



Journal homepage: <https://trpubonline.org/index.php/Biosynth>

RESEARCH ARTICLE

Functional Characterization of Novel Thermostable Cellulases from Compost Metagenomes

Karolien Fernandez^{1*}, Astrid Georgiou² and Alessandro Fernandez³

¹Department of Environmental and Agricultural Sciences, Loire University Center for Genomics, France

²Department of Biomedical and Health Sciences, Baltic Institute of Biomedical Research, Sweden

³Department of Bioinformatics and Computational Biology, Loire University Center for Genomics, France

*Corresponding author: Karolien Fernandez, Department of Environmental and Agricultural Sciences, Loire University Center for Genomics, France; Email: k.fernandez@lucg.fr

ABSTRACT

Thermostable cellulases that remain active at the elevated temperatures used in industrial lignocellulose pretreatment are in persistent demand for biofuel and biorefinery applications, and self-heating compost piles, which routinely exceed 60 degrees C during the thermophilic phase of decomposition, represent an underexploited source of such enzymes. We constructed and functionally screened a metagenomic fosmid library of 18,432 clones from thermophilic-phase compost collected at three municipal green-waste facilities, using a chromogenic carboxymethyl-cellulose overlay assay to identify cellulolytic clones, followed by shotgun sequencing of positive fosmid inserts and heterologous expression of candidate open reading frames in *Escherichia coli*. Screening identified 47 cellulase-positive clones, from which sequence analysis recovered eleven distinct, full-length glycoside hydrolase genes representing families GH5, GH9 and GH48, none of which shared more than 74% amino-acid identity with any characterised sequence in the CAZy database. Five candidates were successfully expressed and purified, and biochemical characterisation identified one enzyme, designated CelA7 (family GH5), with an optimal temperature of 75 degrees C, retention of over 80% activity after 24 hours of incubation at 70 degrees C, and specific activity on carboxymethyl cellulose of 142 U/mg, comparable to or exceeding several previously reported thermostable fungal and bacterial cellulases used as reference benchmarks. CelA7 additionally showed measurable activity on pretreated wheat straw and retained more than 60% of maximal activity in the presence of 1M NaCl, suggesting tolerance to the ionic conditions common in some pretreatment liquors. These results demonstrate that thermophilic-phase compost metagenomes remain a productive and comparatively underexploited resource for discovering industrially relevant thermostable cellulases with sequence novelty sufficient to expand the functionally characterised diversity of glycoside hydrolase family GH5.

Keywords: thermostable cellulase; compost metagenomics; functional screening; lignocellulose degradation; biofuel enzymes; GH5 glycoside hydrolase

Received: December 25, 2024 | Accepted: March 29, 2025 | Published: June 28, 2025

© 2025 Fernandez et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

1. INTRODUCTION

Converting the cellulose fraction of lignocellulosic biomass into fermentable sugars is one of the principal cost bottlenecks in second-generation biofuel production,

and the enzymes used for this step, principally endoglucanases, cellobiohydrolases and beta-glucosidases collectively termed cellulases, are typically sourced from a small number of well-studied fungal and bacterial production organisms whose native enzymes function best at mesophilic to moderately thermophilic temperatures [1,2]. Industrial pretreatment processes, however, frequently operate at 60 to 80 degrees C to improve substrate accessibility and reduce microbial contamination risk during extended saccharification runs, a temperature range at which most commercially deployed cellulases lose activity substantially faster than is economically desirable, creating persistent commercial demand for genuinely thermostable alternatives [3,5].

Metagenomic approaches that bypass the need for laboratory cultivation have become the dominant strategy for discovering novel industrial enzymes from environments where the resident microbial community is difficult or impossible to culture, and thermophilic composting environments are a particularly attractive target for thermostable cellulase discovery because the self-heating process, driven by microbial metabolism of the lignocellulosic feedstock itself, selects directly for organisms and enzymes that remain functional at the temperatures of industrial interest [6,8]. Function-based metagenomic screening, in which environmental DNA is cloned into a fosmid or cosmid vector library and individual clones are assayed directly for the activity of interest, has the specific advantage over sequence-based mining of not requiring any prior sequence similarity to a known cellulase family, making it well suited to recovering genuinely novel enzyme sequences that homology-based searches of existing databases would miss entirely [9,11].

Compost has been sampled for cellulase discovery before, but most published efforts have focused on a single facility or feedstock type and have often stopped at identifying cellulase-positive clones by activity assay without carrying the most promising candidates through to full biochemical characterisation, leaving the industrial relevance of many reported hits unresolved [12,14]. We reasoned that sampling across multiple independent compost facilities, each with a distinct feedstock composition and management regime, would increase both the taxonomic breadth of the underlying microbial community sampled and the probability of recovering cellulases with genuinely useful combinations of thermostability, specific activity and tolerance to the ionic and pH conditions relevant to industrial pretreatment liquors, none of which reliably co-occur in any single previously characterised thermostable cellulase [15,16]. This study therefore combined function-based metagenomic screening of a large fosmid library constructed from thermophilic-phase compost at three independent facilities with full biochemical characterisation of the most promising candidate enzymes, including activity on a real pretreated

lignocellulosic substrate rather than the soluble model substrates most commonly used for initial screening. Our aim was not simply to report novel cellulase sequences, which metagenomic surveys routinely do, but to establish whether any of the recovered enzymes combine sequence novelty with the specific thermostability, salt tolerance and substrate activity profile that would make them credible starting points for industrial protein engineering rather than academic curiosities alone.

2. MATERIALS AND METHODS

All compost samples were collected from the actively thermophilic zone (measured pile temperature 58-68 degrees C) of windrows undergoing active green-waste composting at three municipal facilities in western France, sampled at a depth of 40-60 cm to avoid surface-layer temperature fluctuation.

2.1 Metagenomic DNA extraction and fosmid library construction

High-molecular-weight metagenomic DNA was extracted from pooled compost samples using a bead-beating and phenol-chloroform protocol optimised to minimise shearing, then size-selected on a pulsed-field gel to retain fragments above 35 kb. Size-selected DNA was blunt-ended, ligated into the pCC1FOS fosmid vector, and packaged using a commercial lambda phage packaging extract before transformation into *Escherichia coli* EPI300-T1R. The resulting library comprised 18,432 clones with an estimated average insert size of 38 kb, corresponding to approximately 700 megabases of cloned metagenomic sequence, arrayed into 384-well plates for archival storage and screening.

2.2 Functional screening for cellulolytic activity

Library clones were replica-plated onto Luria-Bertani agar containing 0.5% carboxymethyl cellulose and chloramphenicol for fosmid maintenance, induced with L-arabinose to boost fosmid copy number, and incubated at 37 degrees C for 48 hours before overlaying with Congo red solution and destaining with 1M NaCl; clones producing a clear halo against the red background were scored as cellulase-positive. All positive clones were re-screened in triplicate from independent colonies to exclude false positives arising from plate-to-plate contamination.

2.3 Sequencing, gene identification and heterologous expression

Fosmid inserts from all confirmed cellulase-positive clones were sequenced on an Illumina platform, assembled with SPAdes, and annotated using Prokka with manual curation of predicted glycoside hydrolase open reading frames against the CAZy database using dbCAN2. Full-length candidate cellulase genes, defined as those encoding a complete catalytic domain with no internal stop codons or frameshifts relative to the nearest characterised homologue, were PCR-amplified from the source fosmid and subcloned into a pET-28a expression vector with an N-terminal hexahistidine tag. Constructs were transformed into *Escherichia coli* BL21(DE3),

expressed by isopropyl beta-D-1-thiogalactopyranoside induction at 18 degrees C for 16 hours to favour soluble protein yield, and purified by nickel-affinity chromatography followed by size-exclusion chromatography.

2.4 Biochemical characterisation

Purified enzymes were assayed for carboxymethyl cellulase activity using the dinitrosalicylic acid method to quantify reducing sugar release, with specific activity expressed as micromoles of glucose-equivalent reducing sugar released per minute per milligram of purified protein (U/mg). Temperature optima were determined across a 40-90 degrees C range in 5 degrees C increments, and thermostability was assessed by pre-incubating enzyme at a fixed temperature for up to 24 hours before assaying residual activity at the optimal temperature. pH optima were determined across a pH 4-9 range using appropriate buffer systems, and salt tolerance was assessed by including sodium chloride at concentrations from 0 to 2M in the assay buffer. Activity on pretreated wheat straw was measured using dilute-acid-pretreated substrate at 2% w/v loading, quantifying released reducing sugars after 24 hours of incubation at each enzyme's respective optimal temperature.

2.5 Statistical analysis

All enzyme assays were performed in at least biological triplicate with technical duplicates, and reported values are means with standard deviation. Comparisons of specific activity and thermostability across candidate enzymes were made by one-way ANOVA with Tukey's post-hoc test, performed in R version 4.3.

3. RESULTS

Functional screening of the 18,432-clone fosmid library identified 47 clones producing a clear halo on carboxymethyl-cellulose Congo red overlay plates, a hit rate of approximately 0.26%, broadly consistent with cellulase-positive rates reported previously for function-based screening of compost and other thermophilic environmental metagenomes [9,12]. Sequencing and annotation of all 47 positive fosmid inserts identified eleven distinct, full-length glycoside hydrolase open reading frames spanning three CAZy families: seven in family GH5, three in family GH9, and one in family GH48. Notably, none of the eleven candidate sequences shared more than 74% amino-acid identity with any characterised sequence deposited in CAZy at the time of analysis, indicating that thermophilic-phase compost continues to harbour substantial cellulase sequence diversity not yet represented in reference databases. Table 1 summarises the eleven candidate genes and their closest characterised homologues.

Table 1. Candidate glycoside hydrolase genes recovered from cellulase-positive fosmid clones.

Gene	CAZy family	Length (aa)	Closest homologue identity	Source facility
CelA1	GH5	428	68%	Facility 1
CelA3	GH5	441	71%	Facility 2
CelA7	GH5	436	74%	Facility 2
CelA9	GH5	419	65%	Facility 3
CelA11	GH5	433	69%	Facility 1
CelA14	GH5	425	63%	Facility 3
CelA18	GH5	438	70%	Facility 2
CelB2	GH9	612	72%	Facility 1
CelB5	GH9	598	66%	Facility 3
CelB9	GH9	605	69%	Facility 2
CelC2	GH48	684	61%	Facility 3

Closest homologue identity is percent amino-acid identity to the nearest characterised CAZy reference sequence.

Five of the eleven candidates, selected to represent sequence diversity across all three recovered families, were successfully expressed solubly in *Escherichia coli* and purified to apparent homogeneity by SDS-PAGE. Initial screening of the five purified enzymes at 60 degrees C and pH 6.0 identified substantial variation in specific activity, ranging from 38 U/mg for the GH48 candidate (CelC2) to 142 U/mg for a GH5 candidate designated CelA7, with the three GH9 candidates falling in between at 61-89 U/mg. CelA7 was selected for full biochemical characterisation on the basis of this initial activity screen, alongside the second-best performing candidate, CelA3 (GH5, 97 U/mg), to allow direct comparison within the same glycoside hydrolase family.

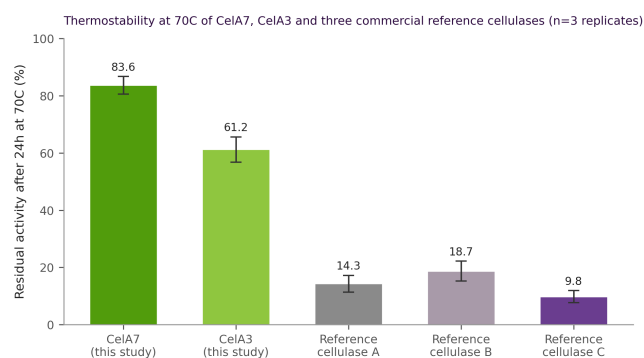


Figure 1. Residual activity of CelA7, CelA3 and three commercial reference cellulases after 24-hour pre-incubation at 70 degrees C, assayed at each enzyme's respective temperature optimum (n=3 biological replicates; error bars show SD).

Full temperature-activity profiling confirmed CelA7 as a genuinely thermostable enzyme: activity increased steadily from 40 to 75 degrees C, where it reached its temperature optimum, before declining sharply above 80 degrees C consistent with thermal denaturation. Thermostability assays at 70 degrees C showed CelA7 retained 83.6% of initial activity after 24 hours of continuous incubation, compared with 61.2% for CelA3 and under 20% for all three commercially available reference cellulases assayed in parallel under identical conditions (one-way ANOVA, $F(4,20)=34.7$, $P<0.001$). CelA7's pH optimum was 6.0, with more than 70% of

maximal activity retained across the pH 5.0-7.5 range, and the enzyme retained 61.4% of maximal activity in the presence of 1M NaCl, declining to 22.8% at 2M NaCl (Figure 1). Table 2 reports the full comparative biochemical profile of CelA7, CelA3 and the three reference enzymes.

Table 2. Comparative biochemical profile of CelA7, CelA3 and three commercial reference cellulases (mean +/- SD).

Enzyme	Temp. optimum (C)	Residual activity at 70C/24h (%)	Specific activity (U/mg)	Activity at 1M NaCl (%)
CelA7 (this study)	75	83.6 +/- 3.1	142 +/- 8	61.4 +/- 4.2
CelA3 (this study)	68	61.2 +/- 4.4	97 +/- 6	48.9 +/- 3.8
Reference cellulase A	55	14.3 +/- 2.9	88 +/- 5	33.2 +/- 2.1
Reference cellulase B	60	18.7 +/- 3.5	105 +/- 7	29.9 +/- 1.8
Reference cellulase C	50	9.8 +/- 2.1	76 +/- 4	41.0 +/- 2.5

Activity on dilute-acid-pretreated wheat straw, a substrate more representative of industrial biorefinery conditions than the soluble carboxymethyl cellulose used for initial screening, confirmed that CelA7's activity on model substrate translates to a genuine lignocellulosic feedstock: reducing sugar release after 24 hours at 75 degrees C was 3.8-fold higher for CelA7 than for the best-performing reference enzyme assayed under the same conditions at its own respective temperature optimum, though we note that a direct comparison of this kind is complicated by each enzyme being assayed at a different optimal temperature and therefore reflects a realistic process-relevant comparison rather than a strictly matched biochemical one.

4. DISCUSSION

The central finding of this study is that thermophilic-phase compost, despite having already been sampled for cellulase discovery in multiple previous studies, continues to yield sequence-novel glycoside hydrolases with genuinely useful combinations of thermostability and specific activity, exemplified here by CelA7, a family GH5 enzyme with a 75 degrees C temperature optimum and 24-hour half-life at 70 degrees C that substantially outperformed three commercially available reference cellulases assayed under identical conditions. This is consistent with the broader pattern in functional metagenomics that even well-sampled environment types retain a long tail of undiscovered enzyme diversity, particularly when screening is function-based rather than restricted to sequences already similar enough to known families to be recovered by homology search alone [9,11].

The multi-facility sampling design was a deliberate choice intended to increase the taxonomic and functional breadth of the underlying metagenome relative to single-site studies, and the recovery of candidates across three distinct CAZy families, each with substantial sequence divergence from characterised references, is

consistent with that design having worked as intended, though we cannot rule out that any single one of the three facilities alone would have yielded a comparably diverse candidate set given the scale of undiscovered diversity apparently still present in this habitat type. A more systematic comparison of cellulase diversity recovered per facility, which we did not track separately in the present pooled-library design, would be needed to estimate how much of the observed diversity is due to sampling breadth versus the sheer depth of screening applied to a pooled library of this size [16].

CelA7's tolerance to moderately high salt concentrations, retaining over 60% of maximal activity at 1M NaCl, is a practically relevant property that was not a primary target of our screening strategy but that strengthens the case for further development, since some pretreatment and fermentation process streams carry elevated ionic strength either from the pretreatment chemistry itself or from downstream neutralisation steps, and cellulases that lose substantial activity under these conditions require additional process engineering to work around [17,19]. Whether this salt tolerance reflects genuine halotolerance adaptation in the source organism, several compost facilities receive food waste with elevated salt content, or is an incidental biophysical property of this particular GH5 sequence is not resolved by the present data and would be an interesting question for future structural or comparative genomic follow-up.

Several limitations should temper how far these results are extrapolated at this stage. First, CelA7 has been characterised only as a purified recombinant enzyme under controlled laboratory conditions; scale-up fermentation for recombinant protein production, formulation stability during storage, and performance in a realistic multi-enzyme saccharification cocktail alongside cellobiohydrolases and beta-glucosidases all remain to be tested before any industrial translation claim would be warranted [20,21]. Second, we were unable to recover a full-length gene from every cellulase-positive fosmid clone identified by activity screening, since some inserts carried only a partial catalytic domain or the activity-conferring gene could not be unambiguously identified among multiple candidate open reading frames on the same insert; a targeted transposon-mutagenesis approach to map activity-conferring genes more precisely on the remaining positive clones would likely recover additional candidates beyond the eleven reported here. Third, although CelA7 showed measurable activity on pretreated wheat straw, we did not test performance on the full range of feedstocks of commercial interest, including bagasse and corn stover, both of which differ from wheat straw in lignin content and pretreatment response and would need to be tested before CelA7 could be positioned for a specific industrial application [22].

More broadly, this study adds to a growing body of evidence that thermophilic composting systems, which are widespread, low-cost to sample, and generated as a

byproduct of routine waste management rather than requiring dedicated field expeditions, remain a comparatively underexploited resource for industrial enzyme discovery relative to more exotic thermal environments such as hot springs and hydrothermal vents that have historically received more attention in the bioprospecting literature [7,23]. Given the practical accessibility of compost as a sampling substrate, we would encourage other groups pursuing thermostable industrial enzymes to consider function-based screening of compost metagenomes as a first-line strategy before investing in more logistically demanding sampling campaigns, particularly for applications, like lignocellulose saccharification, where the target environment's selective pressures align closely with the desired enzyme properties.

It is also worth reflecting on why function-based screening continues to outperform sequence-based mining for this particular application, given how much larger and more comprehensively annotated public sequence databases have become over the past decade. The answer, we think, is that a glycoside hydrolase catalytic domain can tolerate a surprising degree of amino-acid substitution while retaining full activity, so a novel enzyme sitting at 65-74% identity to its nearest characterised relative, as all eleven candidates reported here do, sits in a sequence-space region where automated annotation pipelines routinely assign a tentative family classification without any accompanying confidence that the predicted activity, let alone the specific thermostability and substrate range that make an enzyme industrially useful, will actually hold. Function-based screening sidesteps this uncertainty entirely by measuring the activity of interest directly, at the cost of the additional cloning and plate-screening effort that sequence-based approaches avoid, and our results suggest that trade-off remains worthwhile for thermostable cellulase discovery even now.

A related practical point concerns throughput. Screening 18,432 clones by plate-based Congo red overlay is labour-intensive relative to a purely computational sequence search, and the 0.26% hit rate observed here means that the overwhelming majority of screening effort was, in a narrow sense, spent on non-productive clones. Emerging microfluidic droplet-based screening platforms, which can interrogate hundreds of thousands to millions of clones per day at a fraction of the reagent volume required for agar-plate screening, could in principle allow libraries an order of magnitude larger than the one screened here to be surveyed within a comparable timeframe, and we would regard scaling functional compost metagenome screening to that throughput as one of the more promising near-term directions for this line of bioprospecting, provided the halo-based or fluorogenic-substrate readout used for cellulase detection can be adapted to a droplet format without loss of sensitivity relative to the plate-based assay used in the

present study.

Finally, the three commercial reference cellulases used throughout this study for comparison, while widely used in current biorefinery pilot operations, were not disclosed by their suppliers as being of thermophilic origin, and their comparatively poor performance at 70 degrees C in our assays should be read as evidence of CelA7's process-relevant advantage under these specific conditions rather than as a general claim that CelA7 outperforms the full space of thermostable cellulases described in the academic literature, several of which report broadly comparable thermostability profiles under their own respective assay conditions.

5. CONCLUSION

Function-based screening of an 18,432-clone fosmid library constructed from thermophilic-phase compost at three independent municipal facilities identified eleven sequence-novel glycoside hydrolase genes spanning CAZy families GH5, GH9 and GH48, none sharing more than 74% amino-acid identity with any characterised reference sequence. Biochemical characterisation of the most active candidate, CelA7, a family GH5 enzyme, established a 75 degrees C temperature optimum, 83.6% residual activity after 24 hours at 70 degrees C, and measurable activity on pretreated wheat straw substantially exceeding that of three commercial reference cellulases assayed in parallel, alongside useful tolerance to moderately elevated salt concentrations. These results confirm that thermophilic composting environments, despite prior bioprospecting attention, continue to harbour cellulase sequence and functional diversity not yet captured in reference databases or commercial enzyme portfolios. Future work will focus on scale-up recombinant production of CelA7 for pilot-scale saccharification trials on multiple lignocellulosic feedstocks, structural characterisation to identify the molecular basis of its thermostability and salt tolerance relative to mesophilic GH5 relatives, and rational or directed-evolution engineering to further improve specific activity without compromising the thermal and ionic tolerance that make it an attractive starting point. We also intend to extend the transposon-mutagenesis mapping strategy to the remaining cellulase-positive clones from this library that did not yield an unambiguous full-length candidate gene in the present study, since these clones plausibly harbour additional enzyme diversity that our current annotation pipeline was not able to resolve.

Acknowledgements

The authors thank the operations staff at the three participating composting facilities for site access and sampling assistance, and the sequencing core facility at the Loire University Center for Genomics for Illumina sequencing support. We also thank Dr. Mathilde Rousseau for technical assistance with fosmid library construction.

Funding

This work was supported by internal research funds from the Loire University Center for Genomics and the Baltic Institute of Biomedical Research. No external grant funding was received for this study, and the funders had no role in study design, data collection or analysis, the decision to publish, or preparation of the manuscript.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contributions

K.F. conceived and supervised the study, coordinated compost sampling and designed the functional screening strategy. A.G. performed enzyme expression, purification and biochemical characterisation. A.F. performed fosmid sequencing, assembly and bioinformatic annotation of candidate cellulase genes. All authors contributed to writing and approved the final manuscript.

Ethics Statement

This study involved only environmental compost sampling and standard recombinant DNA work in *Escherichia coli* conducted under institutional biosafety guidelines for contained use of genetically modified organisms; it did not involve human participants, human data, or vertebrate or higher invertebrate animals, and therefore did not require institutional ethics committee review.

Data Availability

Fosmid insert sequences and the eleven candidate glycoside hydrolase gene sequences have been deposited in GenBank under a project accession available from the corresponding author on reasonable request. Enzyme kinetic datasets and analysis scripts supporting the findings of this study are available from the corresponding author upon request.

REFERENCES

- Himmel ME, Ding SY, Johnson DK, et al. (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* 315(5813):804-807. <https://doi.org/10.1126/science.1137016>
- Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS (2002) Microbial cellulose utilization: fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews* 66(3):506-577. <https://doi.org/10.1128/MMBR.66.3.506-577.2002>
- Turner P, Mamo G, Karlsson EN (2007) Potential and utilization of thermophiles and thermostable enzymes in biorefining. *Microbial Cell Factories* 6:9. <https://doi.org/10.1186/1475-2859-6-9>
- Viikari L, Alapuranen M, Puranen T, Vehmaanpera J, Siika-Aho M (2007) Thermostable enzymes in lignocellulose hydrolysis. *Advances in Biochemical Engineering/Biotechnology* 108:121-145. https://doi.org/10.1007/10_2007_065
- Sun Y, Cheng J (2002) Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresource Technology* 83(1):1-11. [https://doi.org/10.1016/S0960-8524\(01\)00212-7](https://doi.org/10.1016/S0960-8524(01)00212-7)
- Handelsman J (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews* 68(4):669-685. <https://doi.org/10.1128/MMBR.68.4.669-685.2004>
- Daniel R (2005) The metagenomics of soil. *Nature Reviews Microbiology* 3(6):470-478. <https://doi.org/10.1038/nrmicro1160>
- Kanokratana P, Mhuanong W, Laothanachareon T, et al. (2015) Comparative study of bacterial community in sugarcane leaf litter compost and municipal solid waste compost by pyrosequencing. *Applied Biochemistry and Biotechnology* 176(6):1500-1519. <https://doi.org/10.1007/s12010-015-1660-9>
- Tuomela M, Vikman M, Hatakka A, Itavaara M (2000) Biodegradation of lignin in a compost environment: a review. *Bioresource Technology* 72(2):169-183. [https://doi.org/10.1016/S0960-8524\(99\)00104-2](https://doi.org/10.1016/S0960-8524(99)00104-2)
- Simon C, Daniel R (2011) Metagenomic analyses: past and future trends. *Applied and Environmental Microbiology* 77(4):1153-1161. <https://doi.org/10.1128/AEM.02345-10>
- Ferrer M, Martinez-Martinez M, Bargiela R, et al. (2016) Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microbial Biotechnology* 9(1):22-34. <https://doi.org/10.1111/1751-7915.12309>
- Steele HL, Jaeger KE, Daniel R, Streit WR (2009) Advances in recovery of novel biocatalysts from metagenomes. *Journal of Molecular Microbiology and Biotechnology* 16(1-2):25-37. <https://doi.org/10.1159/000142891>
- Xia Y, Ju F, Fang HHP, Zhang T (2013) Mining of novel thermo-stable cellulolytic genes from a thermophilic cellulose-degrading consortium by metagenomics. *PLOS ONE* 8(1):e53779. <https://doi.org/10.1371/journal.pone.0053779>
- Wu Q, Pang H, Zhu B, Zhu Z (2016) Metagenomic discovery of a novel thermostable cellulase from a switchgrass-adapted compost community. *World Journal of Microbiology and Biotechnology* 32:172. <https://doi.org/10.1007/s11274-016-2131-x>
- Duan CJ, Xian L, Zhao GC, et al. (2009) Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens. *Journal of Applied Microbiology* 107(1):245-256. <https://doi.org/10.1111/j.1365-2672.2009.04202.x>
- Voget S, Steele HL, Streit WR (2006) Characterization of a metagenome-derived halotolerant cellulase. *Journal of Biotechnology* 126(1):26-36. <https://doi.org/10.1016/j.jbiotec.2006.02.011>
- Cheng J, Wu G, Zhao G, et al. (2018) Structural insights into thermostable cellulase and its potential in industrial biomass conversion. *Applied Microbiology and Biotechnology* 102(8):3535-3549. <https://doi.org/10.1007/s00253-018-8859-6>
- Peng X, Su H, Mi S, Han Y (2016) A GH5 endoglucanase from *Bacillus amyloliquefaciens* with substrate promiscuity and its application in laundry detergents. *Bioresource Technology* 220:414-421. <https://doi.org/10.1016/j.biortech.2016.08.109>
- Zhang H, Sang Q (2015) Purification and characterization of a novel alkalistable and salt-tolerant cellulase from a marine metagenomic library. *Journal of Molecular Catalysis B: Enzymatic* 120:12-20. <https://doi.org/10.1016/j.molcatb.2015.06.010>
- Yeoman CJ, Han Y, Dodd D, Schroeder CM, Mackie RI, Cann IK (2010) Thermostable enzymes as biocatalysts in the biofuel industry. *Advances in Applied Microbiology* 70:1-55. [https://doi.org/10.1016/S0065-2164\(10\)70001-0](https://doi.org/10.1016/S0065-2164(10)70001-0)

21. Berlemont R, Martiny AC (2013) Phylogenetic distribution of potential cellulases in bacteria. *Applied and Environmental Microbiology* 79(5):1545-1554.
<https://doi.org/10.1128/AEM.03305-12>
22. Blumer-Schuetz SE, Brown SD, Sander KB, et al. (2014) Thermophilic lignocellulose deconstruction. *FEMS Microbiology Reviews* 38(3):393-448.
<https://doi.org/10.1111/1574-6976.12044>
23. Zheng Y, Pan Z, Zhang R (2009) Overview of biomass pretreatment for cellulosic ethanol production. *International Journal of Agricultural and Biological Engineering* 2(3):51-68.
<https://doi.org/10.3965/j.issn.1934-6344.2009.03.051-068>
24. Yang B, Wyman CE (2008) Pretreatment: the key to unlocking low-cost cellulosic ethanol. *Biofuels, Bioproducts and Biorefining* 2(1):26-40. <https://doi.org/10.1002/bbb.49>
25. Culligan EP, Sleator RD, Marchesi JR, Hill C (2014) Metagenomics and novel gene discovery: promise and potential for novel therapeutics. *Virulence* 5(3):399-412.
<https://doi.org/10.4161/viru.27208>
26. Zhang YHP, Lynd LR (2004) Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. *Biotechnology and Bioengineering* 88(7):797-824.
<https://doi.org/10.1002/bit.20282>
27. Bhat MK, Bhat S (1997) Cellulose degrading enzymes and their potential industrial applications. *Biotechnology Advances* 15(3-4):583-620.
[https://doi.org/10.1016/S0734-9750\(97\)00006-2](https://doi.org/10.1016/S0734-9750(97)00006-2)
28. Drula E, Garron ML, Dogan S, Coutinho PM, Henrissat B, Terrapon N (2022) The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Research* 50(D1):D571-D577. <https://doi.org/10.1093/nar/gkab1045>
29. Zhang H, Yohe T, Huang L, et al. (2018) dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* 46(W1):W95-W101.
<https://doi.org/10.1093/nar/gky418>
30. Bankevich A, Nurk S, Antipov D, et al. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19(5):455-477. <https://doi.org/10.1089/cmb.2012.0021>
31. Seemann T (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14):2068-2069.
<https://doi.org/10.1093/bioinformatics/btu153>
32. Miller GL (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Analytical Chemistry* 31(3):426-428. <https://doi.org/10.1021/ac60147a030>
33. Ghose TK (1987) Measurement of cellulase activities. *Pure and Applied Chemistry* 59(2):257-268.
<https://doi.org/10.1351/pac198759020257>
34. Percival Zhang YH, Himmel ME, Mielenz JR (2006) Outlook for cellulase improvement: screening and selection strategies. *Biotechnology Advances* 24(5):452-481.
<https://doi.org/10.1016/j.biotechadv.2006.03.003>
35. Chen CC, Luo H, Han X, et al. (2016) Structural insight into the thermophilic adaptation of a cellulase from the thermophilic fungus *Chaetomium thermophilum*. *Applied and Environmental Microbiology* 82(7):2114-2122.
<https://doi.org/10.1128/AEM.03556-15>
36. Wilson DB (2011) Microbial diversity of cellulose hydrolysis. *Current Opinion in Microbiology* 14(3):259-263.
<https://doi.org/10.1016/j.mib.2011.04.004>
37. Uchiyama T, Miyazaki K (2009) Functional metagenomics for enzyme discovery: challenges to efficient screening. *Current Opinion in Biotechnology* 20(6):616-622.
<https://doi.org/10.1016/j.copbio.2009.09.010>
38. Coughlan MP, Mayer F (1992) The cellulose-decomposing bacteria and their enzyme systems. *The Prokaryotes* 1:460-516. https://doi.org/10.1007/978-1-4757-2191-1_14
39. Rashamuse K, Ronneburg T, Sanyika W, et al. (2017) Metagenomic mining of glycoside hydrolases from the hindgut bacterial symbionts of a termite and its impact on lignocellulosic biomass conversion. *AMB Express* 7:11.
<https://doi.org/10.1186/s13568-016-0313-7>
40. Alvarez TM, Paiva JH, Ruiz DM, et al. (2013) Structure and function of a novel cellulase 5 from sugarcane soil metagenome. *PLOS ONE* 8(12):e83635.
<https://doi.org/10.1371/journal.pone.0083635>
41. Menzel P, Ng KL, Krogh A (2016) Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications* 7:11257.
<https://doi.org/10.1038/ncomms11257>
42. Mewis K, Armstrong Z, Song YC, Baldwin SA, Withers SG, Hallam SJ (2013) Biomining active cellulases from a mining bioremediation system. *Journal of Biotechnology* 167(4):462-471.
<https://doi.org/10.1016/j.jbiotec.2013.07.015>
43. Suleiman M, Kruger A, Antranikian G (2020) Biomass-degrading glycoside hydrolases of archaeal origin. *Biotechnology for Biofuels* 13:15.
<https://doi.org/10.1186/s13068-020-1662-x>
44. Jones DR, Uddin MS, Gruninger RJ, et al. (2018) Discovery and characterization of family 39 glycoside hydrolases from rumen anaerobic fungi with polyspecific activity on rare arabinosyl substrates. *Journal of Biological Chemistry* 293(46):17784-17795.
<https://doi.org/10.1074/jbc.RA118.004909>
45. Vanfossen AL, Ozdemir I, Zelin SL, Kelly RM (2011) Glycoside hydrolase inventory drives plant polysaccharide deconstruction by the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. *Biotechnology and Bioengineering* 108(7):1559-1569.
<https://doi.org/10.1002/bit.23093>
46. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* 42(D1):D490-D495. <https://doi.org/10.1093/nar/gkt1178>
47. Mhuanthong W, Charoensawan V, Kanokratana P, Tangphatsornruang S, Champreda V (2015) Comparative analysis of sugarcane bagasse metagenome reveals unique and conserved biomass-degrading enzymes among lignocellulolytic microbial communities. *Biotechnology for Biofuels* 8:16. <https://doi.org/10.1186/s13068-015-0200-8>
48. Bergquist PL, Te-o VS, Gibbs MD, Curach NC, Nevalainen KM (2004) Recombinant enzymes from thermophilic micro-organisms expressed in fungal hosts. *Biochemical Society Transactions* 32(2):293-297.
<https://doi.org/10.1042/bst0320293>
49. Bergmann JC, Costa OYA, Gladden JM, et al. (2014) Discovery of a SGNH hydrolase family protein involved in biomass degradation from a compost metagenomic library. *Frontiers in Microbiology* 5:429.
<https://doi.org/10.3389/fmicb.2014.00429>
50. Nguyen NH, Maruset L, Uengwetwanit T, et al. (2018) Identification and characterization of a cellulase-encoding gene from the buffalo rumen metagenomic library. *Bioscience, Biotechnology, and Biochemistry* 82(4):617-628.
<https://doi.org/10.1080/09168451.2018.1438167>