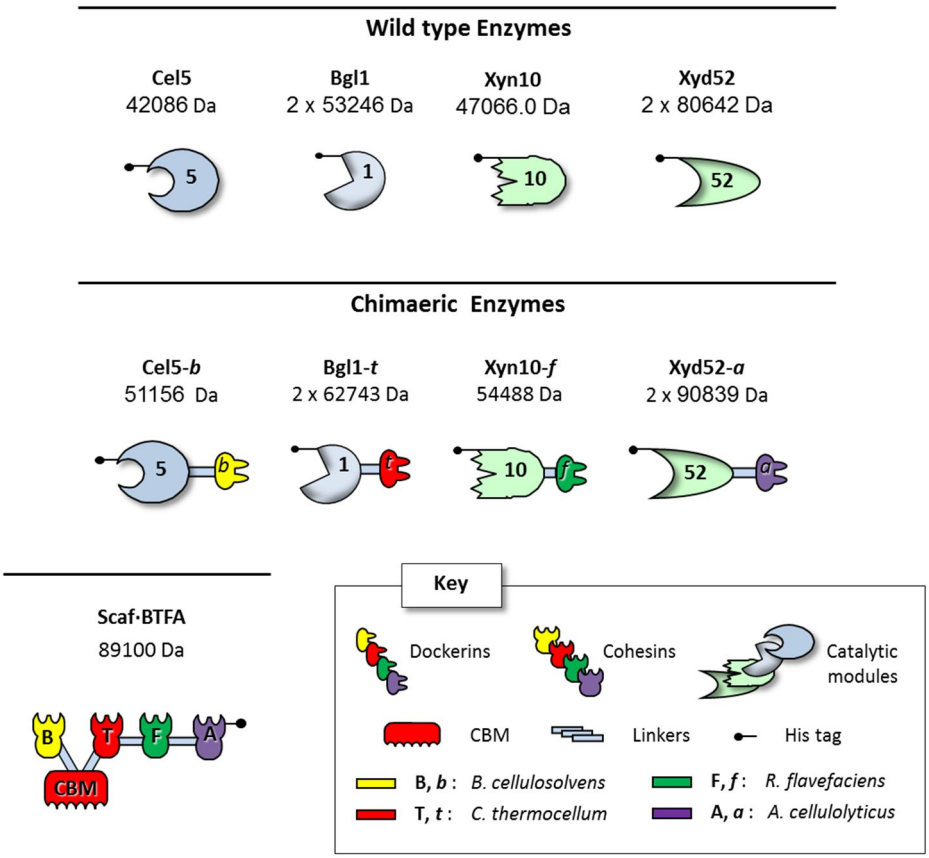


Figure 3.4.1: Recombinant proteins used in this study. The different modules are represented by pictograms as shown in the key. The bacterial source of each protein is indicated as follows: yellow; Bacteroides cellulosovens, red; Clostridium thermocellum, green; Ruminococcus flavefaciens, purple; Acetivibrio cellulolyticus, light red; *Geobacillus kaustophilus* SP50, light green; Caldicellulosiruptor saccharolyticus, light puprle; *G. kaustophilus* SP24. Numbers into the modules refer to the corresponding CAZY family they belong (GH5, GH1, GH10 and GH52). In addition to the color code, upper case letters (B, T, F and A) indicate the source of the cohesin modules and lower case letters (b, t, f and a) indicate the source of the dockerin modules. For every hydrolase the MW of the monomer and the number of subunits of the functional enzyme are indicated (Galanopoulou, Moraïs et al. 2016).

3.4.2 Effect of temperature at activity and stability of the free wild type and chimaeric enzymes

The first step to validate that the free thermoenzymes can be incorporated into the cellulosome complex, was to ensure that the biochemical properties of the enzymes are not altered upon their fusion with the dockerins. As shown on Table 3.4.2, at the majority of the cases the addition of the dockerins resulted in relatively minor alterations of the activity and stability properties of the enzymes. Exceptions were the GH10 cell associated xylanase from *Geobacillus* SP24. Additionally, *β*-glycosidase from *Clostridum thermocellum* was not as thermostable as the rest of

the enzymes so it was excluded from further experiments of incorporation into the thermocellulosome.

Thermoenzymes	Dockerins			
	b	**t**	**f**	**a**
Cel5 SP50	√	√	√	
BglA C.th		√		
BglA C.s		√		
Xyn10 SP50			√	
Xyn10 SP24		√	√	√
Xyd39 SP50				√
Xyd43 SP24				√
Xyd52 SP24				√

Table 3.4.1: Summary Table of the chimaeric enzyme-dockerin combinations constructed. Cel5 SP50: GH5 cellulase isolated from *Geobacillus* strain SP50, BglA C.th: GH1 *β*-glucosidase isolated from *Clostridium thermocellum*, BglA C.s: GH1 *β*-glucosidase from *Caldicellulosiruptor saccharolyticus*, Xyn10 SP50: GH10 xylanase isolated from *Geobacillus* SP50, Xyn10 SP24: GH10 xylanase isolated from *Geobacillus* SP24, Xyd39 SP50, GH39 *β*-xylosidase isolated from *Geobacillus* SP50, Xyd43 SP50:GH43 *β*-xylosidase isolated from *Geobacillus* SP24.

3.4.3 Thermostability of BTFA and monovalent enzymes in complex with BTFA

In order to evaluate the thermostability of Scaf·BTFA, aliquots of the protein were incubated at 50 °C, 55 °C, 60 °C and 65 °C for various time intervals. Each sample was subjected to electrophoresis under non-denaturing conditions, and the relative intensities of the bands were compared (Figure 3.4.2). The same procedure was followed for Scaf·BTFA in complex with one chimaeric hydrolase at a time, at 60 °C and 65 °C (Figure 3.4.3). Scaf·BTFA appeared relatively stable at 50 °C and 55 °C as judged from the relative band densities that were decreased by 50% following approximately 24-h and 6-h incubation at these two temperatures, respectively.

However the scaffoldin did not show significant structural thermostability when incubated at higher temperatures, since the protein band reduced more than 60 % upon 2-h incubation at 60 °C and almost 90 % at 65 °C (Figure 3.4.2).

Table 3.4.2: Biochemical properties of free wild type and chimaeric enzymes .

Hydrolase acronym	**T_{opt}**	**pH_{opt}**	**% Residual activity, 60 °C**		**% Residual activity, 65 °C**		**Specific activity at 60 °C and pH 6.5 (U/nmole)**
			6 h	**12 h**	**6h**	**12 h**	
Bgl1 *C.th*	65	6	75	40	-	-	-
Bgl C.th -t	-	-	-	-	-	-	-
BglA C.s	70	5.5	100	100	100	96	9.7 ± 0.5
BglA C.s-*t*	70	6.5	100	100	93	81	9.4 ± 0.2
Cel5 SP50	65	5	100	100	90	60	2.8 ± 0.2
Cel5 SP50-*b*	65	6	96	89	81	64	3.0 ± 0.3
Cel5 SP50-*t*	65	6	97	90	82	60	3.1 ± 0.4
Cel5 SP50-*f*	65	5.5	97	99	80	62	2.9 ± 0.1
Xyd39 SP24	70	6.5					

Xyd39 SP24-*a*	70	6.5					
Xyd43 SP24	65	6.5					
Xyd43 SP24-*a*	65	6.5					
Xyd52 SP24	70	6	94	62	73	41	5.1 ± 0.6
Xyd52 SP24 –*a*	65	6	82	59	57	34	4.4 ± 0.3
Xyn10 SP24 (cell-associated)	60	7	70	40	30	0	
Xyn10 SP24-*t* (cell-associated)	-	-	0	0	-	-	-
Xyn10 SP24-*f* (cell-associated)	-	-	0	0	-	-	-
Xyn10 SP24-*a* (cell-associated)	-	-	0	0	-	-	-
Xyn10 SP50 (extr)	75	7	100	100	100	100	3.0 ± 0.3
Xyn10 SP50-*f* (extr)	70	6.5	100	100	82	61	3.1 ± 0.4

In contrast, the complex of Scaf·BTFA with each of the chimaeric enzymes revealed significantly higher stability at 60 °C and 65 °C; the monovalent cellulosome of Scaf·BTFA with all the chimaeric Xyn10 sp50-dockerins were the most thermostable, since even after 12-h of incubation at 60 °C the protein bands in non-denaturing PAGE remained unaltered. Moreover,

the complexes proved to be stable at 65 °C since even after 6 h of incubation, the bands retained 60 % of the initial density (Figure 3.4.3). Cel5-b showed significant thermostability as well, since the band intensity was decreased to 80 % of the initial after 12 h at 60 °C and to 50 % after 2 h at 65 °C (Figure 3.4.3).

The complexes of Scaf·BTFA with Bgl1 *C.th*-t or all chimaeric xylosidase-dockerins tested, also appeared more thermostable than the scaffoldin itself. Although they did not reach the high stability levels of the two former complexes. Among all *β*-xylosidases, the chimaeric GH52 *β*-xylosidase revealed the greater improvement on the thermostability of the Scaf·BTFA –enzyme complex and thus it was chosen to be incorporated into the tetravalent complex.

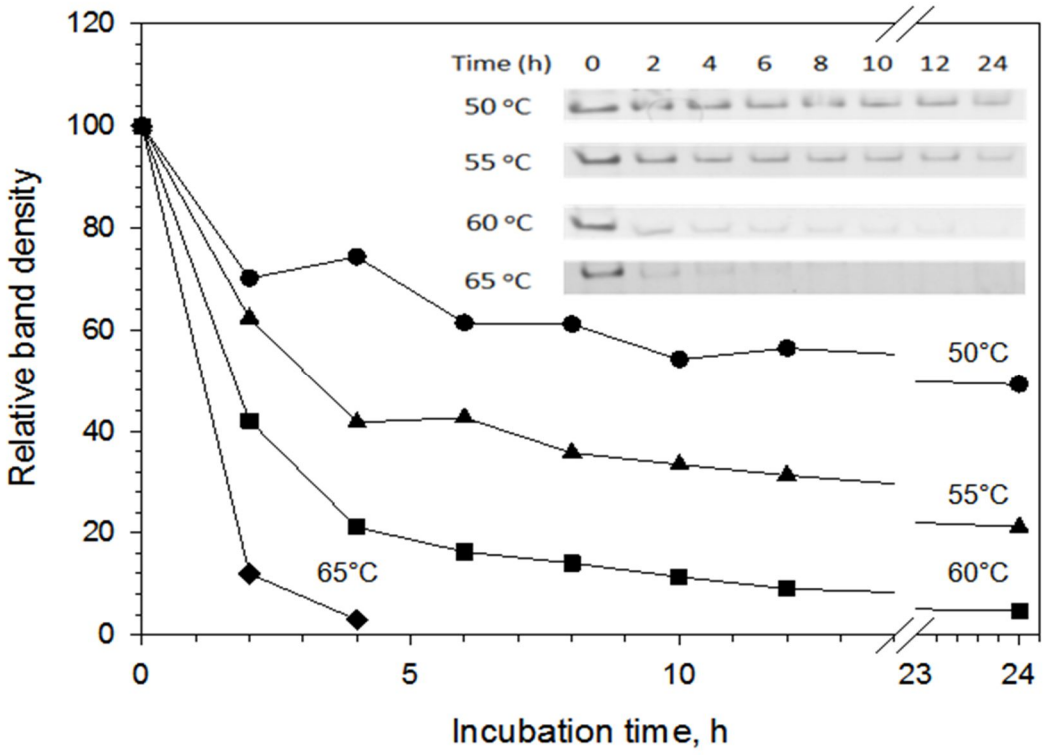


Figure 3.4.2: Densitometric analysis for the determination of of Scaf·BTFA thermal stability. Protein samples were incubated for various time intervals at 50 °C (■), 55 °C (▲), 60 °C (■) and 65 °C (◆) and subjected to non-denaturing PAGE (inset). Band image densities were determined by ImageJ (Galanopoulou, Moraïs et al. 2016).

In particular, both Xyd52-a (as well as all the other Xyd52-dockerins tested) and Bgl1 *C.s*-t, maintained significant levels of thermostability at 60 °C; for the first 2 h at this temperature their

band densities remained practically unaffected and reduced to 50% after 10 and 6 h incubation for Xyd52-a and Bgl1-t, respectively (Figure 3.4.3).

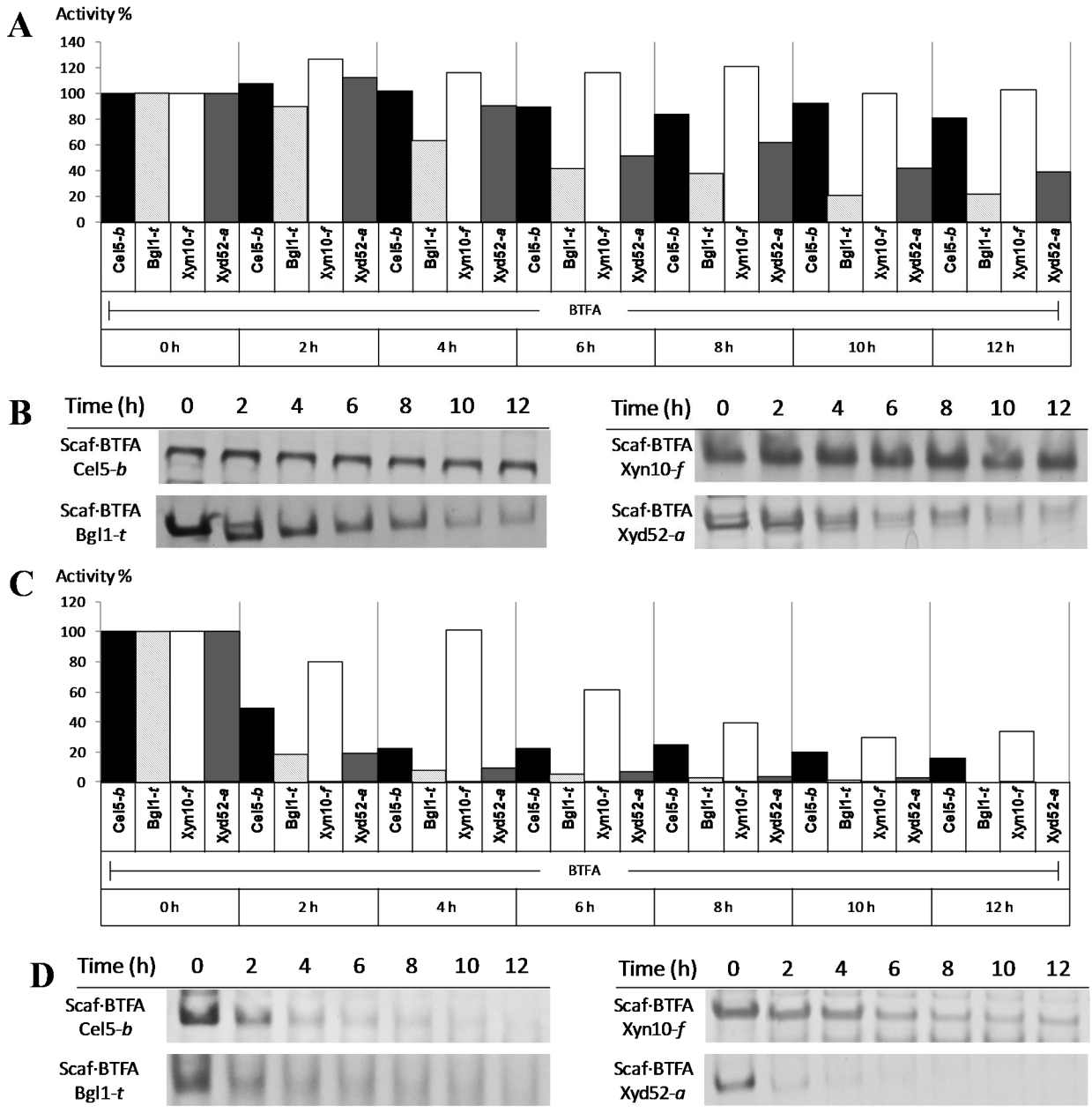


Figure 3.4.3: Structural stability of Scaf·BTFA in complex with individual chimaeric enzymes at 60 °C (A, B) and 65 °C (C, D) as determined by non-denaturing PAGE. A and C are the relative band densities at various time intervals as quantified by ImageJ. B and D are the corresponding gel bands (Galanopoulou, Moraïs et al. 2016).

3.4.3 Thermostability and functionality of CBM

The stability and functionality of the CBM of Scaf•BTFA was determined by evaluating the ability of its complex with Xyn10.sp50-f-since it formed the most stable complex- to bind to microcrystaline cellulose (Avicel) following incubation at 60, 65 and 70 °C. Figure 3.4.4 shows the distribution of measured xylanase activity between the Avicel-bound fraction and its supernatant at various time intervals. At 60 °C and up to 12 h, more than 90% of the total activity was bound to Avicel, indicating a stable complex between the cellulosomal CBM and Avicel. This pattern was more or less maintained at 65 °C but only up to the first 4 to 6 hours of incubation, since after that time, significant xylanase activity was released into the supernatant fluids. At 70 °C, the Avicel-bound activity was reduced at a significantly higher rate with concomitant increase in xylanase levels detected in the supernatant. The total xylanase activity at this temperature was also reduced with time, a fact that infers simultaneous enzyme deactivation.

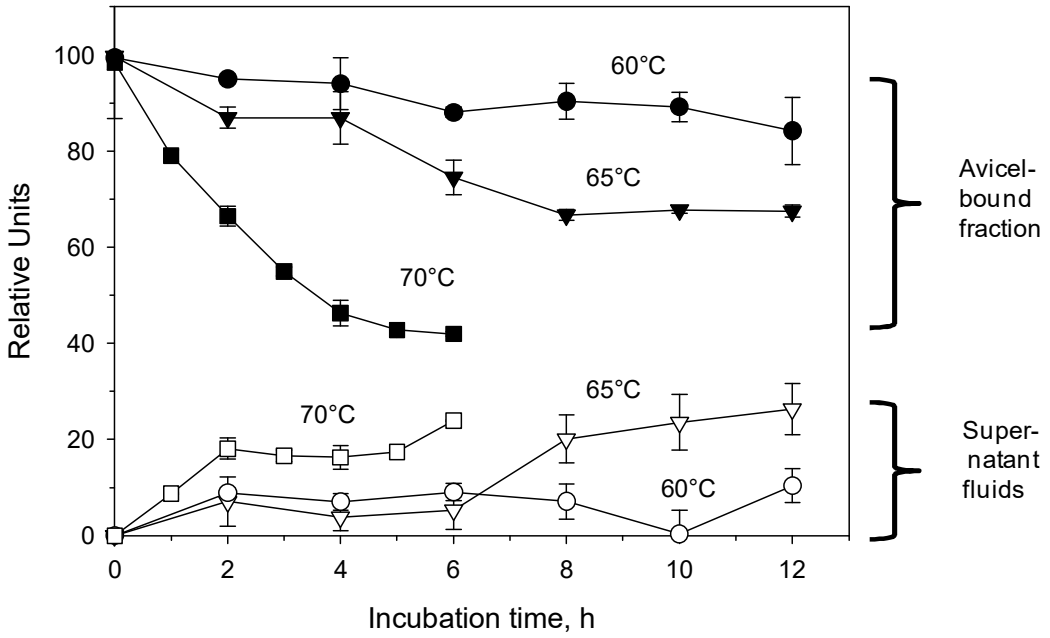


Figure 3.4.4: Functional stability of the CBM. Scaf·BTFA complexed with Xyn10.sp50-*f* was bound to Avicel, and the complex was incubated at 60 °C (circles), 65 °C (triangles) and 70 °C (squares). At specific time intervals, xylanase activity both on the Avicel-bound fraction (closed symbols) as well as in the supernatant fluids (open symbols) was measured. Data represent the mean of triplicate samples ±SD (Galanopoulou, Moraïs et al. 2016).

3.4.4 Thermal stability of the tetravalent complex

Based on all the above, Cel5-b, Bgl1 *C.s*-t, Xyn10sp50-f and Xyd52-a where chosen to be incorporated into the cellulosome complex. The thermal stability of the tetravalent cellulosome was investigated at 60 °C via gel-filtration chromatography on a Sephacryl S-200 column. Aliquots of the fully assembled complex were introduced into the column following 2-, 6-, and 24-h incubation at the above temperature, and the A_{280} of the eluates was followed (Figure 3.4.5). After 2 h of incubation, the chromatogram of the sample was almost identical to that of the control and revealed only one major peak that corresponds to the void volume of the column (MWs > 250 kDa) with only minor indistinguishable peaks appearing at greater elution volumes. The chromatogram of the 6-h incubation sample still showed a major peak at the void volume, accounting for almost two thirds of the total chromatogram area, thus indicating that the structural integrity of the tetravalent complex was largely maintained. In addition, distinct peaks appeared at higher elution volumes that correspond to the cellulosomal components. At 24 h, these dissociation peaks appeared sharper and indicative of the scaffoldin and individual chimeric enzymes (Figure 3.4.5). It is notable that even at this prolonged incubation time almost one third of the total chromatogram area still appears in the void volume of the column.

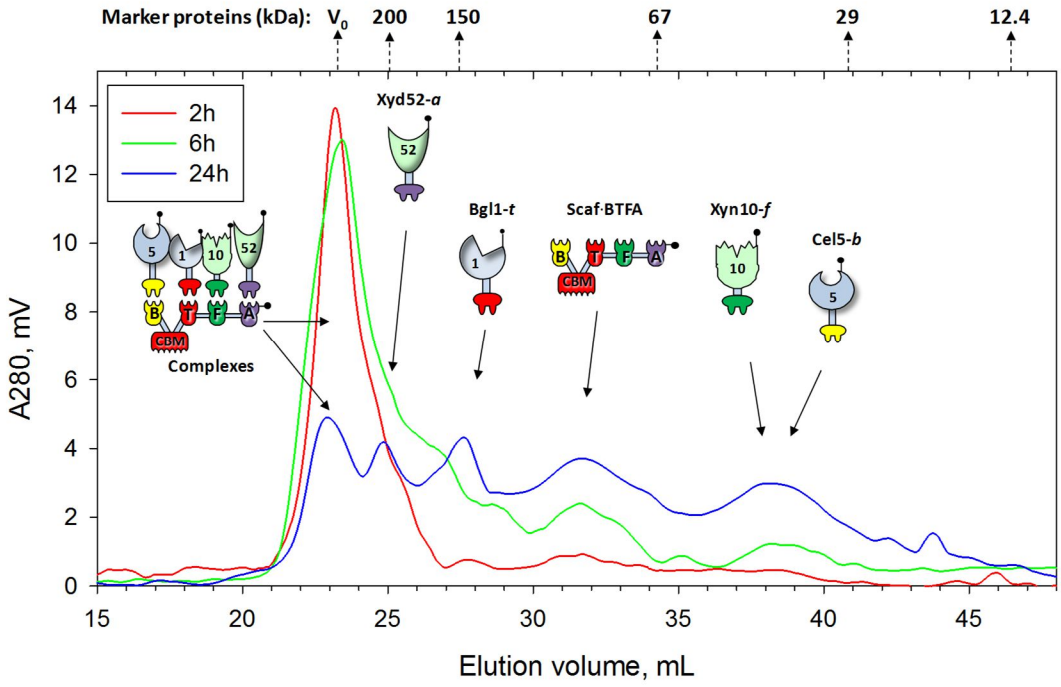


Figure 3.4.5: Analysis of the stability of the fully assembled designer cellulosome complex by gel-filtration chromatography on a Sephacryl S-200 column. Samples of equal volume and initial concentration were introduced into the column following incubation for 2, 6 and 24 h at 60 °C. The expected elution volumes and peak identification of the various cellulosomal modules are shown. The column was calibrated using the following protein markers (Sigma-Aldrich): Blue dextran (void volume, V0), *β*-amylase (200 kDa), alcohol dehydrogenase (150 kDa), bovine serum albumin (66 kDa), carbonic anhydrase (29 kDa) and cytochrome c (12.4 kDa). The 0-h chromatogram almost coincided with the 2 h sample and is omitted for clarity (Galanopoulou, Moraïs et al. 2016).

3.4.5 Hydrolytic performance of the designer thermocellulosome

The overall hydrolytic performance of the tetravalent cellulosome was evaluated on alkali-pretreated corn stover and was compared with that of equimolar amounts of free chimaeric enzymes as well as their wild-type counterparts (Figure 3.4.6). For the first 6 h of the reaction, the hydrolytic productivity of the cellulosome proved to be significantly higher than both the free chimaeric and the wild-type enzymes. After 3 h hydrolysis time, the average release of reducing sugars by the cellulosome was 65% and 46% higher than those of the wild-type and chimaeric enzymes, respectively. The corresponding percentages of enhancement after 6 h of reaction were 40% and 53%. However, no differences were observed after 24 h of incubation, and the concentrations of released reducing sugars among all three enzymatic reactions were very similar.

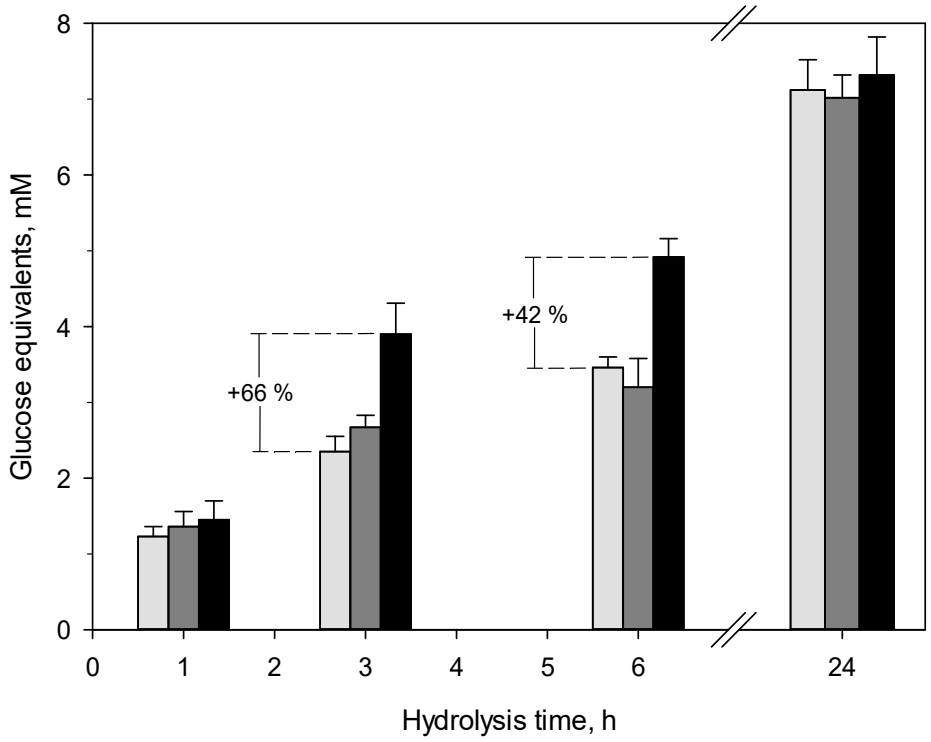


Figure 3.4.6: Hydrolysis of alkaline-pretreated corn stover at 60 °C, pH 6.5. Light gray bars, wild-type uncomplexed enzymes; Dark gray bars, chimaeric enzymes; Black bars, designer cellulosome. Error bars represent the SD among four independent replicates (Galanopoulou, Moraïs et al. 2016).

3.5 GENOME ANALYSIS OF THE CELLULOLYTIC STRAIN '*GEOBACILLUS ICIGIANUS* SP50'

3.5.1 '*Geobacillus icigianus* SP50' is able to grow and hydrolyse *β-D-glucans*

While *Geobacilli* are well known for their xylanolytic potential and the corresponding systems have been extensively described (Shulami, Zaide et al. 2007, De Maayer, Brumm et al. 2014, Shulami, Shenker et al. 2014), there is only a small amount of reports regarding the ability of the genus to grow on *β-D*-glucans and hydrolyze *β*-glycosidic bonds of glucose-based polymers, like cellulose.

Intriguingly enough, '*Geobacillus icigianus* SP50' exibits the ability to grow on CMC as a sole carbon source (Figure 3.5.1) and furthermore, as already discussed in Paragraph 3.2.3, contains a gene encoding for a GH5 broad-specificity endocellulase. This enzyme in particular, acts on the *β*-1,4-D-glucosidic bonds of *β-1,4-D*-glucose polymers of CMC and the *β*-1,3-1,4-D-glucose bonds found in lichenan and barley (Ng, Li et al. 2009).

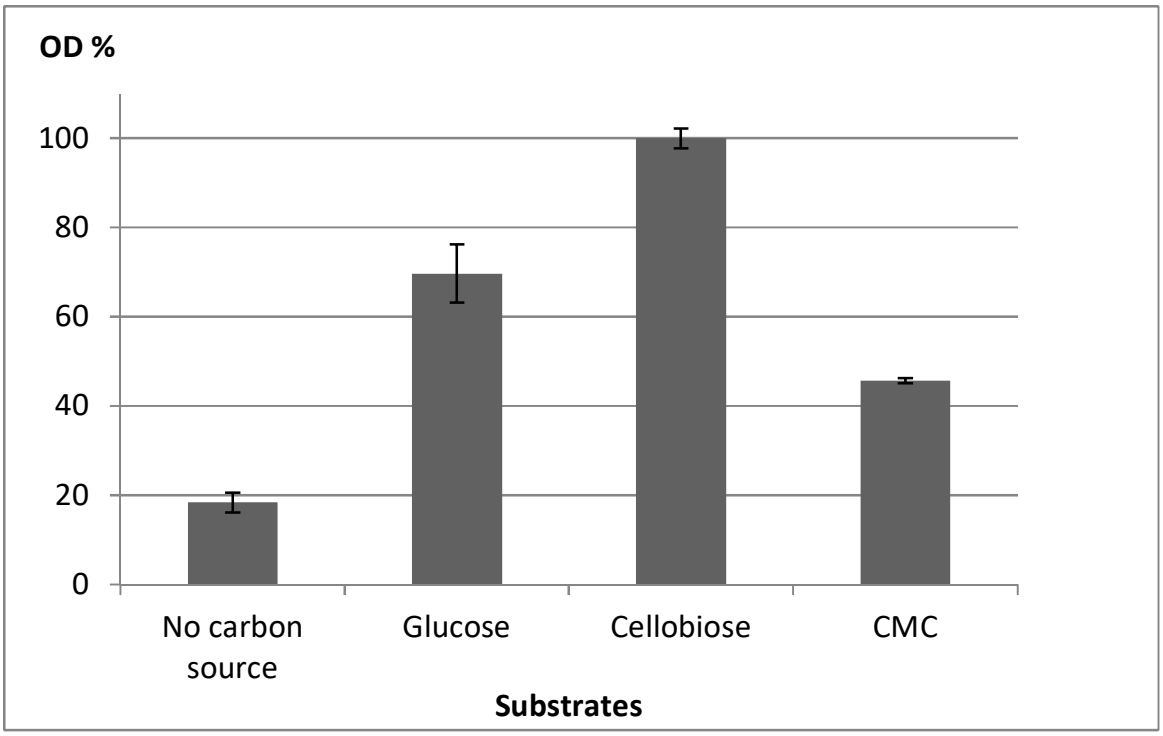


Figure 3.5.1: Growth of '*Geobacillus icigianus* SP50' in the presence of 5 g/L of CMC and its derivatives.

3.5.2 Genome and genomic features of '*Geobacillus icigianus* SP50'

The draft genome of *'Geobacillus icigianus* SP50' consisted of 54 contigs which were further combined to three scaffoldins using the complete genome of *Geobacillus* B4113_201601 (RefSeq assembly accession GCF_001587475.1) as template. The total size of the genome is 3,512,145 bp with 52.4 % GC content. Among the predicted 3,743 genes, 3,596 (96 %) found to encode proteins. Among the non-protein coding genes, are included 15 rRNA genes (8x5S, 4x16S and 3x23S), 71 tRNAs and 26 CRISPR arrays (4 characterized and 22 putative). 1739 Genes, 47 % of the protein-coding genes, were assigned to a SEED subsystem category; among these, 14% of them were related to the degradation utilization and metabolism of carbohydrates and 4% was related with membrane transport (Figure 3.5.2 A). Among the later, a significant portion is associated with ABC transporters, known to be involved in the transfer of proteins and oligosaccharides towards the cytoplasm (Quentin, Fichant et al. 1999, Shulami, Zaide et al. 2007, Shulami, Raz-Pasteur et al. 2011, Tabachnikov and Shoham 2013).

The relatively high percent, 14%, of the protein-coding genes related to the utilization and metabolism of carbohydrates (Figure 3.5.2 B) supplies '*Geobacillus icigianus* SP50' the necessary arsenal for the utilization of a broad range of mono-, di-, and oligosaccharides. Additionally, the strain contains genes that encode for butanol, acetoacetate, lactate and butyrate biosynthesis under anaerobic conditions.

In addition to the classification of the protein-coding genes to a SEED category, the CAZome of the strain was also investigated. As shown in Figure 3.5.3 and also Appendix II, Table 2, the strain contains 98 genes containing domains assigned to a CAZY family: 37 glycoside hydrolases (GH), 27 glycosyl transferases (GT), 16 carbohydrate esterases (CE), 4 auxiliary activity enzymes related to lignin degradation domains (AA) and 14 carbohydrate-binding modules (CBM) .In their majority, the above enzymes form single-domain proteins, something anticipated for proteins of *Geobacillus* strains.

As anticipated, many of the hydrolases and carbohydrate esterases of the strain are related with hemicellulose degradation. For example, the encoded two xylanases of GH10 family, the *β*-xylosidase and α-L-arabinofuranosidase of GH43 family, the two *β*-xylosidases from families GH39 and GH52 are all enzymes extensively characterized and used for various biotechnological applications (Zeigler 2001). Intriguingly, the genome of '*Geobacillus icigianus* SP50' contains also the gene of a GH5 subfamily 2 endoglucanase already characterized in the context of the present thesis and also by Ng and his colleagues (Ng, Li et al. 2009). Additionally, xylan and pectin can also be degraded via deacetylation from the enzymes containing CE1, CE3, CE4, CE7and CE12 domains.

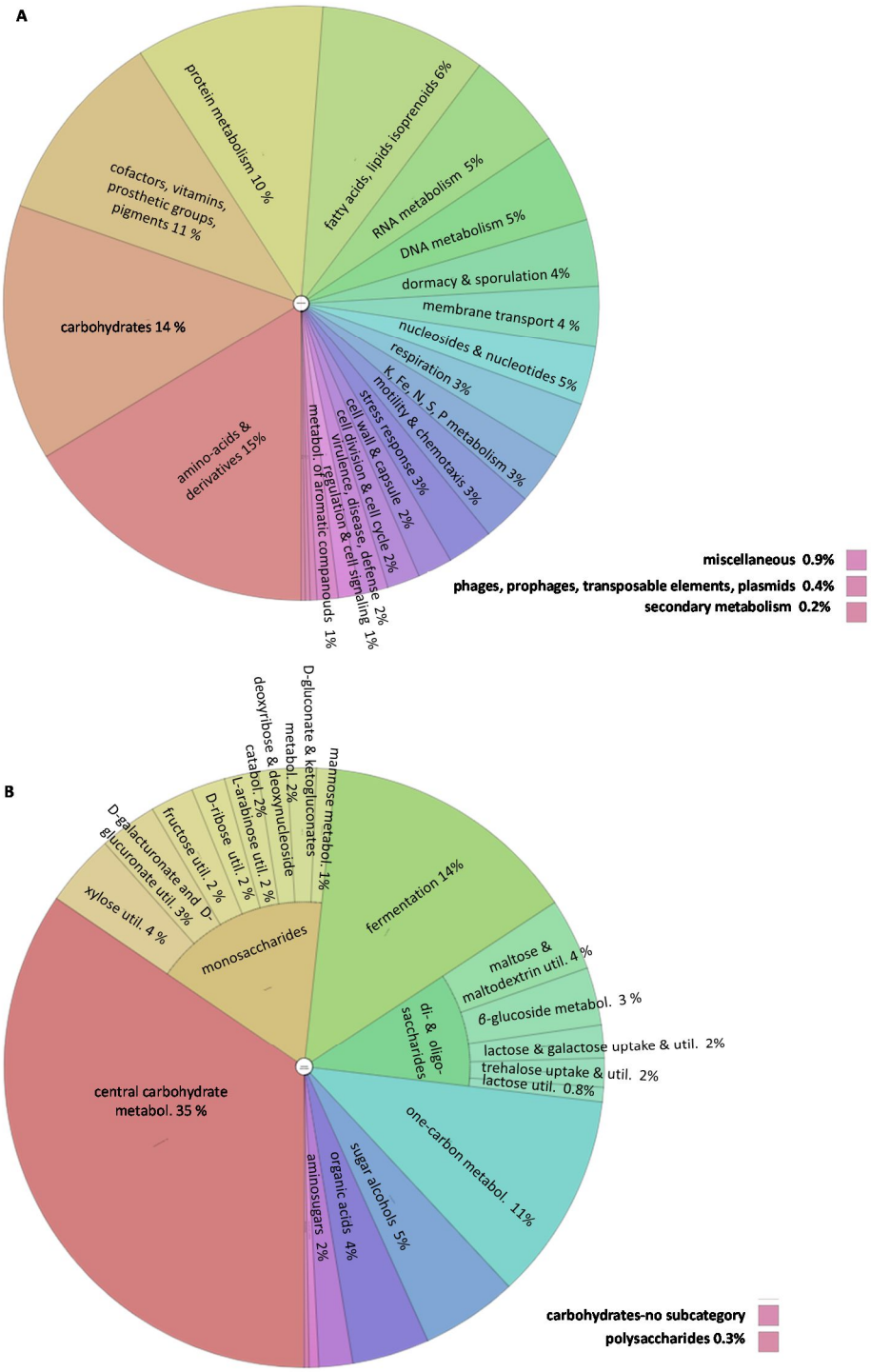


Figure 3.5.2: Distribution pattern of '*Geobacillus icigianus* SP50' proteins into functionally related protein families; Schematic representation of all called proteins into the different subsystem categories (A), Distribution of carbohydrate-related proteins into the different subcategories (B).

GH	1	2	4	5	10	13	18	23	32	36	39	43	51	52	67	73	109	130	133
No of genes	3	1	1	1	2	6	3	2	1	1	1	2	1	1	1	1	7	1	1

GT	2	4	8	28	51	81	83	94
No of genes	8	7	1	3	4	1	2	1

CBM	20	34	50
No of genes	1	2	11

CE	1	3	4	7	9	10	12	14
No of genes	3	3	3	2	1	2	1	1

AA	2	4	6
No of genes	1	2	1

Figure 3.5.3: CAZome of '*Geobacillus icigianus* SP50'. The families of Glycoside Hydrolases (GH),Carbohydrate-binding modules (CBM), Glycosyltransferases (GT), Carbohydrate Esterases (CE) and Auxiliary Activities (AA) are given as well as the number of genes contained in each family.

3.5.3 Systematic position of '*Geobacillus icigianus* SP50' within the genus *Geobacillus*

The systematic position of '*Geobacillus icigianus* SP50' up to the level of genus, was inferred taking into consideration all 16S rRNA sequences of the strain. Further classification of '*Geobacillus icigianus* SP50' within the genus, was achieved through two different approaches; (1) a maximum likelihood phylogenetic tree based on the alignment of *recN*, *rpoB* and *spo0A DNA* sequences, genes broadly used for the classification of *Geobacilli* (Zeigler 2005, Meintanis, Chalkou et al. 2008, Kuisiene, Raugalas et al. 2009, Aliyu, Lebre et al. 2016) (Figure 3.5.4). (2) The average amino acid sequence identity of their shared genes (Konstantinidis and Tiedje 2005) (Figure 3.5.5). Both methods were in agreement, confirming that '*Geobacillus icigianus* SP50', together with *G. icigianus*, *G. subterraneus* PSS2, *Geobacillus* sp. B4113_201601, forms the "*icigianus* clade" which is distinct but closely related to *G. stearothermophilus* and *G.kaustophilus* clades.

Taking into consideration the AAI values within the icigianus clade, all four '*Geobacillus icigianus* SP50', *G. icigianus* G1w1*, G. subterraneus* PSS2 and *Geobacillus* sp.B4113_201601 belong to the same species as they present more than 95% amino-acid identity among their shared genes (Konstantinidis and Tiedje 2005, T Konstantinidis, Rossello-Mora et al. 2017) (Table 3.5.1). Inside the cluster though, '*Geobacillus icigianus* SP50' appears the most distant to the other three -with AAI values slightly higher than 95 %- thus, it could probably form a separate subspecies.

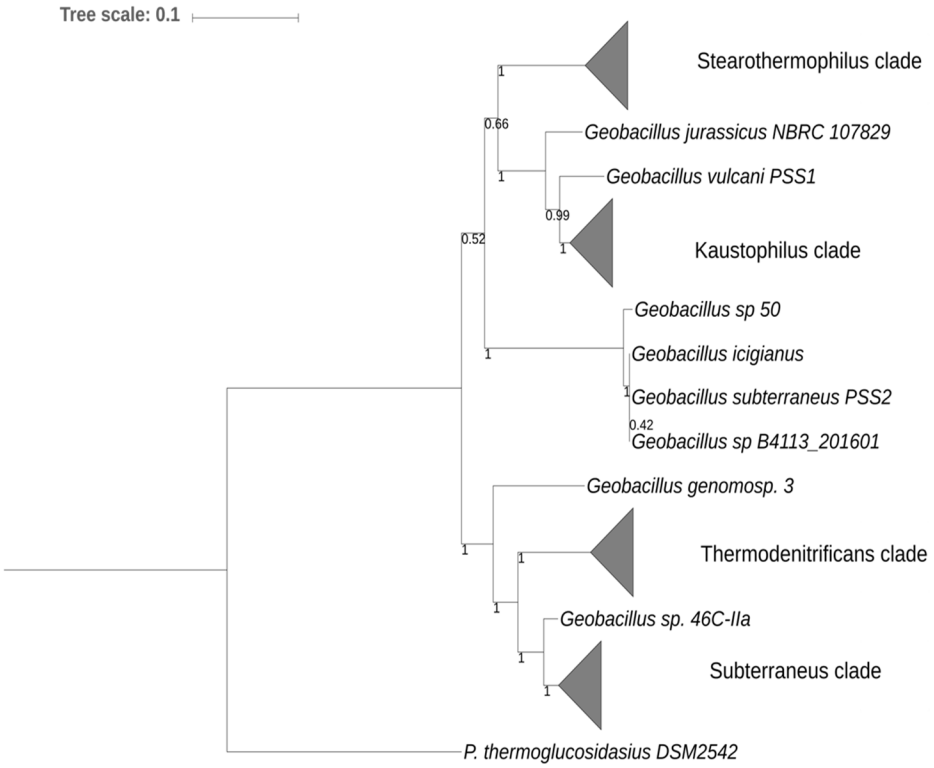


Figure 3.5.4: Maximum likelihood phylogenetic tree of *Geobacillus* strains based on the *spo0A, recA* and *rpoB* sequences. The phylogeny was constructed based on the three genes (total alignment 5718 nt) of 65 taxa deposited in GenBank plus '*Geobacillus icigianus* SP50'. *Parageobacillus thermoglucosidans* was used as outgroup to the genus *Geobacillus*. Values at nodes indicate the bootstrap support as percentages of 100 replicons while the tree scale indicates 0.01 substitutions per site. The analytical list of strains is given at Appendix II, Table 1.

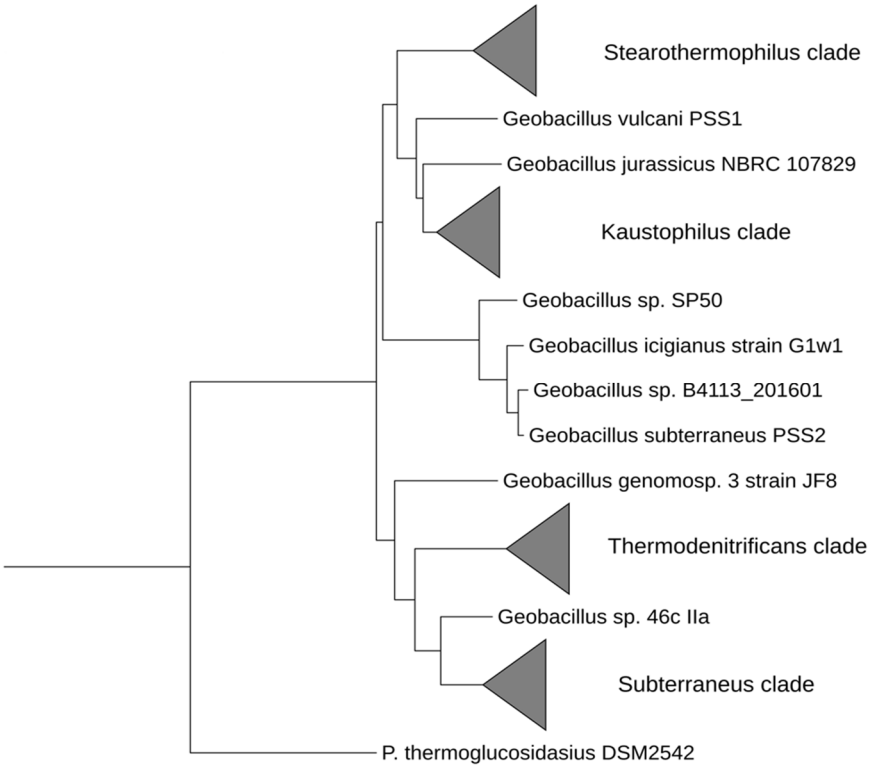


Figure 3.5.5: AAI relationships among *Geobacillus* genomes. The dendrogram was based on the symmetrical matrix with all AAI values of the input genomes.

Table 3.5.1: AAI values between the four sequenced genomes of the icigianus clade. The values in parentheses represent the number of orthologous genes identified in each comparison upon which the AAI estimation is based, while the numbers with an asterisk represent the number of genes form each genome. Numbers in the header correspond to the taxon indicated in the left column.

Taxon	1	2	3	4
Geobacillus sp. Sp50 (1)	100 (3750*)	95.7 (2558)	95.6 (2599)	96.1 (2618)
Geobacillus. icigianous str. G1w1 (2)		100 (3167*)	98.3 (2884)	98.3 (2905)
Geobacillus sp .B4113_201601 (3)			100 (3326*)	99.2 (3092)
Geobacillus subterraneus PSS2 (4)				100 (3361*)

3.5.4 The *β*-Glucan utilization is a unique feature of '*Geobacillus icigianus* SP50' among *Geobacillus* pangenome

As already mentioned, the physiological and biochemical data for *β*-glucanolytic activity of *Geobacilli* are limited. Besides the GH5 endoglucanases detected in the genomic library of *Geobacillus* sp. 70PC53 and the genome of *'Geobacillus icigianus* SP50*'*, there are no reports for any other hydrolase related to the utilization of these substrates. In order to evaluate this literature-based observation, we conducted an extended screening for all GH families exhibiting enzymatic actvity on amorphous cellulose, *β*-1,3-glucan and lichenan within all *Geobacillus* genomes deposited in NCBI in addition to '*Geobacillus icigianus* SP50*'*.

In accordance with the available literature, we were not able to detect any key hydrolytic enzyme. Among the twenty GH families screened, only the three families below, all revealing a broad substrate specificity, were detected in the *Geobacillus* pangenome (Table 3.5.2):

1. Family GH3. This family contains enzymes of eleven different activities including activity against glucan 1,3-*β*-glucosidic bonds. In their majority though, these enzymes are *β*-D-glucosidases and *β*-*D*-xylosidases. The twenty six positive hits of *Geobacilli*, did not present any significant similarity with any of the characterized glucan 1,3-*β*-glucosidases while the presented 34% to 53% amino acid identity (e-value greater than $2.14x10^{-65}$) with characterized *β*-N-acetylhexosaminidases, xylan 1,4-*β*-xylosidases, *β*-glucosidases and exo-1,3-1,4-glucanases (Appendix II, Table 3).
2. Family GH5. The family was recently subdivided into fifty one subfamilies and contains twenty one reported activities, including those of cellulase, cellulose-1,4-*β*-cellobiosidase (non-reducing end), licheninase and glucan 1,3-*β*-glucosidase. Within the genus of *Geobacillus*, we found five positive hits (Appendix II, Table 4):

   -One from '*Geobacillus icigianus* SP50' which belongs to subfamily 2 and presents non-specific endoglucanase activity and

   -Four additional enzymes classified within subfamily 22 and presented over 90% amino-acid similarity (e-value =0) with a characterised arabinoxylan-specific endo *β*-1,4 xylanase isolated from *Geobacillus stearothermophilus* T-1.

3. Family GH51. The majority of the members of the family, are a-L-arabinofuranosidases but three additional activities have been reported, including a cellulolytic one. However, all 37 positive hits from this family, found in *Geobacillus* pangenome revealed more than 90% amino acid identity (e-value =0) with already characterized a-L-arabinofuranosidases (Appendix II, Table 5).

Table 3.5.2: Summary Table of GH families containing cellulolytic and/or *β-1,3*-D-glucanolytic representatives. For each family are given the potential enzymatic activities related with *β*-D-glucan degradation and also the number of each GH family found in *Geobacillus* pangenome.

GH family	Cellulose-acting hydrolases			*β*-1,3-D-glucan-acting hydrolases			No. of positive hits in *Geobacillus* pangenome
	Cellulase (3.2.1.4)	Cellulose-1,4-*β*-cellobiosidase (non-reducing end) (3.2.1.91)	Cellulose-1,4*β*-cellobiosidase (reducing end) (3.2.1.76)	Licheninase (3.2.1.73)	glucan endo-1,3-*β*-D-glucosidase (3.2.1.39)	glucan 1,3-*β*-glucosidase (3.2.1.58)	
3						■	26
5	■	■		■		■	5
6	■	■					0
7	■		■	■			0
8	■			■			0
9	■	■		■			0
12	■			■			0
16				■	■		0
17				■	■	■	0
26				■			0
44	■						0

45							0
48							0
51							37
55							0
64							0
74							0
81							0
124							0
128							0

Overall, we were not able to detect any member of GH family known to specifically act on cellulose nor endo-*β*-1,3-glucanes. In contrast, the enzymes found, belong to broad specificity families and, as already mentioned, all but the GH5 endoglucanase, present higher amino-acid sequence similarities with enzymes not able to act on *β*-1,3 and *β*-1,4 glucans.

3.5.4 *β*-D-Glucan utilization cluster of '*Geobacillus icigianus* SP50'; composition and origin

Within *'Geobacillus icigianus* SP50' genome, the GH5 endoglucanase gene (*cel5*) is located next to a three component operon encoding for an ABC transporter (GluEFG) consisting of (1) an ABC transporter substrate-binding protein (GluE), (2) a binding-protein-dependent transport system inner membrane component (GluF), and (3) a carbohydrate ABC transporter permease (GluG). Between *cel5* and *gluEFG,* there is also a gene encoding for a transcriptional regulator of the LacI family (*lacI)* which is possibly responsible to repress the expression of Cel5 and/or GluEFG in the absence of its substrate. The above three components - *cel5,lacI* and gluEFG - form three autonomous transcriptional units which probably form together a gene cluster facilitating the degradation of *β*-D-glucans and the entrance of the occurring oligomers into the cell (*β*-D-glucan-utilization cluster) (Figure 3.5.5). This cluster is located outside of the

hemicellulose utilization locus of the strain while the genes upstream and downstream of the cluster, encode for proteins of various functionalities, not related to plant biomass, nor even broader carbohydrate, degradation.

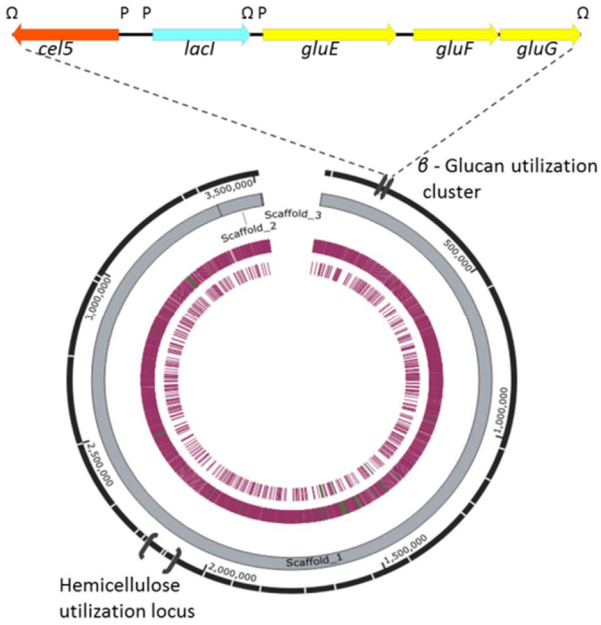


Figure 3.5.5: *β*-D-glucan-utilization cluster of '*Geobacillus icigianus* SP50' (upper part) and its position with respect to the hemicellulose utilization locus within the genome (lower part). *cel5* of the *β*-glucan utilisation locus is shown in orange, *lacI* is designated in light blue and *gluEFG* genes in yellow. 'P' symbols indicate the proposed promoter regions and 'Ω' indicates Rho-independent terminator-like transcription terminators.

Interestingly enough, none of the components of the *β*-D-glucan utilization cluster of '*Geobacillus icigianus* SP50' is conserved in any other *Geobacillus* strain, while the neighboring genomic regions appear to be highly conserved among all four *G. icigianus* members (Fig 3.5.6). Likewise, the GC % content of the genomic region of the *β*-D-glucan utilization cluster (46 %), differs from both the neighboring genomic regions (~50%- 53%) as well as the total GC % content of the '*Geobacillus icigianus* SP50' genome (54.4%) (Figure 3.5.6). Thus, it can be inferred that the *β*-D-glucan utilisation cluster was inserted in '*Geobacillus icigianus* SP50' genome through mechanisms of horizontal gene transfer.

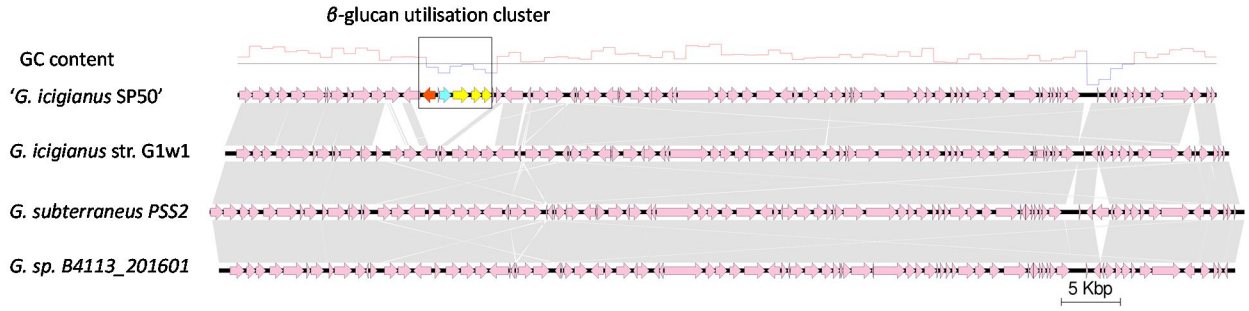


Figure 3.5.6: Comparative diagram of the ~ 85 Kbp homologous segments of the four *G. icigianus* strains where the *β*-glucan utilisation cluster of '*G.icigianus* SP50' is located**.** The aligned regions of the four G. *icigianus* species are shown. *cel5* of the *β*-glucan utilisation locus is shown in orange*, lacI* is designated in light blue and gluEFG genes in yellow. Regions between species presenting a nucleotide similarity greater than 67 % are colored light grey while those presenting lower similarity are colored white. The fluctuation of GC % content through '*G. icigianus* SP50' segment is also shown in the upper part of the graph.

To further examine whether *β*-glucan utilization cluster has been acquired through horizontal transfer, we screened against the RefSeq chromosome records of the NCBI database for all five genes individually and also the three autonomous transcriptional units (*cel5*, *lacI* and *gluEFG)*. We found that the genes of the *β*-D-glucan utilization cluster besides *'Geobacillus icigianus* SP50' are also encountered in various other, phylogenetically distant, bacterial strains (Table 3.5.4). The closest homologues of *cel5* were detected in the genome of various mesophilic strains such as *Bacillus subtilis* str. CW14 (Bacillaceae familly), *Saccharibacillus sacchari* DSM19268 (Paenibacillaceae familly) and *Brenneria goodwinii* (Enterobacteriaceae familly). On the other hand, *lacI* and *gluEFG* were encountered together in the genomes of *Thermicanus aegyptius* (Paenibacilaceae familly) and *Clostridium ultunense* Esp (Clostridiaceae familly) while *gluEFG* without *lacI* was present in various *Paenibacillus* strains, like *P. bouhesdunonensis* (Figure 3.5.7). Interestingly enough, even though the homologues of the *β*-glucan utilization cluster were not detected together in other bacterial strains, they were close to genes encoding for cellulases and glucosidases, analogs of various GH families, a fact revealing the specialization of the individual components of the cluster in *β*-D-glucan uptake.

Table 3.5.4: Refseq genomes with the highest coverage against *cel5* and *lacI-gluEFG* of the *β*-glucan utilization locus of '*G. icigianous* SP50'. The gene coverage and e-value are given for each entry.

Sequence ID	Scientific name	cel5	lacI	gluE	gluF	gluG
gi|1207757276|ref|NZ_CP016767.1|	*Bacillus subtilis* CW14	71% 4.6E-16				
gi|739433897|ref|NZ_KK073875.1|	*Saccharibacillus sacchari* DSM 19268	68% 1.0E-30				
gi|873964046|ref|NZ_CGIG01000001.1|	*Brenneria goodwinii*	67% 4.0E-23				
gi|655940296|ref|NZ_KI783301.1|	*Thermicanus aegyptius* DSM 12793		91% 9.3E-67	87% 2.4E-127	93% 1,0E-125	94% 6.7E-109
gi|560054730|emb|HG764817.1|	*Clostridium ultunense* Esp		91% 8.2E-70	95% 1.1E-131	91% 6.7E-128	94% 4.3E-105
gi|1143639355|emb|LT732548.1|	*Paenibacillus buhesdunonensis*			70% 1.39E-41	91% 1.4E-98	75% 3.1E-56
gi|754884871|ref|NZ_CP009282.1|	*Paenibacillus sp.* FSL R5-0912			61% 1.0E-49	89% 5.5E-72	94% 2E-52
gi|754777894|ref|NZ_CP009280.1|	*Paenibacillus sp*. FSL P4-0081			61% 1.9E-52	89% 1.07E-74	94% 2.4E-51

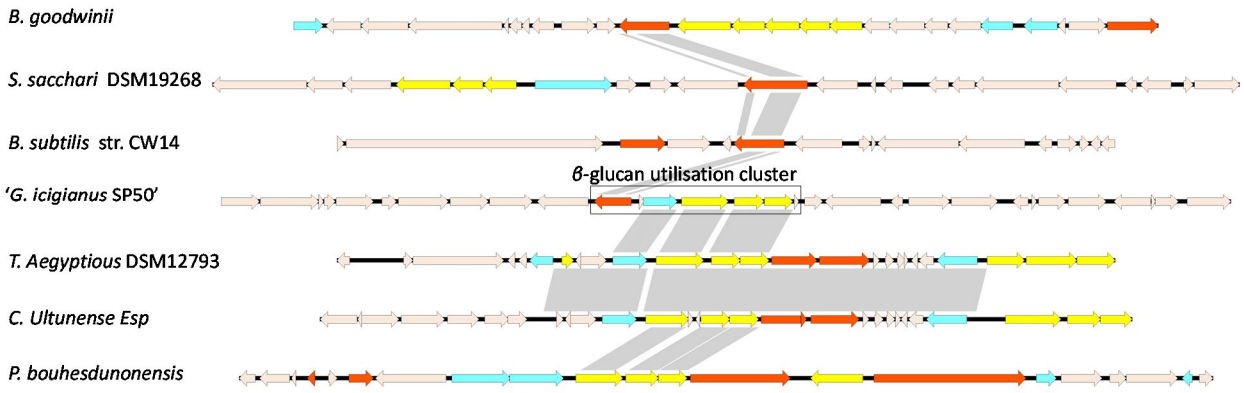


Figure 3.5.7: Comparative diagram of the broader *β*-glucan utilisation cluster genomic region of '*G. icigianus* SP50'(~ 25 Kbp) and the corresponding regions of the genomes bearing the closest homologues of the *β*-glucan utilisation cluster genes. All hydrolases *are shown in orange,* regulatory elements are designated in light blue and ABC transporters in yellow. Regions between species presenting an amino-acid similarity greater than 40 %, are colored light grey while those presenting lower similarity are colored white.

3.6 HETEROLOGOUS EXPRESSION OF THE CEL6A-*T.F* IN THE NON-CELLULOLYTIC *G. STEAROTHERMOPHILUS T-1 STRAIN*

3.6.1 Biochemical characterisation of Cel6A-T.*f*

Endocellulase Cel6A-*T.f,* also known as E2 (Ghangas and Wilson 1988), is an excreted enzyme of *Thermobifida fusca* st. YX. The biochemical properties of the enzyme have been extensively studied and together with the rest cellulases of *T.fusca* YX have become a benchmark for cellulase characterisation and classification (Ghangas and Wilson 1988, Wilson 1988, Walker, Wilson et al. 1992, Barr, Hsieh et al. 1996, Zhang and Wilson 1997, Barr, Wolfgang et al. 1998, Wilson 2004).

The gene *cel6A*, is 1300 bp long and encodes for a two-domain protein; the first domain is the catalytic site and the second is a CBM2 which contributes to the enzyme activity through binding on cellulosic substrates (Figure 3.6.1). Compared to the other cellulases of *T.fusca* YX and the average length of the majority of bacterial cellulases in general, the size of Cel6A-*T.f* is relatively small, a fact favoring genetic manipulations (e.g. plasmid construction, incorporation into the genome, heterologous expression).

Figure 3.6.1: Organization of Cel6A-*T.f.* S.P.; Signal peptide for protein secretion, Cat. site; Catalytic site of the enzyme, Link.; Linker between the catalytic site and CBM2. CBM2; cellulose binding moiety type II.

Based on the available literature, Cel6A-*T.f* is active against several cellulosic substrates including CMC, Avicel, filter paper and microcrystalline cellulose (Ghangas and Wilson 1988, Zhang and Wilson 1997). Being an endocellulase, Cel6A-*T.f* cleaves internal *β*-glucosidic bonds in releasing oligosaccharides of various lengths. The final hydrolysis products are mainly cellobiose and cellotriose while some minor amounts of glucose have also been detected (Walker, Wilson et al. 1992, Harjunpää, Teleman et al. 1996). As will be proven also in the next paragraph (paragraph 3.5.2), *G.stearothermophilus* T-1 is not able to grow on cellulosic polymers. It grows however on their hydrolysis products from the action of Cel6A-*T.f,* namely glucose and cellobiose. Thus it is envisaged that the incorporation of Cel6A-*T.f* into *G.*

stearothermophilus T-1 could result in a transformant strain that could grow on cellulosic substrates.

One equally important aspect concerning the incorporation of Cel6A-*T.f* into *G. stearothermophilus* T-1, is the temperature activity and stability profiles of the enzyme, particularly its performance at 60 ˚C, which is the optimum growth temperature of *G. stearothermophilus* T-1. Since the only information we managed to extract from the available literature was that the enzyme reveals optimum activity at 70 ˚C, we experimentally determined the performance of the enzyme in a broader temperature range. In accordance with Wilson and his colleagues (Wilson 1988, Wilson 2004), Cel6A-*T.f* revealed a maximum activity at 70 ˚C while at 60 ˚C the activity was dropped to 60% of the maximum (Figure 3.6.2). Additionally, Cel6A-*T.f* showed exceptional stability at 60 ˚C; as seen in Figure 3.6.2, the enzyme retained more than 80% of the initial activity even after 48 h of incubation.

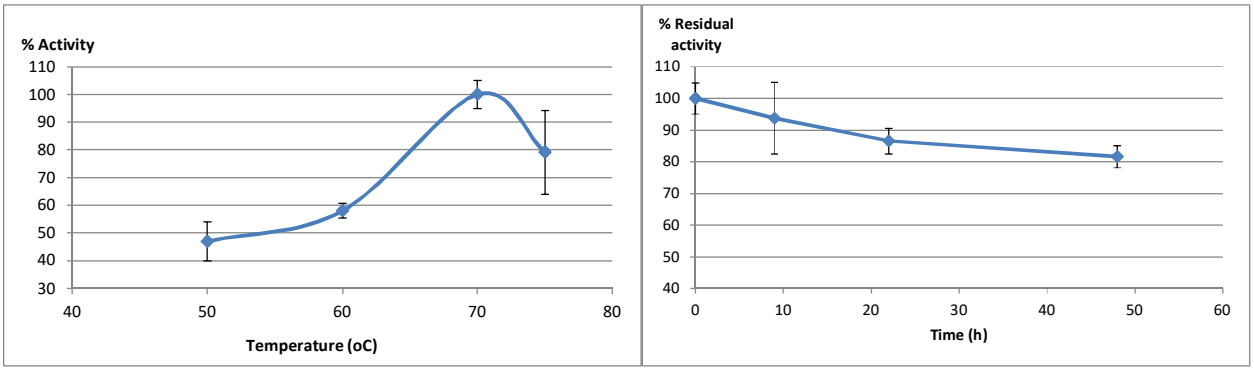


Figure 3.6.2: Activity of Cel6A-*T.f* in correlation with temperature (left) and stability of the enzyme at 60 ˚C (right).

3.6.2 Growth tests of *G. stearothermophilus* T-1 on various carbon sources

The ability of *Geobacillus* strains to utilize mono- and di-saccharides has been experimentally verified (Suzuki, Okazaki et al. 2012) and further supported by the identification of several relevant transporters and hydrolases found within their genomes (Lai and Ingram 1993, Brumm, De Maayer et al. 2015). However, the ability of *G. stearothermophilus* T-1 to grow on glucose-oligomers is unknown (personal communication with Dr. Y. Shoham, University of Haifa, Israel).

Thus, we performed a series of *G. stearothermophilus* T-1 growth tests on various carbon sources, in order to ensure that the hydrolysis products of Cel6A - cellotriose, cellobiose and glucose - could support growth of this specific strain. According to our observations (Figure 3.6.3), *G. stearothermophilus* T-1 did not manage to grow at all on CMC, while the growth on xylose, glucose and cellobiose were at similar levels. Finally, cultures with cellotriose as the sole carbon source showed an approximately 20 % higher growth than the mono- and di-sugars tested.

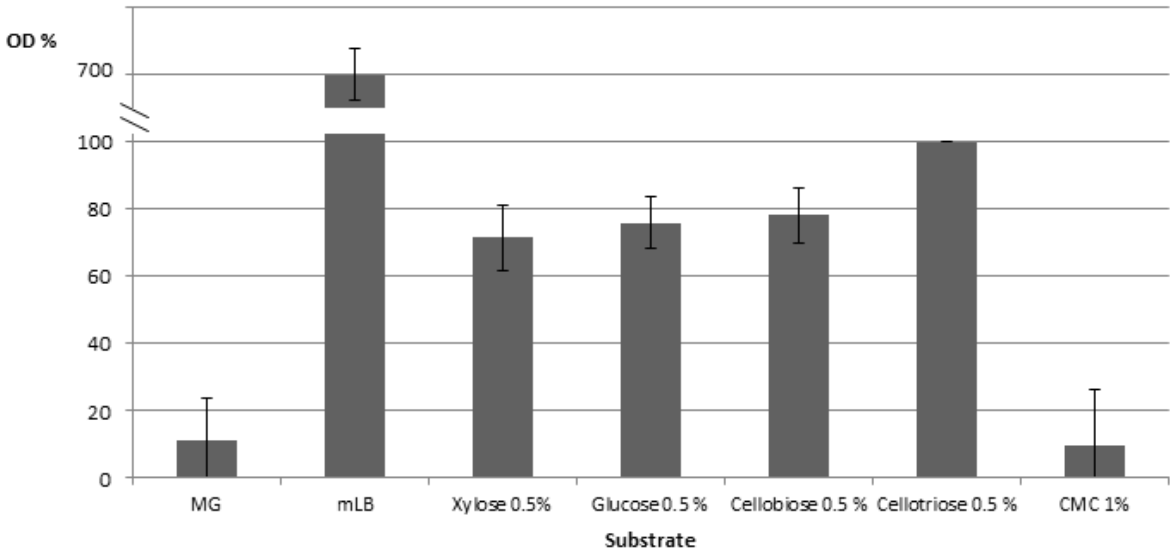


Figure 3.6.3: Relative growth of *G. stearothermophilus* T-1 on different carbon sources; (MG) Basic medium used without any carbon source. (mLB) Luria Bertani broth optimized for *G.stearothermohpilus* strains. The rest of the bars illustrate the growth of *G. stearothermophilus* T-1 in MG medium supplemented with the indicated carbon source.

3.5.3 Expression of Cel6A-T.f* into G. stearothermophilus T-1

For the heterologous expression of *cel*6A-*T.f* * from *G. stearothermophilus* T-1, we employed the regulatory elements of xylanase XynA-T-1. *xynA-T-1* is located within the hemicellulose utilization locus of *G. stearothermophilus* T-1, a DNA region conserved among several *Geobacillus* members, including *G. stearothermophilus* T-6 (Shulami, Shenker et al. 2014). Using genome analysis tools (Solovyev V. 2011), we predicted the possible regions of the promoter and terminator of *xynA*-T-1 and also *xynA*-T-6 (the xylanases of *G. stearothermophilus* T-1 and *G. stearothermophilus* T-6 respectively) which were in accordance with the experimentally found regulatory elements of *xynA* from *G. stearothermophilus* T-6 (Shulami,

Shenker et al. 2014). Despite the differences in gene organization, lengths and sequences of the intragenic regions of *xynA* among the two strains, the -10 and -35 boxes, as well as two out of the three upstream CodY binding sites of *xynA*, appeared significantly conserved (Figure 3.6.4). On the other hand, the catabolite-responsive element (*cre*) of the regulatory region upstream to *xynA*, as well as the predicted terminators did not show any obvious similarity (Figure 3.6.4).

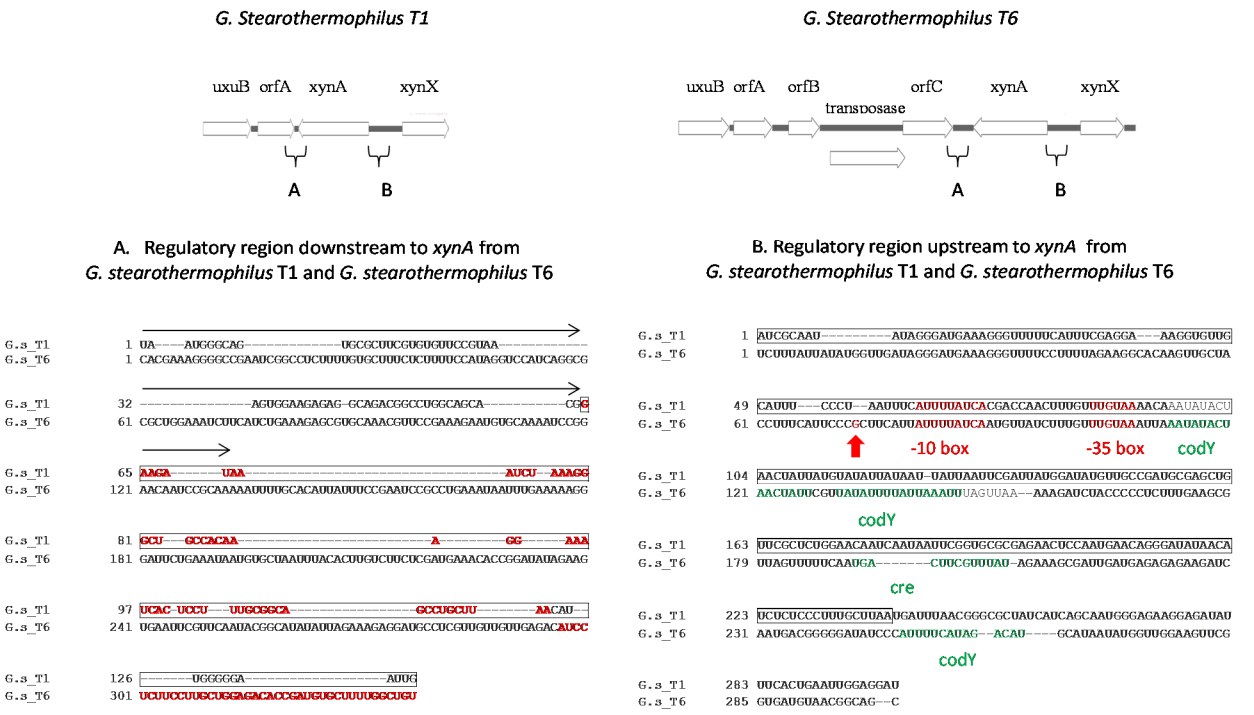


Figure 3.6.4: Graphical representation of the genetic locus of *xynA*-T-1 and the neighbouring genes of *G. stearothermophilus* T-1 and *G. stearothermophilus* T-6 (upper part). Alignment of the sequences downstream to *xynA*-T-1 and *xynA*-T-6. Rho-independent terminators are highlighted in red. For the sequence of G. *stearothermophilus* T-1, the sequence coding for *orfA* is shown with an arrow above, while in rectangle is designated the sequence incorporated in construct pNW35Nii (lower part, A). Alignment of the regulatory region upstream to xynA-T-1 from *G.stearothermophilus* T-1 and *G. stearothermophilus* T-6. The -10 and -35 boxes of both sequences - the proposed ribosome binding sites - are shown in red. For *G. stearothermophilus* T-6, further experimentally determined sites are also shown; (Shulami, Shenker et al. 2014); The red vertical arrow indicates the transcriptional initiation point, while the sites recognized by the regulatory protein CodY and the catabolite-responsive elements (Cre) are designated in green. For the sequence of *G. stearothermophilus* T-1, the sequence incorporated in construct pNW35Ni is shown in rectangle (lower part, A). D-mannonate oxidoreductase (*uxuB*), hypothetical protein (*orfA*), hypothetical protein (*orfC)*,endo-1,4-*β*-xylanase (*xynA-T-1*), potential regulatory protein (*xynX*), transcription factor (*codY)*, binding sites for the catabolite-control protein A(CcpA) (*cre*). Alignemnt was conducting with R-coffee software, taking into consideration the secondary structure of non-coding RNAs (Moretti, Wilm et al. 2008).

Based on the above, we considered that the regions approximately 220 bp upstream and 85 bp downstream to *xynA-T-1*, serve as regulatory elements, and as such, can be used for the

expression of *Cel6A-T.f* * in *G. stearothermophilus* T-1, upon incorporation of the corresponding insert. In an attempt to avoid difficulties in protein translation, we performed codon optimization of *cel6A* based on *G. stearothermophilus* T-1 codon frequencies. Additionally, since Cel6A-*T.f* is a secreted enzyme, we replaced the signal peptide of Cel6A-*T.f* with the one of XynA-T-1 in an effort to limit folding and extracellular secration incompatibilities. The produced pNW35Nii vector, was used for the transformation of *G. stearothermophilus* T-1 (Figure 3.6.5).

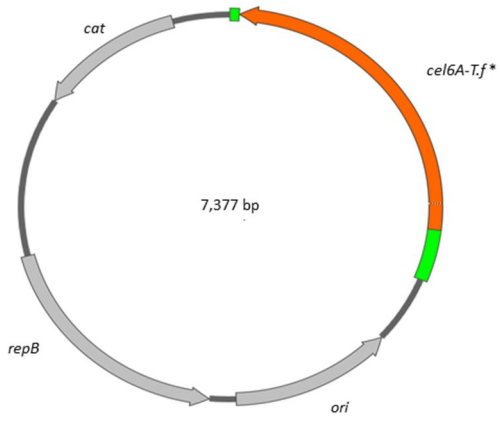


Figure 3.6.5: plasmid pNW35Nii; *cel6A-T.f**, the gene of cellulase from *T. fusca* YX with codon optimization for expression in *G. stearothermophilus* T-1, is shown in orange. The regulatory elements and the signal peptide of *xynA* from *G. stearothermophilus* T-1, used for the expression of *cel6A-T.f* * are designated in green. Genes necessary for the replication and maintenance of pNW35N in the host cells (*ori*, *repB*, *cat*) are designated in light grey.

Upon transformation, we obtained two, morphologically different, types of colonies resistant to chloramphenicol in an aproximate ratio of 1:1. The first type, was characterized by an irregular form and a dense growth with a raised elevation on the petri dish (Figure 3.6.6 E and G, colonies 1-7, 14 and 15). This phenotype was similar with the one observed from the wild type *G. stearothermophilus* T-1, in the absence of antibiotic (Figure 3.6.6, A and C). These cells though, did not produce decolorization zones of CMCase activity, due to the cells' dense growth that prevented Congo red's diffusion below the colonies (Figure 3.6.6 F and H, colonies 1-7, 14 and 15). The same unstained colony imprint was also observed for the wild type strain if was allowed to grow extensively (Figure 3.6.6 B). The same phenotype was also observed in liquid cultures; the strains reached significant OD values while there was no detectable CMC- activity. The above, in combination with the fact that we were not able to isolate any plasmid from these transformants, lead us to hypothesize that this phenotype occurred from a random mutation causing resistance to the antibiotic rather than from the insertion of the resistant genes from the plasmid.

The second type of colonies appeared to be sucesfully transformed; They were tinny, circular-shaped and smooth (Figure 3.6.6. E and G, colonies 8, 9-13). Upon Congo-red staining a decolorization halo was observed indicating the presence of extracellular cellulase activity (Figure 3.6.6 F and H, colonies 8, 9-13). Enzymatic activity was also observed in the supernatant of liquid cultures (Figure 3.6.7 A) and the plasmid extracted from them, had the anticipated molecular weight of pNW35Nii (Figure 3.6.7 B). However, the growth of the transformants in liquid cultures was very low and also the plasmid appeared instable since we were not able to isolate it upon serial recultivations. The above, did not allow us to continue with the growth of the mutants in larger scale liquid cultures and test the ability of the transformants to degrade cellulose substrates.

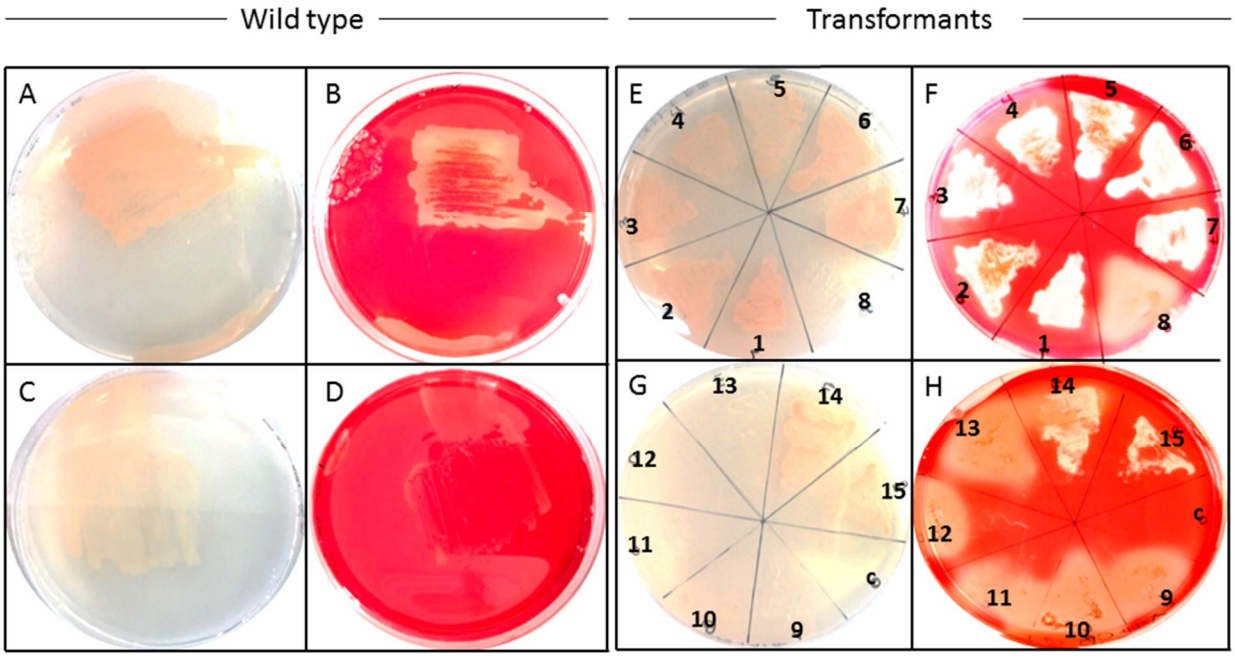


Figure 3.6.6: Agar plate-screening for cellulase activity; wild type G. stearothermophilus T-1 streaked on mLB-CMC plates with high and moderate growth (A and C, respectively). 15 transformed G. stearothermophilus T-1 strains plus the wild type strain -c- streaked on mLB-CMC-chloramphenicol plates (E,G). Plates A, C upon Congo red staining (B, D). Plates E,G upon Congo red staining (F, H). For the colonies 1-7, 14 and 15 at figures F and H, and also for the wild type strain at figure B, the unstained imprint is due to the high cell density. CMase decolorization halo is observed for strains 8 and 9-13 at figures F and H.

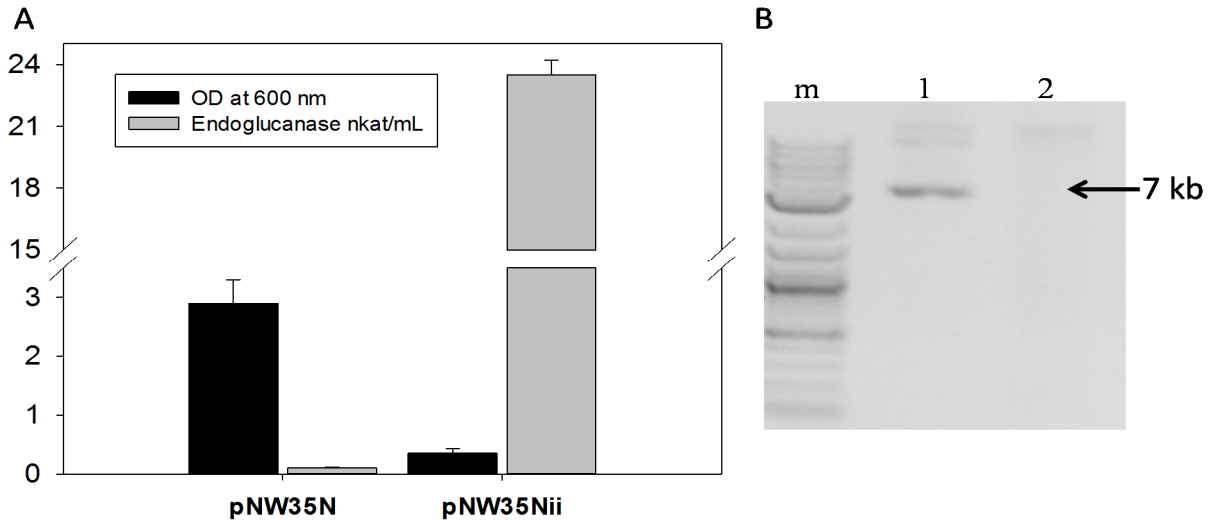


Figure 3.6.7: Liquid cultures of G. stearothermophilus T-1 transformed with either the empty vector pNW35N or pNW35Nii construct. In the present of chloramphenicol, transformants bearing the empty vector pNW35Nii reached high growth levels while transformants with pNW35Nii had endoglucanase activity but low growth (A). Plasmid extracted from a G. stearothermophilus T-1 transformant with pNW35Nii upon digestion with BamHI (lane 1), plasmid extraction from the wild type G. stearothermophilus T-1 (lane 2) (B).

Discussion

Various thermophilic bacilli, including those of the genus *Geobacillus*, are among the most promising candidates for various biotechnological applications. This is due to their relatively high growth rates, the durability of their sporeforms, their robust protein secretion pathways, and the lack of pathogenic strains within the genus. In addition, the derived enzymes are also of great importance, because of their thermophilic characteristics and their compact, singe-domain structure which favors both their bulk production and also performance in highly viscous materials.

Geobacilli are well known for their extensive hemicelulolytic system while there are only a few, not well documented indications for their ability to decompose and grow on cellulose. The hemicellulolytic enzymes of the genus are being extensively exploited over the last few years, while some of them are already employed in industrial applications. In parallel, and in order to exploit *Geobacilli* as whole cells biocatalysts, regulation studies on their hemicellulolytic systems and development of effective transformation techniques are being under way.

Under this framework, the present thesis, aimed in the exploitation with respect to plant biomass degradation of selected thermophilic bacterial strains isolated from the volcanic habitat of Santorini. In particular,

1. The first aim of the present thesis was the detection, production and characterization of the enzymes responsible for plant biomass degradation.
2. Representative enzymes from the first step were incorporated for the first time in designer cellulosomes, leading to multiprotein complexes suitable for use in high-temperature biorefinery applications.
3. Among the isolates from the volcano of Santorini, there were a few strains presenting cellulolytic activity. Since this trait is not well documented in the genus, we sequenced the genome of the cellulolytic strain '*G. icigianus* SP50' and we detected, for the first time within *Geobacilli*, the genomic cluster responsible for cellulose degradation.
4. Finally, since cellulolytic activity is of great biotechnological interest, we also attempted to convert the non-cellulolytic strain *G. tearothermophiluss* T-1 to cellulolytic by transformation with the appropriate plasmid vector.

In more detail, regarding enzyme detection and characterization (1), from the bacterial pool of the 10o bacterial strains, we managed to identify and consequently clone, seven enzymes, classified to six different GH families, revealing four biochemically distinct enzymatic activitites. The two activities are related with xylan degradation while the other two with *β*-D-glucan degradation. All enzymes are single-domain, extracellular or cell-associated. They have activity optima within the range of 60 ˚C to 80 ˚C, a preference for slightly acid pH values, while they remain stable for at least 6 h at 60 ˚C. All isolated enzymes though, present significant similarity with already characterized enzymes isolated from *Geobacilli(Khasin, Alchanati et al. 1993, Shallom, Leon et al. 2004, Ng, Li et al. 2009, Suzuki, Okazaki et al. 2012)*, while all but the GH5 endoglucanase belong to the highly conserved hemicellulolytic locus found in several *Geobacillus* strains.

Besides the main xylanolytic enzymes, *Geobacilli* are known to produce a variety of other hydrolases and esterases with specificity towards the degradation of the hemicellulose side-chains. The majority of the corresponding genes, are located in vicinity with the xylanolytic genes within the hemicellulose utilization locus (De Maayer, Brumm et al. 2014), and their biochemical properties and regulation have been studied (Shulami, Raz-Pasteur et al. 2011, Choi, Lee et al. 2016).Here, we intentionally excluded these hemicellulolytic enzymes from our screening, since their products represent a small fraction of the hemicellulosic hydrolysates. Consequently, their contribution in the hydrolysis process, even though not negligible, seems to be of minor importance.

In the second part of the doctorate thesis (2), three among the isolated enzymes, an additional *β*-glucosidase with similar thermal stability and activity profiles (Hong, Kim et al. 2009), as well as a designer scaffoldin of mainly mesophilic origin (Moraïs, Barak et al. 2011), were combined to form the first designer thermo-cellulosome. Despite the mesophilic origin of the scaffoldin, the holo-complex proved to be functional and stable for up to 6 h at 60 ˚C. This result is of great importance, since it implies that complexing thermostable enzymes on a mesophilic scaffold, somehow conveys thermal stability to the overall complex. Moreover, the hydrolytic performance of the complex, outperformed that of an equimolar free enzyme mixture, proving

that cellulosomal structure offers advantages not only in mesophilic conditions but also in thermophilic ones.

Since there is only a handful of isolated thermophilic organisms known to produce cellulosomes, our knowledge on the performance of cellulosomes under high-temperature conditions were limited (Artzi, Bayer et al. 2016). Thus, the present work proved, for the first time, that designer cellulosomes can be incorporated in processes requiring these conditions. In continuation of this work, a second thermocellulosome was constructed, which was stable and also able to act for extended incubation times at 60 ˚C (24 h) and for shorter time periods at 65 ˚C or even 70 ˚C (Artzi, Bayer et al. 2016, Moraïs, Stern et al. 2016). Since the feasibility of cellulosome used at elevated temperatures was confirmed, we would like to proceed further to test the performance of thermo-cellulosomes in biorefinery processes; thus, one of our ongoing goals is the incorporation of a designer thermocellulosome in Simultaneous Saccharification and Fermentation processes (SSF) towards bioethanol production.

The third part of the thesis concerned the characterization of the strain '*G. icigianus* SP50' which revealed the intriguing ability to degrade cellulose (3). Despite the fact that there were some references reporting cellulolytic activity for some *Geobacillus* strains (Lai and Ingram 1993, Tai, Lin et al. 2004, Rastogi, Bhalla et al. 2010, Assareh, Shahbani Zahiri et al. 2012, Stathopoulou, Galanopoulou et al. 2012, Potprommanee, Wang et al. 2017), there is no genome within the genus for which it has been previously shown to contain the genes necessary for cellulose degradation. Genome sequencing of '*G. icigianus* SP50' allowed us to identify five genes coding for a GH5 broad specificity hydrolase, an ABC transporter and a regulatory protein, composing together the*β*-D-glucan utilization cluster. This cluster sems to be responsible for the degradation of a broad range of *β*-D-glucans including amorphous cellulose and has been probably acquired by '*G. icigianus* SP50' from phylogenetically distant bacteria through horizontal gene transfer.

The presence of the *β*-D glucan utilization cluster does not render '*G. icigianus* SP50' a true cellulose degrader since the strain is still not able to degrade crystalline cellulose. However, the strain might be benefited by the broad specificity *β*-D glucan utilization genes, since they provide cells with the ability to utilize traces of *β*-D-glucan found interlinked with hemicellulose.

From an ecological perspective, this comes in agreement with the proposed opportunistic tactic followed by *Geobacilli* for plant biomass decomposition (Zeigler 2014); According to this hypothesis, *Geobacilli* reach temporarily high cell densities in transiently high-temperature environments like composts. High temperatures are caused by the oxidation of sugar molecules during cellulose degradation by mesophilic microorganisms. Thus, when the required temperature for *Geobacillus* growth is reached, the remaining plant material constitutes of a greatly heterogenous mixture of non-oriented hemicellulose molecules rather than crystalline cellulose.

Since the bioconversion of cellulose by *Geobacilli* is considered of great biotechnological interest, the final part of the thesis (4), was devoted to this direction. In particular, we managed to convert the non-cellulolytic *G. stearothemrophilus* T-1 strain to cellulolytic via the heterologous expression of a cellulase. For the expression of the enzyme from *G. stearothermophilus* T-1, we used established transformation protocols and a derivative of the pNW33N vector (Zeigler 2001, Shulami, Shenker et al. 2014). For the successful expression of the enzyme, the cellulase gene was codon optimized for *G. stearothermophilus* and the regulation of expression was attempted by the already characterized regulatory elements of the xylanase XynA of the same strain.

Indeed, upon transformation we observed transformants with cellulolytic activity in both liquid cultures and agar plates. However, the whole attempt cannot be considered successful since plasmids were not stable and the desired phenotype could not be constantly expressed. The difficulties in *Geobacillus* transformation are known already and according to previous studies several approaches could be followed in order to overcome the observed obstacles. For example, different *Geobacillus* host cells and/or plasmid vectors could be employed (Zeigler 2001, Zeigler 2002, Suzuki, Murakami et al. 2012, Tominaga, Ohshiro et al. 2016). Another alternative could be the incorporation of the construct via homologous recombination into the genome of the strain (Zeigler 2002). From a biotechnological point of view, the later approach is more favorable since it could lead to the construction of a cellulolytic strain which does not require the addition of any antibiotic for its growth.

A summary of the original achievments accomplished in the present PhD thesis are summarized below:

1. Construction of the first designer thermocellulosome and its evaluation in elevated temperatures
2. Sequencing of the first *Geobacillus* genome bearing an endocellulase gene and analysis of the corresponding genetic cluster
3. Introduction of a cellulase into a non-cellulolytic *Geobacillus* strain.

APPENDICES

APPENDIX I: GENETIC SCREENING FOR CELLULOLYTIC AND XYLANOLYTIC GENES

Table 1: Genera of Bacillaceae families. The genera of the ten selected strains are in bold.

Genera of the Bacillaceae family		
Aeribacillus	*Lentibacillus*	*Salsuginibacillus*
Alkalibacillus	*Lysinibacillus*	*Tenuibacillus*
Amphibacillus	*Marinibacillus*	*Terribacillus*
Anoxybacillus	*Oceanobacillus*	*Thalassobacillus*
Bacillus	*Ornithinibacillus*	*Ureibacillus*
Caldalkalibacillus	*Paraliobacillus*	*Virgibacillus*
Cerasibacillus	*Paucisalibacillus*	*Vulcanibacillus*
Exiguobacterium	*Pelagibacillus*	
Filobacillus	*Piscibacillus*	
Geobacillus	*Pontibacillus*	
Gracilibacillus	*Saccharococcus*	
Halalkalibacillus	*Salibacillus*	
Halobacillus	*Salimicrobium*	
Halolactibacillus	*Salinibacillus*	
Jeotgalibacillus	*Salirhabdus*	

Table2: Genera of the Paenibacilaceae family.

Genera of the Paenibaillaceae family		
Paenibacillus	*Brevibacillus*	*Thermicanus*
Ammoniphilus	*Oxalophagus*	*Thermobacillus*
Aneurinibacillus	*Saccharibacillus*	

Table 3: Characterized *β-1,4*-xylanases from members of Bacillaceae and Paenibacillacae families deposited in CAZy database.

I. *β*-1,4 –xylanases (E.C 3.2.18)

Bacillaceae						Paenibacillaceae		
	Microorganim	Uniprot No		Microorganim	Uniprot No		Microorganim	Uniprot No
GH8	*Bacillus sp. KK-1*	O52730	GH11	*Bacillus circulans*	PO9850	GH10	*Paenibacillus barcinonensis BP-23*	O69231
GH10	*Anoxybacillus sp. E2(2009)*	D7NNK8		*Bacilus firmus*	Q6U894		*Paenibacillus barcinonensis BP-23*	O69230
	Bacillus alcalophilus AX2000	Q6TDT4		*Bacilus firmus K-1*	Q71S35		*Paenibacillus barcinonensis B-6*	E3WF08
	Bacillus firmus	Q6U892		*Bacillus haloduransC-125*	Q9KEF3		*Paenibacillus curdlanolyticus B-6*	B1A3N2
	Bacillus halodurans C-125	P07528		*Bacillus licheniformis I5*	A5H0S3		*Paenibacillus sp. E18*	D6BQP4
	Bacillus halodurans MIR32	P07528		*Bacillus MS5-14*	B5SYI8		*Paenibacillus sp HPL-001*	B6VF01

	Bacillus halodurans S7	Q17TM8		*Bacillus pumilus*	P00694		*Paenibacillus sp. JDR-2*	C6CRV0
	Bacillus pumilus TX703	Q8L2X3		*Bacillus pumilus ARA*	E2IHA1		*Paenibacillus sp. JDR-2*	A9QDS0
	Bacillus sp N137	Q45518		*Bacillus pumilus HB030*	Q9AMB5		*Paenibacillus sp KCTC8848P*	Q9F9B8
	Bacillus sp 16-5	D7RA44		Bacillus pumilus HY-20	C8CB65		*Paenibacillus sp W-61*	Q8GHJ4
	Bacillus sp. NG-27	O30700		Bacillus pumilus PJ19	Q06RH9		*Paenibacillus campinasensis BL11*	Q2I6W5
	Geobacillus sp. TC-W7	D0EM78		Bacillus pumilus XY1	P00694		*Paenibacillus curdlanolyticus B-6*	D3GKE3
	Geobacillus sp. WBI	B5M201		Bacillus sp	Q9ZB36		*Paenibacillus polymyxa PPL-3*	E1AHZ6
	Geobacillus stearothermophilus 21	P45703		*Bacillus* sp 41M-1	Q9RC94	GH11	*Paenibacillus sp. DG-22*	A4GG22
	Geobacillus stearothermophilus MT-1	Q3YBZ9		*Bacillus* sp. BP-7	Q84F19		*Paenibacillus HY8*	A3QRI7
	Geobacillus stearothermophilus T-6 NCIMB 40222	Q09LY9		*Bacillus* sp. HBP8	Q58G72		*Paenibacillus sp.JDR-2*	C6D8U8
	Geobacillus stearothermophilus T-6 NCIMB 40222	P40943	GH11	*Bacillus subtilis* cho40	E0YTQ6		*Paenibacillus sp. KCTC8848P*	Q9F9B9

GH10	*Geobacillus thermodenitrificans* T-2	B2Z4E4		*Bacillus subtilis* MW10	P18429		Paenibacillus *sp. W-61*	Q1XGE6
	Bacillus amyloliquefaciens CH51	E0L13		*Bacillus subtilis* R5	B9ZZN9		Thermobacillus xylanolyticus *D3*	Q14RS0
	Bacillus sp YA-14	Q59256		*Bacillus subtilis* subsp subtilis str 168	P18429	GH30	Paenibacillus sp. W-61	P70733
	Bacillus sp YA-335	Q59257		*Geobacillus stearothermophilus* 236	P45705			
GH11	*Bacillus* sp. YJ6	C5MTD6	GH30	B*acillus* sp BP7	E1ACF9			
	Bacillus subtilis AK1	Q3HLJ4						
	Bacillus AMX-4	C7F433						
	Bacillus subtilis AQ1	D7F2D8						
	Bacillus subtilis B230	Q7SID8						

Table 4: Characterized *β-1,4-P-glucosidases* from members of the Bacillaceae family deposited in CAZy database (There are no entries form Paenibacillaceae members).

II. *β*-1,4 P- glucosidases (E.C 3.2.86)

	Microorganism	Uniprot No		Microorganism	Uniprot No
GH1	*Bacillus subtilis subsp.* subtilis str. 168	O05508	GH4	*Bacillus subtilis* subsp. subtilis str. 168	P46320
	Bacillus subtilis subsp. subtilis str. 168	P42973		*Geobacillus stearothermophilus T-1*	P84135
	Geobacillus kaustophilus HTA426	Q5KUY7			
	Geobacillus kaustophilus HTA426	Q5KYU5			

Table 5: Characterized xyloglycan-specific-endo-*β-1,4-P-glucanases* from members of Bacillaceae and Paenibacillaceae families deposited in CAZy database.

III. Xyloglucan-specific endo-beta-1,4-glucanases (E.C 3.2.1.151)

Bacillaceae			Paenibacillacae		
	Microorganism	Uniprot No		Microorganism	Uniprot No
GH12	*Bacillus licheniformis* ATCC 14580/DSM13	Q65FM6	GH5	*Paenibacillus pabuli*	H9KVH3
			Sub. 4	*Paenibacillus* sp. KM21	Q3MUH8
			GH74	*Paenibacillus* sp. KM21	Q3MUH7

Table 6: Characterized cellulose-β-1,4-cellobiosidases from members of the Paenibacillaceae family deposited in CAZy database (There are no entries form Paenibacillaceae members).

IV. Cellulose 1,4-beta-cellobiosidases (reducing end) (E.C 3.2.1.176)

	Microorganism	Uniprot No
GH48	*Paenibacillus barcinonensis* BP-23	Q8KKF7

Table 7: Characterized cellulases from members of Bacillaceae and Paenibacillaceae families deposited in CAZy database.

V.Cellulases (EC 3.2.1.4)

	Bacillaceae					Paenibacillaceae		
	Microorganism	Uniprot No		Microorganism	Uniprot No		Microorganism	Uniprot No
GH5 Subfamily 2	*Bacillus akibai* KSM-1139	P06564	GH5 Subfamily 2	*Bacillus subtilis* CHZ1	Q93LD0	GH9	*Paenibacillus barcinonensis* BP-23	Q9Z4I1
	Bacillus amyloliqefaciens PSM 3.1	D3Y2F0		*Bacillus subtilis* CK-2	G1DE47		*Paenibacillus sp.* BME-14	B7U9C9
	Bacillus licheniformis ATCC 14580/DSM13	A5A671		*Bacillus subtilis* DLG	P07983	GH44	*Paenibacillus lautus* PL236	P29719
	Bacillus licheniformis GXN151/B-41361	Q7X3S6		*Bacillus subtilis* DR	A8D0T0			
	Bacillus licheniformis MD1	Q5QSM2		*Bacillus subtilis* I15	B7UAM4			
	Bacillus sp. 22-	Q4555		*Bacillus*	Q45T7			

28	4		*subtilis* JA18	8	
Bacillus sp. 5H	O83012		*Bacillus subtilis subsp. Subtilis* str 168	P10475	
Bacillus 79-23	O52731		*Geobacillus sp.* 70PC53	C5H6X3	
Bacillus sp. D04	Q45430		*Bacillus agaradhaerens*	O85465	
Bacillus sp. KSM-635	P19424	GH5 Subfamily 4	*Bacillus halodurans* C-125	Q9KF82	
Bacillus sp. KSM-64	Q59241		*Bacillus agardhaerens* DSM 8721	Q6A3Q5	
Bacillus sp. KSM-N252	Q93R81	GH8	*Bacillus circulans* KSM N257	Q93HV0	
Bacillus sp. KSM-S237	Q9F216		*Bacillus licheniformis* GXN151	Q6SYB5	
Bacillus sp.N186-1	Q59232		*Bacillus pumilus* CL16	Q5YLG1	
Bacillus subtilis	Q45532	GH9	*Bacillus pumilus* S-27	Q9R9H6	
Bacillus subtilis BME-15	B6DVJ2		*Bacillus sp.* AC-1	A6Y862	
Bacillus subtilis BSE616	P23549		*Bacillus sp.* KSM-522	P28622	
Bacillus subtilis C-36	Q0PPV7	GH12	*Bacillus licheniformis* GXN151	Q7X4S4	

Table 8: Characterized *β*-xylosidases from members of Bacillaceae and Paenibacillaceae families deposited in CAZy database.

VI. *β*-xylosidases (EC 3.2.1.37)

Bacillaceae			Paenibacillaceae		
	Microorganism	Uniprot No		Microorganism	Uniprot No
GH39	*Bacillus halodurans* C-125	Q9KDZ3	GH52	*Paenibacillus sp.* DG-22	B6C867
	Geobacillus stearothermophilus T-6 NCIMB4022	Q9ZFM2			
GH43	*Bacillus halodurans* C-125	Q9K6P5			
	Bacillus licheniformis ATCC 14580/DSM13	Q65L63			
	Bacillus pumilus IPO	P07129			
	Bacillus pumilus PLS	Q9Z477			
	Bacillus sp. KK-1	O07078			
	Bacillus subtilis subsp. Subtilis str168	P94522			
	Geobacillus stearothermophilus T-6 NCIMB4022	Q09LX0			
	Geobacillus thermoleovorans IT-08	Q2I2N4			
	Geobacillus thermoleovorans IT-08	Q27R94			
GH43	*Geobacillus stearothermophilus* 21	P45702			
	Geobacillus stearothermophilus	P45704			

	236		
	Geobacillus stearothermpophilus T-6 NCIMB 40222	Q09LZ0	

Table 9: Characterized oligosaccharide-reducing-end-xylanases from members of the Bacillaceae family deposited in CAZy database (There are no entries form Paenibacillaceae members).

VII. Oligosaccharide reducing-end xylanase (EC 3.2.11.56)

	Microorganism	Uniprot No
GH8	*Bacillus halodurans* C-125	Q9KB30

Table 10: Summary Table of the DNA sequences from GH39 entries, used to design degenerate primers for GH39 β-xylosidases. In bold is shown the characterized entry.

Strain	GeneBank Accession No (gene)	Uniprot No
Geobacillus sp. C56-T3	ADI26649.1	D7D6C0
Geobacillus sp. GHH01	AGE22453.1	-
Geobacillus sp. Y412MC52	ADU94257.1	C3J6V2
Geobacillus sp. Y412MC61	ACX79287.1	C9RT42
Geobacillus stearothermophilus T-6 NCIMB 40222	**ABI49941.1**	**Q9ZFM2**
Geobacillus thermodenitrificans NG80-2	ABO67133.1	A4IP79
Geobacillus thermoglucosidasius C56-YS93	AEH48187.1	-

Table 11: Microorganism source and Uniprot numbers of the sequences used for degenerate primers for GH43 *β-xylosidases*.

Strain	Uniprot No
Geobacillus stearothermophilus T6	Q09LX0
Geobacillus thermoleovorans IT-08	Q2I2N4
Geobacillus kaustophilus GBlys	U2WUG3

Table 12: Microorganism origin and Uniprot Numbers of the characterized and putative sequences used to design inner degenerate primers for the GH48 family.

	Strain	Uniprot No	Higher order taxa (Family)
Characterized	*Caldicellulosiruptor bescii*	B9MKU9	*Thermoanaerobacterales* Family III
	Thermobifida fusca	Q47NH7	*Nocardiopsaceae*
	Clostridium cellulovorans	O65986	Clostridiaceae
	Clostridium stercocarium	P50900	
	Clostridium phytofermentas	A9KT91	
	Cellulomonas fimi	P50899	*Cellulomonadaceae*
	Myxobacter sp.	Q19VP0	*Myxobacteriaceae*
	Paenibacillus barcinonensis	Q8KK7	
Putative	*Paenibacillus polymyxa*	EORLD5	*Paenibacilaceae*
	Paenibacillus polymyxa	E3EJF1	
	Clostridium thermocellum	A3DH67	*Clostridiaceae*

Table 13: Microorganism origin, Uniprot numbers and GeneBank numbers of the characterized amino acid sequences used to design inner degenerate primers for the GH52 family.

Strain	Uniprot No	GeneBank No
Geobacillus stearothermophilus str. *1A05583*	-	AGE34479.1
Geobacillus stearothermophilus str. *21*	P45702	BAA05667.1
Geobacillus stearothermophilus str. *22*	P45704	AAA50863.1
Geobacillus thermoglucosidasius str. *NBRC 107763*	-	GAJ44942.1
Paenibacillus sp. str. *DG-22*	B6C867	ABV46494.1

APPENDIX II: GENOME ANALYSIS OF THE CELLULOLYTIC STRAIN '*GEOBACILLUS ICIGIANUS* SP50'

Table 1: *Geobacillus* genomes retrieved from GenBank assembly NCBI (August 2017).

	Organism	Sequence name	Submitter	Date of submission	Genome representation		Assembly level
1	*Geobacillus genomosp. 3 strain JF8*	GCA_000445995.2_ASM44599v2_genomic	Shizuoka University	11/18/2013	full	Complete Genome	latest
2	*Geobacillus icigianus strain G1w1*	GCA_000750005.1_ASM75000v1_genomic	the institute of cytology and genetics	9/8/2014	full	Contig	latest
3	*Geobacillus jurassicus NBRC 107829*	GCA_001544315.1_ASM154431v1_genomic	National Institute of Technology and Evaluation	1/28/2016	full	Contig	latest
4	*Geobacillus kaustophilus GBlys*	GCA_000415905.1_Gbl_1.0_genomic	Institute of Genetic Resources, Faculty of Agriculture, Kyushu University	6/25/2013	full	Contig	latest
5	*Geobacillus kaustophilus HTA426*	GCA_000009785.1_ASM978v1_genomic	JAMSTEC	12/4/2004	full	Complete Genome	latest
6	*Geobacillus kaustophilus NBRC 102445*	GCA_000739955.1_ASM73995v1_genomic	National Institute of Technology and Evaluation	8/2/2014	full	Contig	latest
7	*Geobacillus kaustophilus strain Et2/3 LG51_142*	GCA_000948165.1_BFA_3_A_genomic	Los Alamos National Laboratory	3/5/2015	full	Contig	latest
8	*Geobacillus kaustophilus strain Et7/4 LG52*	GCA_000948285.1_BFA_4_A_genomic	Los Alamos National Laboratory	3/5/2015	full	Contig	latest
9	*Geobacillus lituanicus strain N-3*	GCA_002243605.1_ASM224360v1_genomic	Kyungpook National University	8/8/2017	full	Complete Genome	latest

10	*Geobacillus sp. 1017*	GCA_001908025.1_ASM190802v1_genomic	Institute of Bioengineering, Research Center of Biotechnology of RAS	12/19/2016	full	Contig	latest
11	*Geobacillus sp. 15*	GCA_001624615.1	University of Groningen	2016/04/20	full	Scaffold	latest
12	*Geobacillus sp. 46C-IIa*	GCA_002077765.1	Jet Propulsion Laboratory, California Institute of Technology	4/3/2017	full	Scaffold	latest
13	*Geobacillus sp. 47C-IIb*	GCA_002077775.1_ASM207777v1_genomic	Jet Propulsion Laboratory, California Institute of Technology	4/3/2017	full	Scaffold	latest
14	*Geobacillus sp. A8*	GCA_000447395.1_GA8_Ver1_genomic	University of Groningen	4/20/2016	full	Scaffold	latest
15	*Geobacillus sp. B4113_201601*	GCA_001587475.1_B4113_genomic	University of the Free State	8/15/2013	full	Contig	latest
16	*Geobacillus sp. BCO2*	GCA_001294475.1_ASM129447v1_genomic	University of Groningen	3/11/2016	full	Scaffold	latest
17	*Geobacillus sp. C56-T3*	GCA_000092445.1_ASM9244v1_genomic	Microbial Gene Research and Resources Facility	9/18/2015	full	Contig	latest
18	*Geobacillus sp. CAMR12739*	GCA_000691445.1_CAMRasm9_genomic	US DOE Joint Genome Institute	6/1/2010	full	Complete Genome	latest
19	*Geobacillus sp. CAMR5420*	GCA_000691465.1_CAMR5420assem6_genomic	Centre for Microbial Ecology and Genomics	5/15/2014	full	Contig	latest
20	*Geobacillus sp. FW23*	GCA_000617945.1_GeoThermo1	Centre for Microbial Ecology and Genomics	5/15/2014	full	Contig	latest
21	*Geobacillus sp. G11MC16 ctg52*	GCA_000173035.1_ASM17303v1_genomic	Agharkar Research Institute	4/7/2014	full	Contig	latest

22	*Geobacillus sp. GHH01*	GCA_000336445.1_ASM33644v1_genomic	US DOE Joint Genome Institute (JGI-PGF)	8/15/2008	full	Contig	latest
23	*Geobacillus sp. JS12*	GCA_001592395.1_ASM159239v1_genomic	Georg-August-University Goettingen, Genomic and Applied Microbiology, Goettingen Genomics Laboratory	2/4/2013	full	Complete Genome	latest
24	*Geobacillus sp. LEMMY01*	GCA_002042905.1_ASM204290v1_genomic	Korea Polar Research Institute	3/15/2016	full	Complete Genome	latest
25	*Geobacillus sp. MAS1*	GCA_000498995.1_GTMas1.0_genomic	University of Delaware	8/3/2015	full	Complete Genome	latest
26	*Geobacillus sp. PA-3*	GCA_001412125.1_ASM141212v1_genomic	Universidade Federal do Rio de Janeiro	3/20/2017	full	Contig	latest
27	*Geobacillus sp. Sah69*	GCA_001414205.1_ASM141420v1_genomic	University of Georgia	11/25/2013	full	Scaffold	latest
28	*Geobacillus sp. T-6*	GCA_001025095.1_ASM102509v1_genomic	JLU	10/23/2015	full	Contig	latest
29	*Geobacillus sp. WSUCF1*	GCA_000422025.1_DUJ_genomic	INTA	6/10/2015	full	Contig	latest
30	*Geobacillus sp. Y412MC52*	GCA_000174795.2_ASM17479v2_genomic	South Dakota School of Mines and Technology	7/11/2013	full	Contig	latest
31	*Geobacillus sp. Y412MC61*	GCA_000024705.1_ASM2470v1_genomic	US DOE Joint Genome Institute (JGI-PGF)	11/5/2010	full	Complete Genome	latest
32	*Geobacillus sp. ZGt-1*	GCA_001183885.1_ASM118388v1_genomic	Lund Univesrity	6/5/2015	full	Scaffold	latest
33	*Geobacillus stearothermophilus 10*	GCA_001274575.1_ASM127457v1_genomic	THE INSTITUTE OF CYTOLOGY AND GENETICS	8/25/2014	full	Contig	latest

34	*Geobacillus stearothermophilus ATCC 12980*	GCA_001277805.1_ASM127780v1_genomic	THE INSTITUTE OF CYTOLOGY AND GENETICS	9/8/2014	full	Contig	latest
35	*Geobacillus stearothermophilus ATCC 7953*	GCA_000705495.1_GbsDonk1.0_genomic	University of Groningen	3/11/2016	full	Scaffold	latest
36	*Geobacillus stearothermophilus strain 22*	GCA_000743495.1_ASM74349v1_genomic	US DOE Joint Genome Institute (JGI-PGF)	1/7/2011	full	Complete Genome	latest
37	*Geobacillus stearothermophilus strain 53*	GCA_000749985.1_ASM74998v1_genomic	US DOE Joint Genome Institute (JGI-PGF)	10/15/2009	full	Complete Genome	latest
38	*Geobacillus stearothermophilus strain A1*	GCA_001183895.1_ASM118389v1_genomic	New Zealand Genomics Ltd	7/21/2015	full	Scaffold	latest
39	*Geobacillus stearothermophilus strain B4109*	GCA_001587495.1_B4109_genomic	Griffith University	4/13/2016	full	Contig	latest
40	*Geobacillus stearothermophilus strain B4114*	GCA_001587395.1_B4114_genomic	New Zealand Genomics Ltd	7/21/2015	full	Scaffold	latest
41	*Geobacillus stearothermophilus strain D1*	GCA_001183885.1	New Zealand Genomics Ltd	7/21/2015	full	Scaffold	latest
42	*Geobacillus stearothermophilus strain GS27*	GCA_001651555.1_ASM165155v1_genomic	NIZO food research	5/25/2016	full	Scaffold	latest
43	*Geobacillus stearothermophilus strain P3*	GCA_001183915.1_ASM118391v1_genomic	New Zealand Genomics Ltd	7/21/2015	full	Scaffold	latest

44	*Geobacillus subterraneus PSS2*	GCA_000744755.1_ASM74475v1_genomic	Bacillus Genetic Stock Center	9/1/2015	full	Scaffold	latest
45	*Geobacillus subterraneus strain K*	GCA_001632595.1_ASM163259v1_genomic	The University of Oklahoma	8/26/2015	full	Complete Genome	latest
46	*Geobacillus subterraneus strain KCTC 3922*	GCA_001618685.1_ASM161868v1_genomic	University of Groningen	3/11/2016	full	Scaffold	latest
47	*Geobacillus thermocatenulatus GS-1*	GCA_000612265.1_GeoThe1.0_genomic	Jet Propulsion Laboratory, California Institute of Technology	4/3/2017	full	Scaffold	latest
48	*Geobacillus thermocatenulatus strain BGSC 93A1*	GCA_002217655.1_ASM221765v1_genomic	Jet Propulsion Laboratory, California Institute of Technology	4/3/2017	full	Scaffold	latest
49	*Geobacillus thermocatenulatus strain KCTC 3921*	GCA_002243665.1_ASM224366v1_genomic	DOE Joint Genome Institute	8/28/2014	full	Contig	latest
50	*Geobacillus thermocatenulatus strain SURF-114*	GCA_002077855.1_ASM207785v1_genomic	Research Center of Biotechnology RAS	4/26/2016	full	Contig	latest
51	*Geobacillus thermocatenulatus strain SURF-189*	GCA_002077835.1_ASM207783v1_genomic	Kyungpook National University	4/12/2016	full	Complete Genome	latest
52	*Geobacillus thermocatenulatus strain SURF-48B*	GCA_002077815.1_ASM207781v1_genomic	MRIGlobal	6/12/2014	full	Contig	latest
53	*Geobacillus thermodenitrificans NG80-2*	GCA_000015745.1_ASM1574v1_genomic	Kyungpook National University	8/8/2017	full	Complete Genome	latest

54	*Geobacillus thermodenitrificans strain KCTC3902*	GCA_002072065.1_ASM207206v1_genomic	University of Pretoria	7/13/2017	full	Contig	latest
55	*Geobacillus thermodenitrificans strain T12*	GCA_002119625.1_ASM211962v1_genomic	Jet Propulsion Laboratory, California Institute of Technology	4/3/2017	full	Contig	latest
56	*Geobacillus thermodenitrificans subsp. thermodenitrificans DSM 465*	GCA_000496575.1_DSM465V1.0_genomic	The First Affiliated Hospital, School of Medicine, Zhejiang University	4/2/2014	full	Contig	latest
57	*Geobacillus thermoleovorans B23 DNA*	GCA_000474195.1_ASM474199v1_genomic	Nankai university	11/12/2013	full	Contig	latest
58	*Geobacillus thermoleovorans CCB_US3_UF5*	GCA_000236605.1_ASM23660v1_genomic	Kyungpook National University	4/5/2016	full	Complete Genome	latest
59	*Geobacillus thermoleovorans strain FJAT-2391*	GCA_001719205.1_ASM171920v1_genomic	TEDA School of Biological Sciences and Biotechnology, Nankai University	3/29/2007	full	Complete Genome	latest
60	*Geobacillus thermoleovorans strain KCTC 3570*	GCA_001610955.1_ASM161095v1_genomic	Kyungpook National University	3/23/2017	full	Complete Genome	latest
61	*Geobacillus thermoleovorans strain N7*	GCA_001707765.1_ASM170776v1_genomic	Wageningen University & Research	5/9/2017	full	Complete Genome	latest
62	*Geobacillus uzenensis strain BGSC 92A1*	ASM221766v1	University of Pretoria	7/13/2017	full	Contig	latest
63	*Geobacillus vulcani PSS1*	ASM73384v1	DOE Joint Genome Institute	7/24/2014	full	Contig	latest

64	*Geobacillus zalihae NBRC 101842*	ASM154413v1	National Institute of Technology and Evaluation	1/26/2016	Full	Contig	latest
65	*Geobacillus sp. WCH70* [1]	GCA_000023385.1_ASM2338v1_genomic	University of Pretoria	10/26/2015	full	Contig	latest
66	*Geobacillus sp. Y4.1MC1*[1]	GCA_000166075.1_ASM16607v1_genomic	US DOE Joint Genome Institute (JGI-PGF)	6/11/2009	full	Complete Genome	latest
67	*Geobacillus sp. 44B*[1]	GCA_002077755.1_ASM207775v1_genomic	University of Groningen	4/20/2016	full	Scaffold	latest
68	*Geobacillus sp. 44C*[1]	GCA_002077865.1_ASM207786v1_genomic	Jet Propulsion Laboratory, California Institute of Technology	4/3/2017	full	Contig	latest
69	*Geobacillus galactosidasius strain DSM 18751*[1]	GCA_002217735.1	University of Pretoria	7/13/2017	full	Contig	latest

([1]) These entries did not contain all three *spo0A*, *rpoB*, *recN* genes. Thus, they were excluded from the concatenated tree built on these markers.

Table 2: Complete list of the CAZome of *'Geobacillus icigianus SP50' as* identified by running local HMMER v3.1b2 against the dbCAN database.

GH family	HMM length	Query ID	Query length	E value	HMM start	HMM end	Query start	Query end	Coverage
GH67	669	fig|6666666.283630.peg.1	682	0	4	669	7	678	0.994021
CBM50	40	fig|6666666.283630.peg.1009	512	1.60E-16	1	40	4	48	0.975
GT51	177	fig|6666666.283630.peg.1109	923	3.10E-59	4	177	85	271	0.977401

CE10	341	fig|6666666.283630.peg.12	487	4.30E-45	76	313	69	408	0.695015
GH43_27	285	fig|6666666.283630.peg.1202	604	1.10E-121	1	285	285	580	0.996491
GH32	293	fig|6666666.283630.peg.1202	604	1.50E-07	78	172	8	125	0.320819
GH51	630	fig|6666666.283630.peg.1203	502	1.60E-152	82	542	4	500	0.730159
GH109	126	fig|6666666.283630.peg.1204	331	1.80E-10	18	121	12	116	0.81746
SLH	42	fig|6666666.283630.peg.1212	825	0.00049	19	41	699	721	0.52381
SLH	42	fig|6666666.283630.peg.1212	825	5.70E-06	2	41	612	650	0.928571
GT2	168	fig|6666666.283630.peg.1239	320	1.00E-29	1	167	5	167	0.988095
GT83	540	fig|6666666.283630.peg.1240	680	3.10E-31	37	226	35	229	0.35
GT83	540	fig|6666666.283630.peg.1240	680	4.70E-09	259	444	338	516	0.342593
GH18	296	fig|6666666.283630.peg.1276	420	2.10E-44	69	278	171	401	0.706081
CBM50	40	fig|6666666.283630.peg.1276	420	6.10E-12	1	40	4	46	0.975
GT2	168	fig|6666666.283630.peg.1308	319	1.90E-24	3	156	7	170	0.910714
CBM50	40	fig|6666666.283630.peg.1365	216	2.80E-14	1	40	32	74	0.975

AA2	255	fig|6666666.283630.peg.1408	736	3.90E-11	18	251	459	719	0.913725
AA2	255	fig|6666666.283630.peg.1408	736	5.80E-18	7	117	76	188	0.431373
CE3	194	fig|6666666.283630.peg.1454	263	8.60E-11	3	189	58	251	0.958763
GT4	160	fig|6666666.283630.peg.1465	394	1.10E-38	5	160	194	358	0.96875
GT2	168	fig|6666666.283630.peg.1473	295	1.20E-24	1	167	10	168	0.988095
CBM50	40	fig|6666666.283630.peg.1481	340	1.20E-13	1	39	147	190	0.95
CBM50	40	fig|6666666.283630.peg.1481	340	4.50E-15	1	39	81	124	0.95
CBM50	40	fig|6666666.283630.peg.1481	340	5.90E-17	1	39	26	67	0.95
GT51	177	fig|6666666.283630.peg.154	681	6.40E-72	2	176	62	237	0.983051
CE4	130	fig|6666666.283630.peg.1542	328	4.70E-36	3	124	125	248	0.930769
GT2	168	fig|6666666.283630.peg.1780	468	5.80E-24	1	134	63	220	0.791667
CBM50	40	fig|6666666.283630.peg.1822	194	2.10E-18	1	40	28	69	0.975
CE10	341	fig|6666666.283630.peg.1844	672	1.80E-24	70	335	414	665	0.777126
GH133	372	fig|6666666.283630.peg.1891	698	2.20E-18	14	369	269	626	0.954301

CE1	227	fig|6666666.283630.peg.1935	240	2.30E-31	3	222	15	236	0.964758
GH23	135	fig|6666666.283630.peg.1969	198	5.90E-37	14	119	90	194	0.777778
GH39	431	fig|6666666.283630.peg.2	504	1.20E-189	3	431	12	437	0.993039
CE1	227	fig|6666666.283630.peg.2043	255	7.70E-12	11	199	16	226	0.828194
GH13_1	304	fig|6666666.283630.peg.2072	511	1.00E-82	1	304	65	363	0.996711
CBM34	120	fig|6666666.283630.peg.2076	587	3.20E-40	1	119	6	124	0.983333
GH13_20	285	fig|6666666.283630.peg.2076	587	7.40E-136	1	284	173	469	0.992982
GH13_31	350	fig|6666666.283630.peg.2176	555	2.40E-172	1	349	28	372	0.994286
AA6	195	fig|6666666.283630.peg.2205	202	1.30E-61	1	193	4	194	0.984615
GT8	257	fig|6666666.283630.peg.2261	276	5.20E-49	31	256	3	252	0.875486
GH109	126	fig|6666666.283630.peg.2318	388	1.10E-08	3	110	9	114	0.849206
GT94	283	fig|6666666.283630.peg.2320	395	1.80E-13	41	280	111	376	0.844523
GH109	126	fig|6666666.283630.peg.2333	384	2.80E-07	2	98	2	96	0.761905
CE4	130	fig|6666666.283630.peg.2371	265	6.00E-38	2	123	58	182	0.930769

GT28	157	fig|6666666.283630.peg.2396	353	8.60E-10	2	151	181	329	0.949045
GT2	168	fig|6666666.283630.peg.2402	302	4.40E-24	1	166	9	166	0.982143
GT51	177	fig|6666666.283630.peg.2522	720	1.10E-70	2	176	87	261	0.983051
GH109	126	fig|6666666.283630.peg.2576	328	1.80E-15	1	123	2	115	0.968254
AA4	522	fig|6666666.283630.peg.2584	455	8.40E-26	35	249	20	225	0.409962
GH1	429	fig|6666666.283630.peg.2585	478	9.40E-174	3	428	7	475	0.990676
GH4	179	fig|6666666.283630.peg.2587	447	4.80E-81	1	179	6	186	0.994413
GH73	128	fig|6666666.283630.peg.2617	875	9.90E-23	2	125	574	700	0.960938
GH1	429	fig|6666666.283630.peg.2651	455	2.90E-153	4	428	8	452	0.988345
CE9	373	fig|6666666.283630.peg.2717	406	2.20E-146	1	373	6	384	0.997319
CBM50	40	fig|6666666.283630.peg.272	165	5.20E-12	1	40	4	47	0.975
CBM50	40	fig|6666666.283630.peg.2750	179	1.30E-11	1	40	122	167	0.975
GH130	296	fig|6666666.283630.peg.277	352	7.90E-95	2	294	4	348	0.986486
CE14	124	fig|6666666.283630.peg.2814	237	3.30E-29	1	124	8	114	0.991935

GT4	160	fig|6666666.283630.peg.2815	378	3.80E-44	7	160	195	345	0.95625
GT51	177	fig|6666666.283630.peg.2829	886	8.10E-69	2	176	82	257	0.983051
GH36	688	fig|6666666.283630.peg.2848	728	0	2	687	12	709	0.99564
GH13_5	343	fig|6666666.283630.peg.2853	232	4.30E-50	238	342	2	106	0.303207
CBM34	120	fig|6666666.283630.peg.2856	1640	2.20E-20	3	119	268	392	0.966667
GH13_20	285	fig|6666666.283630.peg.2856	1640	6.70E-78	1	285	459	835	0.996491
CBM34	120	fig|6666666.283630.peg.2856	1640	0.00015	32	93	1090	1153	0.508333
CBM20	90	fig|6666666.283630.peg.2856	1640	2.10E-06	13	82	52	127	0.766667
CBM20	90	fig|6666666.283630.peg.2856	1640	2.70E-08	1	66	1256	1321	0.722222
GH18	296	fig|6666666.283630.peg.2928	430	2.80E-40	71	284	184	415	0.719595
CBM50	40	fig|6666666.283630.peg.2928	430	2.60E-14	1	40	53	96	0.975
CBM50	40	fig|6666666.283630.peg.2928	430	3.10E-14	1	40	4	47	0.975
GH2	752	fig|6666666.283630.peg.2939	1013	7.00E-227	20	751	39	896	0.972074
GT4	160	fig|6666666.283630.peg.2988	387	1.40E-28	6	146	197	332	0.875

GT4	160	fig|6666666.283630.peg.2989	362	7.90E-32	8	158	189	332	0.9375
CE7	313	fig|6666666.283630.peg.3044	352	2.80E-07	53	216	59	200	0.520767
GT2	168	fig|6666666.283630.peg.3059	342	1.20E-19	2	165	8	174	0.970238
AA4	522	fig|6666666.283630.peg.3085	482	1.30E-28	16	261	11	248	0.469349
GH109	126	fig|6666666.283630.peg.3244	359	9.40E-19	1	126	7	124	0.992063
GH109	126	fig|6666666.283630.peg.3246	384	1.50E-13	1	121	4	115	0.952381
GH13_29	345	fig|6666666.283630.peg.3294	563	2.90E-200	1	345	28	378	0.997101
CE3	194	fig|6666666.283630.peg.3322	267	6.40E-16	3	194	50	250	0.984536
CE7	313	fig|6666666.283630.peg.3373	250	8.30E-06	172	297	87	226	0.399361
CE1	227	fig|6666666.283630.peg.3373	250	8.60E-09	29	138	23	123	0.480176
GT2	168	fig|6666666.283630.peg.3378	273	2.80E-20	2	168	7	182	0.988095
GT2	168	fig|6666666.283630.peg.3379	336	2.60E-34	1	168	7	175	0.994048
GT4	160	fig|6666666.283630.peg.3383	446	1.50E-22	7	157	238	386	0.9375
GT4	160	fig|6666666.283630.peg.3385	386	2.40E-40	5	147	203	344	0.8875

GT4	160	fig|6666666.283630.peg.3388	407	5.40E-16	7	158	223	379	0.94375
GH109	126	fig|6666666.283630.peg.3392	337	4.40E-12	1	119	2	108	0.936508
GT28	157	fig|6666666.283630.peg.3441	380	8.00E-21	2	137	204	339	0.859873
CE4	130	fig|6666666.283630.peg.3443	242	5.10E-33	6	113	36	145	0.823077
GT28	157	fig|6666666.283630.peg.3444	357	3.70E-42	2	149	191	339	0.936306
GH23	135	fig|6666666.283630.peg.3613	877	7.80E-11	14	121	465	625	0.792593
GH1	429	fig|6666666.283630.peg.587	470	2.20E-162	2	428	10	469	0.993007
GH43_11	292	fig|6666666.283630.peg.606	535	1.30E-154	1	292	4	302	0.996575
CE3	194	fig|6666666.283630.peg.612	217	1.10E-15	53	194	58	211	0.726804
GH10	303	fig|6666666.283630.peg.614	407	2.50E-111	2	302	48	405	0.990099
GH10	303	fig|6666666.283630.peg.617	331	2.20E-118	1	302	7	330	0.993399
GH52	415	fig|6666666.283630.peg.618	705	1.10E-216	1	415	20	439	0.99759
CE12	210	fig|6666666.283630.peg.643	278	5.90E-84	2	210	41	249	0.990476
GH18	296	fig|6666666.283630.peg.663	470	2.20E-35	71	280	226	452	0.706081

CBM50	40	fig|6666666.283630.peg.663	470	2.00E-14	1	40	100	143	0.975
CBM50	40	fig|6666666.283630.peg.663	470	4.60E-13	1	40	4	47	0.975
CBM50	40	fig|6666666.283630.peg.663	470	8.50E-13	1	40	53	96	0.975
CBM50	40	fig|6666666.283630.peg.691	387	0.00078	6	40	8	46	0.85
CBM50	40	fig|6666666.283630.peg.713	103	0.00011	3	40	49	99	0.925
GH5_2	197	fig|6666666.283630.peg.805	368	8.30E-70	1	196	21	271	0.989848
GT81	293	fig|6666666.283630.peg.924	253	1.30E-22	19	204	2	178	0.631399

Table 3: Summary Table of all GH3 hydrolases found in *Geobacillus* genomes. For each entry are shown: The GenBank ID (Query ID), the GenBank ID, GH family and activity of the entry with the highest similarity as found by local BLASTp, the length of the alignment, the amino-acid identity, E-value and bit-scores produced by BLASTp.

	Query ID	Subject ID| GH family| E.C.	length	Identity %	E-value	Bit-score
1	KZM54368.1	AFK20866.1|GH3|3.2.1.-	350	36.857	2.14E-65	219
2	OQP05151.1	AFK20866.1|GH3|3.2.1.-	343	36.152	3.14E-69	233
3	GAD14306.1	AFK20866.1|GH3|3.2.1.-	340	36.471	8.39E-71	233
4	OQP08380.1	AFK20866.1|GH3|3.2.1.-	343	37.026	2.10E-72	241

5	EDY04875.1	AFK20866.1|GH3|3.2.1.-	343	37.026	1.49E-72	241
6	ABO68582.1	AFK20866.1|GH3|3.2.1.-	350	36.857	1.46E-72	241
7	ARA98339.1	AFK20866.1|GH3|3.2.1.-	343	37.026	5.52E-73	242
8	KQB91707.1	AFK20866.1|GH3|3.2.1.-	343	37.026	5.52E-73	242
9	ASS88724.1	AFK20866.1|GH3|3.2.1.-	340	37.941	3.94E-74	242
10	AGT33650.1	AFK20866.1|GH3|3.2.1.-	343	38.192	2.04E-74	246
11	ABO66659.1	BAD39285.1|GH3|3.2.1.52	306	51.961	5.54E-92	299
12	ARA96980.1	BAD39285.1|GH3|3.2.1.52	306	51.961	2.94E-92	300
13	KQB93628.1	BAD39285.1|GH3|3.2.1.52	306	51.961	2.46E-92	300
14	OQP09840.1	BAD39285.1|GH3|3.2.1.52	306	51.961	2.46E-92	300
15	ARP42411.1	BAD39285.1|GH3|3.2.1.52	306	51.961	2.35E-92	300
16	EDY06966.1	BAD39285.1|GH3|3.2.1.52	306	51.961	1.79E-92	300
17	ACS24203.1	BAD39285.1|GH3|3.2.1.52	306	52.614	2.09E-94	305
18	ADP75106.1	BAD39285.1|GH3|3.2.1.52	306	52.614	9.74E-95	306

19	OQP01011.1	BAD39285.1|GH3|3.2.1.52	306	52.288	6.12E-95	307
20	AGT32204.1	ABN52488.1|GH3|3.2.1.21	721	35.229	3.79E-109	351
21	KQB93111.1	ACK41548.1|GH3|3.2.1.21	725	34.621	1.11E-113	363
22	AGE23865.1	CAA91219.1|GH3|3.2.1.21|3.2.1.37	717	41.841	4.21E-176	526
23	OQP03417.1	CAA91219.1|GH3|3.2.1.21|3.2.1.37	717	42.12	4.99E-178	531
24	AMX83398.1	CAC97071.1|GH3|3.2.1.-|3.2.1.-	720	50.278	0	736
25	KZS25548.1	CAC97071.1|GH3|3.2.1.-|3.2.1.-	720	50.278	0	736
26	OXB90422.1	CAC97071.1|GH3|3.2.1.-|3.2.1.-	720	50.278	0	736

Table 4: Summary Table of all GH5 hydrolases found in *Geobacillus* genomes. For each entry are shown: The GenBank ID (Query ID), the GenBank ID, GH damily and activity of the entry with the highest similarity as found by local BLASTp, the length of the alignment, the amino-acid identity, E-value and bit-scores produced by BLASTp.

	Query ID	Subject ID| GH family| E.C.	length	Identity %	E-value	bitscore
1	AGT32208.1	ACH69873.1|GH5|3.2.1.4	505	37.03	1.39E-99	311
2	ESU73063.1	ACH69873.1|GH5|3.2.1.4	502	37.45	1.07E-102	320
3	fig:2.517.peg.3457	ACJ60856.1|GH5|3.2.1.4	392	99.75	0	810
4	KDE47535.1	ACH69873.1|GH5|3.2.1.4	237	45.99	1.26E-66	218
5	KQB93199.1	ACH69873.1|GH5|3.2.1.4	503	38.37	1.29E-105	327

Table 5: Glycoside hydrolases from *Geobacillus* genomes containing a N-terminus GH51 domain(pf00933). For each entry, the closest entry from DBcan, the length, the percentage identity, evalue and bitscore of the alignment are given. The description of the closest characterized entry as well as the source organism of the query, are additionally given.

	Query ID	Subject ID| GH family| E.C.	length	Identity %	E-value	Bit-score
1	ABO67151.1	EDY06090.1|GH51|3.2.1.55	502	99.602	0	1043
2	ACX79307.1	AAD45520.2|GH51|3.2.1.55	502	97.809	0	1027
3	ACX79309.1	ACE73681.1|GH51|3.2.1.55	505	99.802	0	1060
4	ADI26630.1	AAD45520.2|GH51|3.2.1.55	502	97.41	0	1023
5	ADU94277.1	AAD45520.2|GH51|3.2.1.55	502	97.809	0	1027
6	ADU94279.1	ACE73681.1|GH51|3.2.1.55	505	99.802	0	1060
7	AGE22472.1	AAD45520.2|GH51|3.2.1.55	501	97.006	0	1021
8	AGT32226.1	AAD45520.2|GH51|3.2.1.55	502	97.211	0	1018
9	AMQ20687.1	AAD45520.2|GH51|3.2.1.55	502	97.211	0	1024
10	AMQ20688.1	ACE73681.1|GH51|3.2.1.55	495	99.596	0	1039
11	AOL34702.1	AAD45520.2|GH51|3.2.1.55	501	97.405	0	1024
12	ARA99616.1	EDY06090.1|GH51|3.2.1.55	502	99.801	0	1046
13	ARP42915.1	EDY06090.1|GH51|3.2.1.55	502	100	0	1047
14	EDY06090.1	EDY06090.1|GH51|3.2.1.55	502	100	0	1047
15	EPR27557.1	ACE73681.1|GH51|3.2.1.55	505	99.604	0	1057
16	EPR27560.1	ABD48560.1|GH51|3.2.1.55	401	97.257	0	819
17	EQB97033.1	AAD45520.2|GH51|3.2.1.55	501	97.206	0	1022
18	ESU72549.1	AAD45520.2|GH51|3.2.1.55	502	97.41	0	1025
19	fig:2.517.peg.3056	ABD48560.1|GH51|3.2.1.55	502	98.207	0	1032
20	GAD13935.1	EDY06090.1|GH51|3.2.1.55	495	97.172	0	1014
21	KDE47551.1	EDY06090.1|GH51|3.2.1.55	495	96.97	0	1014

22	KDE48495.1	ABD48560.1|GH51|3.2.1.55	502	97.41	0	1025
23	KFL17040.1	AAD45520.2|GH51|3.2.1.55	501	97.605	0	1025
24	KFL17041.1	ACE73681.1|GH51|3.2.1.55	505	99.208	0	1055
25	KFX35278.1	ACE73681.1|GH51|3.2.1.55	505	99.208	0	1055
26	KFX35279.1	AAD45520.2|GH51|3.2.1.55	501	97.605	0	1025
27	KJE25695.1	ACE73681.1|GH51|3.2.1.55	505	99.604	0	1058
28	KJE26225.1	AAD45520.2|GH51|3.2.1.55	502	98.008	0	1031
29	KPC99743.1	AAD45520.2|GH51|3.2.1.55	327	96.942	0	664
30	KQB93193.1	AAD45520.2|GH51|3.2.1.55	502	97.61	0	1022
31	KQB93195.1	ACE73681.1|GH51|3.2.1.55	497	90.141	0	964
32	KZM53195.1	ABD48560.1|GH51|3.2.1.55	502	96.813	0	1021
33	OQP08299.1	EDY06090.1|GH51|3.2.1.55	502	99.602	0	1045
34	OQP13834.1	ACE73681.1|GH51|3.2.1.55	505	98.416	0	1053
35	OQP13835.1	AAD45520.2|GH51|3.2.1.55	502	97.012	0	1024
36	OQP24143.1	ABD48560.1|GH51|3.2.1.55	502	97.41	0	1025
37	OQP24145.1	ACE73681.1|GH51|3.2.1.55	505	99.604	0	1057

References

Ahmetović, E., M. Martín and I. E. Grossmann (2010). "Optimization of Energy and Water Consumption in Corn-Based Ethanol Plants." Industrial & Engineering Chemistry Research **49**(17): 7972-7982.

Aliyu, H., P. Lebre, J. Blom, D. Cowan and P. De Maayer (2016). "Phylogenomic re-assessment of the thermophilic genus Geobacillus." Systematic and Applied Microbiology **39**(8): 527-533.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). "Basic local alignment search tool." Journal of Molecular Biology **215**(3): 403-410.

Arantes, V., A. Milagres, T. Filley and B. Goodell (2014). "Lignocellulosic polysaccharides and lignin degradation by wood decay fungi: the relevance of nonenzymatic Fenton-based reactions." Journal of Industrial Microbiology & Biotechnology **38**(4): 541-555.

Arevalo-Gallegos, A., Z. Ahmad, M. Asgher, R. Parra-Saldivar and H. M. N. Iqbal (2017). "Lignocellulose: A sustainable material to produce value-added products with a zero waste approach—A review." International Journal of Biological Macromolecules **99**: 308-318.

Artzi, L., E. A. Bayer and S. Moraïs (2016). "Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides." Nature Reviews Microbiology **15**: 83.

Ash, C., J. A. E. Farrow, W. Sally and M. D. Collins (1991). "Phylogenetic heterogeneity of the genus Bacillus revealed by comparative analysis of small-subunit-ribosomal RNA sequences." Letters in Applied Microbiology **13**(4): 202-206.

Assareh, R., H. Shahbani Zahiri, K. Akbari Noghabi, S. Aminzadeh and G. Bakhshi khaniki (2012). "Characterization of the newly isolated Geobacillus sp. T1, the efficient cellulase-producer on untreated barley and wheat straws." Bioresource Technology **120**: 99-105.

Aziz, R. K., D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke and O. Zagnitko (2008). "The RAST Server: Rapid Annotations using Subsystems Technology." BMC Genomics **9**: 75-75.

Bacon, L. F., C. Hamley-Bennett, M. J. Danson and D. J. Leak (2017). "Development of an efficient technique for gene deletion and allelic exchange in Geobacillus spp." Microbial Cell Factories **16**: 58.

Bairoch, A. and R. Apweiler (2000). "The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000." Nucleic Acids Research **28**(1): 45-48.

Balan, V. (2014). "Current Challenges in Commercially Producing Biofuels from Lignocellulosic Biomass." ISRN biotechnology **2014**: 463074.

Balat, M. (2011). "Production of bioethanol from lignocellulosic materials via the biochemical pathway: A review." Energy Conversion and Management **52**(2): 858-875.

Barr, B. K., Y.-L. Hsieh, B. Ganem and D. B. Wilson (1996). "Identification of Two Functionally Different Classes of Exocellulases." Biochemistry **35**(2): 586-592.

Barr, B. K., D. E. Wolfgang, K. Piens, M. Claeyssens and D. B. Wilson (1998). "Active-Site Binding of Glycosides by Thermomonospora fusca Endocellulase E2." Biochemistry **37**(26): 9220-9229.

Bayer, E. A., J.-P. Belaich, Y. Shoham and R. Lamed (2004). "The cellulosomes: Multi-enzyme machines for degradation of plant cell wall polysaccharides." Ann ual Review of Microbiology **58**: 521-554.

Beg, Q. K., M. Kapoor, L. Mahajan and G. S. Hoondal (2001). "Microbial xylanases and their industrial applications: a review." Applied Microbiology and Biotechnology **56**(3): 326-338.

Bezuidt, O. K., R. Pierneef, A. M. Gomri, F. Adesioye, T. P. Makhalanyane, K. Kharroub and D. A. Cowan (2016). "The Geobacillus Pan-Genome: Implications for the Evolution of the Genus." Frontiers in Microbiology **7**: 723.

Bhalla, A., N. Bansal, S. Kumar, K. M. Bischoff and R. K. Sani (2013). "Improved lignocellulose conversion to biofuels with thermophilic bacteria and thermostable enzymes." Bioresource Technology **128**: 751-759.

Björdal, C. G. (2012). "Evaluation of microbial degradation of shipwrecks in the Baltic Sea." International Biodeterioration & Biodegradation **70**: 126-140.

Blum, H., H. Beier and H. J. Gross (1987). "Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels." Electrophoresis **8**(2): 93-99.

Blumer-Schuette, S. E., S. D. Brown, K. B. Sander, E. A. Bayer, I. Kataeva, J. V. Zurawski, J. M. Conway, M. W. W. Adams and R. M. Kelly (2014). Thermophilic lignocellulose deconstruction.

Blumer-Schuette, S. E., I. Kataeva, J. Westpheling, M. W. W. Adams and R. M. Kelly (2008). "Extremely thermophilic microorganisms for biomass conversion: status and prospects." Current Opinion in Biotechnology **19**(3): 210-217.

Bosi, E., B. Donati, M. Galardini, S. Brunetti, M.-F. Sagot, P. Lió, P. Crescenzi, R. Fani and M. Fondi (2015). "MeDuSa: a multi-draft based scaffolder." Bioinformatics **31**(15): 2443-2451.

Brumm, P. J., P. De Maayer, D. A. Mead and D. A. Cowan (2015). "Genomic analysis of six new Geobacillus strains reveals highly conserved carbohydrate degradation architectures and strategies." Frontiers in Microbiology **6**: 430.

Brux, C., A. Ben-David, D. Shallom-Shezifi, M. Leon, K. Niefind, G. Shoham, Y. Shoham and D. Schomburg (2006). "The Structure of an Inverting GH43 *β*-Xylosidase from *Geobacillus stearothermophilus* with its Substrate Reveals the Role of the Three Catalytic Residues." Journal of Molecular Biology **359**(1): 97-109.

Bugg, T. D. H., M. Ahmad, E. M. Hardiman and R. Rahmanpour (2011). "Pathways for degradation of lignin in bacteria and fungi." Natural Product Reports **28**(12): 1883-1896.

Bugg, T. D. H., M. Ahmad, E. M. Hardiman and R. Singh (2011). "The emerging role for bacteria in lignin degradation and bio-product formation." Current Opinion in Biotechnology **22**(3): 394-400.

Canakci, S., Z. Cevher, K. Inan, M. Tokgoz, F. Bahar, M. Kacagan, F. A. Sal and A. O. Belduz (2012). "Cloning, purification and characterization of an alkali-stable endoxylanase from thermophilic Geobacillus sp. 71." World J Microbiol Biotechnol **28**(5): 1981-1988.

Capolupo, L. and V. Faraco (2016). "Green methods of lignocellulose pretreatment for biorefinery development." Applied Microbiology and Biotechnology **100**(22): 9451-9467.

Chakdar, H., M. Kumar, K. Pandiyan, A. Singh, K. Nanjappan, P. L. Kashyap and A. K. Srivastava (2016). "Bacterial xylanases: biology to biotechnology." 3 Biotech **6**(2): 150.

Chen, Z. F., S. F. Wojcik and N. E. Welker (1986). "Genetic analysis of Bacillus stearothermophilus by protoplast fusion." Journal of Bacteriology **165**(3): 994-1001.

Chew, K. W., J. Y. Yap, P. L. Show, N. H. Suan, J. C. Juan, T. C. Ling, D.-J. Lee and J.-S. Chang (2017). "Microalgae biorefinery: High value products perspectives." Bioresource Technology **229**: 53-62.

Choi, J. M., Y.-J. Lee, T.-P. Cao, S.-M. Shin, M.-K. Park, H.-S. Lee, E. di Luccio, S.-B. Kim, S.-J. Lee, S. J. Lee, S. H. Lee and D.-W. Lee (2016). "Structure of the thermophilic l-Arabinose isomerase from Geobacillus kaustophilus reveals metal-mediated intersubunit interactions for activity and thermostability." Archives of Biochemistry and Biophysics **596**: 51-62.

Choi, Y.-J. and S.-G. Cho (1995). "Nucleotide sequence analysis of an endo-xylanase gene (xynA) from Bacillus stearothermophilus." Journal of Microbiology and Biotechnology **5**(3): 117-124.

Collins, T., C. Gerday and G. Feller (2005). "Xylanases, xylanase families and extremophilic xylanases." FEMS Microbiology Reviews **29**(1): 3-23.

Consortium, U. (2017). "UniProt: the universal protein knowledgebase." Nucleic Acids Research **45**(D1): D158-D169.

Coordinators, N. R. (2016). "Database resources of the National Center for Biotechnology Information." Nucleic Acids Research **44**(Database issue): D7-D19.

Corpet, F. (1988). "Multiple sequence alignment with hierarchical clustering." Nucleic Acids Research **16**(22): 10881-10890.

Cragg, S. M., G. T. Beckham, N. C. Bruce, T. D. H. Bugg, D. L. Distel, P. Dupree, A. G. Etxabe, B. S. Goodell, J. Jellison, J. E. McGeehan, S. J. McQueen-Mason, K. Schnorr, P. H. Walton, J. E. M. Watts and M. Zimmer (2015). "Lignocellulose degradation mechanisms across the Tree of Life." Current Opinion in Chemical Biology **29**: 108-119.

Cripps, R. E., K. Eley, D. J. Leak, B. Rudd, M. Taylor, M. Todd, S. Boakes, S. Martin and T. Atkinson (2009). "Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production." Metabolic Engineering **11**(6): 398-408.

Darling, A. C. E., B. Mau, F. R. Blattner and N. T. Perna (2004). "Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements." Genome Research **14**(7): 1394-1403.

De Maayer, P., P. J. Brumm, D. A. Mead and D. A. Cowan (2014). "Comparative analysis of the Geobacillus hemicellulose utilization locus reveals a highly variable target for improved hemicellulolysis." BMC Genomics **15**(1): 836.

De Souza, W. R. (2013). "Microbial degradation of lignocellulosic biomass." Chandel A, Da Silva, S. Sustainable degradation of lignocellulosic biomass-techniques, applications and commercialization. Brazil: InTech.

Elliott, D. C., P. Biller, A. B. Ross, A. J. Schmidt and S. B. Jones (2015). "Hydrothermal liquefaction of biomass: Developments from batch to continuous process." Bioresource Technology **178**: 147-156.

Eppink, M. H. M., G. Olivieri, H. Reith, C. van den Berg, M. J. Barbosa and R. H. Wijffels (2017). From Current Algae Products to Future Biorefinery Practices: A Review. Berlin, Heidelberg, Springer Berlin Heidelberg**:** 1-25.

European-Biomass-Association (2016). European Bioenergy outlook AEBIOM statistical report. E. B. Association. Place du Champ de Mars 2, 1050 Brussels, European Biomass Association.

Floudas, D., M. Binder, R. Riley, K. Barry, R. A. Blanchette, B. Henrissat, A. T. Martínez, R. Otillar, J. W. Spatafora, J. S. Yadav, A. Aerts, I. Benoit, A. Boyd, A. Carlson, A. Copeland, P. M. Coutinho, R. P. de Vries, P. Ferreira, K. Findley, B. Foster, J. Gaskell, D. Glotzer, P. Górecki, J. Heitman, C. Hesse, C. Hori, K. Igarashi, J. A. Jurgens, N. Kallen, P. Kersten, A. Kohler, U. Kües, T. K. A. Kumar, A. Kuo, K. LaButti, L. F. Larrondo, E. Lindquist, A. Ling, V.

Lombard, S. Lucas, T. Lundell, R. Martin, D. J. McLaughlin, I. Morgenstern, E. Morin, C. Murat, L. G. Nagy, M. Nolan, R. A. Ohm, A. Patyshakuliyeva, A. Rokas, F. J. Ruiz-Dueñas, G. Sabat, A. Salamov, M. Samejima, J. Schmutz, J. C. Slot, F. St. John, J. Stenlid, H. Sun, S. Sun, K. Syed, A. Tsang, A. Wiebenga, D. Young, A. Pisabarro, D. C. Eastwood, F. Martin, D. Cullen, I. V. Grigoriev and D. S. Hibbett (2012). "The Paleozoic Origin of Enzymatic Lignin Decomposition Reconstructed from 31 Fungal Genomes." Science **336**(6089): 1715-1719.

Fox, G., E. Stackebrandt, R. Hespell, J. Gibson, J. Maniloff, T. Dyer, R. Wolfe, W. Balch, R. Tanner, L. Magrum, L. Zablen, R. Blakemore, R. Gupta, L. Bonen, B. Lewis, D. Stahl, K. Luehrsen, K. Chen and C. Woese (1980). "The phylogeny of prokaryotes." Science **209**(4455): 457-463.

Galanopoulou, A. P. (2012). Καθαρισμός, βιοχημικός και κινητικός χαρακτηρισμός μίας β-ξυλοζιδάσης από ένα θερμόφιλο βακτηριακό στέλεχος Geobacillus sp. απομονωμένο από το ηφαιστειακό σύμπλεγμα της Σαντορίνης, Μεταπτυχιακή εργασία, ΕΚΠΑ.

Galanopoulou, A. P., S. Moraïs, A. Georgoulis, E. Morag, E. A. Bayer and D. G. Hatzinikolaou (2016). "Insights into the functionality and stability of designer cellulosomes at elevated temperatures." Applied Microbiology and Biotechnology **100**(20): 8731-8743.

Gasteiger, E., C. Hoogland, A. Gattiker, S. e. Duvaud, M. Wilkins, R. Appel and A. Bairoch (2005). Protein Identification and Analysis Tools on the ExPASy Server. The Proteomics Protocols Handbook. J. Walker, Humana Press**:** 571-607.

Gat, O., A. Lapidot, I. Alchanati, C. Regueros and Y. Shoham (1994). "Cloning and DNA sequence of the gene coding for Bacillus stearothermophilus T-6 xylanase." Appl Environ Microbiol **60**(6): 1889-1896.

Ghangas, G. S. and D. B. Wilson (1988). "Cloning of the Thermomonospora fusca Endoglucanase E2 Gene in Streptomyces lividans: Affinity Purification and Functional Domains of the Cloned Gene Product." Applied and Environmental Microbiology **54**(10): 2521-2526.

Gordon, R. E., W. C. Haynes and C. H. N. Pang (1973). The Genus Bacillus, Agricultural Research Service, U.S. Department of Agriculture.

Gordon, R. E. and N. R. Smith (1949). "aerobic sporeforming bacteria capable of growth at high temperatures." Journal of Bacteriology **58**(3): 327-341.

Grigoriev, I. V., D. Cullen, S. B. Goodwin, D. Hibbett, T. W. Jeffries, C. P. Kubicek, C. Kuske, J. K. Magnuson, F. Martin, J. W. Spatafora, A. Tsang and S. E. Baker (2011). "Fueling the future with fungal genomics." Mycology **2**(3): 192-209.

Hall, B. G. (2013). "Building Phylogenetic Trees from Molecular Data with MEGA." Molecular Biology and Evolution **30**(5): 1229-1235.

Harjunpää, V., A. Teleman, A. Koivula, L. Ruohonen, T. T. Teeri, O. Teleman and T. Drakenberg (1996). "Cello-Oligosaccharide Hydrolysis by Cellobiohydrolase II from Trichoderma Reesei." European Journal of Biochemistry **240**(3): 584-591.

Henrissat, B. (1991). "A classification of glycosyl hydrolases based on amino acid sequence similarities." Biochemical Journal **280**(2): 309-316.

Henrissat, B. and G. Davies (1997). "Structural and sequence-based classification of glycoside hydrolases." Current Opinion in Structural Biology **7**(5): 637-644.

Hiras, J., Y.-W. Wu, K. Deng, C. D. Nicora, J. T. Aldrich, D. Frey, S. Kolinko, E. W. Robinson, J. M. Jacobs, P. D. Adams, T. R. Northen, B. A. Simmons and S. W. Singer (2016). "Comparative Community Proteomics Demonstrates the Unexpected Importance of Actinobacterial Glycoside Hydrolase Family 12 Protein for Crystalline Cellulose Hydrolysis." mBio **7**(4).

Ho, D. P., H. H. Ngo and W. Guo (2014). "A mini review on renewable sources for biofuel." Bioresource Technology **169**: 742-749.

Hong, M.-R., Y.-S. Kim, C.-S. Park, J.-K. Lee, Y.-S. Kim and D.-K. Oh (2009). "Characterization of a recombinant β-glucosidase from the thermophilic bacterium Caldicellulosiruptor saccharolyticus." Journal of Bioscience and Bioengineering **108**(1): 36-40.

Hoornweg D., B.-T. P. (2012). What a waste-A Global Review of Solid Waste Management. Urban Development & Local Government Unit World Bank 1818 H Street, NW Washington, DC 20433 USA, World Bank. **15**.

Huang, Z., X. Liu, S. Zhang and Z. Liu (2014). "GH52 xylosidase from Geobacillus stearothermophilus: characterization and introduction of xylanase activity by sitedirected mutagenesis of Tyr509." J Ind Microbiol Biotechnol **41**(1): 65-74.

Hussein, A. H., B. K. Lisowska and D. J. Leak (2015). Chapter One - The Genus Geobacillus and Their Biotechnological Potential. Advances in Applied Microbiology. S. Sima and G. Geoffrey Michael, Academic Press. **Volume 92:** 1-48.

Kananaviciute, R. and D. Citavicius (2015). "Genetic engineering of Geobacillus spp." J Microbiol Methods **111**: 31-39.

Katoh, K. and D. M. Standley (2013). "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." Molecular Biology and Evolution **30**(4): 772-780.

Kersten, P. and D. Cullen (2007). "Extracellular oxidative systems of the lignin-degrading Basidiomycete Phanerochaete chrysosporium." Fungal Genetics and Biology **44**(2): 77-87.

Khasin, A., I. Alchanati and Y. Shoham (1993). "Purification and characterization of a thermostable xylanase from Bacillus stearothermophilus T-6." Appl Environ Microbiol **59**(6): 1725-1730.

Kitts, P. A., D. M. Church, F. Thibaud-Nissen, J. Choi, V. Hem, V. Sapojnikov, R. G. Smith, T. Tatusova, C. Xiang, A. Zherikov, M. DiCuccio, T. D. Murphy, K. D. Pruitt and A. Kimchi (2016). "Assembly: a resource for assembled genomes at NCBI." Nucleic Acids Research **44**(Database issue): D73-D80.

Koeck, D. E., A. Pechtl, V. V. Zverlov and W. H. Schwarz (2014). "Genomics of cellulolytic bacteria." Current Opinion in Biotechnology **29**: 171-183.

Konstantinidis, K. T. and J. M. Tiedje (2005). "Towards a Genome-Based Taxonomy for Prokaryotes." Journal of Bacteriology **187**(18): 6258-6264.

Kuisiene, N., J. Raugalas and D. Chitavichius (2009). "Phylogenetic, Inter, and Intraspecific Sequence Analysis of spo0A Gene of the Genus Geobacillus." Current Microbiology **58**(6): 547.

Kumar, R., S. Singh and O. Singh (2008). "Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives." Journal of Industrial Microbiology & Biotechnology **35**(5): 377-391.

Lai, X. and L. O. Ingram (1993). "Cloning and sequencing of a cellobiose phosphotransferase system operon from Bacillus stearothermophilus XL-65-6 and functional expression in Escherichia coli." Journal of Bacteriology **175**(20): 6441-6450.

Lane, D. J. (1991). 16S/23S rRNA Sequencing. Nucleic acid techniques in bacterial systematics. E. Stackebrandt and M. Goodfellow. New York, John Wiley & Sons Ltd**:** 115-147.

Lee, H. V., S. B. A. Hamid and S. K. Zain (2014). "Conversion of Lignocellulosic Biomass to Nanocellulose: Structure and Chemical Process." The Scientific World Journal **2014**: 631013.

Leisola, M., O. Pastinen and D. Axe (2012). Lignin—Designed Randomness.

Letunic, I. and P. Bork (2016). "Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees." Nucleic Acids Research **44**(W1): W242-W245.

Lin, P. P., K. S. Rabe, J. L. Takasumi, M. Kadisch, F. H. Arnold and J. C. Liao (2014). "Isobutanol production at elevated temperatures in thermophilic Geobacillus thermoglucosidasius." Metabolic Engineering **24**: 1-8.

Lombard, V., H. Golaconda Ramulu, E. Drula, P. M. Coutinho and B. Henrissat (2014). "The carbohydrate-active enzymes database (CAZy) in 2013." Nucleic Acids Research **42**(Database issue): D490-D495.

Lynd, L. R., M. S. Laser, D. Bransby, B. E. Dale, B. Davison, R. Hamilton, M. Himmel, M. Keller, J. D. McMillan, J. Sheehan and C. E. Wyman (2008). "How biotech can transform biofuels." Nat Biotech **26**(2): 169-172.

Lynd, L. R., P. J. Weimer, W. H. van Zyl and I. S. Pretorius (2002). "Microbial Cellulose Utilization: Fundamentals and Biotechnology." Microbiol. Mol. Biol. Rev. **66**(3): 506-577.

Marchler-Bauer, A., M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz, L. Y. Geer, R. C. Geer, J. He, M. Gwadz, D. I. Hurwitz, C. J. Lanczycki, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, Z. Wang, R. A. Yamashita, D. Zhang, C. Zheng and S. H. Bryant (2015). "CDD: NCBI's conserved domain database." Nucleic Acids Research **43**(Database issue): D222-D226.

McMullan, G., J. M. Christie, T. J. Rahman, I. M. Banat, N. G. Ternan and R. Marchant (2004). "Habitat, applications and genomics of the aerobic, thermophilic genus Geobacillus." Biochem. Soc. Trans. **32**(Pt 2): 214-217.

Meintanis, C., K. Chalkou, K. Kormas and A. D. Karagouni (2006). "Biodegradation of crude oil by thermophilic bacteria isolated from a volcano island." Biodegradation **17**(2): 105-111.

Meintanis, C., K. Chalkou, K. A. Kormas, D. Lymperopoulou, E. Katsifas, D. Hatzinikolaou and A. Karagouni (2008). "Application of rpoB sequence similarity analysis, REP-PCR and BOX-PCR for the differentiation of species within the genus *Geobacillus*." Letters in Applied Microbiology **46**(3): 395-401.

Menon, V. and M. Rao (2012). "Trends in bioconversion of lignocellulose: Biofuels, platform chemicals & biorefinery concept." Progress in Energy and Combustion Science **38**(4): 522-550.

Miller, R. G. and S. R. Sorrell (2014). "The future of oil supply." Philosophical transactions. Series A, Mathematical, physical, and engineering sciences **372**(2006): 20130179.

Minana-Galbis, D., D. L. Pinzon, J. G. Loren, A. Manresa and R. M. Oliart-Ros (2010). "Reclassification of *Geobacillus pallidus* (Scholz et al. 1988) Banat et al. 2004 as *Aeribacillus pallidus gen. nov., comb. nov*." International Journal of Systematic and Evolutionary Microbiology **60**(7): 1600-1604.

Moraïs, S., Y. Barak, J. Caspi, Y. Hadar, R. Lamed, Y. Shoham, D. B. Wilson and E. A. Bayer (2010). "Cellulase-Xylanase Synergy in Designer Cellulosomes for Enhanced Degradation of a Complex Cellulosic Substrate." mBio **1**(5).

Moraïs, S., Y. Barak, Y. Hadar, D. B. Wilson, Y. Shoham, R. Lamed and E. A. Bayer (2011). "Assembly of Xylanases into Designer Cellulosomes Promotes Efficient Hydrolysis of the Xylan Component of a Natural Recalcitrant Cellulosic Substrate." mBio **2**(6).

Moraïs, S., J. Stern, A. Kahn, A. P. Galanopoulou, S. Yoav, M. Shamshoum, M. A. Smith, D. G. Hatzinikolaou, F. H. Arnold and E. A. Bayer (2016).

"Enhancement of cellulosome-mediated deconstruction of cellulose by improving enzyme thermostability." Biotechnology for Biofuels **9**: 164.

Moretti, S., A. Wilm, D. G. Higgins, I. Xenarios and C. Notredame (2008). "R-Coffee: a web server for accurately aligning noncoding RNA sequences." Nucleic Acids Research **36**(Web Server issue): W10-W13.

Nazina, T. N., T. P. Tourova, A. B. Poltaraus, E. V. Novikova, A. A. Grigoryan, A. E. Ivanova, A. M. Lysenko, V. V. Petrunyaka, G. A. Osipov, S. S. Belyaev and M. V. Ivanov (2001). "Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus gen. nov.,* sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus, Bacillus thermoglucosidasius and Bacillus thermodenitrificans to Geobacillus as the new combinations G. stearothermophilus, G. thermocatenulatus, G. thermoleovorans, G. kaustophilus, G. thermoglucosidasius* and *G. thermodenitrificans*." Int J Syst Evol Microbiol **51**(2): 433-446.

Nehring, R. (2009). "Traversing the mountaintop: world fossil fuel production to 2050." Philosophical Transactions of the Royal Society B: Biological Sciences **364**(1532): 3067-3079.

Ng, I. S., C.-W. Li, Y.-F. Yeh, P. Chen, J.-L. Chir, C.-H. Ma, S.-M. Yu, T.-h. Ho and C.-G. Tong (2009). "A novel endo-glucanase from the thermophilic bacterium Geobacillus sp. 70PC53 with high activity and stability over a broad range of temperatures." Extremophiles **13**(3): 425-435.

Overbeek, R., R. Olson, G. D. Pusch, G. J. Olsen, J. J. Davis, T. Disz, R. A. Edwards, S. Gerdes, B. Parrello, M. Shukla, V. Vonstein, A. R. Wattam, F. Xia and R. Stevens (2014). "The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST)." Nucleic Acids Research **42**(Database issue): D206-D214.

Peralta-Yahya, P. P., F. Zhang, S. B. del Cardayre and J. D. Keasling (2012). "Microbial engineering for the production of advanced biofuels." Nature **488**(7411): 320-328.

Petersen, T. N., S. Brunak, G. von Heijne and H. Nielsen (2011). "SignalP 4.0: discriminating signal peptides from transmembrane regions." Nat Meth **8**(10): 785-786.

Pogrebnyakov, I., C. B. Jendresen and A. T. Nielsen (2017). "Genetic toolbox for controlled expression of functional proteins in Geobacillus spp." PLoS ONE **12**(2): e0171313.

Potprommanee, L., X.-Q. Wang, Y.-J. Han, D. Nyobe, Y.-P. Peng, Q. Huang, J.-y. Liu, Y.-L. Liao and K.-L. Chang (2017). "Characterization of a thermophilic cellulase from Geobacillus sp. HTA426, an efficient cellulase-producer on alkali pretreated of lignocellulosic biomass." PLOS ONE **12**(4): e0175004.

Quentin, Y., G. Fichant and F. Denizot (1999). "Inventory, assembly and analysis of Bacillus subtilis ABC transport systems11Edited by G. Von Heijne." Journal of Molecular Biology **287**(3): 467-484.

Rastogi, G., A. Bhalla, A. Adhikari, K. M. Bischoff, S. R. Hughes, L. P. Christopher and R. K. Sani (2010). "Characterization of thermostable cellulases produced by *Bacillus* and *Geobacillus* strains." Bioresource Technology **101**(22): 8798-8806.

Rodriguez-R, L. M., S. Gunturu, W. T. Harvey, R. Rosselló-Mora, J. M. Tiedje, J. R. Cole and K. T. Konstantinidis (2018). "The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level." Nucleic Acids Research: gky467-gky467.

Rodriguez-R, L. M. and K. T. Konstantinidis (2016). "The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes." PeerJ Preprints **4**: e1900v1901.

Saini, J. K., R. Saini and L. Tewari (2015). "Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: concepts and recent developments." 3 Biotech **5**(4): 337-353.

Schneider, C. A., W. S. Rasband and K. W. Eliceiri (2012). "NIH Image to ImageJ: 25 years of image analysis." Nat Meth **9**(7): 671-675.

Scoma, A., S. Rebecchi, L. Bertin and F. Fava (2016). "High impact biowastes from South European agro-industries as feedstock for second-generation biorefineries." Critical Reviews in Biotechnology **36**(1): 175-189.

Shallom, D., M. Leon, T. Bravman, A. Ben-David, G. Zaide, V. Belakhov, G. Shoham, D. Schomburg, T. Baasov and Y. Shoham (2004). "Biochemical Characterization and Identification of the Catalytic Residues of a Family 43 β-d-Xylosidase from *Geobacillus stearothermophilus* T-6." Biochemistry **44**(1): 387-397.

Shida, O., H. Takagi, K. Kadowaki and K. Komagata (1996). "Proposal for two new genera, Brevibacillus gen. nov. and Aneurinibacillus gen. nov." International Journal of Systematic Bacteriology **46**(4): 939-946.

Shoham, Y., R. Lamed and E. A. Bayer (1999). "The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides." Trends Microbiol. **7**: 275-281.

Shulami, S., O. Gat, A. L. Sonenshein and Y. Shoham (1999). "The Glucuronic Acid Utilization Gene Cluster from Bacillus stearothermophilus T-6." Journal of Bacteriology **181**(12): 3695-3704.

Shulami, S., A. Raz-Pasteur, O. Tabachnikov, S. Gilead-Gropper, I. Shner and Y. Shoham (2011). "The l-Arabinan Utilization System of Geobacillus stearothermophilus." Journal of Bacteriology **193**(11): 2838-2850.

Shulami, S., O. Shenker, Y. Langut, N. Lavid, O. Gat, G. Zaide, A. Zehavi, A. L. Sonenshein and Y. Shoham (2014). "Multiple regulatory mechanisms control the expression of the Geobacillus stearothermophilus gene for extracellular xylanase." J Biol Chem **289**(37): 25957-25975.

Shulami, S., G. Zaide, G. Zolotnitsky, Y. Langut, G. Feld, A. L. Sonenshein and Y. Shoham (2007). "A Two-Component System Regulates the Expression of an ABC Transporter for Xylo-Oligosaccharides in *Geobacillus stearothermophilus*." Appl. Environ. Microbiol. **73**(3): 874-884.

Solovyev V., S. A. (2011). Automatic Annotation of Microbial Genomes and Metagenomic Sequences. Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies. R. W. Li**:** 61-78.

Stathopoulou, P. M., A. P. Galanopoulou, G. E. Anasontzis, A. D. Karagouni and D. Hatzinikolaou (2012). "Assessment of the biomass hydrolysis potential in bacterial isolates from a volcanic environment: biosynthesis of the corresponding activities." World Journal of Microbiology and Biotechnology **28**(9): 2889-2902.

Stathopoulou, P. M., A. L. Savvides, A. D. Karagouni and D. G. Hatzinikolaou (2013). "Unraveling the Lipolytic Activity of Thermophilic Bacteria Isolated from a Volcanic Environment." BioMed Research International **2013**: Article ID 703130.

Stephanopoulos, G. (2007). "Challenges in Engineering Microbes for Biofuels Production." Science **315**(5813): 801-804.

Subramaniyan, S. and P. Prema (2002). "Biotechnology of Microbial Xylanases: Enzymology, Molecular Biology, and Application." Critical Reviews in Biotechnology **22**(1): 33 - 64.

Sullivan, M. J., N. K. Petty and S. A. Beatson (2011). "Easyfig: a genome comparison visualizer." Bioinformatics **27**(7): 1009-1010.

Suzuki, H., A. Murakami and K.-i. Yoshida (2012). "Counterselection System for Geobacillus kaustophilus HTA426 through Disruption of pyrF and pyrR." Applied and Environmental Microbiology **78**(20): 7376-7383.

Suzuki, H., F. Okazaki, A. Kondo and K.-i. Yoshida (2012). "Genome mining and motif modifications of glycoside hydrolase family 1 members encoded by <i>Geobacillus kaustophilus HTA426 provide thermostable 6-phospho-β-glycosidase and β-fucosidase." Applied Microbiology and Biotechnology: 1-10.

Suzuki, H., K.-i. Yoshida and T. Ohshima (2013). "Polysaccharide-Degrading Thermophiles Generated by Heterologous Gene Expression in Geobacillus kaustophilus HTA426." Applied and Environmental Microbiology **79**(17): 5151-5158.

T Konstantinidis, K., R. Rossello-Mora and R. Amann (2017). Uncultivated microbes in need of their own taxonomy.

Tabachnikov, O. and Y. Shoham (2013). "Functional characterization of the galactan utilization system of Geobacillus stearothermophilus." FEBS Journal **280**(3): 950-964.

Tai, S.-K., H.-P. P. Lin, J. Kuo and J.-K. Liu (2004). "Isolation and characterization of a cellulolytic Geobacillus thermoleovorans T4 strain from sugar refinery wastewater." Extremophiles **8**(5): 345-349.

Talavera, G. and J. Castresana (2007). "Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments." Systematic Biology **56**(4): 564-577.

Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar (2013). "MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0." Molecular Biology and Evolution **30**(12): 2725-2729.

Teather, R. M. and P. J. Wood (1982). "Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen." Applied and Environmental Microbiology **43**(4): 777-780.

Thompson, J., F. W. Lichtenthaler, S. Peters and A. Pikis (2002). "β-Glucoside Kinase (BglK) from Klebsiella pneumoniae : Purification,,properties, and preparative synthesis of 6-phospho-β-d-glucosides." Journal of Biological Chemistry **277**(37): 34310-34321.

Tominaga, Y., T. Ohshiro and H. Suzuki (2016). "Conjugative plasmid transfer from Escherichia coli is a versatile approach for genetic transformation of thermophilic Bacillus and Geobacillus species." Extremophiles: 1-7.

Van Zyl, L. J., M. P. Taylor, K. Eley, M. Tuffin and D. A. Cowan (2014). "Engineering pyruvate decarboxylase-mediated ethanol production in the thermophilic host Geobacillus thermoglucosidasius." Applied Microbiology and Biotechnology **98**(3): 1247-1259.

Vos, P. D., W. Ludwig, K.-H. Schleifer and W. B. Whitman (2015). “Paenibacillaceae” fam. nov. Bergey's Manual of Systematics of Archaea and Bacteria, John Wiley & Sons, Ltd.

Walker, L. P., D. B. Wilson, D. C. Irvin, C. McQuire and M. Price (1992). "Fragmentation of cellulose by the major Thermomonospora fusca cellulases, Trichoderma reesei CBHI, and their mixtures." Biotechnology and Bioengineering **40**(9): 1019-1026.

Wang, Q., G. M. Garrity, J. M. Tiedje and J. R. Cole (2007). "Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy." Applied and Environmental Microbiology **73**(16): 5261-5267.

White, D., R. J. Sharp and F. G. Priest (1993). "A polyphasic taxonomic study of thermophilic bacilli from a wide geographical area." Antonie van Leeuwenhoek **64**(3): 357-386.

William, S., H. Feil and A. Copeland (2012). "Bacterial genomic DNA isolation using CTAB." Sigma **50**: 6876.

Wilson, D. B. (1988). Cellulases of Thermomonospora fusca. Methods in Enzymology, Academic Press. **Volume 160:** 314-323.

Wilson, D. B. (2004). "Studies of Thermobifida fusca plant cell wall degrading enzymes." Chemical Record **4**(2): 72-82.

World-Energy-Council (2016). World Energy Resources 2016, Biomass. World Energy Resources 2016. 62–64 Cornhill, London EC3V 3NH, United Kingdom, World Energy Resources 2016.

Wu, L. and N. E. Welker (1989). "Protoplast Transformation of Bacillus stearothermophilus NUB36 by Plasmid DNA." Microbiology **135**(5): 1315-1324.

Wu, S., B. Liu and X. Zhang (2006). "Characterization of a recombinant thermostable xylanase from deep-sea thermophilic Geobacillus sp. MT-1 in East Pacific." Applied Microbiology and Biotechnology **72**(6): 1210-1216.

Yeoman, C. J., Y. Han, D. Dodd, C. M. Schroeder, R. I. Mackie and I. K. O. Cann (2010). Chapter 1 - Thermostable Enzymes as Biocatalysts in the Biofuel Industry. Advances in Applied Microbiology. I. L. Allen, S. Sima and M. G. Geoffrey, Academic Press. **Volume 70:** 1-55.

Yin, Y., X. Mao, J. Yang, X. Chen, F. Mao and Y. Xu (2012). "dbCAN: a web resource for automated carbohydrate-active enzyme annotation." Nucleic Acids Research **40**(W1): W445-W451.

Zeigler, D. (2001). The Genus Geobacillus: Introduction & Strain Catalog. Bacillus Genetic Stock Center Catalog of Strains, Volume 3. Columbus, Ohio State University**:** 27-27.

Zeigler, D. (2002). Integration Vectors for Gram-Positive Organisms. Bacillus Genetic Stock Center Catalog of Strains, Volume 4. Columbus, Ohio State University.

Zeigler, D. R. (2005). "Application of a recN sequence similarity analysis to the identification of species within the bacterial genus *Geobacillus*." Int J Syst Evol Microbiol **55**(3): 1171-1179.

Zeigler, D. R. (2013). The Family Paenibacillaceae. The Bacillus Genetic Stock Center, Catalogue of strains. T. B. G. S. Center, D. o. Biochemistry and T. O. S. University. **5**.

Zeigler, D. R. (2014). "The Geobacillus paradox: why is a thermophilic bacterial genus so prevalent on a mesophilic planet?" Microbiology **160**(1): 1-11.

Zhang, S. and D. B. Wilson (1997). "Surface residue mutations which change the substrate specificity of Thermomonospora fusca endoglucanase E2." Journal of Biotechnology **57**(1): 101-113.