

**BỘ GIÁO DỤC
VÀ ĐÀO TẠO**

**VIỆN HÀN LÂM KHOA HỌC
VÀ CÔNG NGHỆ VIỆT NAM**

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ



Đào Trọng Khoa

**XÂY DỰNG CƠ SỞ DỮ LIỆU DNA METAGENOME
HỆ VI KHUẨN DA CỔ DÊ VÀ KHAI THÁC, NGHIÊN CỨU
TÍNH CHẤT CỦA ENDO-XYLANASE**

LUẬN ÁN TIẾN SĨ SINH HỌC

Hà Nội - 2024

BỘ GIÁO DỤC
VÀ ĐÀO TẠO

VIỆN HÀN LÂM KHOA HỌC
VÀ CÔNG NGHỆ VIỆT NAM

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

Đào Trọng Khoa

**XÂY DỰNG CƠ SỞ DỮ LIỆU DNA METAGENOME
HỆ VI KHUẨN ĐẠ CỎ ĐÊ VÀ KHAI THÁC, NGHIÊN CỨU
TÍNH CHẤT CỦA ENDO-XYLANASE**

LUẬN ÁN TIẾN SĨ SINH HỌC

Ngành: Hóa sinh học

Mã số: 9.42.01.16

**Xác nhận của Học viện
Khoa học và Công nghệ**

Người hướng dẫn 1
(Ký, ghi rõ họ tên)

Người hướng dẫn 2
(Ký, ghi rõ họ tên)

GS.TS. Trương Nam Hải PGS. TS. Đỗ Thị Huyền

Hà Nội – 2024

LỜI CAM ĐOAN

Tôi xin cam đoan luận án: "**Xây dựng cơ sở dữ liệu DNA metagenome hệ vi khuẩn dạ cỏ dê và khai thác, nghiên cứu tính chất của endo-xylanase**" là công trình nghiên cứu của chính mình dưới sự hướng dẫn khoa học của tập thể hướng dẫn. Luận án sử dụng thông tin trích dẫn từ nhiều nguồn tham khảo khác nhau và các thông tin trích dẫn được ghi rõ nguồn gốc. Các kết quả nghiên cứu của tôi được công bố chung với các tác giả khác đã được sự nhất trí của đồng tác giả khi đưa vào luận án. Các số liệu, kết quả được trình bày trong luận án là hoàn toàn trung thực và chưa từng được công bố trong bất kỳ một công trình nào khác ngoài các công trình công bố của tác giả. Luận án được hoàn thành trong thời gian tôi làm nghiên cứu sinh tại Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam.

Hà Nội, ngày tháng năm 2024

Tác giả luận án



Đào Trọng Khoa

LỜI CẢM ƠN

Tôi xin bày tỏ lòng biết ơn sâu sắc tới **GS. TS. Trương Nam Hải** và **PGS. TS. Đỗ Thị Huyền**, Viện Công nghệ sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam đã định hướng nghiên cứu, hướng dẫn tận tình và tạo mọi điều kiện cho tôi hoàn thành luận án này.

Tôi xin bày tỏ lòng cảm ơn sâu sắc tới các cán bộ, thầy cô tại Học viện Khoa học và Công nghệ và Viện Công nghệ sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam đã giảng dạy và chỉ bảo cho tôi kiến thức chuyên môn và các kỹ năng cần thiết, cũng như tạo mọi điều kiện thuận lợi cho tôi trong quá trình học tập và làm thủ tục bảo vệ luận án.

Tôi xin chân thành cảm ơn tập thể Phòng Kỹ thuật di truyền, Viện Công nghệ sinh học - nơi tôi thực hiện luận án này - đã giúp đỡ, chỉ bảo tận tình cũng như chia sẻ những kinh nghiệm chuyên môn quý báu.

Tôi xin cảm ơn sự hỗ trợ kinh phí từ Đề tài Nghị định thư cấp Nhà nước, hợp tác giữa Việt Nam và Cộng hòa Liên bang Đức (mã số NĐT.50.GER/18) do GS. TS. Trương Nam Hải làm chủ nhiệm và một phần kinh phí từ đề tài ĐTĐLCN.15/14 do PGS. TS Đỗ Thị Huyền làm chủ nhiệm.

Cuối cùng, tôi xin cảm ơn gia đình, bạn bè đã giúp đỡ, động viên và chia sẻ khó khăn cùng tôi trong suốt thời gian qua!

Hà Nội, ngày tháng năm 2024

Tác giả luận án



Đào Trọng Khoa

Mục lục

LỜI CAM ĐOAN	i
LỜI CẢM ƠN	ii
DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT.....	vii
DANH MỤC BẢNG.....	x
DANH MỤC HÌNH	xi
MỞ ĐẦU.....	1
CHƯƠNG 1. TỔNG QUAN.....	4
1.1. Kỹ thuật Metagenomics nhằm khai thác hiệu quả gene tiềm năng	4
1.1.1. Sơ lược chung về kỹ thuật Metagenomics	4
1.1.2. Tình hình nghiên cứu ứng dụng kỹ thuật Metagenomics trong khai thác gene	6
1.1.2.1. Nghiên cứu ứng dụng Metagenomics trên thế giới.....	6
1.1.2.2. Nghiên cứu ứng dụng kỹ thuật Metagenomics ở Việt Nam	8
1.1.3. Chú giải gene trong tin sinh học.....	10
1.1.3.1. Một số công cụ chú giải gene thông dụng	10
1.1.3.2. Mô hình đại diện HMM: công cụ mới khai thác hiệu quả dữ liệu metagenome.....	12
1.2. Tổng quan về lignocellulose	13
1.2.1. Vai trò của lignocellulose trong nền kinh tế sinh học	13
1.2.1.1. Lignocellulose là nguồn năng lượng tái tạo phong phú.....	13
1.2.1.2. Lignocellulose là nguồn tiềm năng sản xuất vật liệu, hóa chất mới.....	15
1.2.2. Enzyme phân giải lignocellulose nói chung.....	16
1.2.2.1. Enzyme thủy phân pectin, lignin và các enzyme hỗ trợ khác.....	16
1.2.2.2. Cellulase và hemicellulase	18
1.2.2.3. Vi khuẩn tham gia chuyển hóa lignocellulose trong dạ cỏ	20
1.2.3. Xylanase	21
1.2.3.1. Vai trò của xylanase	21
1.2.3.2. Phân loại xylanase.....	22
1.2.3.3. Xylanase trong tự nhiên	23

1.2.3.4. Ứng dụng của xylanase	24
1.2.3.5. Tình hình biểu hiện xylanase trong nước và trên thế giới	25
CHƯƠNG 2. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU	27
2.1. Đối tượng, vật liệu hóa chất và thiết bị máy móc	27
2.1.1. Đối tượng và vật liệu nghiên cứu	27
2.1.2. Một số dung dịch hóa chất chính:	27
2.1.3. Máy móc và thiết bị.....	29
2.2. Phương pháp nghiên cứu	29
2.2.1. Các phương pháp sinh học phân tử, vi sinh vật	29
2.2.1.1. Tách chiết, tinh chế DNA metagenome	29
2.2.1.2. Phương pháp giải trình tự đa hệ gene	30
2.2.1.3. Tổng hợp gene, thiết kế vector biểu hiện mang gene exl	30
2.2.1.4. Biến nạp DNA plasmid vào vi khuẩn E. coli.....	30
2.2.1.5. Tách chiết DNA plasmid từ tế bào vi khuẩn E. coli	31
2.2.1.6. Điện di trên gel agarose	31
2.2.1.7. Tinh chế DNA từ gel agarose	31
2.2.2. Các phương pháp hóa sinh protein.....	31
2.2.2.1. Biểu hiện protein tái tổ hợp trong E. coli.....	31
2.2.2.2. Điện di protein trên gel polyacrylamide	32
2.2.2.3. Tách chiết protein tái tổ hợp và tinh chế bằng sắc ký ái lực His-tag33	
2.2.2.4. Xác định độ sạch của enzyme tái tổ hợp bằng phần mềm Quantity One.....	34
2.2.2.5. Định lượng protein bằng phương pháp Bradford	34
2.2.2.6. Xác định hoạt tính enzyme.....	35
2.2.2.7. Xác định ảnh hưởng của nhiệt độ, pH, các ion kim loại và một số hóa chất lên hoạt tính enzyme	38
2.2.2.8. Xác định độ bền nhiệt của enzyme	39
2.2.2.9. Xác định thông số động học của enzyme.....	39
2.2.3. Các phương pháp tin sinh học	40
2.2.3.1. Lắp ráp DNA đa hệ gene, chú giải các gene chức năng	40
2.2.3.2. Phương pháp nghiên cứu Pfam của các trình tự	41

2.2.3.3. Nghiên cứu vùng bảo thủ và dự đoán cấu trúc bậc ba của các trình tự.....	42
2.2.3.4. Dự đoán khả năng chịu kiềm/acid.....	42
2.2.3.5. Dự đoán khả năng chịu nhiệt của enzyme	43
2.2.3.6. Định loại loài các trình tự ORF.....	43
2.2.3.7. Tối ưu mã và tổng hợp gene mã hóa enzyme thủy phân xylan được khai thác từ dữ liệu giải trình tự DNA metagenome vi khuẩn dạ cỏ dê.....	43
2.2.3.8. Phương pháp xử lý số liệu	44
CHƯƠNG 3. KẾT QUẢ VÀ THẢO LUẬN	45
3.1. Nghiên cứu giải mã xây dựng bộ dữ liệu và đánh giá đa dạng vi khuẩn trong dạ cỏ dê.....	45
3.1.1. Tách chiết DNA đa hệ gene của vi khuẩn	45
3.1.2. Giải trình tự, đánh giá chất lượng bộ dữ liệu và chú giải gene	46
3.1.3. Đánh giá đa dạng vi khuẩn trong mẫu DNA metagenome	48
3.1.3.1. Đa dạng vi khuẩn trong dạ cỏ dê được đánh giá dựa trên bộ dữ liệu 8,6 Gb	48
3.1.3.2. Đa dạng vi khuẩn trong dạ cỏ dê được đánh giá dựa trên bộ dữ liệu giải mã sâu.....	49
3.2. Khai thác gene và thiết lập công cụ HMM cho chú giải gene, khai thác gene mã hóa protein/enzyme tham gia thủy phân lignocellulose trong dạ cỏ dê.....	50
3.2.1. Khai thác gene mã hóa enzyme thủy phân lignocellulose dựa trên cơ sở dữ liệu KEGG	50
3.2.1.1. Khai thác gene từ dữ liệu giải trình tự 8,6 Gb	51
3.2.1.2. Khai thác gene từ dữ liệu giải trình tự sâu 48,6 Gb.....	54
3.2.2. Phân tích đa dạng vi khuẩn mang gene thủy phân lignocellulose.....	57
3.2.2.1. Đa dạng vi khuẩn mang gene phân giải lignocellulose khai thác được từ dữ liệu 8,6 Gb.....	57
3.2.2.2. Đa dạng vi khuẩn mang gene phân giải lignocellulose khai thác từ kết quả giải trình tự sâu	59
3.2.2.3. Vai trò của chi Prevotella trong phân giải lignocellulose	60
3.2.2.4. Vai trò của chi Prevotella trong hỗ trợ tiêu hóa các nguồn thức ăn khác trong dạ cỏ dê.....	62

3.2.3. Xây dựng công cụ mới khai thác hiệu quả protein/enzyme tham gia chuyển hóa và tiền xử lý lignocelullose	68
3.2.3.1. Xây dựng mô hình HMM đại diện cho các enzyme/protein cần khai thác.....	68
3.2.3.2. Khai thác các enzyme tham gia thủy phân lignocelullose từ dữ liệu DNA đa hệ gene vi khuẩn trong dạ cỏ dê.....	70
3.3. Nghiên cứu lựa chọn, biểu hiện và xác định đặc điểm của endo-xylanase ..	71
3.3.1. Nghiên cứu lựa chọn gene mã hóa endo-xylanase cho biểu hiện	71
3.3.1.1. Nghiên cứu đa dạng vi khuẩn mang enzyme endo-xylanase	71
3.3.1.2. Nghiên cứu đa dạng cấu trúc endo-xylanase	72
3.3.1.3. Nghiên cứu lựa chọn trình tự xylanase để biểu hiện.....	74
3.3.2. Nghiên cứu biểu hiện endo-xylanase	77
3.3.2.1. Phân tích tối ưu mã bộ ba của trình tự gene exl.....	77
3.3.2.2. Thiết kế vector biểu hiện mang gene exl	80
3.3.2.3. Biểu hiện gene endo-xylanase [denovogenes]_5086.....	81
3.3.3. Tinh chế protein endo-xylanase tái tổ hợp	87
3.3.3.1. Tinh chế endo-xylanase bằng sắc ký ái lực His-tag.....	87
3.3.3.2. Loại muối protein endo-xylanase sau khi tinh sạch.....	89
3.3.4. Nghiên cứu đặc tính enzyme xylanase và các thông số động học enzyme	91
3.3.4.1. Nghiên cứu xác định nhiệt độ tối ưu cho enzyme hoạt động.....	91
3.3.4.2. Nghiên cứu xác định pH tối ưu cho enzyme hoạt động.....	92
3.3.4.3. Nghiên cứu tính bền nhiệt của enzyme	93
3.3.4.4. Nghiên cứu ảnh hưởng của một số ion kim loại và hóa chất lên hoạt tính xylanase	94
3.3.4.5. Nghiên cứu tính đặc hiệu cơ chất của enzyme.....	95
3.3.4.6. Nghiên cứu thông số động học của enzyme endo xylanase tái tổ hợp	96
KẾT LUẬN VÀ KIẾN NGHỊ.....	98
KẾT LUẬN	98
DANH MỤC TÀI LIỆU THAM KHẢO	102

DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT

Tên viết tắt	Tên tiếng Anh	Nghĩa tiếng Việt
BLAST	Basic local alignment search tool	Công cụ so sánh mức độ tương đồng về trình tự được phát hành mở bởi Trung tâm Thông tin Công nghệ sinh học của Mỹ
bp	Base pair	Cặp base
CBH	Cellobio-hydrolase	Enzyme cellobiohydrolase
CBD	Carbohydrate binding domain	Vùng bám carbohydrate
CBM	Cellulose binding module	Vùng bám cellulose
CE	Carbohydrate esterase	Enzyme carbohydrate-esterase
cDNA	Complementary DNA	DNA tương ứng với trình tự khuôn mRNA được tổng hợp nhờ enzyme phiên mã ngược
CMC	Carboxymethyl cellulose	Cơ chất carboxymethyl cellulose
COG	Clusters of orthologous groups	Tập hợp dữ liệu trình tự protein của một số sinh vật nhân sơ và nhân chuẩn đơn bào
CSDL		Cơ sở dữ liệu
DNA	Deoxyribonucleic acid	Axit deoxyribonucleic
DNS/DNSA	3,5-Dinitrosalicylic acid	Axit 3,5-dinitrosalicylic
dNTP	2'-deoxyribonucleoside 5'-triphosphate	2'-deoxyribonucleoside 5'-triphosphate
EDTA	Ethylene diamine tetraacetic acid	Axit ethylene diamine tetraacetic
EG	Endoglucanase	Endoglucanase
eggNOG	Evolutionary genealogy of genes: Non-supervised orthologous groups	Cơ sở dữ liệu chứa các nhóm có chung nguồn gốc tổ tiên
Fn3	Fibronectin type III module	Module Fn3
Gb	Gigabase	Đơn vị di truyền Gigabase
His (H)	Histidine	Amino acid histidin
HMM	Hidden Markov Model	Mô hình Markov ẩn
HTS	High throughput sequencing	Giải trình tự thông lượng cao
Ig	Immunoglobulin module	Module Ig
IPTG	Isopropyl- β -D-thiogalactosidase	Isopropyl- β -D-thiogalactosidase

Kb	Kilobase	Đơn vị di truyền kilobase
kDa	Kilodalton	Đơn vị đo khối lượng protein kilodalton
KEGG	Kyoto Encyclopedia of genes and genomes	Cơ sở dữ liệu về bộ gen, con đường sinh học, bệnh tật và các chất hóa học
Km	Michaelis constant	Nồng độ cơ chất để tốc độ phản ứng đạt $\frac{1}{2}$ tốc độ tối đa
KOG	Eukaryotic orthologous groups	Tập hợp dữ liệu trình tự protein, phiên bản COG tập trung về hệ gene của một số sinh vật nhân chuẩn điển hình
LB	Luria-Betani	Môi trường LB
LBA	Luria-Betani ampicillin	Môi trường LB bổ sung ampicillin
MB	Megabyte	Đơn vị thông tin megabyte
MCP	Microbial crude protein	Protein thô của vi khuẩn
MEGAN	MEtaGeneomic ANalyser	Phần mềm phân tích tối ưu các trình tự đa hệ gen
NCBI	National center for Biotechnology information	Trung tâm thông tin Công nghệ Sinh học của Mỹ
NGS	Next-generation sequencing	Giải trình tự thế hệ mới
NR	Non-redundant	Cơ sở dữ liệu chứa các trình tự non-redundant từ ngân hàng gen
OD	Optimal density	Mật độ quang học
ORF	Open reading frame	Khung đọc mở
PCR	Polymarase chain reaction	Phản ứng chuỗi trùng hợp
PFAM	Protein families	Cơ sở dữ liệu các họ protein
pH	Hydrogen power	Chỉ số đo nồng độ ion H ⁺ , ion OH ⁻ trong dung dịch
pNPG	p-nitrophenol- β -glucoside	p-nitrophenol- β -glucoside
pNPX	4-nitrophenyl β -D-xylopyranoside	4-nitrophenyl β -D-xylopyranoside
PUL	polysaccharide utilization loci	vùng sử dụng polysaccharide
RNA	Ribonucleic acid	Axit ribonucleic
SDS	Sodium dodecyl sulphate	Sodium dodecyl sulphate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis	Điện di trên gel polyacrylamide có SDS

TEMED	Tetramethylethylenediamine	Tetramethylethylenediamine
V _{max}	Maximum velocity	Vận tốc tối đa
VFA	Volatile fatty acid	Axit béo bay hơi
w/v	Weight/volume	Khối lượng/thể tích

DANH MỤC BẢNG

Bảng 2. 1. Thành phần gel polyacrylamide biến tính	32
Bảng 3. 1. Các thông tin về bộ dữ liệu DNA metagenome của vi khuẩn trong dạ cỏ dê sau khi giải trình tự và lắp ráp	47
Bảng 3. 2. Kết quả chú giải gene chức năng từ dữ liệu giải trình tự DNA metagenome vi khuẩn trong dạ cỏ dê dựa trên các cơ sở dữ liệu khác nhau	48
Bảng 3. 3. Kết quả khai thác gene mã hóa enzyme tiền xử lý lignocellulose từ dữ liệu giải trình tự 8,6 Gb dựa trên dữ liệu KEGG	51
Bảng 3. 4. Danh sách enzyme thủy phân hemicellulose mã hóa từ gene trong dữ liệu giải trình tự 8,6 Gb dựa trên dữ liệu KEGG.....	52
Bảng 3. 5. Bảng tổng hợp các enzyme lignocellulase khai thác từ hai bộ dữ liệu giải trình tự	54
Bảng 3. 6. Các gene của <i>Prevotella</i> mã hóa enzyme tham gia thủy phân lignocellulose	61
Bảng 3. 7. Các gene từ <i>Prevotella</i> mã hóa các enzyme tham gia phân giải các chất dinh dưỡng khác được phân tích từ dữ liệu giải trình tự sâu metagenomic của vi khuẩn trong dạ cỏ dê Việt Nam.....	63
Bảng 3. 8. Thống kê các mô hình HMM được sử dụng cho khai thác nhóm protein/enzyme mới trong đề tài	68
Bảng 3. 9. Thống kê độ dài các vùng bảo thủ dùng để thiết lập mô hình HMM cho khai thác gene họ CBM.....	70
Bảng 3. 10. Khai thác một số enzyme hiệu quả từ dữ liệu DNA metagenome vi khuẩn trong dạ cỏ dê bằng mô hình HMM đã được xây dựng	71
Bảng 3. 11. Đa dạng vi khuẩn mang gene mã hóa endo-xylanase.....	71
Bảng 3. 12. Các trình tự tương đồng với trình tự denovogenes_5086.....	74
Bảng 3. 13. Các thông số các lần tinh chế enzyme endo-xylanase.....	91

DANH MỤC HÌNH

Hình 1. 1. Một phần mô hình đại diện HMM của họ GH10.....	13
Hình 1. 2. Hệ enzyme phân hủy cellulose.....	18
Hình 1. 3. Hệ enzyme phân hủy hemicellulose.....	19
Hình 2. 1. Đường chuẩn biểu thị sự liên hệ giữa OD595 và hàm lượng BSA.....	35
Hình 2. 2. Đường chuẩn xylose.....	36
Hình 2. 3. Đường chuẩn glucose.....	37
Hình 2. 4. Đường chuẩn pNP	38
Hình 2. 5. Sự phụ thuộc của tốc độ phản ứng và nồng độ cơ chất theo Lineweaver-Burk [187].....	40
Hình 2. 6. Quá trình phân tích và khai thác gene từ dữ liệu DNA đa hệ gene	41
Hình 3. 1. Điện di đồ phân tích DNA đa hệ gene vi khuẩn được tách chiết từ 10 mẫu dạ cỏ dê (A) và sản phẩm khuếch đại gene 16S rDNA của vi khuẩn từ các mẫu DNA được tách chiết (B); DNA metagenome được gửi giải trình tự (C).	45
Hình 3. 2. Biểu đồ đánh giá chất lượng giải trình tự đa hệ gene của vi khuẩn bằng FastQC trên bộ dữ liệu giải mã sâu	46
Hình 3. 3. Biểu đồ phân bố đa dạng phân loại học ở mức độ ngành và mức độ chi của vi khuẩn trong dạ cỏ dê khai thác được từ dữ liệu 8,6 Gb.....	49
Hình 3. 4. Biểu đồ phân bố đa dạng phân loại học ở mức độ ngành và mức độ chi của vi khuẩn trong dạ cỏ dê khai thác được từ dữ liệu giải mã sâu.....	50
Hình 3. 5. Bức tranh tổng quan về các họ GH/CE/PL liên quan đến quá trình phân giải lignocellulose của vi khuẩn trong dạ cỏ dê.....	57
Hình 3. 6. Đa dạng phân loại vi khuẩn mang gene lignocellulase trong dạ cỏ dê Việt Nam đã được KEGG chú thích và phân loại bởi MEGAN	60
Hình 3. 7. Các gene nguồn gốc từ <i>Prevotella</i> tham gia tiêu hóa trong dạ cỏ dê.....	62
Hình 3. 8. Các locus gene phân giải celluloses/hemicelluloses trong những contig tiềm năng xây dựng từ dữ liệu giải trình tự sâu DNA metagenome vi khuẩn dạ cỏ dê.	66
Hình 3. 9. Kết quả phân tích trình tự vùng CBM có cấu trúc beta-sandwich bằng Cytoscape với giới hạn cut-off của độ tương đồng trình tự là 0,55.....	69
Hình 3. 10. Tóm lược các cấu trúc domain của enzyme nhóm endo-xylanase có mang vùng hoạt tính GH, được vẽ bằng ngôn ngữ R trên phần mềm drawProteins	73
Hình 3. 11. Kết quả minh họa trình tự tương đồng với trình tự denovogenes_5086 trên NCBI bằng BLASTP.....	75

Hình 3. 12. Cây phát sinh chủng loài trình tự amino acid gene denovogenes_508676	
Hình 3. 13. Dự đoán các vùng bảo thủ trên trình tự denovogenes_5086.....	76
Hình 3. 14. Mô hình cấu trúc không gian của protein được dự đoán bằng Phyre2 ..	77
Hình 3. 15. Khả năng sử dụng các mã bộ ba trên gene (A) và mức độ phù hợp các mã bộ ba (B) của gene <i>exl</i> mã hóa được so sánh với trước và sau khi cải biến để biểu hiện trên chủng chủ <i>E. coli</i>	78
Hình 3. 16. Trình tự gene [denovogenes]_5086 trước và sau khi tối ưu các mã bộ ba phù hợp cho việc biểu hiện trong <i>E. coli</i>	79
Hình 3. 17. Kết quả điện di kiểm tra vector pJET mang gene endo-xylanase.....	80
Hình 3. 18. Điện di đồ protein tổng số (A), pha protein tủa, pha protein tan (B) và hoạt độ endo-xylanase tổng số (C) của protein được biểu hiện trong các chủng <i>E. coli</i> mang gene <i>exl</i>	82
Hình 3. 19. Kết quả điện di protein tổng số (A), protein pha tan (B) và hoạt endo-xylanase (C) từ các chủng <i>E. coli</i> Rosetta1 mang gene mã hóa endo-xylanase được nuôi trong các môi trường dinh dưỡng khác nhau.	83
Hình 3. 20. Kết quả điện di protein (A) và hoạt độ endo-xylanase tổng hợp từ chủng <i>E. coli</i> Rosetta1 mang gene mã hóa endo-xylanase được nuôi cấy trong môi trường cảm ứng ở các nhiệt độ khác nhau (B)	84
Hình 3. 21. Kết quả điện di protein các mẫu biểu hiện endo-xylanase ở các nồng độ chất cảm ứng.....	85
Hình 3. 22. Kết quả điện di protein pha tan và kiểm tra hoạt tính thô các mẫu biểu hiện endo-xylanase ở các thời điểm thu mẫu sau cảm ứng	86
Hình 3.23. Điện di đồ kiểm tra sản phẩm trong các phân đoạn tinh chế lần 1.	87
Hình 3. 24. Điện di đồ kiểm tra sản phẩm trong các phân đoạn tinh chế lần 2	88
Hình 3. 25. Xác định độ tinh khiết của protein bằng phần mềm Quantity One.....	89
Hình 3. 26. Điện di đồ kiểm tra sản phẩm trong các phân đoạn loại muối.....	89
Hình 3. 27. Kết quả kiểm tra độ sạch của mẫu bằng phần mềm Quantity One	90
Hình 3.28. Biểu đồ thể hiện sự ảnh hưởng của nhiệt độ đến hoạt tính xylanase	91
Hình 3. 29. Biểu đồ thể hiện sự ảnh hưởng của pH đến hoạt tính xylanase	92
Hình 3. 30. Biểu đồ nghiên cứu sự bền nhiệt của xylanase	93
Hình 3. 31. Biểu đồ ảnh hưởng của ion kim loại và hóa chất đến hoạt độ xylanase	95
Hình 3. 32. Biểu đồ thể hiện tính đặc hiệu cơ chất của xylanase	96
Hình 3. 33. Đồ thị tương quan giữa nồng độ cơ chất và tốc độ phản ứng của enzyme endo xylanase tái tổ hợp.	96

MỞ ĐẦU

Thực tế hiện nay, nhiên liệu hóa thạch (than đá, dầu mỏ, khí đốt...) vẫn đóng vai trò lớn chi phối sự phát triển của kinh tế thế giới nhiều năm qua. Qua nhiều năm khai thác, các nguồn nhiên liệu này đang ngày càng cạn kiệt và việc sử dụng chúng gây ra nhiều hậu quả tiêu cực về môi trường sống. Nhằm điều chỉnh thích ứng với biến đổi khí hậu, kiềm chế sự nóng lên toàn cầu, bảo đảm phát triển bền vững, nhiều quốc gia trên thế giới đã đi theo hướng sử dụng nhiên liệu tái tạo thay thế cho nhiên liệu hóa thạch và trong đó phải kể đến nhiên liệu sinh học [1].

Lignocellulose, một trong những nguồn nguyên liệu tái tạo dồi dào trên Trái Đất tích trữ trong các phụ phẩm sản xuất nông lâm nghiệp như thân cây, lá cây... Phần lớn các loại rác thải này bị đốt bỏ, gây lãng phí và ảnh hưởng nghiêm trọng đến chất lượng môi trường sống cũng như sức khỏe của người dân. Với sự phát triển của công nghệ sinh học, lignocellulose có thể được chuyển hóa thành nhiên liệu sinh học và các chất có giá trị khác thay thế cho nguồn nguyên, nhiên liệu không thay thế, phục vụ phát triển kinh tế xã hội. Điển hình, furfural và 5-hydroxymethylfurfural - một trong những sản phẩm được chuyển hóa từ lignocellulose - được dùng để điều chế rất nhiều loại dung môi, polymer, nhiên liệu và các hợp chất hữu ích khác [2]. Vì vậy, việc tận dụng nguồn nguyên liệu dư thừa này không những làm giảm thiểu ô nhiễm môi trường mà còn góp phần giải quyết nhu cầu năng lượng quốc gia, tạo nguồn thu nhập tại chỗ cho nông dân.

Trong những năm gần đây, chúng ta đã có nhiều giải pháp để tận dụng được nguồn sinh khối trên. Tuy nhiên, trên thực tế, lignocellulose là sinh khối rắn chắc khó chuyển hóa và đường hóa. Việc chuyển hóa lignocellulose trong công nghiệp hiện nay chủ yếu bằng các phương pháp vật lý và hóa học rất phức tạp, có giá thành cao và chưa thực sự hiệu quả và khó xử lý các chất thải hóa học. Hướng xử lý phân giải lignocellulose bằng phương pháp sinh học thân thiện với môi trường ngày càng được xem trọng và ứng dụng rộng rãi. Vấn đề khai thác những enzyme phân giải lignocellulose (gọi chung là lignocellulase) có hoạt tính cao đã và đang là một trong những hướng nghiên cứu trọng tâm của nhiều nhà khoa học trên thế giới.

Vi sinh vật nói chung và vi khuẩn nói riêng đã từ lâu được nghiên cứu chứng minh có rất nhiều ứng dụng thực tiễn trong các lĩnh vực của đời sống con người bao gồm sản xuất công nghiệp, nông nghiệp, xử lý môi trường, khai thác vật liệu mới, sản xuất các chế phẩm y dược học... Vi khuẩn khu trú trong các khu hệ giàu lignocellulose như mùn đất [3], đường tiêu hóa của động vật nhai lại [4],... được xác định là những nguồn tiềm năng để khai thác gene nói chung và gene phân giải

lignocellulose nói riêng vì sự đa dạng và phong phú của chúng. Tuy nhiên, thực tế hiện tại 99% các loài vi sinh vật vẫn chưa thể phân lập và nuôi cấy được. Để khắc phục hạn chế đó, kỹ thuật Metagenomics cho phép nghiên cứu và đánh giá trực tiếp và tổng thể tất cả các loài vi sinh vật trong mẫu mà không cần nuôi cấy [5]. Vì vậy đây là nguồn gene quan trọng cho khai thác các gene mã hóa lignocellulase.

Trong tự nhiên, nhiều loài sinh vật có khả năng thủy phân lignocellulose nhờ cấu trúc hệ cellulosome gồm nhiều loại enzyme thủy phân lignocellulose tạo cấu trúc theo trật tự phân cắt nhất định gắn trên bề mặt tế bào. Ví dụ cellulosome của *C. cellulovorans* có mang các enzyme cellulase, hemicellulase, pectinase, chitinase và nhiều enzyme phụ trợ khác để phân giải thành tế bào thực vật [6]. Các vi khuẩn thuộc ngành Bacteroidetes đặc biệt là chi *Prevotella* lại có các vùng sử dụng polysaccharide (PUL-polysaccharide utilization loci) mang một chuỗi gene mã hóa nhiều loại enzyme khác nhau giúp tế bào chuyển hóa lignocellulose thành đường đơn. Cellulosome và PUL được tìm thấy rất phong phú, đa dạng trong dạ cỏ dê, tuần lộc nhưng cellulosome chưa tìm thấy trong dạ cỏ của động vật nhai lại khác như trâu, bò. Vì vậy, hệ sinh thái mini của dạ cỏ dê nuôi ở Việt Nam là một trong những hệ rất tiềm năng, chưa được nghiên cứu nhiều. Vì vậy, nghiên cứu này được thực hiện để giải mã DNA đa hệ gene vi khuẩn trong dạ cỏ dê (giải mã tạo bộ dữ liệu nhỏ, thông thường (dung lượng khoảng 8-10 Gb) và giải mã sâu (dung lượng trên 40 Gb) để đánh giá khả năng khai thác gene của cả hai bộ dữ liệu) và tìm cách tiếp cận mới nhằm khai thác hiệu quả enzyme phân giải lignocellulose, bao gồm các nhóm enzyme tiền xử lý, enzyme phân giải cellulose, hemicellulose và lignin. Do đó, chúng tôi đã thực hiện đề tài luận án: “**Xây dựng cơ sở dữ liệu DNA metagenome hệ vi khuẩn dạ cỏ dê và khai thác, nghiên cứu tính chất của endo-xylanase**”.

- **Mục tiêu nghiên cứu:**

- Xây dựng được bộ dữ liệu DNA metagenome của hệ vi khuẩn dạ cỏ dê;
- Biểu hiện và nghiên cứu được tính chất của một endo-xylanase mã hóa từ gene trong bộ dữ liệu gene chức năng phân giải lignocellulose từ mẫu DNA metagenome vi khuẩn dạ cỏ dê.

- **Nội dung nghiên cứu:**

Để đạt được mục tiêu của đề tài, chúng tôi đã thực hiện các nội dung nghiên cứu chính sau:

1. Nghiên cứu giải mã DNA đa hệ gene của vi khuẩn trong dạ cỏ dê với dung lượng

thông thường (8-10 Gb) và dung lượng lớn (giải mã sâu, 45-50 Gb), xây dựng bộ dữ liệu và đánh giá đa dạng vi khuẩn trong dạ cỏ dê;

2. Khai thác gene và thiết lập công cụ HMM chú giải chức năng gene cho khai thác gene mã hóa enzyme/protein tham gia chuyển hóa lignocellulose.
3. Nghiên cứu lựa chọn gen, biểu hiện và xác định đặc điểm endo-xylanase mã hóa bởi gene trong dữ liệu DNA metagenome của vi khuẩn trong dạ cỏ dê.

• **Ý nghĩa khoa học thực tiễn của Luận án:**

* **Ý nghĩa khoa học:**

- Bộ dữ liệu giải mã sâu DNA metagenome của vi khuẩn trong dạ cỏ dê là bộ dữ liệu rất lớn, với sự đa dạng của vi khuẩn và gene mã hóa các loại enzyme tham gia vào nhiều con đường chuyển hóa khác nhau. Bộ dữ liệu là nguồn gene giá trị cho việc khai thác tìm kiếm các enzyme/protein mới có thể ứng dụng trong công nghiệp, nông nghiệp và y dược.

- Đã làm chủ phương pháp xây dựng mô hình Markov ẩn (HMM), thiết lập thành công quy trình khai thác, chú giải gene dựa trên các công cụ tin sinh để khai thác nhanh và chính xác các enzyme/protein mới từ dữ liệu DNA đa hệ gen.

- Xác định được vai trò quan trọng của chi vi khuẩn *Prevotella* sinh enzyme/protein phân giải lignocellulose giúp tăng cường chuyển hóa thức ăn của dê dựa trên kết quả khai thác gene từ dữ liệu DNA đa hệ gen.

* **Ý nghĩa thực tiễn:**

- Đã biểu hiện được enzyme endo-xylanase có hoạt tính cao, có thể sử dụng để phối trộn với các enzyme khác để làm tăng chuyển hóa lignocellulose thành đường, phục vụ cho thử nghiệm sản xuất ethanol.

• **Những đóng góp mới của Luận án:**

1. Đã xây dựng được bộ dữ liệu DNA metagenome vi khuẩn dạ cỏ dê với dung lượng 48,66 Gb và lần đầu tiên vai trò của *Prevotella* trong việc tăng cường tiêu hóa thức ăn trong dạ cỏ dê đã được phân tích sâu và gợi ý làm rõ.
2. Lần đầu tiên, luận án đã xây dựng được công cụ mô hình Markov ẩn (HMM) cho chú giải chức năng của nhóm gene mã hóa vùng liên kết carbohydrate (CBM) và một số enzyme tham gia tiền xử lý lignocellulose, cellulase, hemicellulase.
3. Endo-xylanase EXL mã hóa từ gene của vi khuẩn *Prevotella* trong dạ cỏ dê đã được biểu hiện, tinh sạch thành công với hoạt tính cao.

CHƯƠNG 1. TỔNG QUAN

1.1. Kỹ thuật Metagenomics nhằm khai thác hiệu quả gene tiềm năng

1.1.1. Sơ lược chung về kỹ thuật Metagenomics

Thuật ngữ "Metagenomics" được sử dụng lần đầu vào năm 1998 để chỉ các nghiên cứu về toàn bộ hệ gene của tất cả các vi sinh vật trong một khu hệ môi trường, ngược với hướng tiếp cận chỉ phân lập và nuôi cấy một loài vi sinh vật duy nhất *in vitro* [7]. Kỹ thuật Metagenomics khắc phục được nhược điểm của những phương pháp nghiên cứu truyền thống về phân lập, với thực tế là hiện nay, vẫn chưa có phương pháp thích hợp để nuôi cấy và phân lập thành công hơn 99% số vi sinh vật trong sinh quyển [5]. Về nguyên tắc, bất kỳ nghiên cứu nào hướng đến đối tượng là tất cả các cá thể trong một quần xã vi sinh vật như là một nguồn DNA duy nhất đều được coi là nghiên cứu metagenomics [8]. Hiện nay kỹ thuật Metagenomics đã và đang được sử dụng ngày càng nhiều để nghiên cứu đa dạng vi sinh vật của nhiều khu hệ sinh thái như môi trường suối nước nóng [9], môi trường biển [10], môi trường đất [11], môi trường bên trong cơ thể người và động vật - hệ vi sinh vật khu trú trong hệ tiêu hóa của người [12] hay các động vật nhai lại [13]. Mối tương tác giữa các loài sinh vật cộng sinh trong đường tiêu hóa của động vật nhai lại được thể hiện mạnh mẽ trong quá trình phân giải các hợp chất lignocellulose, vì vậy đây là đối tượng tiềm năng cho việc nghiên cứu khai thác các gene mới mã hóa cho các enzyme tham gia chuyển hóa sinh khối thực vật [14].

Phương pháp nghiên cứu metagenomics gồm hai hướng tiếp cận chính, bao gồm hướng tiếp cận theo chức năng và hướng tiếp cận theo trình tự. Hướng tiếp cận theo chức năng dựa vào việc xây dựng một thư viện vật lý các đoạn trình tự DNA từ mẫu DNA metagenome sau khi đã được tách chiết, trong đó các đoạn trình tự này được tách dòng trong các vector có thể chèn những đoạn kích thước lớn (vector cosmid hoặc fosmid) [15], biểu hiện trong hệ vật chủ phổ biến sau đó tiến hành sàng lọc để tìm những dòng thể hiện hoạt tính sinh học mong muốn. Hạn chế lớn của phương pháp này là hầu hết các gene không được biểu hiện, tạo sản phẩm protein hoàn chỉnh trong những chủng vi khuẩn xác định chuyên dùng để tách dòng, ví dụ như *E. coli*, mặc dù có rất nhiều gene từ các nguồn vi sinh vật khác nhau đã được biểu hiện thành công trong đối tượng này [16]. Đây vẫn là rào cản cực lớn khiến cho việc thu nhận thông tin nghiên cứu metagenome bằng phương pháp này bị hạn chế. Một hạn chế nữa của phương pháp này là tần số dòng có hoạt tính mong

muốn từ thư viện DNA metagenome là rất thấp. Ví dụ, nghiên cứu của Henne và các cộng sự chỉ thu được 1 dòng trong 730.000 dòng trong thư viện DNA metagenome của vi sinh vật trong đất thể hiện hoạt tính lipolytic [17].

Một hướng tiếp cận khác trong nghiên cứu Metagenomics đó là giải trình tự toàn bộ DNA metagenome và phân tích dữ liệu. Trong những năm gần đây, với sự phát triển vượt bậc của khoa học công nghệ, việc giải trình tự toàn bộ hệ gene đã trở nên đơn giản, tiết kiệm thời gian và kinh phí, cung cấp một lượng lớn thông tin vật chất di truyền của các loài vi sinh vật trong môi trường, đó là nguồn nguyên liệu để khai thác các gene mới phục vụ cho mục đích của con người. Kỹ thuật giải trình tự thông lượng cao (High Throughput Sequencing - HTS) phân biệt đặc thù với kỹ thuật giải trình tự truyền thống (giải trình tự Sanger) về mặt dung lượng giải trình tự với mỗi lần tiến hành. Cụ thể, với mẫu giải trình tự Sanger thường tạo ra được 10^2 trình tự (với độ dài 600-900 bp) thì HTS có thể giải được $10^6 - 10^9$ trình tự (với độ dài 100-700 bp) cho mỗi lần thực hiện [18]. Với ưu điểm giải trình tự nhanh và chính xác, các hệ thống máy giải trình tự thế hệ mới - Next Generation Sequencing (NGS) được sử dụng phổ biến như máy giải trình tự của Roche 454 của 454 Life Sciences, HiSeq 2000 của Illumina và AB SoLid của Life Technologie [19].

Kết quả giải trình tự DNA đa hệ gene là tập hợp các đoạn trình tự riêng rẽ. Sau đó, những đoạn DNA đã được giải trình tự sẽ được tập hợp và lắp ghép lại thành những đoạn trình tự dài bằng cách sử dụng một trong hai cách sau, lắp ghép dựa trên trình tự tham chiếu được so sánh hoặc lắp ghép *de novo*. Thông tin trình tự được phân tích xử lý bằng phần mềm để phân chia các đoạn trình tự DNA vào các nhóm phân loại có thể đại diện cho một loài hoặc một nhóm loài. Dựa trên những cơ sở dữ liệu tham khảo có sẵn như KEGG là cơ sở dữ liệu trực tuyến liên quan đến hệ gen, các con đường enzyme và các sản phẩm sinh học [20], eggNOG là cơ sở dữ liệu chứa các trình tự từ những nguồn khác biệt nhưng có chung nguồn gốc tổ tiên được xếp chung vào một nhóm [21], COG là cơ sở dữ liệu protein của sinh vật nhân sơ, nhân chuẩn đơn bào [22], PFAM/TIGRFAM -protein families là cơ sở dữ liệu các họ protein [23, 24], với một mức độ tương đồng nhất định khi so sánh trình tự cần chú giải với trình tự tham chiếu, những trình tự này có thể được dự đoán chức năng. Tuy nhiên, chú giải chức năng của dữ liệu metagenome luôn là một thách thức lớn, vẫn chưa có công cụ nào có thể chú giải hoàn toàn tất cả các trình tự một cách chính xác tuyệt đối. Những trình tự không thể chú giải có thể là kết quả của quá trình giải trình tự đưa đến kết quả lỗi, hoặc có thể là gene thật nhưng mã hóa cho những chức năng sinh học chưa xác định, hoặc là gene mới không có tính tương đồng với những gene đã biết, đã công bố. Vì vậy, nhu cầu xác định được những

phương pháp mới và hiệu quả để chú giải và khai thác gene tiềm năng theo nhu cầu thực tế luôn được đề cao.

1.1.2. Tình hình nghiên cứu ứng dụng kỹ thuật Metagenomics trong khai thác gene

1.1.2.1. Nghiên cứu ứng dụng Metagenomics trên thế giới

Hướng nghiên cứu đa hệ gene theo chức năng đã được sử dụng để phát hiện nhiều gene chức năng mới như gene mã hóa polyketide synthase, gene kháng kháng sinh như aminoglycoside và tetracycline, gene mã hóa các enzyme như oxidoreductase, esterase và glycosidase hay các gene mã hóa chất xúc tác sinh học khác [25]. Tuy nhiên có khá nhiều thách thức trong nghiên cứu Metagenomics theo hướng chức năng. Trước hết, điều kiện sống của các quần xã vi sinh vật trong môi trường cực kỳ phức tạp, DNA của những loài vi sinh vật có mật độ thấp có nguy cơ không được tách chiết và phân lập. Thứ hai, có khả năng các đoạn DNA nhỏ bị mất trong quá trình tách dòng gen. Thứ ba, khả năng biểu hiện của các gene ngoại lai trong các chủng tách dòng là khá thấp. Thứ tư, những phương pháp sàng lọc hiện nay vẫn còn hạn chế, không thể đáp ứng toàn bộ yêu cầu sàng lọc enzyme. Thứ năm, quá trình sàng lọc có hiệu suất thấp, ước tính từ hàng chục ngàn dòng phân lập, chỉ có một vài dòng là có hoạt tính và được lựa chọn. Thứ sáu, với những hạn chế về nhiệt độ, pH và những yếu tố khác, chỉ có một phần rất nhỏ enzyme trong số các enzyme được biểu hiện có hoạt tính có thể được ứng dụng vào thực tiễn. Ngoài ra, nghiên cứu giải trình tự metagenome nói chung và nghiên cứu theo hướng chức năng nói riêng chỉ được tập trung ở những nước phát triển, nguồn vật chất di truyền cực lớn ở các nước đang phát triển vẫn còn bị bỏ ngỏ.

Giải trình tự metagenome đã được áp dụng với mẫu vi khuẩn, nấm, virus và các loài khác. Những phương pháp phân tích tin sinh học được sử dụng để xác định trình tự genome của vi sinh vật dựa trên hệ gene tham chiếu. Trong trường hợp không có sẵn những thông tin cơ bản về trình tự gene tham chiếu, kết quả giải trình tự được lắp ráp *de novo*. Bằng phương pháp này, thông tin di truyền của virus mới như COVID-19 đã được công bố [26]. Nghiên cứu Metagenomics đã được ứng dụng trong nghiên cứu vi sinh vật từ rất nhiều vùng sinh thái đa dạng về điều kiện môi trường sống, và được áp dụng trong việc nhận dạng các vi sinh vật gây bệnh trong đường tiêu hóa, hệ miễn dịch. Aron-Wisnewsky và cộng sự đã sử dụng dữ liệu giải trình tự metagenome đường tiêu hóa của người để phát hiện hệ vi sinh liên quan đến bệnh viêm gan của nhóm người không nghiện rượu [27]. Tương tự, hệ vi sinh vật liên quan đến bệnh tự miễn [28] và bệnh ung thư [29] cũng đã được phát

hiện. Giải trình tự metagenome góp phần phát hiện nhiều vi khuẩn kháng kháng sinh [30] và gene kháng kháng sinh mới [31]. Năm 2021, Chi nhánh Phòng thí nghiệm Y dược của Hội Y học Trung Quốc công bố một báo cáo chuyên gia, khuyên dùng phương pháp giải trình tự thế hệ mới để áp dụng trên đối tượng mẫu nghi nhiễm bệnh nhưng không được phát hiện bằng phương pháp thường quy như phân tích hóa sinh, nuôi cấy vi sinh vật và việc sử dụng các liệu trình chống nhiễm khuẩn thông thường không có tác dụng. Hiện nay, giải trình tự metagenome đã được ứng dụng rộng rãi trong việc phát hiện các tác nhân gây bệnh, ví dụ như năm loài sinh vật (*Neisseria meningitidis*, *Streptococcus agalactiae*, *Candida albicans*, *Mycobacterium fortuitum*, *Mycobacterium abscessus*) đã được phát hiện bằng giải trình tự metagenome mẫu dịch não tủy [32] gây bệnh viêm màng não mô cầu và bệnh viêm não Nhật Bản. Nghiên cứu metagenome cũng đã được ứng dụng trong việc sản xuất rượu, cụ thể đa dạng các thành phần quần xã vi sinh vật và đặc tính của quần xã trong hàm rượu ảnh hưởng rất lớn đến khả năng cải thiện chất lượng và hương vị của rượu [33]. Nghiên cứu metagenome cũng cho thấy giá trị dược liệu của các loại thuốc Trung Quốc từ các vùng dược liệu khác nhau có sự khác biệt đáng kể, liên quan đến thành phần vi sinh vật của môi trường dưới điều kiện khí hậu và thổ nhưỡng đặc trưng, và sự gia tăng của sản lượng dược liệu liên quan mật thiết đến quần xã vi sinh vật trong đất và rễ cây [34].

Hệ vi sinh vật nói chung và hệ vi khuẩn nói riêng trong đường tiêu hóa của động vật nhai lại đóng vai trò cực kỳ quan trọng như là một nguồn xử lý và cung cấp năng lượng cho vật chủ. Vì vậy, nghiên cứu đa dạng hệ gene trong vi hệ sinh thái này để đánh giá và dự đoán những thay đổi có thể gây nên những rối loạn chuyển hóa của vật chủ đóng vai trò quan trọng. Việc nghiên cứu đa hệ gene của toàn bộ hệ sinh vật đường ruột được tiến hành bởi công nghệ giải trình tự và những hướng tiếp cận “omic”, bao gồm Metabolomic, Metaproteomic, Metatranscriptomic, Metagenomic [25]. Công nghệ giải trình tự đưa ra những kết quả về định danh của hệ vi sinh vật, còn những hướng nghiên cứu “omic” làm rõ chức năng của những vi sinh vật có trong dạ cỏ và khẳng định vai trò tương hỗ của các vi sinh vật trong đường tiêu hóa đối với vật chủ. Cụ thể, nhiều nghiên cứu cho thấy vi khuẩn ngành Bacteroidetes chiếm tỷ lệ lớn hơn đáng kể trong đường tiêu hóa so với ngành Firmicutes. Các vi khuẩn nội sinh trong dạ cỏ có thể chuyển hóa các thành phần carbohydrate trong thức ăn thành các VFA (volatile fatty acid - axit béo bay hơi), cung cấp đến 80% năng lượng của vật chủ động vật nhai lại. Protein của bản thân vi khuẩn MCP (microbial crude protein – protein thô của vi khuẩn) bị vật chủ tiêu hóa và hấp thụ ở ruột non, chiếm 50-90 lượng protein đi vào tá tràng và

cung cấp hầu hết các amino acid cần thiết cho sự sinh trưởng và tạo sữa, vì vậy đây cũng là một nguồn dinh dưỡng quan trọng [35]. Hơn nữa, một số loài vi khuẩn dạ cỏ có thể tổng hợp nhiều loại vitamin, như vitamin nhóm B và vitamin K, bổ sung vào thành phần dinh dưỡng của vật chủ [36]. Vai trò lớn nhất của vi sinh vật dạ cỏ nói chung là khả năng phân giải các polysaccharide và đạm vô cơ trong thức ăn, chuyển hóa thành những thành phần có thể được hấp thụ bởi vật chủ. Hệ vi sinh vật dạ cỏ đóng vai trò thiết yếu trong quá trình sử dụng hiệu quả thức ăn, quá trình tiết sữa và chất lượng sữa [36]. Để tiêu hóa hiệu quả chất dinh dưỡng trong thức ăn, sự cân bằng giữa quá trình chuyển hóa carbohydrate và nitrogen là cực kỳ quan trọng. Dư thừa nitrogen trong ruột non gây áp lực lên sự sinh trưởng của vi sinh vật chuyển hóa carbohydrate, thay đổi cấu trúc hệ vi sinh vật và ảnh hưởng đến mối tương tác chung giữa vật chủ và hệ vi sinh [37]. Nghiên cứu của Wang và các cộng sự cho thấy *Fibrobacter succinogenes*, *Butyrivibrio fibrisolvens* và *Ruminococcus_sp._HUN007* là những loài vi khuẩn dạ cỏ tối thiết trong quá trình chuyển hóa nitrogen cung cấp cho vật chủ [38]. Vi khuẩn thuộc ngành Proteobacteria, Tenericutes, các chi *Anaeroplasma*, *Campylobacter*, *Clostridium* có liên quan chặt chẽ đến quá trình phân giải chất xơ trong đường tiêu hóa của lợn [39]. Kỹ thuật Metagenomics đã góp phần làm rõ quá trình chuyển hóa chất dinh dưỡng trong dạ cỏ, mở rộng cơ sở dữ liệu về hệ vi sinh vật, giải thích mối quan hệ tương hỗ giữa vật chủ và hệ vi sinh vật đường ruột.

1.1.2.2. Nghiên cứu ứng dụng kỹ thuật Metagenomics ở Việt Nam

Hướng nghiên cứu sử dụng kỹ thuật Metagenomic đã được một số nhóm nghiên cứu áp dụng ở Việt Nam trong giai đoạn gần đây. Nhóm nghiên cứu của PGS. TS. Trần Đình Mẫn đã giải trình tự DNA metagenome vi sinh vật suối nước nóng Bình Châu, tỉnh Bà Rịa – Vũng Tàu, và xây dựng được bộ dữ liệu DNA metagenome dung lượng 9,4 Gb với 106.903 gene được chú giải trên tổng số 156.093 gene dự đoán. Qua phân tích nhóm tác giả đã xác định được đa dạng vi sinh vật gồm 41 ngành, 57 lớp, 128 bộ, 245 họ, 825 chi và 2.250 loài khác nhau, với ngành đa dạng nhất là ngành Proteobacteria, tiếp đến là Firmicutes, đều thuộc nhóm vi khuẩn ưa nhiệt [40]. Nhóm nghiên cứu của TS. Kim Thị Phương Oanh từ Viện Nghiên cứu hệ gene đã phân tích dữ liệu metagenome từ tu hài (*Lutraria rhynchaena*) bị nhiễm bệnh sùng vôi ở vùng nuôi tu hài Cát Bà, Hải Phòng và Vân Đồn, Quảng Ninh. Metagenome của tu hài nhiễm bệnh sùng vôi đã được giải mã bằng hệ thống NGS của Illumina và máy MinION của Oxford Nanopore. Kết quả hệ gene của vật chủ tu hài đã được lắp ráp và chú giải thành công. Bên cạnh đó, dữ liệu metagenome của các loài vi sinh vật trong dịch lọc vôi tu hài nhiễm bệnh đã

được xác định [41], và phát hiện một hệ gene virus mới [42]. PGS. TS. Hồ Huỳnh Thùy Dương và cộng sự đã sử dụng công nghệ giải trình tự Oxford Nanopore với mẫu vi sinh vật từ nước đầm nuôi tôm ở Ninh Thuận và xác định thành phần quần xã vi sinh vật và khả năng kháng kháng sinh, từ đó định hướng ứng dụng trong việc kiểm soát sức khỏe tôm nuôi trong ngành công nghiệp thủy sản [43]. TS. Trần Bình Minh và cộng sự đã xây dựng cơ sở dữ liệu quần xã vi sinh vật trong mẫu nước mặt thu thập từ sông Serepok của vườn Quốc gia Yok Don, Tây Nguyên. Cơ sở dữ liệu cung cấp thông tin về đa dạng quần xã vi sinh vật và chức năng, có thể được sử dụng để tăng cường công tác bảo tồn cũng như tăng khả năng sản xuất nguyên liệu tái tạo [44]. Nhóm nghiên cứu của TS. Nguyễn Trung Nam đã thu thập và giải trình tự DNA ribosome 16S của mẫu vi khuẩn trong đường tiêu hóa của bệnh nhân đái tháo đường type 2 của Việt Nam bằng công nghệ giải trình tự của Illumina và so sánh với đối chứng không bị bệnh. Kết quả cho thấy có sự khác biệt rõ ràng về tỷ lệ ngành Firmicutes và lớp Clostridia giữa bệnh nhân và đối chứng, gợi ý rằng sự suy giảm tỷ lệ của ngành và lớp trên có liên quan đến chứng béo phì và bệnh tiểu đường type 2 [45]. Nhóm nghiên cứu của GS. TS. Chu Hoàng Hà đã nghiên cứu giải trình tự metagenome của vi sinh vật trong các đầm nuôi tôm ở Việt Nam trên hệ thống Illumina. Kết quả giải trình tự cho thấy định đa dạng vi khuẩn khu trú trong hệ tiêu hóa của tôm nuôi và môi trường xung quanh [46] cũng như vai trò của chi vi khuẩn *Pseudomonas* trong việc chuyển hóa nitrogen của quần xã vi khuẩn trong mẫu nước từ ba đầm nuôi tôm ở Sóc Trăng. Kết quả này hứa hẹn khả năng sử dụng chi *Pseudomonas* trong việc xử lý nước đầm nuôi tôm [47].

Nghiên cứu metagenomic cũng được ứng dụng để khai thác các cấu trúc gene mới, đánh giá đa dạng cấu trúc gene mới mà các hướng tiếp cận khác khó thực hiện được. Điển hình, từ năm 2014 đến năm 2017, bằng nguồn kinh phí của Đề tài độc lập cấp Nhà nước (Mã số: ĐTĐLCN.15/14) do PGS. TS Đỗ Thị Huyền làm chủ nhiệm, DNA đa hệ gene của vi khuẩn trong dạ cỏ dê đã được giải mã, khai thác các nguồn gene khác nhau tham gia thủy phân lignocellulose và đào tạo hai nghiên cứu sinh. Trong đó NCS Nguyễn Khánh Hoàng Việt đã thực hiện và bảo vệ thành công đề tài luận án “Nghiên cứu đánh giá sự đa dạng và vai trò của một số module trong cấu trúc enzyme thủy phân cellulose từ khu hệ vi sinh vật trong dạ cỏ của dê” năm 2022. Luận án đã tập trung khai thác các gene mã hóa nhóm cellulase từ bộ dữ liệu DNA đa hệ gene 8,4 kb và đã khai thác được 816 trình tự mã hóa cho cellulase trong đó phát hiện cấu trúc đặc thù của beta-glucosidase trong dạ cỏ dê có mang chủ yếu vùng FN3 nằm phía sau vùng xúc tác. Trong khi đó, các endo-glucanase chủ yếu có mang vùng Ig và nằm ở phía trước của vùng xúc tác. Duy nhất, một trình tự

mã hóa cho endo-glucanase nhưng có mang vùng FN3 và nằm ở phía trước vùng xúc tác. Các vùng module này đã được tách dòng, tái tổ hợp và nghiên cứu phát hiện được vai trò của vùng FN3 trong việc làm ổn định cấu trúc phân tử của enzyme, làm tăng ái lực của enzyme với cơ chất, giúp ổn định vùng xúc tác cho sự hoạt động chuyển hóa của enzyme [48]. Luận án chưa nghiên cứu tạo bộ dữ liệu DNA metagenome. Đồng thời, luận án cũng chưa đánh giá cấu trúc, đa dạng vi khuẩn trong toàn bộ dữ liệu, chưa nghiên cứu vai trò của vi khuẩn trong thủy phân lignocellulose nói chung, vai trò của *Prevotella* nói riêng. Bộ dữ liệu 8,46 Gb được xác định là có chỉ số độ bao phủ (mapping rate) chỉ đạt khoảng 27%, vì vậy cần giải trình tự với dung lượng cao hơn với hy vọng bao phủ hoàn toàn, từ đó phân tích sâu để so sánh trên khía cạnh đa dạng vi khuẩn, độ đa dạng, phong phú của các gene liên quan đến chuyển hóa lignocellulose ở hai bộ dữ liệu, vai trò của vi khuẩn *Prevotella* trong thủy phân lignocellulose và tìm kiếm các gene mã hóa endo-xylanase mới để sản xuất ở dạng tái tổ hợp. Do vậy, luận án này được tiến hành song song đề giải quyết các vấn đề được nêu ở trên.

1.1.3. Chú giải gene trong tin sinh học

1.1.3.1. Một số công cụ chú giải gene thông dụng

Chú thích chức năng của trình tự protein với độ chính xác cao (AC) đã trở thành một trong những vấn đề quan trọng nhất trong việc tìm hiểu cơ chế phân tử của sự sống và có ảnh hưởng lớn trong nghiên cứu sinh học cơ bản, bệnh lý và dược phẩm. Với sự tích lũy nhanh chóng của số lượng trình tự protein được giải trình tự và công bố, việc chú thích chức năng của protein ngày càng trở nên khó khăn và cấp thiết. Những phương pháp chú giải gene hiện tại có thể chia thành ba loại, dựa trên tương đồng trình tự, dựa trên tương đồng cấu trúc chức năng và dựa trên bối cảnh gene (gene context).

Các công cụ chú giải dựa trên sự tương đồng trình tự bằng cách so sánh trình tự của các gene cần xác định với trình tự của cơ sở dữ liệu biết trước, từ đó kết luận chức năng của gene nếu độ tương đồng lớn hơn ngưỡng giới hạn với độ tin cậy cho phép. Phương pháp chú giải này là phương pháp thông dụng nhất nhờ khả năng tiếp cận dễ dàng các thư viện trình tự số lượng lớn cũng như khả năng dự đoán chức năng của những protein cùng nguồn gốc. Công cụ chú giải dựa trên sự tương đồng trình tự phổ biến nhất là BLAST (Basic Local Alignment Search Tool). Tuy nhiên công cụ này mất khá nhiều thời gian để dự đoán chức năng nếu số lượng gen/protein cần chú giải khá lớn. Một số công cụ chú giải hiện đại dựa trên sự tương đồng trình tự sử dụng phương pháp so sánh tối ưu có thể rút ngắn thời gian phân

tích và chú giải gen. Công cụ eggNOG-mapper chú giải chức năng bằng cách so sánh trình tự với những cụm trình tự tương đồng có chung nguồn gốc tổ tiên (ortholog), dựa trên cơ sở dữ liệu eggNOG về các nhóm ortholog, có thể đưa ra kết quả nhanh và chính xác hơn so với so sánh từng cặp [49]. Công cụ PANNZER là một công cụ chú giải tương tác trực tuyến có khả năng xử lý phân tích đến 100000 trình tự đầu vào cùng lúc và đưa ra thông tin chú giải về gene ontology (GO – bản thể gen) [50]. BlastKOALA và GhostKOALA là những công cụ chú giải trình tự hệ gene và đa hệ gene dựa trên cơ sở dữ liệu KEGG về các gene chức năng và con đường trao đổi chất [51].

Các công cụ chú giải dựa trên sự tương đồng cấu trúc chức năng có ưu điểm cao về độ chính xác, tuy nhiên phương pháp này có những giới hạn do bị phụ thuộc chặt chẽ vào thông tin của những protein đã được nghiên cứu sâu về cấu trúc tinh thể và cấu trúc không gian, trong một số trường hợp mối quan hệ giữa sự tương đồng về cấu trúc không gian và tương đồng về chức năng vẫn chưa được làm sáng tỏ hoàn toàn. FunctionAnnotator là một công cụ chú giải trực tuyến một cách hiệu quả, có thể chú giải chi tiết về nhiều khía cạnh, bao gồm chú giải chức năng của enzyme, phân loại GO, nhận dạng vùng hoạt tính domain/motif và dự đoán vị trí nội bào [52]. Server dbCAN3 là một công cụ trực tuyến có khả năng chú giải tự động protein, với đối tượng tập trung là các enzyme hoạt động trên carbohydrate, đây là phiên bản nâng cấp của công cụ dbCAN và dbCAN2 với khả năng phân tích sâu và chính xác hơn [53]. Phyre2 là một bộ công cụ có sẵn trên web để dự đoán và phân tích cấu trúc, chức năng và các đột biến của protein, cung cấp cho các nhà sinh học một giao diện đơn giản và trực quan cho các công cụ tin sinh học protein hiện đại. Phyre2 sử dụng các phương pháp phát hiện tương đồng tiên tiến để xây dựng mô hình 3D, dự đoán các vị trí liên kết phối tử và phân tích tác động của các biến thể đơn amino acid đối với trình tự protein [54]. Cơ sở dữ liệu InterPro (<https://www.ebi.ac.uk/interpro/>) cung cấp thông tin phân loại các trình tự protein tích hợp thành các họ và xác định các vùng chức năng quan trọng cũng như các vùng bảo thủ trong trình tự. InterPro (phiên bản 90.0) có những cập nhật và phát triển mở rộng cơ sở dữ liệu, đồng thời cung cấp quyền truy cập dữ liệu thân thiện hơn với người dùng. Các tính năng của công cụ Pfam bao gồm thông tin chung về các họ protein, trình tự và mô hình đại diện chung của họ cũng được cập nhật vào phiên bản mới của InterPro [55].

Phương pháp chú giải dựa trên bối cảnh gene xem xét đến vị trí, tương tác của gene cần chú giải với các gene khác, vì vậy có thể hỗ trợ bổ sung cho những trường hợp các gene mới không tương đồng với gene đã biết, không thể chú giải bằng

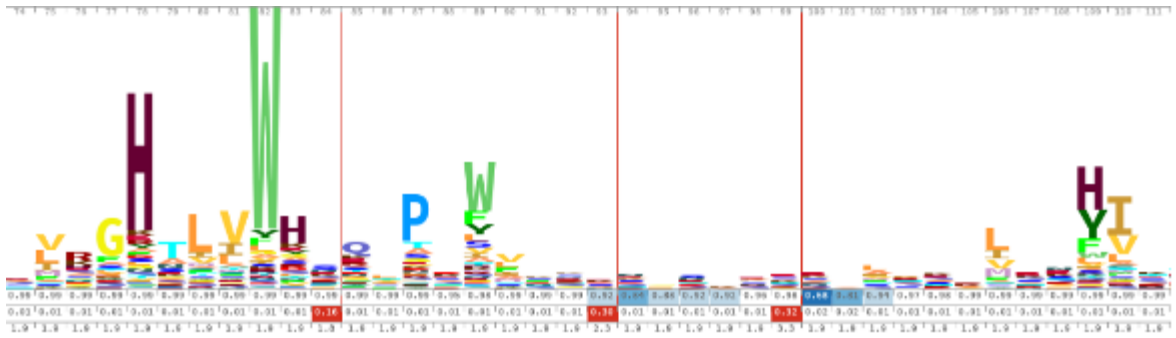
những cách tiếp cận trên. Phương pháp này đặc biệt phù hợp trong việc chú giải chức năng của những protein tương tác với những protein khác, những protein được đồng biểu hiện trong tế bào, tuy nhiên bị hạn chế bởi những thông tin gây nhiễu trong dữ liệu tương tác protein [56]. Công cụ FunGeCo là một công cụ phân tích trực tuyến cho phép chú giải chức năng metagenome vi khuẩn dựa trên bối cảnh gen. Cụ thể, các đơn vị chức năng GCM (Gene Context-based Module) được xác định từ các hệ gene cụ thể, xây dựng một hệ thống bản đồ chi tiết thông tin bối cảnh gen, ghi nhớ vị trí trên genome [57]. Công cụ GeConT cho phép người sử dụng mô hình hóa bối cảnh của một gene hoặc một nhóm gene trong genome và mối quan hệ tương đồng nguồn gốc với những trình tự đã được báo cáo. Công cụ này kết hợp thông tin từ cơ sở dữ liệu COG, Pfam và KEGG, cho phép người dùng có biểu diễn đồ họa tích hợp về chức năng của gene ở nhiều cấp độ, cây phát sinh chủng loại và bối cảnh gene của chúng [58].

1.1.3.2 Mô hình đại diện HMM: công cụ mới khai thác hiệu quả dữ liệu metagenome

Số lượng các protein đã được xác định cấu trúc không gian và hoạt tính luôn nhỏ hơn rất nhiều so với số lượng protein đã biết trình tự cấu trúc bậc 1. Để dự đoán cấu trúc không gian, phương pháp phân tích sử dụng mô hình đại diện dựa trên so sánh trình tự khá phổ biến. Cụ thể, nếu một trình tự mới được phát hiện có độ tương đồng nhất định với trình tự của một protein đã biết, chúng ta có thể dự đoán rằng hai protein có cấu trúc không gian và hoạt tính tương đồng [59]. Phương pháp phân tích sử dụng mô hình đại diện là phương pháp nhận biết những trình tự ít tương đồng bằng cách so sánh trình tự bậc 1. Nhóm trình tự tương đồng nằm trong cùng một họ sẽ được so sánh tương đồng để tạo ra kết quả so sánh đa trình tự. Thuật toán so sánh tương đồng sắp xếp các trình tự sao cho kết quả so sánh có số lượng vị trí tương đồng là tối đa và số lượng khoảng trống (gap) là tối thiểu. Mỗi vị trí tương đồng được đánh giá bằng điểm số, độ tương đồng càng lớn thì điểm số càng cao, nếu là khoảng trống thì mang giá trị âm. Kết quả so sánh đa trình tự cuối cùng là kết quả mà tổng giá trị điểm tương đồng là cao nhất. Thông tin từ kết quả so sánh đa trình tự được lượng hóa bằng một ma trận giá trị tương ứng với từng vị trí trên trình tự, được gọi là mô hình đại diện (profile) [60]. Sử dụng mô hình đại diện để tìm kiếm trong cơ sở dữ liệu là phương pháp mang độ nhạy và độ đặc hiệu cao hơn so với phương pháp so sánh cặp đôi sử dụng hệ thống tính điểm độc lập cho từng vị trí.

Mô hình đại diện dựa trên mô hình Markov ẩn (Hidden Markov Model - HMM) là một trong những phương pháp có độ nhạy và độ chính xác cao nhất đại

diện cho các trình tự tương đồng trong họ trình tự [61]. Ưu thế của mô hình đại diện HMM so với các mô hình đại diện khác được thể hiện ở việc giá trị điểm đại diện cho khả năng bắt hoặc chèn trình tự được xây dựng trên cơ sở lý thuyết thống kê, trái với các mô hình đại diện khác chỉ bằng kết quả suy diễn thực nghiệm. Cụ thể, mô hình HMM áp dụng nguyên lý xác suất thống kê để ước tính tần số thật mà một vị trí trình tự có thể nhận từ kết quả so sánh thực nghiệm trong khi các mô hình thông thường sử dụng chính giá trị thực nghiệm [62].



Hình 1. 1. Một phần mô hình đại diện HMM của họ GH10 [23]

Mô hình đại diện HMM đã và đang được sử dụng ngày càng phổ biến để phân tích và nghiên cứu các họ protein. Hiện nay, nhiều cơ sở dữ liệu protein đã ứng dụng mô hình đại diện HMM bao gồm Pfam [23], TIGRFAM [24], SUPERFAMILY [63], CATH/Gene3D [64]. Khi số lượng dữ liệu giải trình tự chi tiết của các loài sinh vật càng ngày càng nhiều, thông tin trình tự trong các cơ sở dữ liệu trên cũng tăng lên nhanh chóng và ngày càng đồ sộ. Cụ thể, khi mới được thiết lập vào năm 1996, cơ sở dữ liệu Pfam chỉ có 100 họ protein xây dựng từ 10.431 trình tự, phiên bản Pfam 35.0 được công bố năm 2021 đã chứa đến 19.632 họ protein.

Ưu thế của việc sử dụng mô hình đại diện HMM so với so sánh đơn lẻ từng đôi trình tự là tiền đề để nhiều kết quả nghiên cứu phát triển các thuật toán giải quyết các vấn đề giữa so sánh trình tự với mô hình đại diện và giữa các mô hình đại diện với nhau, trong đó có các thuật toán như SAM, Prof_Sim, COMPASS, HHsearch, PRC, AlignHUSH và HMMER. Tính chính xác của các thuật toán được tăng cường bằng cách kết hợp với phân tích so sánh thông tin về cấu trúc, đặc tính lý hóa, nguồn gốc tiến hóa với mục tiêu cuối cùng là để các trình tự được so sánh một cách chính xác nhất [65].

1.2. Tổng quan về lignocellulose

1.2.1. Vai trò của lignocellulose trong nền kinh tế sinh học

1.2.1.1. Lignocellulose là nguồn năng lượng tái tạo phong phú

Nhu cầu năng lượng là nhu cầu bức thiết của nhân loại, nhu cầu này không ngừng tăng lên theo sự tăng trưởng dân số thế giới. Để thỏa mãn những tiện nghi trong đời sống, nhu cầu năng lượng của con người càng trở nên cao gấp nhiều lần so với nhu cầu tối thiểu [66]. Sử dụng các nguồn năng lượng truyền thống để đáp ứng nhu cầu năng lượng đi kèm với những thách thức về môi trường. Nguồn năng lượng từ nhiên liệu hóa thạch (than đá, dầu mỏ, khí đốt) có những hạn chế lớn: (i) sự phát thải khí CO₂ cực lớn từ quá trình sử dụng nhiên liệu vào môi trường, dẫn đến biến đổi khí hậu; (ii) một lượng lớn các loại khí độc hại, chất độc, các tác nhân gây ung thư cũng được giải phóng vào khí quyển, gây ô nhiễm môi trường; (iii) nguồn nhiên liệu hóa thạch là tài nguyên giới hạn, không thể tái tạo [67]. Sử dụng năng lượng của nước để làm thủy điện có thể có tác động xấu đến môi trường trong trường hợp các hồ thủy điện gặp trục trặc, cần phải xả tràn dẫn đến ngập lụt ở khu vực hạ du và ảnh hưởng đến toàn bộ quần xã sinh vật. Tương tự, sử dụng năng lượng hạt nhân cũng có nguy cơ dẫn đến thảm họa sinh thái trong trường hợp lò phản ứng của nhà máy điện hạt nhân có sự cố.

Khai thác năng lượng từ sinh khối là một trong những nguồn năng lượng mới, thay thế, bên cạnh khai thác năng lượng gió, năng lượng mặt trời, năng lượng địa nhiệt. Nguồn năng lượng từ sinh khối khắc phục những nhược điểm của nguồn năng lượng hóa thạch vì đây là nguồn năng lượng có khả năng tái tạo thường xuyên, trong quá trình sử dụng ít phát thải các loại hợp chất độc hại cho môi trường, đồng thời lượng khí cacbonic thải ra được tái hấp thụ bởi sinh vật sản xuất [1].

Ba nhóm sinh khối chính có thể được dùng để làm nguồn nhiên liệu chuyển hóa thành nhiên liệu xanh là: (i) dầu thực vật, mỡ động vật; (ii) tinh bột và (iii) lignocellulose. Dựa trên cách sử dụng nguồn sinh khối, nhiên liệu sinh học có thể được chia thành nhiên liệu sơ cấp và nhiên liệu thứ cấp [68]. Nhiên liệu sơ cấp là sinh khối được sử dụng trực tiếp mà không qua bất kỳ quá trình xử lý bổ sung nào, ví dụ như gỗ, chất béo động thực vật, phế phụ phẩm nông lâm nghiệp được đốt trực tiếp nhằm cung cấp nhiệt. Cách sử dụng này rất phổ biến ở các nước đang phát triển. Nhiên liệu sơ cấp được sử dụng đơn giản, không yêu cầu những quy trình công nghệ phức tạp, tuy nhiên tính ứng dụng rất hạn chế. Để khắc phục nhược điểm trên, nhiên liệu thứ cấp được sản xuất từ sinh khối bằng cách chiết xuất những hợp chất giàu năng lượng bằng quá trình xử lý công nghệ cao.

Dựa trên nguồn sinh khối, phương pháp xử lý và sản xuất, nhiên liệu thứ cấp hiện tại được chia thành 3 thế hệ. Nhiên liệu sinh học thế hệ thứ nhất sử dụng nguồn sinh khối là từ dầu mỡ động thực vật và tinh bột, đường (mía đường, củ cải đường) để sản xuất cồn và dầu diesel sinh học. Quá trình sản xuất loại nhiên liệu này có ưu

điểm là khá đơn giản và chi phí tương đối thấp nhờ sử dụng phương pháp lên men vi sinh vật, tuy nhiên cũng đi kèm với những hạn chế lớn. Cụ thể là, năng suất tạo ra nhiên liệu này khá thấp vì chỉ có một phần rất nhỏ của sinh khối có thể được sử dụng để sản xuất nhiên liệu, đồng thời lại gây ra sự cạnh tranh với công nghiệp chế biến thực phẩm, hoặc thậm chí làm mất cân bằng cung-cầu lương thực [69].

Để giải quyết những vấn đề tồn tại trong thế hệ đầu tiên của nhiên liệu sinh học, các nhà khoa học đã chuyển hướng sử dụng nguồn sinh khối lignocellulose, là nguồn sinh khối có trữ lượng dồi dào hơn hẳn. Một số loài thực vật có khả năng sinh trưởng rất nhanh, thu hoạch hàng năm, đồng thời các loài cây này có thể được trồng ở các khu đất ngoại vi nên không ảnh hưởng đến diện tích đất sản xuất nông nghiệp, đảm bảo an ninh lương thực, đồng thời đất trồng cần rất ít phân bón nên hạn chế khả năng làm đất trở nên hoang hóa. Ví dụ, cây bạch đàn có thể trồng ở trên đất nhiễm mặn đến mức không thể trồng các loại cây lương thực hay hoa màu, thậm chí còn giúp phục hồi trạng thái của đất để sản xuất trong tương lai [70]. Một nguồn lignocellulose lớn khác mà có thể được khai thác là từ chất thải trong quá trình sản xuất và sinh hoạt, đặc biệt là chất thải trong quá trình chế biến gỗ (mạt gỗ, mùn cưa) và rác thải hữu cơ trong sinh hoạt. Dựa trên một báo cáo của Cơ quan Năng lượng Quốc tế (IEA) năm 2007, chúng ta có thể khai thác đến 10^{20} Jun đơn vị năng lượng từ nguồn chất thải này, chiếm đến 20% tổng nhu cầu năng lượng thế giới. Tuy nhiên, để khai thác hiệu quả nguồn sinh khối lignocellulose, cần có những phương pháp xử lý khá phức tạp [71].

1.2.1.2. Lignocellulose là nguồn tiềm năng sản xuất vật liệu, hóa chất mới

Chuyển hóa lignocellulose có thể tạo ra các hợp chất trung gian quan trọng như các monosaccharide (glucose, xylose), polyol (glycerol, arabitol, sorbitol), furan (furfural, 5-hydroxymethylfurfural) và axit hữu cơ (2,5-furandicarboxylic acid, levulinic acid, và lactic acid), trong đó furfural và 5-hydroxymethylfurfural có vai trò nổi bật là cầu nối giữa nguồn nguyên liệu sinh khối thô và ngành công nghiệp tinh lọc sinh học (biorefinery) [2].

Furfural có thể được chuyển hóa thành furfuryl alcohol, tetrahydrofurfuryl alcohol, 2-methylfuran, 2-methyltetrahydrofuran, tetrahydrofuran, 1,2-pentanediol, 1,5-pentanediol, cyclopentanone, furfuryl acetate, furanic ester, 1-octanol, dioctyl ether, và nhiên liệu lỏng. Furfuryl alcohol có vai trò quan trọng nhất là nguyên liệu sản xuất nhựa dẻo poly(furfuryl alcohol) từ đó sản xuất plastic, xi-măng, chất đệm, chất kết dính [72]. Nhựa poly(furfuryl alcohol) có ưu điểm là trơ với hóa chất và khó cháy. Tetrahydrofurfuryl alcohol được sử dụng như chất tẩy rửa, tẩy màu sơn,

dung môi thay thế cho các dung môi gốc chlorine, trong dung dịch sơn tan trong nước, để sản xuất dược phẩm và chất dẻo [72]. 2-Methylfuran được sử dụng làm chất tạo hương vị hoặc tá dược trong dược phẩm [72]. 2-Methyltetrahydrofuran là thành phần phản ứng sản xuất nhiều hóa chất, là dung môi hòa tan nhiều hợp chất hữu cơ chứa kim loại, và làm dung dịch điện giải trong pin lithium [72]. Tetrahydrofuran cũng là dung môi của nhiều loại polymer và chất dẻo bao gồm nhiều loại mực, sơn, chất kết dính, là hóa chất sử dụng trong ngành dệt may [72]. 1,2-Pentanediol là hóa chất trong ngành hóa mỹ phẩm, được sử dụng trong các sản phẩm tẩy rửa, làm sạch, bao gồm xử lý ô nhiễm không khí, cũng như sản phẩm mực và nước tẩy trang trong ngành trang điểm [72]. 1,5-Pentanediol là chất hóa dẻo và phụ gia giảm tốc; trong y học, chất này đã được xác định là có hoạt tính kháng khuẩn và kháng nấm [73]. Cyclopentanone là chất trung gian sản xuất sinh dược phẩm, thuốc trừ sâu và cao su [72]. Furfuryl acetate đã được Cục quản lý Thực phẩm và Dược phẩm Hoa Kỳ phê chuẩn là một chất phụ gia tạo vị [72]. 1-Octanol được sử dụng trong sản xuất nước hoa, mỹ phẩm, dầu gội, chất nhũ hóa, chất tẩy rửa, chất điều vị, chất chống tạo bọt [72]. Dioctyl ether cũng là một loại dung môi sử dụng trong sản xuất mỹ phẩm, bao gồm tẩy trang và dưỡng ẩm [72].

5-Hydroxymethylfurfural có thể được chuyển hóa thành 2,5-bis(hydroxymethyl)furan, 2,5-bis(hydroxymethyl)-tetrahydrofuran, 2,5-dimethylfuran, furfuryl alcohol, 1,6-hexanediol, 3-hydroxymethylcyclopentanone, 1-hydroxyhexane-2,5-dione, 1,2,6-hexanetriol, 3-(hydroxymethyl)cyclopentanol, furfuryl ether. 2,5-Bis(hydroxymethyl)furan có hoạt tính kháng nấm, có thể được sử dụng trong y dược. 2,5-Dimethylfuran được Cục quản lý Thực phẩm và Dược phẩm Hoa Kỳ phê chuẩn là một loại tá dược và chất tạo vị. 1,6-Hexanediol được sử dụng như dung môi trung gian sản xuất nylon và polyester, chất phủ, chất hóa dẻo, chất kết dính và trong công nghiệp dệt may [72].

1.2.2. Enzyme phân giải lignocellulose nói chung

Với cấu trúc phức tạp của thành tế bào thực vật, để phân giải hoàn toàn cấu trúc này cần phải trải qua 3 giai đoạn chính: (i) phân giải các hợp chất gốc pectin nằm trong phiến giữa đóng vai trò kết dính giữa các tế bào với nhau; (ii) tiền xử lý bằng tác nhân hóa, lý hoặc sinh học nhằm giải phóng hai thành phần chính là cellulose và hemicellulose ra khỏi lignin - chất kết dính làm cấu trúc lignocellulose bền vững; (iii) phân giải cellulose, hemicellulose và lignin thành các đơn phân cấu trúc.

1.2.2.1. Enzyme thủy phân pectin, lignin và các enzyme hỗ trợ khác

Trong tự nhiên, pectin bị phân hủy bởi một hỗn hợp các enzyme pectinase, trong đó bao gồm các pectate lyase [74] (họ PL1, PL2, PL3, PL9 theo CAZy), polygalacturonase [75], rhamnogalacturonase (họ glycoside hydrolase 28) [76], pectin methylesterase (họ carbohydrate esterase 8) [77] và pectin acetylerase (họ carbohydrate esterase 12) [78]. Cụ thể, pectin methylesterase và pectin acetylerase phân hủy liên kết este trong cấu trúc pectin, giúp cho enzyme polygalacturonase và pectate lyase có thể tiếp cận cơ chất để phân giải cấu trúc pectin thành các đơn phân [79]. Các enzyme phân hủy pectin được tiết bởi một loạt tác nhân gây bệnh thực vật, bao gồm cả vi khuẩn và nấm, đóng vai trò quan trọng gây nên độc lực của các tác nhân này.

Để giúp cho khối lignocellulose được phân giải một cách hiệu quả hơn, có một nhóm protein đóng vai trò hỗ trợ trong việc làm nới lỏng cấu trúc tinh thể, từ đó các enzyme phân hủy cellulose, hemicellulose và lignin có thể tiếp cận với cơ chất và xúc tác phản ứng phân hủy. Các protein này bao gồm expansin, swollenin, loosinin [80]. Expansin làm giảm độ bền vững của thành tế bào thông qua một cơ chế phụ thuộc pH, tuy nhiên bản thân protein không có hoạt tính thủy phân cấu trúc lignocellulose [81]. Loosening là một protein tương tự expansin, có hoạt tính nới lỏng vi sợi cellulose, dẫn đến xáo trộn cấu trúc tinh thể cellulose [82], được tìm thấy ở nấm mục trắng *Phanerochaete carnosus* [83]. Swollenin được tìm thấy ở chi nấm *Trichoderma* giúp tăng cường khả năng tiếp cận của các enzyme phân hủy cellulose với sợi cellulose bằng cách phá vỡ cấu trúc tương tự như expansin, nhưng đồng thời cũng có thể có hoạt tính endoglucanase, cellobiohydrolase, hoặc xylanase [84].

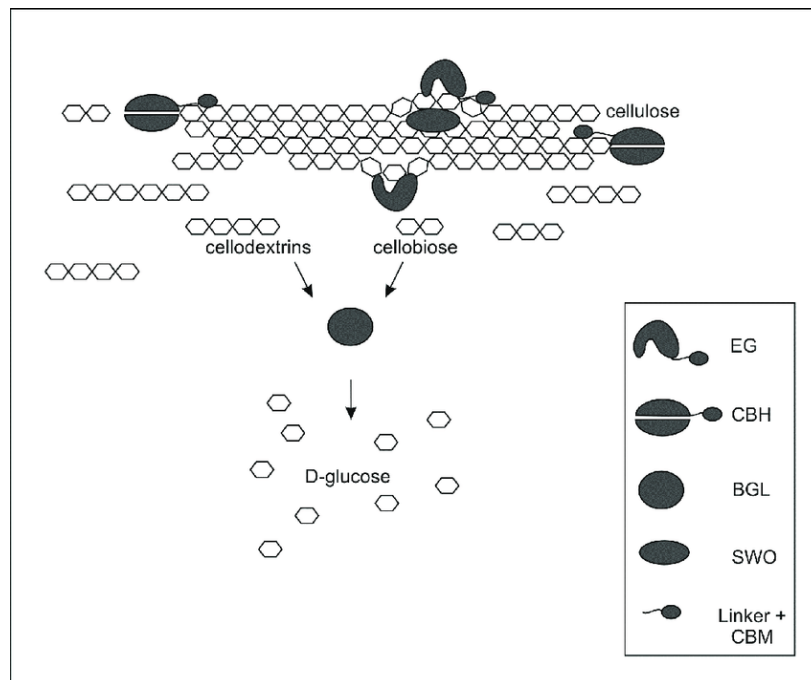
Cấu trúc phức tạp của lignin, cùng với khối lượng phân tử khá lớn và tính chất không tan khiến cho việc phân hủy lignin rất khó khăn. Những enzyme oxy hóa ngoại bào ít đặc hiệu có khả năng giải phóng những sản phẩm kém bền và dễ dàng tham gia các quá trình oxy hóa khác nhau, là những enzyme khởi đầu của quá trình thủy phân lignin. Hai họ enzyme chính liên quan đến quá trình thủy phân lignin là peroxidase và laccase [85]. Lignin peroxidase (LiP) và peroxidase phụ thuộc mangan (manganese-dependent peroxidase, MnP) là hai nhóm peroxidase đã được nghiên cứu khá kỹ. Enzyme peroxidase lý tưởng là enzyme có cả hai hoạt tính của LiP và MnP (versatile peroxidase, VP). Laccase là phenoloxidase xúc tác phản ứng oxy hóa các hợp chất phenolic và non-phenolic dưới sự có mặt của chất trung gian [86]. Nấm mục thủy phân gỗ là nguồn chính tạo ra laccase, tuy nhiên enzyme này cũng đã được phân lập từ nhiều loài nấm khác bao gồm *Aspergillus*, *Myceliophora thermophila* và *Chaetomium thermophilum*. Có những nghiên cứu gần đây cho thấy có những loài vi khuẩn cũng có khả năng tiết laccase [87].

1.2.2.2. Cellulase và hemicellulase

* Cellulase

Vi sinh vật thủy phân cellulose tạo ra một hỗn hợp các enzyme với đặc tính và cơ chất khác nhau, phối hợp với nhau để thủy phân hoàn toàn cấu trúc này. Về cơ bản, enzyme cellulase thủy phân liên kết β -1,4-glycoside trong cấu trúc phân tử cellulose. Enzyme cellulase được chia thành 2 loại là endoglucanase và cellobiohydrolase. Endoglucanase (endo-1,4- β -glucanase, EC 3.2.1.4) thủy phân liên kết nội phân tử (đặc biệt là trong vùng vô định hình) để tạo ra các đầu khử và không khử mới. Cellobiohydrolase (exo-1,4- β -glucanase, EC 3.2.1.91) hoạt động trên các đầu tận cùng của phân tử cơ chất cellulose. Cả hai loại enzyme này hoạt động tốt nhất với phân tử cellulose không nằm trong cấu trúc tinh thể, giải phóng sản phẩm là cellobiose.

Việc thủy phân hoàn toàn và hiệu quả cellulose còn cần thêm một loại enzyme nữa là β -glucosidase (EC 3.2.1.21), với khả năng phân cắt liên kết đôi trong phân tử cellobiose để tạo ra hai gốc glucose [88]. Để hoạt động bình thường, endoglucanase, exoglucanase và β -glucosidase cần phải giữ được cấu trúc ổn định trong môi trường ngoại bào và hình thành phức hệ cấu trúc không gian chính xác với cơ chất.



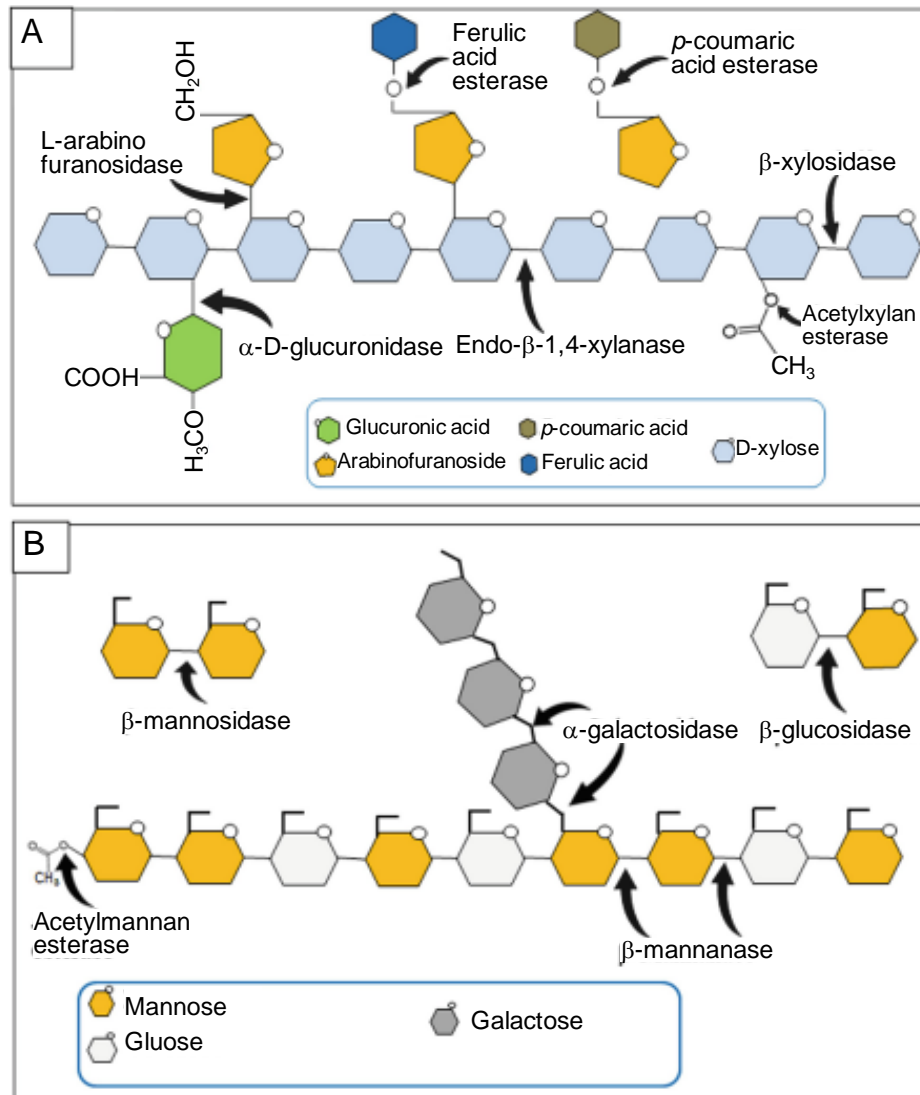
Hình 1. 2. Hệ enzyme phân hủy cellulose [89]

EG: endo-glucanase; CBH: cellobiose hydrolase; BGL: β -glucosidase; SWO: swollenin;
CBM: carbohydrate-binding module

* Hemicellulase

Như đã đề cập ở phần trên, cấu trúc của hemicellulose rất phức tạp và không đồng nhất, vì vậy để thủy phân hoàn toàn hemicellulose, cần sự phối hợp của rất nhiều loại enzyme [90].

Riêng để thủy phân xylan, cần sự có mặt của endo-1,4- β -xylanase (EC 3.2.1.8) phân cắt mạch dài của xylan tạo ra các oligosaccharide và xylan-1,4- β -xylosidase (EC 3.2.1.37) thủy phân xylan oligosaccharide tạo ra xylose. Tương tự với trường hợp mạch khung của phân tử hemicellulose không phải là xylan thì cần các enzyme tương ứng khác để phân hủy. Cụ thể, cần có sự có mặt của β -glucosidase và β -mannosidase (EC 3.2.1.25) để phân cắt mạch glucomannan, α -galactosidase (EC 3.2.1.22), β -glucosidase và β -mannosidase với mạch galactoglucomannan, α -arabinanase (EC 3.2.1.99) với mạch arabinan, α -arabinanase và α -galactosidase đối với mạch arabinogalactan.



Hình 1. 3. Hệ enzyme phân hủy hemicellulose [91]

A: enzyme phân hủy xylan; B: enzyme phân hủy mannan

Ngoài ra, quá trình phân hủy hemicellulose có thể cần các enzyme khác như acetyl xylan esterase (EC 3.1.1.72), acetyl mannan esterase (EC 3.1.1.6) phân giải liên kết este của các gốc acetyl với mạch xylan hay mannan, α -glucuronidase (EC 3.2.1.139) thủy phân liên kết với gốc α -D-glucoside, ferulic acid esterase (EC 3.1.1.73) và p -coumaric acid esterase (EC 3.1.1.10) thủy phân liên kết este giải phóng gốc acid tương ứng, α -L-arabinofuranosidase (EC 3.2.1.55) thủy phân các liên kết với gốc α -L-arabinofuranoside ở đầu không khử của chuỗi polysaccharide, tương tác bổ sung lẫn nhau để thủy phân hiệu quả [91].

Động vật sử dụng lignocellulose trong chế độ ăn tự nhiên thường luôn phụ thuộc vào khả năng tiết enzyme phân giải cellulose, hemicellulose do bản thân vật chủ thiếu hệ enzyme tương ứng để chuyển hóa những cấu trúc này. Vì vậy vật chủ và vi sinh vật đã hình thành mối quan hệ cộng sinh trong đó vi sinh vật khu trú trong đường tiêu hóa của vật chủ sẽ phân giải thức ăn cung cấp năng lượng cho toàn bộ vật chủ và bản thân vi sinh vật [92]. Hệ tiêu hóa của động vật nhai lại (trâu, bò, dê...) là nơi khu trú lý tưởng của vô số dạng sống bao gồm nấm, động vật nguyên sinh, vi khuẩn cổ và vi khuẩn trong một phức hệ có mối tương tác chặt chẽ lẫn nhau.

1.2.2.3. Vi khuẩn tham gia chuyển hóa lignocellulose trong dạ cỏ

Thành phần vi khuẩn khu trú trong dạ cỏ cực kỳ đa dạng, trong đó hầu hết các loài đều trực tiếp tham gia vào quá trình thủy phân sinh khối thực vật nói chung và thành tế bào thực vật nói riêng. Cụ thể, số lượng vi khuẩn trực tiếp liên quan đến quá trình thủy phân lignocellulose chiếm đến 50-75% tổng số quần thể vi khuẩn [93]. Dựa theo chức năng và đối tượng cơ chất chuyển hóa, vi khuẩn được chia thành vi khuẩn phân giải các cấu trúc sợi (cellulose, hemicellulose, pectin), lactic acid, tinh bột, chất béo, tannin, protein.... Trong dạ cỏ có những loài vi khuẩn thể hiện khả năng phân giải cellulose chủ yếu như *Fibrobacter succinogenes*, *Butyrivibrio fibrisolvens*, *Ruminococcus flavefaciens*, *Ruminococcus albus*, *Clostridium cellobioparum*, vi khuẩn có hoạt tính phân giải tinh bột như *Bacteriodes ruminicola*, *Ruminobacter amylophilus*, *Selenomonas ruminantium*, *Succinomonas amylolitica*, *Streptococcus bovis*, vi khuẩn phân giải lipid là *Anaerovibrio lipolytica*, vi khuẩn phân giải lactate như *Selenomonas lactilytica*, *Megashaera elsdonii*; bên cạnh đó loài *Lachnospira multiparus* là loài vi khuẩn được xác định có hoạt tính thủy phân pectin mạnh [94].

Chi vi khuẩn *Prevotella* thường được báo cáo là chi chiếm ưu thế của hệ vi sinh vật dạ cỏ [95]. Cụ thể, trong nghiên cứu toàn diện của Henderson và các cộng sự từ hơn 700 cá thể thuộc 32 loài động vật nhai lại từ 35 quốc gia khác nhau, các

đơn vị phân loại được xác định nhiều nhất tương ứng với các thành viên của các chi *Prevotella*, *Ruminococcus* và *Butyrivibrio*, cũng như các thành viên chưa được phân loại của các bộ Clostridiales và Bacteroidales và của các họ Ruminococcaceae và Lachnospiraceae [96]. Trong một nghiên cứu khác với 695 mẫu từ 8 vị trí trong 4 ngăn của dạ dày trâu cũng cho thấy rằng *Prevotella* nằm trong số các chi phổ biến nhất [97]. *Prevotella* cũng được chứng minh là một trong những chi phổ biến nhất trong một nghiên cứu tập hợp dữ liệu của 10.000 mẫu metagenome dạ cỏ [4]. Hơn nửa thế kỷ trước, Robert Hungate và các cộng sự đã đưa ra quan điểm rằng hệ vi sinh vật dạ cỏ bị ảnh hưởng nhiều bởi nguồn thức ăn hấp thụ, một quá trình chọn lọc được thúc đẩy bởi hiệu quả quá trình trao đổi chất của vi sinh vật, các vi sinh vật có khả năng cạnh tranh cao sẽ thay thế những vi sinh vật kém cạnh tranh hơn. Do đó, sự phong phú của chi *Prevotella* trong các nghiên cứu metagenomic trên một số loài động vật nhai lại có lẽ phản ánh hiệu quả chuyển hóa của chúng và sự thích nghi của chúng với môi trường dạ cỏ.

Prevotella đóng vai trò quan trọng trong quá trình chuyển hóa carbohydrate và nitrogen [98]. Tương tự như một số chi khác trong ngành vi khuẩn Bacteroidetes và nói chung là ở vi khuẩn phân hủy lignocellulose, hệ gene của *Prevotella* có các vùng PUL là các cụm gene mã hóa các protein chuyên phân giải carbohydrate phức tạp [99]. *Prevotella* có thể phân hủy nhiều loại polysaccharide và có khả năng tổng hợp propionate, là tiền chất quan trọng nhất cho quá trình tổng hợp phân tử đường glucose trong gan của động vật nhai lại [100]. Chi *Prevotella* thể hiện vai trò thiết yếu trong duy trì cân bằng hệ vi sinh đường ruột, giúp duy trì sức khỏe và hạn chế bệnh đối với hệ tiêu hóa của người [101]. *Prevotella* cũng có vai trò trong quá trình chuyển hóa hydro, giảm phát sinh methane trong quá trình tiêu hóa thức ăn lignocellulose của dạ cỏ động vật nhai lại [102].

1.2.3. Xylanase

1.2.3.1. Vai trò của xylanase

Do cấu trúc hóa học phức tạp và không đồng nhất của xylan tự nhiên, để thủy phân hoàn toàn xylan cần có sự phối hợp của một hỗn hợp các enzyme tương hỗ lẫn nhau. Hệ enzyme phân giải xylan bao gồm β -1,4-endoxylanase, β -xylosidase, α -L-arabinofuranosidase, α -glucuronidase, acetyl xylan esterase, phenolic acid (ferulic và p-coumaric acid) esterase. Thông thường, hệ enzyme phân giải xylan được tiết ra đồng thời bởi những chủng có khả năng phân hủy xylan, bao gồm vi khuẩn [103], nấm [104], xạ khuẩn [105]. “Xylanosome” là một cấu trúc tập hợp đa enzyme có nhiều chức năng, được tìm thấy trên bề mặt của một số vi sinh vật ví dụ như phức

hệ xylanosome tìm thấy ở *Cellulosimicrobium funkei* và *Cellulosimicrobium terreum* phân giải hemicellulose một cách hiệu quả [106]. Phức hệ xylanosome từ *Clostridium papyrosolvens* C7 có khối lượng trong khoảng 500-660 kDa với 7 tiểu phân protein [107].

Trong số những enzyme thủy phân xylan, xylanase (β -1,4-endo-xylanase) đóng vai trò quan trọng bậc nhất vì enzyme giữ chức năng xúc tác phân cắt mạch chính của phân tử xylan, tạo thành các xylooligosaccharide kích thước ngắn [108], từ đó tạo điều kiện để các enzyme khác tiếp tục thủy phân để giải phóng các đơn phân. Các xylanase trong phức hệ xylanosome thường có tính đặc hiệu cơ chất khác nhau nhưng trùng lặp lên nhau, tương tác, hỗ trợ nhau giúp vi sinh vật có thể phân hủy xylan một cách hiệu quả nhất. Trong quá trình thủy phân, tương tác giữa các enzyme trong phức hệ sẽ giải phóng acetic acid nhờ hoạt động của acetyl xylan esterase giúp làm tăng hoạt tính của endo-xylanase, đồng thời hoạt động của endo-xylanase tạo ra nhiều đoạn nhỏ oligosaccharide chứa gốc acetyl, giúp hoạt động của enzyme esterase hiệu quả hơn nữa. β -xylosidase tăng cường hoạt động của endo-xylanase bằng cách giải phóng ức chế ở các đầu tận cùng mạch xylan, tương tự, hoạt động của α -arabinofuranosidase giúp tăng cường hoạt động của endo-xylanase trong phân cắt mạch arabinoxylan [91].

1.2.3.2. Phân loại xylanase

Dựa trên khối lượng phân tử và điểm đẳng điện, xylanase được chia thành hai nhóm đó là nhóm cao phân tử (khối lượng lớn hơn 30 kDa) nhưng có điểm đẳng điện thấp (pI acid) và nhóm có khối lượng phân tử thấp (nhỏ hơn 30 kDa) nhưng có điểm đẳng điện cao (pI kiềm). Tuy nhiên có một số trường hợp enzyme có hoạt tính xylanase không phù hợp với cách phân loại này [109], vì vậy, một cách phân loại chi tiết và chính xác hơn dựa trên cả yếu tố cấu trúc tinh thể và vùng trung tâm xúc tác cùng với các đặc tính động học xúc tác, đặc hiệu cơ chất và sản phẩm đã được đề xuất [110]. Cách phân loại mới này phân chia enzyme xylanase về các họ glycoside hydrolase (GH) theo cơ sở dữ liệu CAZy.

Những họ GH có thể hiện hoạt tính xylanase là các họ GH 5, 7, 8, 9, 10, 11, 12, 16, 26, 30, 43, 44, 51, 62. Các enzyme thuộc họ GH 9, 12, 26, 30, 44 chỉ thể hiện hoạt tính xylanase như là hoạt tính thứ cấp. Dựa trên phân tích cấu trúc và trình tự của vùng trung tâm hoạt động, xylanase thuộc hai họ GH10 và GH11 đã được phân tích chi tiết, trong khi thông tin liên quan đến xylanase thuộc các họ còn lại vẫn còn khá hạn chế [108]. Theo cơ sở dữ liệu CAZy, những họ GH quan trọng nhất có hoạt tính xylanase là GH 5, 7, 8, 10, 11 và 43 [111].

Cách thức thủy phân xylan của xylanase với các họ GH được thực hiện theo hai cơ chế: cơ chế giữ nguyên hoặc cơ chế đảo ngược. Với cơ chế giữ nguyên, gốc amino acid glutamate của enzyme thủy phân đóng vai trò rất quan trọng. Cơ chất của phản ứng ở dạng β -glycoside sẽ hình thành sản phẩm trung gian dạng α với enzyme xúc tác trước khi được giải phóng ở dạng β ở sản phẩm cuối cùng [112]. Các họ GH 5, 7, 10 và 11 hầu hết đều thủy phân theo cơ chế này. Ngược lại, với cơ chế đảo ngược, sản phẩm thủy phân cuối cùng có sự thay đổi về cấu trúc không gian so với cơ chất ban đầu, từ dạng β thành dạng α và ngược lại; các amino acid có vòng thơm như glutamate và acid aspartic đóng vai trò là các gốc xúc tác trọng yếu [113]. Enzyme thuộc họ GH 8 và 43 thủy phân theo cơ chế này.

Enzyme xylanase thuộc họ GH 10 thường là các protein có khối lượng phân tử lớn với điểm đẳng điện acid và có cấu trúc không gian dạng hộp hình thành từ 8 phiến α/β [114]. Các xylanase trong họ GH10 có tính đặc hiệu cơ chất tương đối thấp, có thể xúc tác thủy phân một phổ cơ chất rộng hơn so với xylanase của họ GH11 [115]. Xylanase GH10 thủy phân những liên kết gắn với các gốc xylose gần với mạch bên của phân tử xylan [116]. Mức độ biến đổi của mạch bên phân tử xylan ảnh hưởng đến sự tương tác cấu trúc của enzyme với cơ chất, dẫn đến thay đổi khả năng phân giải xylan của các loại xylanase.

1.2.3.3. Xylanase trong tự nhiên

Xylanase rất phổ biến trong tự nhiên, sự có mặt của enzyme này đã được phát hiện ở rất nhiều loài sinh vật, bao gồm các vi khuẩn khu trú trong đại dương, trong đất và trong dạ cỏ của động vật [117], nấm ưa nhiệt độ cao [118] và trung bình [119], thực vật bậc cao [120], động vật nguyên sinh [121], động vật ngành Thân mềm như sên có vỏ [122], côn trùng [123], tảo [124]. Xylanase từ vi khuẩn và nấm đã và đang được nghiên cứu và ứng dụng rộng rãi trong nhiều ngành công nghiệp.

Trong các loài vi khuẩn, chi *Bacillus* đã được xác định là nguồn sinh enzyme phân giải xylan quan trọng nhất, cụ thể là *Bacillus* sp., *B. halodurans* [125], *B. pumilus* [126], *B. subtilis* [127], *B. amyloliquefaciens*, *B. circulans*, *B. stearothermophilus* [117]. Xylanase với khả năng chịu nhiệt cao (60-70°C) đã được phân lập từ chủng *B. pumilus*, *B. halodurans* TSEV1 [128], *Clostridium thermocellum* [129], *Rhodothermus marinus* [130], *Streptomyces* sp. [131], *Stenotrophomonas maltophila* [132], *Thermotoga thermarum* [133]. Ngược lại, xylanase có thể chịu lạnh tuy không quá phổ biến nhưng đã được phân lập từ một số chủng vi khuẩn như *Clostridium* sp. PXYLY1 [134], *Flavobacterium* sp. MSY-2 [135], *Pseudoalteromonas haloplanktis* [136], *Flavobacterium frigidarium* [137].

Một vài xylanase chịu kiềm đã được phân lập từ vi khuẩn thuộc ngành Firmicutes như *B. pumilus*, *B. halodurans* TSEV1, *Geobacillus thermoleovorans* [138], hoặc xạ khuẩn như *Actinomadura* sp. Cpt20, *Streptomyces althioticus* LMZM [139].

1.2.3.4. Ứng dụng của xylanase

Các enzyme phân hủy xylan nói chung và xylanase nói riêng có tiềm năng rất lớn để ứng dụng trong rất nhiều ngành công nghiệp, ví dụ như công nghiệp thực phẩm, thức ăn chăn nuôi, sản xuất gỗ và giấy. Ngay từ khoảng những năm 1980, xylanase đã bắt đầu được ứng dụng trong ngành sản xuất thức ăn chăn nuôi. Xylanase được sử dụng trong quá trình xử lý nguyên liệu lignocellulose để làm tăng khả năng tiêu hóa thức ăn [140]. Xylanase được bổ sung vào bã rơm để giúp phân giải những thành phần chất xơ phức tạp, giúp làm tăng giá trị dinh dưỡng của thành phần này khi sử dụng làm thức ăn chăn nuôi [141]. Xylanase tăng tỷ lệ chuyển hóa thức ăn của vật nuôi, giúp làm tăng khả năng tăng trọng, đặc biệt đối với động vật có dạ dày đơn [142].

Trong công nghiệp thực phẩm, xylanase được sử dụng trong quá trình lên men bánh mì. Cụ thể, xylanase phân hủy arabinoxylan trong khối bột, giúp phân bố đều nước trong quá trình nhào bột [143], cải thiện sự hình thành mạng gluten, dẫn đến tăng độ mềm xốp, độ dẻo, độ đàn hồi, thể tích khối bột, độ cứng và giòn của sản phẩm [144]. Sản phẩm thủy phân của arabinoxylan, cụ thể là các arabinoxyloligosaccharide trong khối bánh mì có nhiều lợi ích với sức khỏe [145]. Ngoài ra, xylanase cũng được sử dụng trong quá trình xử lý sản phẩm nước ép hoa quả, giúp làm giảm độ nhớt của dịch ép, hạn chế hình thành vón cục, tăng độ trong suốt, mùi thơm và màu sắc của sản phẩm [146]. Xylanase từ *Streptomyces* sp. được sử dụng trong dịch ép nước cam, chanh ngọt, dứa [147]. Xylanase từ *B. pumilus* VLK-1 được sử dụng trong nước ép cam và nho [148]. Xylanase từ *Pediococcus acidilactici* GC25 có tác dụng làm tăng lượng đường cũng như làm giảm độ đục của dịch chiết các loại quả [149].

Trong công nghiệp sản xuất giấy, xylanase được sử dụng để làm trắng bột giấy. Cụ thể, xylanase được sử dụng trong quá trình tiền xử lý để làm tăng hiệu quả của hóa chất, bản thân xylanase không trực tiếp tác động thủy phân các gốc cấu thành lignin mà xylanase thủy phân mạng xylan mà lignin bao quanh. Việc thủy phân ở mức độ nhất định mạng xylan giúp tăng hiệu quả tiếp xúc và tác động của các tác nhân hóa học trong các loại hóa chất tẩy trắng. Trong nghiên cứu của Nagar và cộng sự năm 2021, xylanase bền nhiệt và chịu kiềm có nguồn gốc từ *B. pumilus* SV-205 được biểu hiện lượng lớn và sử dụng trong quá trình tẩy trắng bột giấy, với

kết quả thu được làm tăng độ sáng 1,08 lần so với đối chứng, giảm 19,01% lượng chlorine được sử dụng để thu được bột giấy có độ sáng tương đương [150]. Quá trình làm giấy và xử lý bột giấy diễn ra ở điều kiện phổ pH và nhiệt độ rất rộng, vì vậy những enzyme xylanase có khả năng chịu kiềm và chịu nhiệt đóng vai trò rất quan trọng trong quá trình sản xuất. Chủng *Bacillus* chịu kiềm có khả năng tiết enzyme xylanase hoạt động ở nhiệt độ cao (60°C) và pH cao (6-10) đã được ứng dụng trong quá trình tẩy trắng bột gỗ kraft để làm bìa [151]. Xylanase từ *Scytalidium thermophilum* hoạt động ở nhiệt độ 50-70°C được sử dụng để tẩy trắng bột bã mía [152]. Mặt khác, quá trình tẩy mực từ các sản phẩm giấy đã qua sử dụng là bước rất quan trọng trong quá trình tái sử dụng và tái chế. Những cách xử lý bằng hóa học như sử dụng các chất tẩy rửa gốc ClO⁻, NaOH, NaCO₃, H₂O₂, Na₂SiO₂ thường gây tác động lớn đến môi trường [153] nên những biện pháp xử lý bằng enzyme với ưu điểm thân thiện với môi trường hơn đang được nghiên cứu thay thế [154].

Xylanase cũng được sử dụng trong xử lý sản phẩm sợi bông, làm tăng độ bền chắc của sản phẩm [155] và làm trắng vải thay vì sử dụng biện pháp hóa học ảnh hưởng đến môi trường [156]. Các chủng *B. stearothermophilus* SDX, *B. pumilus* ASH [157], *Trichoderma longibrachiatum* KT693225 [158] là những chủng sản sinh xylanase được sử dụng trong quá trình xử lý vải.

Sản phẩm phân giải xylan của xylanase, các xylo-oligosaccharide có nhiều ứng dụng trong công nghệ dược phẩm. Các xylo-oligosaccharide không được hấp thụ và chuyển hóa bởi hệ tiêu hóa của động vật, vì vậy chúng đóng vai trò quan trọng như là nguồn prebiotic – nguồn dinh dưỡng của probiotic, những vi sinh vật có lợi cho tiêu hóa của động vật và người [159]. Xylo-oligosaccharide còn có tiềm năng ứng dụng trong việc điều hòa hệ miễn dịch [160], kháng ung thư [161], kháng khuẩn, chống oxy hóa [162], chống dị ứng, chống viêm [163].

1.2.3.5. Tình hình biểu hiện xylanase trong nước và trên thế giới

Xylanase là một trong những enzyme phân giải lignocellulose được nghiên cứu và biểu hiện nhiều trong các nghiên cứu của các nhà khoa học ở khắp nơi trên thế giới. Xylanase được khai thác từ nhiều nguồn khác nhau bao gồm có nấm mốc *Aspergillus niger* [164], *A. oryzae* [165], nấm sợi *Cryptovalsa mangrovei* [166], *Myceliophthora thermophila* [167], côn trùng như loài bọ thuộc họ Xén tóc *Apriona japonica* [168], một đục quả cà phê *Hypothenemus hampei* [169], động vật không xương sống như ốc sên *Ampullaria crossean* [170], thực vật bậc cao như hạt đại mạch *Hordeum vulgare* nảy mầm [171], động vật nguyên sinh *Polyplastron*

multivesiculatum [172]. Xylanase được khai thác từ rất nhiều loài vi khuẩn thuộc các chi *Acidobacterium*, *Bacillus*, *Bacteroides*, *Caldicellulosiruptor*, *Cellulomonas*, *Cellulosimicrobium*, *Clostridium*, *Demequina*, *Dictyoglomus*, *Fibrobacter*, *Geobacillus*, *Kocuria*, *Lechevalieria*, *Paenibacillus*, *Paraglaciecola*, *Planococcus*, *Ruminiclostridium*, *Sorangium*, *Sphingobacterium*, *Streptomyces*, *Thermobacillus*, *Thermotoga*, trong đó các chi *Bacillus*, *Geobacillus*, *Paenibacillus*, *Streptomyces* là các chi được nghiên cứu và khai thác nhiều. Xylanase đã được biểu hiện trực tiếp từ chủng tự nhiên hoặc biểu hiện tái tổ hợp trong một số vật chủ phổ biến như vi khuẩn *Escherichia coli*, nấm men *Pichia pastoris*, nấm mốc *Aspergillus oryzae*. Xylanase thương mại đã được sản xuất và phân phối bởi hãng Sigma-Aldrich/Merck, trong đó xylanase có nguồn gốc từ *Thermomyces lanuginosus* được biểu hiện tái tổ hợp trong *Aspergillus oryzae*, sản phẩm xylanase tinh sạch tồn tại ở dạng bột với hoạt tính ít nhất là 2500 U/g. Sản phẩm được bán với giá 234 SGD/10 g hoặc 561 SGD/ 50 g (thông tin truy cập trên trang chủ của hãng Merck vào ngày 22 tháng 5 năm 2023).

Tại Việt Nam, một số nghiên cứu liên quan đến biểu hiện xylanase đã được tiến hành tuy nhiên vẫn còn hạn chế về số lượng. Xylanase tái tổ hợp nguồn gốc từ chủng *Aspergillus oryzae* VTCC-F187 trong *A. niger* VTCC-F017 đã được biểu hiện bởi nhóm nghiên cứu của PGS. TS. Đỗ Thị Tuyên nhằm tạo ra xylanase có hoạt tính cao [165]. Xylanase tái tổ hợp có khối lượng 21 kDa với hoạt tính riêng là 1025 IU/mg đối với cơ chất xylan có nồng độ 0,5% (khối lượng/thể tích). Trong một nghiên cứu khác từ nhóm nghiên cứu của PGS.TS. Vũ Nguyên Thành, gene xylanase có nguồn gốc từ chủng nấm *Bispora* sp. MEY-1 được biểu hiện tái tổ hợp trong nấm men *P. pastoris* với hoạt tính 96 IU/ml dịch nuôi cấy [173]. Nghiên cứu của Trần Liên Hà và cộng sự đã tách dòng gene xylanase từ *Aspergillus niger* C1 và biểu hiện tái tổ hợp trong *E. coli* BL21, thu được dịch enzyme thô có hoạt tính 381 U/ml [174].

CHƯƠNG 2. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Đối tượng, vật liệu hóa chất và thiết bị máy móc

2.1.1. Đối tượng và vật liệu nghiên cứu

Nghiên cứu này được thực hiện ở phòng thí nghiệm của Viện Công nghệ sinh học – Viện Hàn lâm Khoa học và Công nghệ Việt Nam, với các hóa chất và thiết bị uy tín.

✓ **Đối tượng nghiên cứu:** Mẫu dạ cỏ của 3 dê Cỏ và 2 dê Bách Thảo thu thập tại tỉnh Ninh Bình (tọa độ GPS 20.269002 105.893267), 2 dê Cỏ, và 3 Bách Thảo thu tại Thanh Hóa (tọa độ GPS 19.897450 105.795899). Dê được lựa chọn là dê ăn cỏ, lá cây và cành cây trên núi vào ban ngày, ban đêm cho ăn thêm phế phụ phẩm nông nghiệp khác nhau, không cho ăn cám.

✓ **Các chủng vi sinh vật:** Chủng vi khuẩn *E. coli* DH10B (F⁻ *mcrA* Δ (*mrr-hsdRMS-mcrBC*) Φ 80*lacZ* Δ M15 Δ *lacX74* *recA1* *endA1* *araD139* Δ (*ara leu*) 7697 *galU* *galK* *rpsL* *nupG* λ -) của hãng Invitrogen (Mỹ) được sử dụng làm thể nhận trong thí nghiệm tách dòng gene; chủng *E. coli* BL21(DE3) (F⁻*ompT* *hsdS_B*(*r_B⁻m_B⁻*) *gal* *dcm* (DE3)), Rosetta1 (F⁻ *ompT* *gal* *dcm* *lon* *hsdS_B*(*r_B⁻m_B⁻*) λ (DE3 [*lacI* *lacUV5-T7p07* *ind1* *sam7* *nin5*]) [*malB⁺*]_{K-12}(λ^S)), JM109 (*endA1* *glnV44* *thi-1* *relA1* *gyrA96* *recA1* *mcrB⁺* Δ (*lac-proAB*) *e14-* [F' *traD36* *proAB⁺* *lacI^q* *lacZ* Δ M15] *hsdR17*(*r_K⁻m_K⁺*)), SoluBL21 (BL21 Soluble), Origami (F⁻ *ompT* *gal* *dcm* *lon* *hsdS_B*(*r_B⁻m_B⁻*) *trxB* *gor* λ (DE3 [*lacI* *lacUV5-T7p07* *ind1* *sam7* *nin5*]) [*malB⁺*]_{K-12}(λ^S) được sử dụng làm thể nhận để biểu hiện gene.

✓ **Plasmid:** pET22b(+) được sử dụng làm vector biểu hiện gene (Thermo Scientific, Mỹ).

2.1.2. Một số dung dịch hóa chất chính:

* Các dung dịch sử dụng trong tách chiết DNA plasmid từ *E. coli*:

Dung dịch I (50 mM glucose, 25 mM Tris-HCl, pH 8,0; 10 mM EDTA, pH 8,0);
Dung dịch II (0,2 N sodium hydroxide, 1% SDS); Dung dịch III (3 M potassium acetate, 11,5% acetic acid); Dung dịch phenol: chloroform: isoamylalcohol tỷ lệ theo thể tích tương ứng là 25: 24: 1.

* Dung dịch sử dụng trong điện di DNA:

Dung dịch TAE 50X (Tris-base 2 M; acetic acid 1 M; EDTA 0,05 M, pH 8,0); Dung dịch nhuộm gel: ethidium bromide 0,5 µg/ml.

** Dung dịch sử dụng trong điện di protein:*

Đệm xử lý mẫu protein 6X (Tris-HCl 0,64 M, pH 6,8; glycerol 20%; SDS 10%; 2-mercaptoethanol 6%; bromophenol 0,12%); Đệm chạy điện di (Tris-HCl 50 mM pH 8,4; glycine 192 mM; SDS 0,1%); Dung dịch nhuộm gel polyacrylamide (coomassie brilliant blue 0,1%; methanol 30%; acetic acid 10%); Dung dịch tẩy nhuộm coomassie (methanol 40%; acetic acid 10%).

** Dung dịch sử dụng để tinh chế protein:*

Dung dịch cân bằng giá thể (Tris-HCl 20 mM, NaCl 100 mM, imidazole 10 mM); Dung dịch rửa mẫu (Tris-HCl 20 mM, NaCl 100 mM, 50-100 mM imidazole); Dung dịch thu mẫu (Tris-HCl 20 mM, NaCl 100 mM, 250 mM imidazole).

** Dung dịch đo hàm lượng protein:* Dung dịch Bradford reagent (Sigma), BSA chuẩn (Sigma).

** Dung dịch thử hoạt tính enzyme:* Dung dịch PBS 10X (NaCl 1,37 M, KCl 27 mM, Na₂HPO₄ 100 mM, K₂HPO₄ 20 mM); Dung dịch Tris-HCl 1 M; Dung dịch đệm citrate 1 M; Dung dịch xylan 1% (khối lượng/thể tích); Dung dịch pNPG/pNPX (Sigma) 100 mM; Dung dịch CMC (Trung Quốc) 1%; Dung dịch Congo red 0,1% (khối lượng/thể tích); Dung dịch màu DNS (DNS 1%; NaOH 1%; K-Na-tartrate 18,2%) và một số dung môi hữu cơ khác.

** Một số bộ kit được sử dụng:* DNeasy Blood & Tissue và QIAquick Spin; QIAquick gel (Qiagen); cột PD-10 loại muối (Amersham Biosciences, Mỹ), PSP®Spin Stool DNA (Invitex GmbH); QIAamp® DNA Stool Mini Kit (Qiagen) dùng cho tách chiết, tinh chế DNA metagenome.

** Môi trường nuôi cấy:*

Môi trường LB lỏng (bacto-trypton 1%; cao nấm men 0,5%; NaCl 1%) được pha trong nước lọc RO; Môi trường LB đặc: Môi trường LB lỏng có bổ sung thêm 1,6% agar; Môi trường PE: Bao gồm 2% bacto-peptone, 1% cao nấm men pha trong nước lọc RO; Môi trường TB: 1,2% bacto tryptone, 0,5% cao nấm men, 0,4% glycerol, 72 mM K₂HPO₄, 17 mM KH₂PO₄, pH 7 - 7,2; Môi trường SB: 2% Bacto™ yeast extract, 3% Bacto™ tryptone, 1% MOPS; Môi trường YT: 1% Bacto™ yeast extract, 1,6% Bacto™ tryptone, 0,5% NaCl; Môi trường SOB: 0,5% Bacto™ yeast extract, 2% Bacto™ tryptone, 10 mM NaCl, 2,5 mM KCl, 20 mM MgSO₄.

Các môi trường sau khi pha được khử trùng ở 121°C trong 30 phút, để nguội

và bổ sung ampicillin đến nồng độ cuối cùng là 100 µg/ml.

2.1.3. Máy móc và thiết bị

Một số máy móc và thiết bị chính sử dụng như máy quang phổ UV 1650C (Shimadzu, Nhật Bản), máy giải trình tự thế hệ mới Hiseq 2000 (Illumina, Mỹ), máy PCR (Applied Biosystems, Mỹ), máy điện di (Bio-Rad, Mỹ), cân phân tích (Precisa, Thụy Sĩ), cân điện tử (Precisa, Thụy Sĩ), máy đo nồng độ quang học nanophotometer (Implen, Đức), máy đo pH (Metler, Thụy Sĩ), tủ nuôi cấy tế bào (Multitron, Đức), máy sắc kí FPLC (Amersham Pharmacia, Mỹ), máy UV (Bio-Rad, Mỹ), máy ly tâm lạnh (Sorvall RC5B, Mỹ và Sorvall RC5B, Mỹ), máy lắc ổn nhiệt (New Jersey, Mỹ), máy hút chân không – speed Vac Sc 110^a (Savant, Mỹ), cột sắc kí ái lực Hitrap (GE, Healthcare, Thụy Điển), tủ lạnh sâu -80°C (Panasonic, Nhật), máy đọc ELISA ELx800 (BioTek, Mỹ).

2.2. Phương pháp nghiên cứu

2.2.1. Các phương pháp sinh học phân tử, vi sinh vật

2.2.1.1. Tách chiết, tinh chế DNA metagenome

Dịch dạ cỏ dê mỗi con dê được lọc qua 4 lớp vải màn. Phần cặn không qua lớp vải được hòa vào 2 lit dung dịch PBS (137 mM NaCl, 2,0 mM KH₂PO₄, 10 mM Na₂HPO₄ và 2,7 mM KCl, pH 7,4) sau đó lọc lại bằng 4 lớp vải. Dịch lỏng thu được được dồn lại và ly tâm ở tốc độ 700 vòng / phút (v/p) (khoảng 150-200 xg) trong 10 phút để tách riêng protozoa và xác, bã thực vật còn sót lại. Bước này được lặp lại hai lần. Các vi khuẩn trong dịch nổi được thu lại bằng ly tâm 4500 xg trong 5 phút, rửa hai lần với dung dịch PBS, và huyền phù vào 500 ml dung dịch PBS pH 7,4 có chứa 20% glycerol và giữ trong -80°C. Quy trình tách chiết DNA metagenome được dựa trên phương pháp của Sambrook và cộng sự [175]. DNA metagenome của vi khuẩn được tách chiết bằng phương pháp tách chiết DNA vi khuẩn (để loại trừ DNA của nấm và eukaryot khác) cải biến theo phương pháp của Sambrook và cộng sự (2001) sau đó được tinh chế bằng bộ sinh phẩm PSP Spin stool DNA kit (Strattec, Đức) theo hướng dẫn của nhà sản xuất.

Hệ vi khuẩn trong dạ cỏ dê có tính đặc thù theo hệ gene, theo cá thể, theo vị trí địa lý. Vì vậy, với mục đích khai thác gene mã hóa enzyme thủy phân lignocellulose và tìm hiểu vai trò của vi khuẩn tham gia chuyển hóa lignocellulose trong dạ cỏ dê, yếu tố dinh dưỡng là yếu tố đóng vai trò quan trọng nhất. Cùng với giá thành giải trình tự, thời gian cũng như nguồn lực cho phân tích bộ dữ liệu lớn, việc giải trình tự DNA metagenome của từng cá thể dê hoặc từng nhóm dê ở thời

điểm thực hiện luận án gặp nhiều khó khăn. Vì vậy mẫu DNA metagenome đã tinh sạch có hàm lượng như nhau từ 10 con dế được trộn chung thành một mẫu DNA metagenome của vi khuẩn dạ cỏ dế Việt Nam. Mẫu này được gửi sang công ty BGI Hồng Kông để giải mã tạo hai bộ dữ liệu là bộ dữ liệu có dung lượng khoảng 8-10 Gb (bộ dữ liệu thông thường - conventional metagenomic data) và bộ dữ liệu có dung lượng 40-50 Gb (bộ dữ liệu giải mã sâu - deep metagenomic sequencing data). Bộ dữ liệu thông thường được giải mã trước và nhận thấy độ bao phủ gene trong dữ liệu này chỉ đạt 27%. Vì vậy chúng tôi đã tiến hành giải mã sâu để hy vọng có được bộ dữ liệu tốt cho phân tích đa dạng cũng như khai thác gen.

2.2.1.2. Phương pháp giải trình tự đa hệ gene

Mẫu DNA metagenome sau khi tách chiết được kiểm tra để đảm bảo đạt yêu cầu về độ sạch, nồng độ DNA và không có mặt của chất ức chế phản ứng PCR. Mẫu đạt yêu cầu được gửi sang công ty BGI (Hồng Kông, Trung Quốc) để giải trình tự đa hệ gene trên hệ thống Illumina [176]. Về cơ bản, nguyên lý của phương pháp giải trình tự này tương tự như giải trình tự DNA theo phương pháp Sanger, trong đó DNA polymerase tổng hợp chuỗi DNA hình thành bằng cách sử dụng dNTP gắn vào đầu 3' của chuỗi DNA đang tổng hợp theo nguyên tắc bổ sung. Tuy nhiên, đối với giải trình tự thế hệ mới, thay vì giải trình tự một đoạn đơn lẻ, kỹ thuật này cho phép giải trình tự với một lượng lớn các đoạn DNA khác nhau song song tại cùng một thời điểm, từ đó tiết kiệm thời gian và cho lượng dữ liệu đầu ra vô cùng lớn so với phương pháp Sanger cũ.

2.2.1.3. Tổng hợp gene, thiết kế vector biểu hiện mang gene *exl*

Sau khi được phân tích lựa chọn bằng các phần mềm tin sinh học, gene mã hóa endo-xylanase được nghiên cứu tối ưu mã bộ ba để biểu hiện trong vi khuẩn *E. coli*, và loại bỏ trình tự tín hiệu tiết tự thân của gene. Sau đó, gene được tổng hợp nhân tạo tại công ty Genscript (Mỹ). Gene phân lập và tổng hợp được đưa vào vector pET22b(+) dựa trên điểm cắt của cặp enzyme giới hạn *NcoI* và *XhoI*. Các phương pháp cắt, nối gene được thực hiện theo phương pháp của Sambrook và đồng tác giả [177].

2.2.1.4. Biến nạp DNA plasmid vào vi khuẩn *E. coli*

Tế bào *E. coli* khả biến được chuẩn bị và biến nạp theo phương pháp sốc nhiệt của Froger và Hall [178]. DNA plasmid (1 μ l) được trộn đều với 50 μ l tế bào khả biến (sử dụng cho sốc nhiệt) đã được lấy từ tủ bảo quản -80°C , để sẵn trên đá vụn trong 30 phút. Sau đó mẫu được ủ ở 42°C trong 90 giây, tiếp theo để trên đá

trong 2 phút trước khi bổ sung 500 μ l môi trường LB, nuôi lắc 180 vòng/phút (v/p) ở 37°C trong 45 phút. Dịch tế bào được cấy trải trên đĩa môi trường LB có bổ sung 100 μ g/ml (môi trường LBA) đặc, ủ ở 37°C trong 16 giờ. Khuẩn lạc mọc trên đĩa là chủng mang plasmid.

2.2.1.5. Tách chiết DNA plasmid từ tế bào vi khuẩn *E. coli*

DNA plasmid trong tế bào *E. coli* được tách chiết theo phương pháp của Sambrook và cộng sự [179]. Mẫu DNA plasmid sau khi tách chiết được kiểm tra bằng điện di trên gel agarose 0,8%.

2.2.1.6. Điện di trên gel agarose

Kích thước đoạn DNA và plasmid được kiểm tra bằng phương pháp điện di trên gel agarose theo Sambrook và cộng sự [180]. Gel agarose nồng độ thích hợp được chuẩn bị trong khuôn chuyên dụng. Sau khi gel thành hình, các mẫu DNA được trộn với đệm xử lý mẫu và được nhỏ vào các giếng. Toàn bộ bản gel ngâm trong đệm điện di được đặt trong điện trường của dòng điện một chiều 100 V trong thời gian thích hợp để phân tách các đoạn DNA. Kết thúc điện di, dung dịch ethidium bromide nồng độ 0,5 μ g/ml được sử dụng để nhuộm toàn bộ bản gel trong khoảng 10 phút. Băng DNA trên gel được quan sát khi được kích thích bằng tia tử ngoại ở bước sóng 302 nm.

2.2.1.7. Tinh chế DNA từ gel agarose

Sản phẩm cắt DNA từ gel agarose hoặc sản phẩm PCR được tinh chế theo các bước chỉ dẫn của bộ sinh phẩm tinh sạch DNA QIAgene - QIAquick Gel Extraction Kit.

2.2.2. Các phương pháp hóa sinh protein

2.2.2.1. Biểu hiện protein tái tổ hợp trong *E. coli*

Quy trình biểu hiện protein tái tổ hợp được thực hiện dựa trên quy trình của Sambrook và cộng sự [181] có điều chỉnh. Sau khi tạo được các chủng *E. coli* chứa vector biểu hiện mang gene cần nghiên cứu, dòng khuẩn lạc mang các gene được lựa chọn cấy chuyển vào 5 ml môi trường LBA và nuôi lắc 200 vòng/phút ở 37°C qua đêm. Dịch tế bào nuôi cấy qua đêm được chuyển sang môi trường mới sao cho OD₆₀₀ đạt khoảng 0,1 (hoặc tỷ lệ 1/100). Dịch nuôi cấy tiếp tục được lắc 200 vòng/phút ở 37°C cho đến khi OD₆₀₀ đạt khoảng từ 0,4 đến 0,6 (khoảng 1,5-2 giờ) thì được cảm ứng sinh tổng hợp protein tái tổ hợp bằng cách bổ sung IPTG đến nồng độ cuối cùng 0,1 mM và tiếp tục nuôi trong thời gian 5 giờ ở nhiệt độ 25°C.

Các điều kiện ảnh hưởng đến sinh tổng hợp protein tái tổ hợp như nồng độ IPTG, nhiệt độ, môi trường nuôi cấy cũng sẽ được khảo sát. Sau thời gian cảm ứng biểu hiện, các tế bào được thu lại bằng cách ly tâm 6000 vòng/phút, trong 10 phút và đưa về cùng giá trị OD₆₀₀ bằng 10 với đệm thích hợp. Protein được phân tích bằng phương pháp điện di biến tính trên gel polyacrylamide.

Để đánh giá khả năng tan của protein tái tổ hợp, tế bào biểu hiện được phá vỡ bằng siêu âm với chu kỳ 10 giây siêu âm và 20 giây nghỉ trong tổng khoảng 30 đến 40 xung ở tần số 20 kHz. Sau đó, dịch siêu âm được ly tâm 13.000 vòng/phút trong 10 phút để tách protein pha tan và pha không tan. Pha tan là phần dịch nổi được thu lại, còn phần cặn là pha không tan được hòa trở lại trong nước cất vô trùng và điện di trên gel polyacrylamide 12,6% để kiểm tra so sánh với mẫu protein tổng số, xác định khả năng tan, tủa của enzyme.

2.2.2.2. Điện di protein trên gel polyacrylamide

Quy trình điện di protein trên gel polyacrylamide được dựa trên quy trình chuẩn Laemmli và đồng tác giả [182].

* *Chuẩn bị gel*: Tùy theo mục đích kiểm tra mà 2 loại gel polyacrylamide biến tính và không biến tính (không có SDS) được sử dụng. Gel được chuẩn bị với thành phần và nồng độ như mô tả trong Bảng 2.1. Các thành phần được bổ sung theo thứ tự và trộn đều trước khi cho vào giá đỡ bản gel. Gel tách được chuẩn bị trước và tra vào khuôn cho đến khi đông hoàn toàn (khoảng 30 phút) mới chuẩn bị tiếp gel cô. Bản gel ổn định sau khi gel cô được đưa vào bản gel sau khoảng 30 phút có thể được sử dụng để điện di kiểm tra protein.

Bảng 2. 1. Thành phần gel polyacrylamide biến tính

Thành phần	Gel biến tính	
	Gel tách 12,6% (µl)	Gel cô 5% (µl)
Tris-HCl 0,5 M pH 6,8	-	200
Tris-HCl 1,5 M pH 8,8	1125	-
Glycerol 50%	900	-
Bis-Acrylamide 30%	1890	140
SDS 10%	45	4
APS 10%	30	8
TEMED	3	1
dH ₂ O	550	450
Tổng	4543	803

* *Quy trình điện di trên gel polyacrylamide-SDS*: Hỗn hợp protein cần điện di được bổ sung đệm xử lý mẫu (sample buffer 6X) với tỷ lệ một phần năm thể tích mẫu ban đầu, trộn đều trước khi được xử lý ở nhiệt độ cao (100°C) trong 10 phút. Sau đó, mẫu được lấy ra, đặt lên đá và tra vào giếng điện di. Khi mẫu điện di (có mang chất chỉ thị màu) vẫn đang ở trong phần gel có nồng độ acrylamide thấp, toàn bộ bản gel và đệm điện di được đặt dưới cường độ dòng điện 10 mA. Khi vạch màu chỉ thị mẫu đã tiến đến phần gel có nồng độ cao, mỗi bản gel được điện di với cường độ dòng điện lớn hơn (20 mA) đến khi vạch chỉ thị đã đi hết chiều dài bản gel. Tiếp đó, bản gel được gỡ ra khỏi khung kính và được ngâm trong dung dịch thuốc nhuộm khoảng 60 phút, trước khi loại bỏ thuốc nhuộm bám không đặc hiệu trên bản gel bằng dung dịch tẩy nhuộm. Mẫu protein đã được phân tách và gắn thuốc nhuộm có thể quan sát bằng mắt thường.

2.2.2.3. Tách chiết protein tái tổ hợp và tinh chế bằng sắc ký ái lực His-tag

Theo như chiến lược đưa gene *exl* vào vector biểu hiện pET22b(+) và biểu hiện trong vật chủ *E. coli*, gene *exl* không chứa mã kết thúc đã được nối trực tiếp với một đoạn DNA ngắn mã hóa 6 gốc Histidine ở tận cùng đầu 3'. Do đó, enzyme tái tổ hợp endo-xylanase được biểu hiện trong hệ vector này theo lý thuyết sẽ được biểu hiện và đưa ra khoang chu chất của tế bào *E. coli*, đồng thời protein đều có chứa đoạn trình tự 6His (His-tag) ở đầu C. Các gốc His này sẽ liên kết với Ni²⁺ trên cột Hitrap giúp cho quá trình tinh sạch protein thuận tiện. Để tách chiết protein tái tổ hợp, màng tế bào vật chủ được phá vỡ bằng sóng siêu âm cường độ cao để giải phóng vật chất nội bào, sau đó dịch phá tế bào được tinh chế bằng sắc ký.

Quá trình tách chiết và tinh chế protein tái tổ hợp được thực hiện theo các bước như sau: Chuyển 15 ml dịch tế bào có OD₆₀₀ = 10 vào ống falcon loại 25 ml. Tiến hành phá tế bào bằng sóng siêu âm với chu trình phóng 10 giây, nghỉ 5 giây, tổng thời gian trong 9 phút trên đá. Ly tâm dịch phá tế bào ở tốc độ 8000 vòng/phút trong 10 phút, thu pha tan chứa protein tái tổ hợp. Dịch sau khi thu được bơm toàn bộ (15 ml) lên cột HisTrap 5 ml trước đó đã cân bằng với đệm Tris-HCl 20 mM (pH 7), 10 mM imidazole. Dịch chảy qua cột (F) được thu hồi để kiểm tra mức độ bám của protein tái tổ hợp. Sau đó, protein tạp được rửa lần lượt bằng 50 ml đệm Tris-HCl 20 mM, NaCl 100 mM chứa imidazole có nồng độ 50 mM, 100 mM, 150 mM imidazole. Thu các dịch rửa để kiểm tra protein đi ra khỏi cột. Tiếp theo, protein tái tổ hợp được đẩy ra khỏi cột bằng 25 ml đệm chứa 250 mM imidazole. Dịch chứa protein được thu vào các ống Eppendorf (2 ml/ống). Các phân đoạn chứa protein tinh sạch được gộp lại và loại muối bằng cột PD-10 (Amersham Biosciences, Mỹ)

theo hướng dẫn của nhà sản xuất.

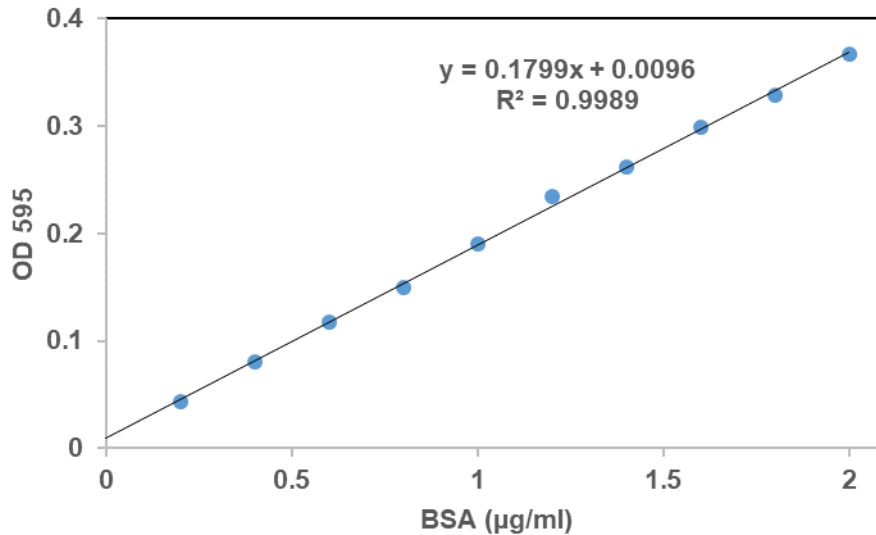
2.2.2.4. *Xác định độ sạch của enzyme tái tổ hợp bằng phần mềm Quantity One*

Phần mềm Quantity One (<http://www.biorad.com/en-us/product/quantity-one-1-danalysis-software>) là phần mềm online của Bio-Rad cho phép đánh giá độ tinh sạch của một mẫu protein được nhuộm màu. Các bước thực hiện chính bao gồm điện di trên gel polyacrylamide các mẫu protein cần kiểm tra cùng với thang chuẩn protein. Sau đó, gel được nhuộm bằng thuốc nhuộm comassie, rửa sạch và đưa lên máy scan để quét. Chế độ quét được cài đặt sao cho chất lượng ảnh đạt được tốt nhất. Sau đó, ảnh quét được đưa về chế độ đen trắng và chuyển vào phần mềm Quantity One version 4.6. Lượng protein có trên giếng chạy sẽ được phần mềm nhận biết và quét để định lượng tương đối. Độ sạch của protein tái tổ hợp chính là tỷ lệ của khối lượng xác định được của băng protein tái tổ hợp so với tổng khối lượng của protein trên mỗi đường chạy ở bản gel.

2.2.2.5. *Định lượng protein bằng phương pháp Bradford*

Hàm lượng protein trong mẫu được xác định bằng phương pháp Bradford [183] có cải biến phù hợp với dung dịch Bradford thương mại 5X (Sigma). BSA chuẩn (Sigma) được hòa trong nước cất thành 11 nồng độ cách đều nhau lần lượt từ từ 0 đến 2 $\mu\text{g}/\mu\text{l}$. Mỗi phản ứng có tổng thể tích là 250 μl gồm 200 μl mẫu dịch BSA chuẩn và 50 μl Bradford reagent. Mẫu được đo OD_{595} bằng máy đọc ELISA. Kết quả đo độ hấp thụ ở bước sóng 595 nm của các mẫu BSA với nồng độ chính xác như trên được sử dụng để thiết lập đường chuẩn biểu thị mối quan hệ giữa độ hấp thụ quang học ở bước sóng này và nồng độ protein trong mẫu đã bổ sung dung dịch Bradford (Hình 2.1).

Để đo nồng độ protein của mẫu cần quan tâm, mẫu được pha loãng sao cho mật độ quang học đo được nằm trong khoảng đường chuẩn. Sau khi xác định được giá trị OD_{595} của mẫu cần đo (y), lượng protein tương ứng (x) được xác định dựa vào phương trình tuyến tính đã được thiết lập. Hàm lượng protein trong mẫu được xác định là tích của lượng protein (x) và hệ số pha loãng mẫu với nước cất trong quá trình chuẩn bị mẫu đo.



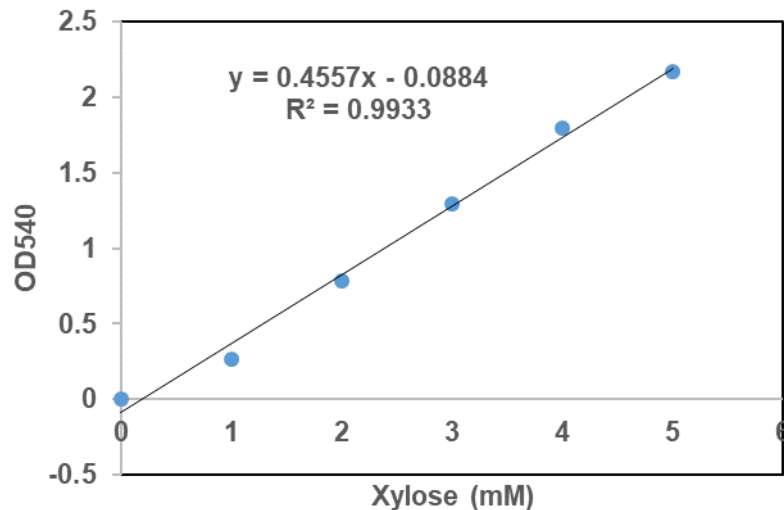
Hình 2. 1. Đường chuẩn biểu thị sự liên hệ giữa OD₅₉₅ và hàm lượng BSA

2.2.2.6. Xác định hoạt tính enzyme

(i) Xác định hoạt tính xylanase

- Nguyên lý: Lượng đường khử có mặt trong dung dịch được xác định dựa vào phương pháp DNS của Miller [184] có điều chỉnh để phù hợp với mẫu thực nghiệm. DNS là chất phản ứng chứa dinitrosalicylic acid (DNS). Cấu trúc hóa học của DNS có hai nhóm nitro và khi có mặt đường khử thì một nhóm nitro bị khử thành amino dẫn đến DNS vốn có màu vàng trở thành 3-amino-5-nitrosalicylic có màu nâu đỏ. Màu nâu đỏ của 3-amino-5-nitrosalicylic hấp thụ cao nhất ở bước sóng 540 nm.

- Thiết lập đường chuẩn: Đồ thị chuẩn được thiết lập với thang chuẩn 6 nồng độ đường D-xylose từ 0 đến 5 mM. Mỗi ống phản ứng có thể tích 75 µl D-xylose được pha loãng và 75 µl chất phản ứng DNS (DNS 1%, NaOH 0,25 M, Na-K tartrate 0,645 M). Tất cả các hỗn hợp được ủ ở 100°C trong 15 phút. Sau đó bổ sung 25 µl K-Na-tartrate 40% và để mẫu nguội ở nhiệt độ phòng. Mẫu được đo OD ở bước sóng 540 nm bằng máy đọc ELISA. Kết quả đo OD₅₄₀ được sử dụng để thiết lập đường chuẩn biểu thị mối quan hệ giữa OD₅₄₀ và nồng độ đường khử.



Hình 2. 2. Đường chuẩn xylose

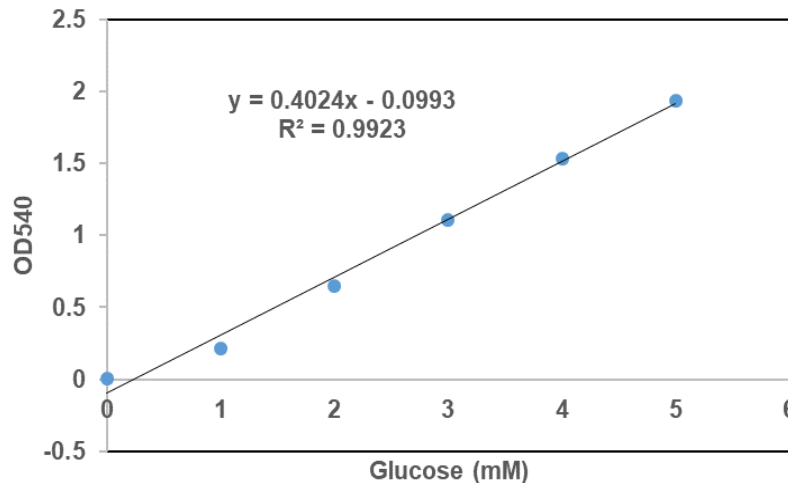
- *Xác định hoạt tính xylanase trên cơ chất tan xylan*: Hỗn hợp được chuẩn bị gồm có 100 μ l enzyme thô dịch pha tan, hoặc 100 μ l enzyme sau tinh chế loại muối (tương đương 5 μ g enzyme), 100 μ l đệm PBS 5x pH 6,8 (hoặc ở giá trị pH tối ưu), 100 μ l xylan 1% pha trong đệm PBS 1x và nước với tổng thể tích là 500 μ l. Hỗn hợp sau đó được ủ ở 37°C (hoặc nhiệt độ tối ưu) trong 30 phút. Sau thời gian ủ, hỗn hợp được bổ sung 500 μ l chất phản ứng DNS và đun sôi hỗn hợp 15 phút. Sau đó, bổ sung tiếp 125 μ l K-Na-tartrate 40%. Mẫu được đo OD ở bước sóng 540 nm. Lượng đường khử tạo ra được tính dựa vào đường chuẩn về mối quan hệ giữa OD₅₄₀ và nồng độ xylose đã được thiết lập. Mỗi phép thử được lặp lại 3 lần. Hoạt tính xylanase (tính theo đơn vị IU) là số μ mol xylose được giải phóng ra trong mỗi phút ở điều kiện đo như nhiệt độ và pH [185]. Mẫu đối chứng có thành phần giống hệt như mẫu phản ứng nhưng enzyme chỉ được bổ sung sau khi đã kết thúc quá trình ủ.

(ii) Xác định hoạt tính cellulase

- Thiết lập đường chuẩn glucose

Lượng đường khử glucose có mặt trong dung dịch sẽ được xác định dựa vào phương pháp DNS của Miller có thay đổi phù hợp với mẫu thực nghiệm. Đồ thị chuẩn được thiết lập với thang chuẩn 6 nồng độ đường D-glucose từ 0 mM đến 10 mM (Hình 2.3). Mỗi ống phản ứng có thể tích 75 μ l D-glucose được pha loãng và 75 μ l chất phản ứng DNS (DNS 1%, NaOH 0,25 M, Na-K tartrate 0,645M). Tất cả các hỗn hợp được đun sôi 15 phút. Bổ sung 25 μ l K-Na-Tartrate 40% và để mẫu nguội ở nhiệt độ phòng. Tương tự như đường chuẩn xylose, mối quan hệ giữa nồng độ đường glucose trong phản ứng và độ hấp thụ quang học ở bước sóng 540 nm

được thiết lập, từ đó là cơ sở tính toán lượng glucose trong mẫu phản ứng enzyme.



Hình 2. 3. Đường chuẩn glucose.

* *Xác định hoạt tính cellulase thủy phân cơ chất CMC*: Dịch pha tan của enzyme thô hoặc enzyme tinh sạch được sử dụng để xác định hoạt độ với cơ chất là CMC. Tổng thể tích phản ứng là 0,5 ml gồm có 0,25 ml CMC 1%, 0,05 ml đệm PBS 10X, pH 6,8 (hoặc pH tối ưu) và 5 μ g enzyme tinh sạch. Đối chứng âm có enzyme và cơ chất được ủ riêng rẽ, trộn trước khi bổ sung DNS. Phản ứng được tiến hành ở 40°C (hoặc nhiệt độ thích hợp) trong 30 phút. Sau đó, hỗn hợp phản ứng được bổ sung thêm 0,5 ml dung dịch DNS và được đun sôi trong 15 phút. Hỗn hợp dung dịch phản ứng được bổ sung 125 μ l Na-K tartrate 40%, để nguội ở nhiệt độ phòng và đo độ hấp phụ tại bước sóng 540 nm. Một đơn vị hoạt độ (IU) cellulase được định nghĩa là lượng enzyme cần thiết để phân giải cơ chất CMC thành đường khử tương đương với 1 μ mol glucose trong 1 phút ở điều kiện thí nghiệm.

Xác định hoạt tính của enzyme thông qua đo OD ở bước sóng 540 nm dựa trên lượng đường khử được giải phóng so sánh với đường chuẩn glucose [185].

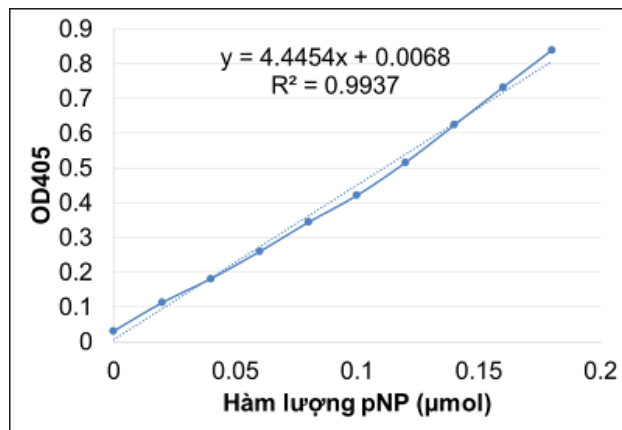
* *Xác định hoạt tính cellulase thủy phân cơ chất giấy lọc*: Dịch pha tan của enzyme thô hoặc enzyme tinh sạch được sử dụng để kiểm tra hoạt tính với cơ chất là giấy lọc. Giấy lọc (Whatman No. 1) được cắt bằng kéo thành các miếng rất nhỏ và đều nhau theo phương pháp của Camassola và cộng sự [186]. Tổng thể tích phản ứng là 0,5 ml gồm có 10 mg giấy lọc, 0,05 ml đệm PBS 10X, pH 6,8 (hoặc pH tối ưu) và 5 μ g enzyme tinh sạch. Đối chứng âm có enzyme và cơ chất được ủ riêng rẽ, trộn trước khi bổ sung DNS. Phản ứng được tiến hành ở 40°C (hoặc nhiệt độ thích hợp) trong 30 phút. Sau đó, hỗn hợp phản ứng được bổ sung thêm 0,5 ml dung dịch DNS và được đun sôi trong 15 phút. Hỗn hợp dung dịch phản ứng được bổ sung 125 μ l Na-K tartrate 40% để nguội ở nhiệt độ phòng và đo độ hấp phụ tại bước sóng

540 nm. Một đơn vị hoạt độ được định nghĩa là lượng enzyme cần thiết để phân giải cơ chất giấy lọc thành đường khử tương đương với 1 μmol glucose trong 1 phút ở điều kiện thí nghiệm.

(iii) Xác định hoạt tính beta-glucosidase và beta-xylosidase

- Thiết lập đường chuẩn p-nitrophenol (pNP)

Tương tự với việc xây dựng đường chuẩn ở trên, chất chuẩn pNP được pha bằng đệm phosphate 20 mM với 10 nồng độ chất chuẩn từ 0 mM đến 10 mM. Kết quả đo độ hấp thụ ở 405 nm được sử dụng để thiết lập đường chuẩn biểu thị mối quan hệ giữa OD_{405} và nồng độ pNP (Hình 2.4).



Hình 2. 4. Đường chuẩn pNP

* *Xác định hoạt tính beta-glucosidase và beta-xylosidase*: Dịch enzyme tinh sạch được sử dụng để kiểm tra hoạt tính beta-glucosidase và beta-xylosidase với cơ chất tương ứng là pNPG và pNPX. Enzyme có hoạt tính beta-glucosidase và beta-xylosidase sẽ xúc tác phản ứng phân cắt hai cơ chất trên và tạo ra pNP, lượng pNP tỷ lệ thuận với hoạt tính của enzyme. Mỗi phản ứng chứa 100 μl dịch enzyme sau khi tinh sạch (tương đương 5 μg enzyme), 100 μl cơ chất 10 mM được pha trong đệm PBS 5X, pH 6,0 và 300 μl nước deion. Phản ứng được thực hiện ở nhiệt độ 37°C trong 30 phút. Phản ứng được kết thúc bằng cách bổ sung 500 μl Na_2CO_3 1 M, trộn đều. Mẫu được đo OD_{405} . Lượng pNP tạo ra do xúc tác của enzyme được tính dựa vào đường chuẩn. Các thí nghiệm đều được lặp lại 3 lần và có bố trí đối chứng âm để tính toán hiệu quả thủy phân. Một đơn vị hoạt độ beta-glucosidase/beta-xylosidase được định nghĩa là lượng enzyme cần thiết để phân giải cơ chất tương ứng pNPG/ pNPX thành 1 μmol pNP trong 1 phút ở điều kiện thí nghiệm.

2.2.2.7. *Xác định ảnh hưởng của nhiệt độ, pH, các ion kim loại và một số hóa chất lên hoạt tính enzyme*

* *Xác định ảnh hưởng của nhiệt độ và pH:* Để xác định ảnh hưởng của nhiệt độ đối với hoạt tính enzyme, phản ứng thủy phân xylan của enzyme được tiến hành ở dải nhiệt độ: 40°C, 50°C và 60°C. Ảnh hưởng của pH đối với hoạt tính enzyme được xác định khi thay đổi pH của đệm PBS 10X (hoặc thay bằng đệm citrate, đệm Tris) ở các pH khác nhau là pH 4,0; pH 4,5; pH 5,0; pH 5,5; pH 6,0; pH 6,5, pH 7,0; pH 7,5; pH 8,0; pH 8,5; pH 9,0; pH 10,0; pH 11,0. Xác định lượng đường khử được giải phóng sau phản ứng để đánh giá hiệu quả thủy phân cơ chất ở các điều kiện nhiệt độ và pH khác nhau, từ đó xác định được ảnh hưởng của nhiệt độ và pH đến enzyme. Các thí nghiệm đều được lặp lại 3 lần và có bố trí đối chứng âm để tính toán hiệu quả thủy phân.

* *Xác định ảnh hưởng của ion kim loại và hóa chất:* Xác định ảnh hưởng của các ion kim loại, hóa chất đến hoạt tính của enzyme được thực hiện theo phương pháp đánh giá hoạt tính của endoglucanase với cơ chất CMC được miêu tả ở phần trên nhưng trong dung dịch có bổ sung các ion kim loại và hóa chất với nồng độ khác nhau. Các ion kim loại phổ biến được sử dụng với nồng độ 10 mM là Mn^{2+} , Fe^{3+} , Ca^{2+} , Mg^{2+} , Cu^{2+} , Ni^{2+} , Zn^{2+} , K^+ , Na^+ , Co^{2+} . Các hóa chất được sử dụng với nồng độ khác nhau là 1 μ M urea, 1 μ M EDTA, 1 μ M 2-mercaptoethanol, 1 μ M triton X-100, 1% SDS và 1 mM tween 80. Các thí nghiệm đều được lặp lại 3 lần, cùng với đối chứng âm là các thành phần đệm, enzyme, cơ chất được xử lý độc lập nhau nhưng ở cùng điều kiện như mẫu cần đo, sau khi kết thúc thời gian ủ phản ứng thì được phối trộn và bổ sung dung dịch DNS.

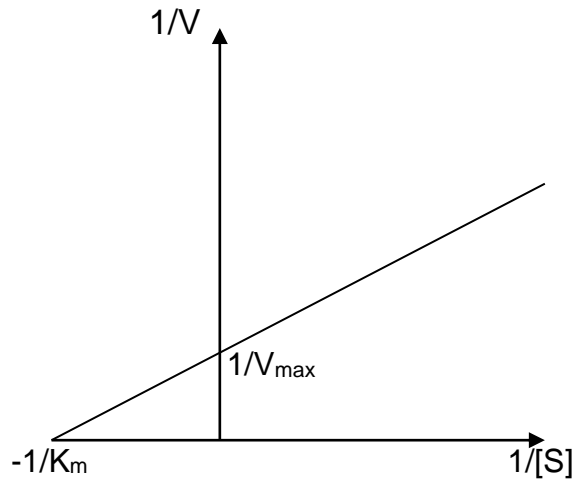
2.2.2.8. *Xác định độ bền nhiệt của enzyme*

Enzyme được ủ ở các nhiệt độ khác nhau từ 40°C, 50°C đến 60°C trong các khoảng thời gian khác nhau (1 giờ; 2 giờ, 4 giờ, 6 giờ và 24 giờ) trong đệm PBS, pH 5,5. Sau khi kết thúc thời gian ủ, hỗn hợp enzyme được bổ sung thêm cơ chất xylan 1% để thực hiện phản ứng với cơ chất. Thành phần, quy trình phản ứng được thực hiện như đã miêu tả ở trên. Hoạt tính của enzyme sau khi xử lý nhiệt được xác định và so sánh với hoạt tính của enzyme được thực hiện cùng thời điểm nhưng không trải qua quá trình ủ nhiệt độ.

2.2.2.9. *Xác định thông số động học của enzyme*

Giá trị K_m , V_{max} , được tính toán dựa trên phương trình tương quan giữa tốc độ phản ứng và nồng độ cơ chất được thiết lập theo phương pháp của Lineweaver-Burk, trong đó V được coi là số micromole glucose được giải phóng ra trong 1 phút, $1/[S]$ được tính là $1/[\text{nồng độ cơ chất}]$, ở đó nồng độ cơ chất được tính là mg/ml

(Hình 2.5). Hoạt tính xylanase được xác định ở dải nồng độ xylan khác nhau để xây dựng lên đường chuẩn làm cơ sở để tính toán giá trị K_m , V_{max} . Thí nghiệm xác định thông số động học của enzyme sau tinh chế thông qua phản ứng thủy phân cơ chất xylan ở dải nồng độ khác nhau. Kết quả thu được có thể được biểu diễn trên đồ thị có phương trình dạng $y = ax/(b+x)$, trong đó a là giá trị V_{max} , b là giá trị K_m ; kết quả được xử lý và đồ thị đường cong được xây dựng qua phần mềm trên web theo đường dẫn.

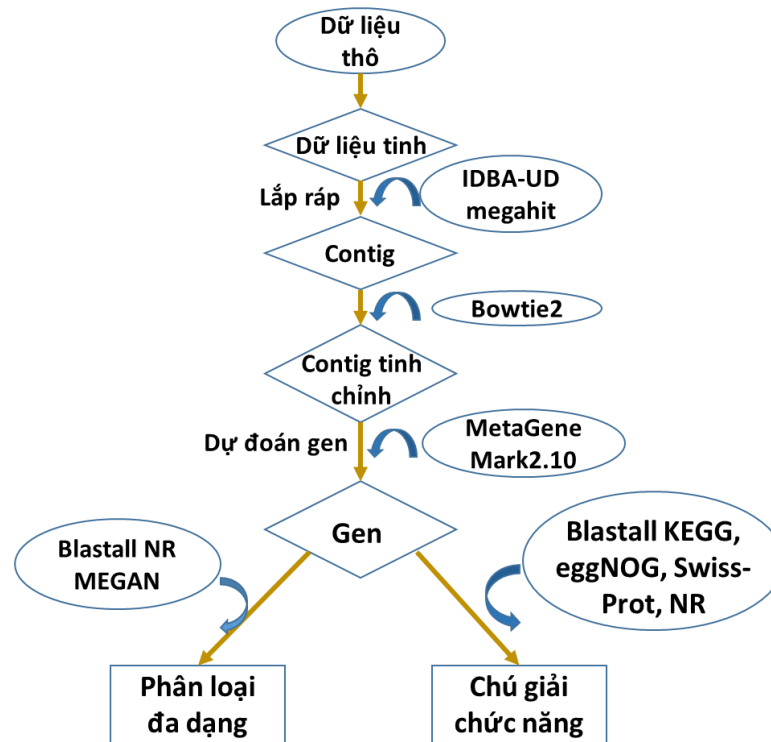


Hình 2. 5. Sự phụ thuộc của tốc độ phản ứng và nồng độ cơ chất theo Lineweaver-Burk [187]

2.2.3. Các phương pháp tin sinh học

2.2.3.1. Lắp ráp DNA đa hệ gene, chú giải các gene chức năng

Kết quả giải trình tự DNA đa hệ gene được phân tích, kiểm tra chất lượng, lắp ghép thành các contig, sau đó được dự đoán trình tự gene, từ đó là cơ sở để đánh giá phân loại về đa dạng cũng như chú giải chức năng gene. Cụ thể, dữ liệu DNA thô gồm hàng triệu đoạn trình tự ngắn (gọi là các read) được kiểm tra chất lượng bằng phần mềm, từ đó loại bỏ những trình tự có chất lượng thấp, gồm các trình tự có chứa số base không xác định (ký hiệu N) lớn hơn hoặc bằng 5%, các trình tự chứa tối thiểu 50% số base có chất lượng Q nhỏ hơn 20 và các trình tự adapter. Dữ liệu thô sau khi được xử lý trở thành bộ dữ liệu tinh, được đưa vào phần mềm chuyên dụng như IDBA-UD, megahit để lắp ráp thành các contig, là các đoạn trình tự dài chứa các trình tự read có những vùng chồng lấp, gối nhau. Phần mềm Bowtie2 được sử dụng để kiểm tra lại các contig với các thành phần read tạo thành và chỉnh sửa các contig, đồng thời đánh giá sự bao phủ của các read. Phần mềm MetaGene Mark2.0 được sử dụng để dự đoán các gene từ các contig đã tinh chỉnh.



Hình 2. 6. Quá trình phân tích và khai thác gene từ dữ liệu DNA đa hệ gene

Các gene trên được đưa vào các phần mềm tương ứng để dự đoán về phân loại đa dạng học và chú giải chức năng. Phần mềm MEGAN (MEtaGenomic ANalyser) phiên bản 4.6 được sử dụng để dự đoán các mức độ phân loại. Các trình tự gene được so sánh với trình tự của cơ sở dữ liệu NR (trình tự không trùng lặp non-redundant từ ngân hàng gene NCBI cùng với các trình tự từ các dữ liệu ngân hàng khác như Refseq, PDB, SwissProt, PIR and PRF). Thuật toán cơ sở LCA (Least Common Ancestors) của phần mềm MEGAN được dùng để xác định cấp độ phân loại của từng gene. Trình tự được xếp vào nhóm phân loại sao cho cấp độ phân loại của nhóm đó phản ánh được mức độ bảo thủ của trình tự gene. Việc phân loại, đánh giá đa dạng vi khuẩn ở các cấp độ được mô hình hóa bằng công cụ Krona tích hợp Excel (<http://krona.source>).

2.2.3.2. Phương pháp nghiên cứu Pfam của các trình tự

Pfam là cơ sở dữ liệu các họ protein được chú giải vùng chức năng dựa trên so sánh giống hàng trình tự có sử dụng mô hình Markov ẩn. Dựa trên kết quả ước đoán chức năng gene bằng các CSDL CAZy, KEGG, eggNOG, chúng tôi lọc riêng các trình tự mã hóa cho enzyme có hoạt tính cellulase, hemicellulase từ dữ liệu DNA metagenome của vi sinh vật trong dạ cỏ dê. Các trình tự gene được dịch mã sang trình tự amino acid và giữ nguyên mã trình tự. Tập hợp các trình tự amino acid được format dưới dạng FASTA và chuyển vào tìm kiếm Pfam từ CSDL PFAM theo đường dẫn trên web <http://pfam.janelia.org/search>. Kết quả phân tích cung cấp

thông tin dự đoán về các vùng bảo thủ, các vùng chức năng của gene dựa theo cơ sở dữ liệu.

2.2.3.3. Nghiên cứu vùng bảo thủ và dự đoán cấu trúc bậc ba của các trình tự

Để ước đoán vùng bảo thủ của trình tự, chúng tôi đã sử dụng phần mềm BLASTP để so sánh trình tự amino acid suy diễn từ các ORF được lựa chọn với dữ liệu sẵn có trên NCBI theo đường dẫn <http://blast.ncbi.nlm.nih.gov/Blast.cgi>. Bên cạnh đó để dự đoán cấu trúc bậc ba của enzyme, chúng tôi sử dụng hai phần mềm khác nhau bao gồm Phyre2 và Swiss model để phân tích và so sánh kết quả. So với phần mềm Phyre2 ước đoán cấu trúc không gian của protein dựa trên việc so sánh các vị trí amino acid bảo tồn, thì Swiss-Prot cho phép ước đoán chi tiết được cấu trúc không gian của hầu hết các trình tự protein, dựa trên toàn bộ trình tự. Trình tự amino acid của các ORF được xử lý trực tuyến bằng chương trình trực tuyến Phyre2 theo đường dẫn được thiết kế trên web có địa chỉ <http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>. Phần mềm sẽ so sánh mức độ tương đồng của các trình tự amino acid cần nghiên cứu với các trình tự có sẵn trên CSDL và đưa ra được cấu trúc 3D thích hợp nhất. Phyre2 là phần mềm dựa trên các nguyên tắc tương đồng ở các vùng bảo tồn cao của protein, cho phép dự đoán cấu trúc, chức năng, phân loại, tiến hóa,... các cấu trúc tinh thể protein. Để phân tích, các trình tự được so sánh tương đồng cấu trúc với cơ sở dữ liệu có sẵn của Phyre2. Kết quả so sánh trình tự bao gồm khả năng ước đoán mức độ tin cậy, mô hình 3D được dự đoán dựa trên những mô hình đã được công bố, thông tin thu được từ một trong hai cấu trúc theo cơ sở dữ liệu Protein (Scop) hoặc Ngân hàng dữ liệu Protein (PDB) tùy thuộc vào nguồn gốc của các mẫu phát hiện [54]. Cấu trúc bậc ba của các trình tự có độ tương đồng thấp có thể được ước đoán bằng Swiss-Prot trên web theo đường dẫn <https://swissmodel.expasy.org/interactive>.

2.2.3.4. Dự đoán khả năng chịu kiềm/acid

Để dự đoán khả năng chịu kiềm hoặc acid của các enzyme dựa trên trình tự, chúng tôi đã sử dụng phần mềm để dự đoán là AcalPred [188]. Đây là công cụ mở được phát triển bởi một nhóm nhà khoa học Trung Quốc. Công cụ này được xây dựng nhằm phân biệt các enzyme chịu kiềm và chịu acid chỉ dựa vào trình tự. Phần mềm này dựa trên các thông tin theo thứ tự tổ hợp nhiều chỉ số khác nhau của các protein đã nghiên cứu thực nghiệm bao gồm: thành phần các amino acid, chỉ số GO, nhóm các amino acid được bảo tồn, giá trị của điện tích,... Trang web trực tuyến miễn phí AcalPred được xây dựng dựa trên phần mềm để cung cấp một công cụ dự đoán tính chất của trình tự protein cho các nhà khoa học.

2.2.3.5. Dự đoán khả năng chịu nhiệt của enzyme

Dựa trên thành phần và trình tự của các amino acid, liên kết hydro, liên kết Van der Waals, tương tác kỵ nước và đặc điểm của các enzyme từ các sinh vật sống trong điều kiện môi trường có nhiệt độ cao [189], phần mềm của TBI được xây dựng trên số liệu của 150.000 protein chịu nhiệt độ khác nhau trong ngân hàng NCBI đã được sử dụng để dự đoán khả năng chịu nhiệt, dựa trên nguyên tắc tương đồng. Phần mềm dự đoán khả năng chịu nhiệt sẽ dự đoán kết quả T_m thể hiện ở ba mức độ: khả năng chịu nhiệt độ trên 65°C, dưới 55°C và trong khoảng nhiệt độ 55°C-65°C tương ứng với chỉ số tính toán từ phần mềm là lớn hơn 1, nhỏ hơn 0 hay trong khoảng từ 0 đến 1. Phần mềm này cho phép dự đoán khả năng chịu nhiệt của một lượng lớn trình tự amino acid trong cùng một lần thực hiện.

2.2.3.6. Định loại loài các trình tự ORF

Các ORF được định loại bằng phần mềm MEGAN (version 4.6) sử dụng thuật toán tìm tổ tiên chung gần nhất (LCA – lowest common ancestor) [190]. Thuật toán này định loại gene về các bậc phân loại mà thể hiện được mức độ bảo thủ của gene. Điều này có nghĩa là các trình tự được phân loại ở các mức độ cao hơn trong cây phân loại trình tự có độ bảo thủ cao hơn so với các trình tự gene nằm ở các nhánh lớn hơn.

Phần mềm MEGAN sẽ phân loại các trình tự ORF về các bậc phân loại từ giới, ngành, lớp, bộ, họ, chi, loài. Sau đó, số lượng và sự đa dạng của mỗi bậc phân loại được tập hợp.

2.2.3.7. Tối ưu mã và tổng hợp gene mã hóa enzyme thủy phân xylan được khai thác từ dữ liệu giải trình tự DNA metagenome vi khuẩn dạ cỏ dê

Mức độ biểu hiện của gene phụ thuộc vào nhiều yếu tố trong đó sự phù hợp mã bộ ba của gene ngoại lai và chủng chủ nhận gene đóng vai trò rất quan trọng. Vì vậy, gene mã hóa enzyme tham gia thủy phân cellulose được khai thác từ dữ liệu giải trình tự DNA metagenome của vi sinh vật dạ cỏ dê được kiểm tra sự phù hợp mã bộ ba bằng phần mềm trực tuyến (Rare Codon Analysis Tool) của hãng Genscript để đảm bảo các gene được biểu hiện phù hợp nhất với hệ biểu hiện *E. coli* (<https://www.genscript.com/tools/rare-codon-analysis>). Một số thông số quyết định sự phù hợp là chỉ số phù hợp mã bộ ba CAI, tỷ lệ nucleotide guanine và cytosine cũng như sự phân bố của các nucleotide loại này trên gene, và sự phân bố các mã bộ ba cùng với tỷ lệ của chúng trên toàn bộ gene. Sau đó, bằng phần mềm tối ưu mã bộ ba của Genscript, các gene này đã được tối ưu mã để tăng chỉ số CAI lên gần 1

nhưng không làm thay đổi trình tự amino acid. Gene *exl* sau khi tổng hợp tại công ty Genscript (Mỹ) được đưa vào vector biểu hiện tại vị trí *NcoI*+*XhoI* trên vector pET22b(+).

2.2.3.8. Phương pháp xử lý số liệu

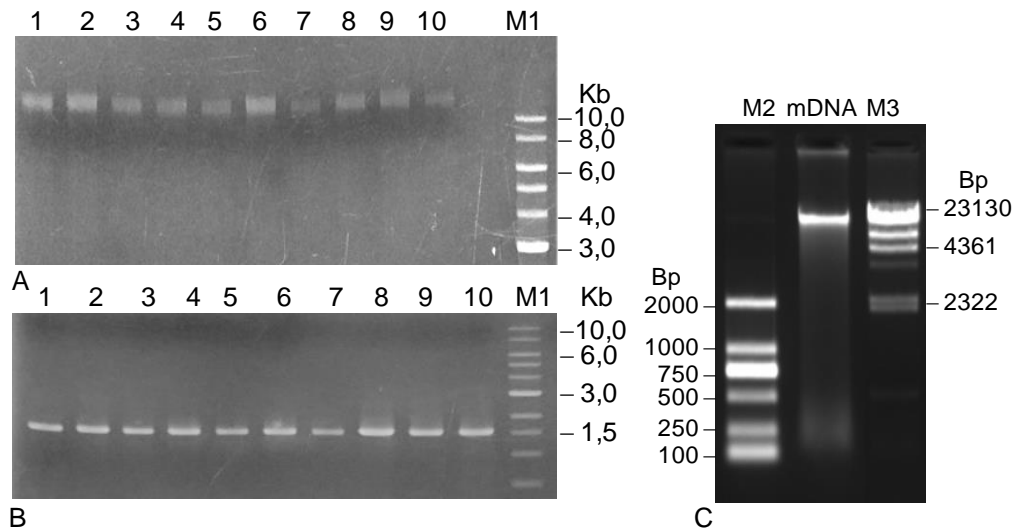
Các số liệu thực nghiệm trong Báo cáo tổng hợp được xử lý bằng phần mềm Microsoft Excel. Kết quả là trung bình cộng của 3 lần thí nghiệm được trình bày dưới dạng \pm SE (Standard Error). Các số liệu lắp ráp, chú giải gene được xử lý theo giá trị $E \leq 10^{-5}$.

CHƯƠNG 3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Nghiên cứu giải mã xây dựng bộ dữ liệu và đánh giá đa dạng vi khuẩn trong dạ cỏ dê

3.1.1. Tách chiết DNA đa hệ gene của vi khuẩn

Đối tượng khai thác gene của đề tài là vi khuẩn trong dạ cỏ, vì vậy trước khi tách chiết DNA, hệ vi khuẩn trong dạ cỏ dê được phân tách riêng rẽ khỏi các thành phần khác nhờ ly tâm phân đoạn. Sau đó, DNA vi khuẩn trong dạ cỏ dê đã được tách chiết theo phương pháp của Sambrook và cộng và được tinh chế bằng PSP Spin stool DNA kit. Dựa trên kết quả tách chiết được thể hiện bằng hình ảnh điện di DNA tổng số (Hình 3.1A), cả 10 mẫu vi khuẩn từ dạ cỏ dê đã được tách chiết, tinh chế thành công với DNA có kích thước lớn, đáp ứng được yêu cầu nghiên cứu.



Hình 3. 1. Điện di đồ phân tích DNA đa hệ gene vi khuẩn được tách chiết từ 10 mẫu dạ cỏ dê (A) và sản phẩm khuếch đại gene 16S rDNA của vi khuẩn từ các mẫu DNA được tách chiết (B); DNA metagenome được gửi giải trình tự (C).

M1: DNA chuẩn 1 kb (Fermentas); M2: DNA chuẩn D2000 (Tiangen), mDNA: DNA metagenome của vi khuẩn dạ cỏ dê sau khi gộp để giải trình tự (2 μ l); M3: DNA lambda cắt bằng *HindIII* (Takara); 1-10: Mẫu có nguồn gốc từ 10 mẫu dạ cỏ dê;

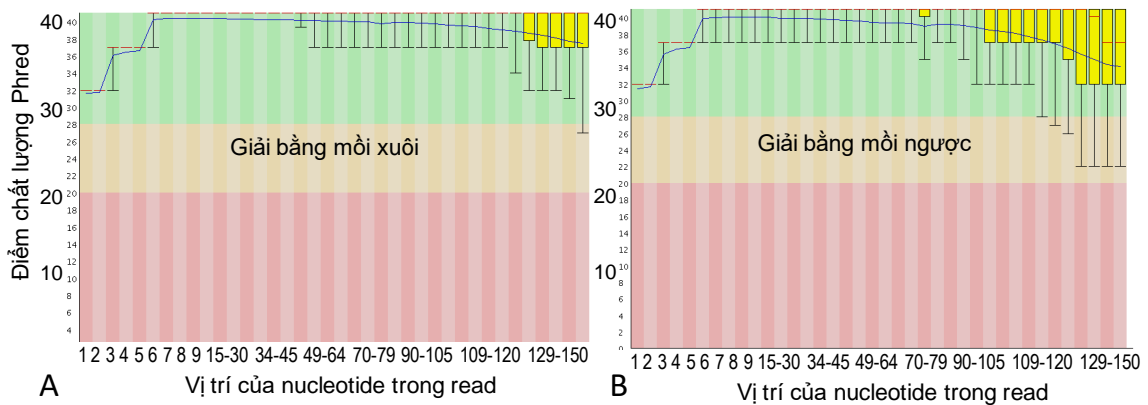
Để đáp ứng yêu cầu cho giải trình tự DNA metagenome, nồng độ DNA trong mẫu cần phải đạt 50 ng/ μ l và chỉ số A260/280 \geq 1,8. Kết quả kiểm tra chất lượng DNA và hàm lượng DNA bằng máy nanodrop cho thấy, nồng độ DNA đạt từ 53,5 đến 137 ng/ μ L và chỉ số A260/280 đạt từ 1,921 đến 2,028. Như vậy, nồng độ và chất lượng mẫu sau tách chiết và tinh chế đáp ứng yêu cầu cho giải trình tự.

Để đảm bảo cho việc giải trình tự dựa trên nguyên lý tổng hợp gene nên các mẫu sau tách chiết đã được kiểm tra khả năng ức chế DNA polymerase của mẫu. Do

vậy, cả 10 mẫu DNA metagenome từ vi khuẩn dạ cỏ dê được sử dụng làm khuôn cho khuếch đại đoạn gene mã hóa ribosome 16S (16S rRNA) của vi khuẩn. Kết quả (Hình 3.1B) cho thấy, các gene 16S rRNA có kích thước ~1,5 Kb đã được khuếch đại đặc hiệu. Điều này chứng tỏ trong mẫu không chứa chất ức chế DNA polymerase để có thể phát hiện được.

Các mẫu DNA đa hệ gene tách chiết từ dạ cỏ dê được gộp lại thành một mẫu, sao cho lượng DNA từ 10 mẫu là như nhau. Hình 3.1C cho thấy lượng DNA từ 2 μ l mẫu trên giếng được thể hệ là bằng rất sắc nét. Sử dụng máy Qubit fluorometer, nồng độ DNA đo được là 48,4 ng/ μ l. Với nồng độ DNA này, 45 μ g DNA metagenome từ vi khuẩn dạ cỏ dê đã được gửi sang Công ty BGI để giải trình tự.

3.1.2. Giải trình tự, đánh giá chất lượng bộ dữ liệu và chú giải gene



Hình 3. 2. Biểu đồ đánh giá chất lượng giải trình tự đa hệ gene của vi khuẩn bằng FastQC trên bộ dữ liệu giải mã sâu

A: Bộ dữ liệu giải bằng môi xuôi; B: Bộ dữ liệu giải bằng môi ngược

Dữ liệu trình tự nucleotide DNA metagenome ban đầu ở dạng thô đã được đánh giá điểm chất lượng Phred (ký hiệu là Q) với mức điểm từ 0 đến 39. Nếu giá trị Phred của một base là $Q=30$ sẽ tương ứng với khả năng sai hỏng cho base được gọi là 1/1000 hay độ chính xác của base đó là 99,9%. Tương tự với giá trị $Q=20$ và $Q=10$, độ chính xác của các base lần lượt là 99% và 95%. Kết quả cho thấy bộ dữ liệu DNA metagenome có kích thước nhỏ (metagenomic conventional sequencing - MCS) có Q30 đạt 90,01%, bộ dữ liệu DNA metagenome giải mã sâu (Metagenomic deep sequencing - MDS) có Q30 đạt 94,59%. Như vậy, cả hai bộ dữ liệu đều có chất lượng tốt. Tổng dung lượng dữ liệu giải trình tự metagenome của vi khuẩn trong dạ cỏ dê lần lượt là 89,96 triệu read và 392,63 triệu read. Sau khi lọc bỏ các trình tự lỗi và trình tự adapter, số lượng read thu được là 84,62 triệu read từ dữ liệu MCS và 324 triệu read từ dữ liệu MDS, dữ liệu tinh tương đương 8,6 Gb và 48,66 Gb (Bảng 3.1).

Sau khi lắp ghép thành contig, số lượng các contig của bộ dữ liệu là khá lớn. Bộ dữ liệu DNA metagenome MDS của vi khuẩn trong dạ cỏ dê được lắp ghép thành 3.411.867 contig với tổng độ dài là 3.164 Mb. Trong đó, 50% trình tự có kích thước lớn hơn 1.162 bp, độ dài trung bình của các contig là 927 bp và contig có kích thước lớn nhất là 295.214 bp. Các contig bao phủ khoảng 64,22% read.

Bảng 3. 1. Các thông tin về bộ dữ liệu DNA metagenome của vi khuẩn trong dạ cỏ dê sau khi giải trình tự và lắp ráp

Thông tin bộ dữ liệu	Dữ liệu 8,6 GB (MCS)	Dữ liệu giải trình tự sâu (MDS)
Số lượng read thô (M)	87127	39263
Số lượng base tính (Gb)	8,46	48,66
Chất lượng read dùng để phân tích Q20 (%)	96,25	98,18
Chất lượng read dùng để phân tích Q30 (%)	90,01	94,59
Số lượng read tính dùng lắp ghép contig	84.625.346	324.080.208
Số lượng contig	172.918	3.411.867
Contig có kích thước lớn nhất (bp)	124.798	295.214
Kích thước trung bình của contig (bp)	1.032	927
Độ dài contig N50 (bp)	1.879	1.162
Mức độ bao phủ của contig lên các read (%)	27,3	64,22
Số lượng gene thu được (gene)	164.644	5.367.270
Tổng độ dài của các gene (bp)	-	2.828.583.591
Độ dài của gene N50 (bp)	-	645
Độ dài trung bình của gene (bp)	-	527
Số gene được chú giải dựa trên Nr	122.304 (74,28%)	4.360.747 (81,25%)
Số gene được chú giải dựa trên SwissProt	33.471 (20,33%)	2.169.982 (40,43%)
Số gene được chú giải dựa trên KEGG	56.571 (34,36%)	2.700.371 (50,31%)
Số gene được chú giải dựa trên eggNOG	134.843 (81,9%)	3.414.550 (63,62%)

- : kết quả không được thống kê

So với bộ dữ liệu MCS, mẫu DNA metagenome vi khuẩn dạ cỏ dê từ bộ dữ liệu MDS có dung lượng gấp 5,4 lần. Bộ dữ liệu MCS dữ liệu tinh của vi khuẩn trong dạ cỏ dê có kích thước 8,46 Gb và các contig chỉ bao phủ được 27,3% các read, trong khi đó giải trình tự sâu MDS, dung lượng giải trình tự của hệ vi khuẩn trong dạ cỏ dê tăng lên 48,66 Gb và mức độ bao phủ của các contig lên các read đạt 64,22%. Với mức độ bao phủ dưới 65% thì tỷ lệ trình tự gene không hoàn thiện dự đoán vẫn sẽ còn tương đối cao, vì vậy, để có được bộ dữ liệu vi khuẩn đầy đủ hơn thì cần phải giải trình tự DNA của mẫu này đạt mức ít nhất 90-95 Gb.

Dựa trên các contig thu được, phần mềm MetaGeneMark đã ước đoán được hơn 164644 gene khác nhau từ dữ liệu NCS và khoảng 5 triệu ORF từ bộ dữ liệu MDS của vi khuẩn trong dạ cỏ dê. Số liệu gene được chú giải cụ thể dựa trên các bộ dữ liệu khác nhau là khác nhau (Bảng 3.2), trong đó số lượng gene được chú giải chức năng ở bộ dữ liệu MDS bởi ít nhất một cơ sở dữ liệu là 4.385.296 gene (chiếm 81,7%).

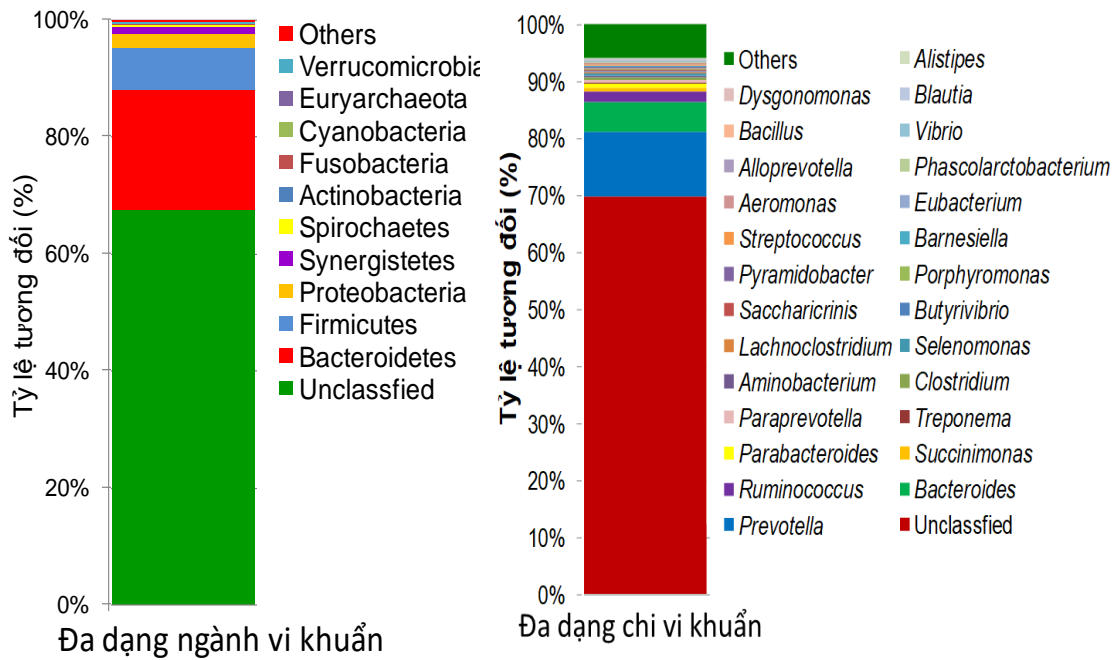
Bảng 3. 2. Kết quả chú giải gene chức năng từ dữ liệu giải trình tự DNA metagenome vi khuẩn trong dạ cỏ dê dựa trên các cơ sở dữ liệu khác nhau

	Tổng gene	Nr	Swissprot	KEGG	eggNOG	Tổng quát
Dữ liệu giải trình tự có dung lượng 8,6 Gb						
Số lượng (gene)	164.644	122.304	33.471	56.751	134.843	-
Tỉ lệ %	100%	74,28%	20,33%	34,47%	81,90%	-
Dữ liệu giải trình tự có dung lượng 48,7 Gb						
Số lượng (gene)	5.367.270	4.360.747	2.169,982	2.700.371	3.414.550	4.385.296
Tỉ lệ %	100%	81,25%	40,43%	50,31%	63,62%	81,70 %

3.1.3. Đánh giá đa dạng vi khuẩn trong mẫu DNA metagenome

3.1.3.1. Đa dạng vi khuẩn trong dạ cỏ dê được đánh giá dựa trên bộ dữ liệu 8,6 Gb

Từ dữ liệu giải trình tự 8,6 Gb, 164.644 gene đã được xác định, trong đó 99,8% số gene có nguồn gốc từ vi khuẩn, còn lại chỉ 0,02% có nguồn gốc từ sinh vật nhân thật, 0,005% từ virus và 0,17% gene từ vi khuẩn cổ. Trong số đó, 39.579 ORF đã được xác định và định loại, trong đó 99,8% thuộc về vi khuẩn, cụ thể, 39501 ORF vi khuẩn được xác định thuộc về 28 loài, 41 lớp, 95 bộ, 181 họ, 571 chi và 1634 loài. Ngành vi khuẩn chiếm số lượng đông nhất là ngành Bacteroidetes (63,5%), tiếp đó là ngành Firmicutes (22,6%), Proteobacteria (7,5%), Synergistetes (3,1%). Tỷ lệ điển hình của Firmicutes so với Bacteroidetes trong hệ vi khuẩn là 0,36: 1. Phân tích sâu hơn cho thấy rằng, trong ngành Bacteroidetes, bộ đông nhất là Bacteroidales (60,8%). Bộ đông nhất của ngành Firmicutes là Clostridiales (16,4%), tiếp theo là Selenomonadales (2,8%). Bộ Synergistales thuộc ngành Synergistetes chiếm khoảng 3,1% ORF và Aeromonadales – một bộ của ngành Proteobacteria, cũng chiếm tỷ lệ nhất định trong dữ liệu. Ở cấp độ chi, *Prevotella* (35,3%) và *Bacteroides* (16%) thuộc về bộ Bacteroidales là phổ biến nhất. Phân tích 13 loài phong phú nhất cho kết quả là tám loài thuộc chi *Prevotella*, một loài thuộc chi *Bacteroides*, hai loài thuộc bộ Clostridiales và chỉ một loài thuộc bộ Synergistales.

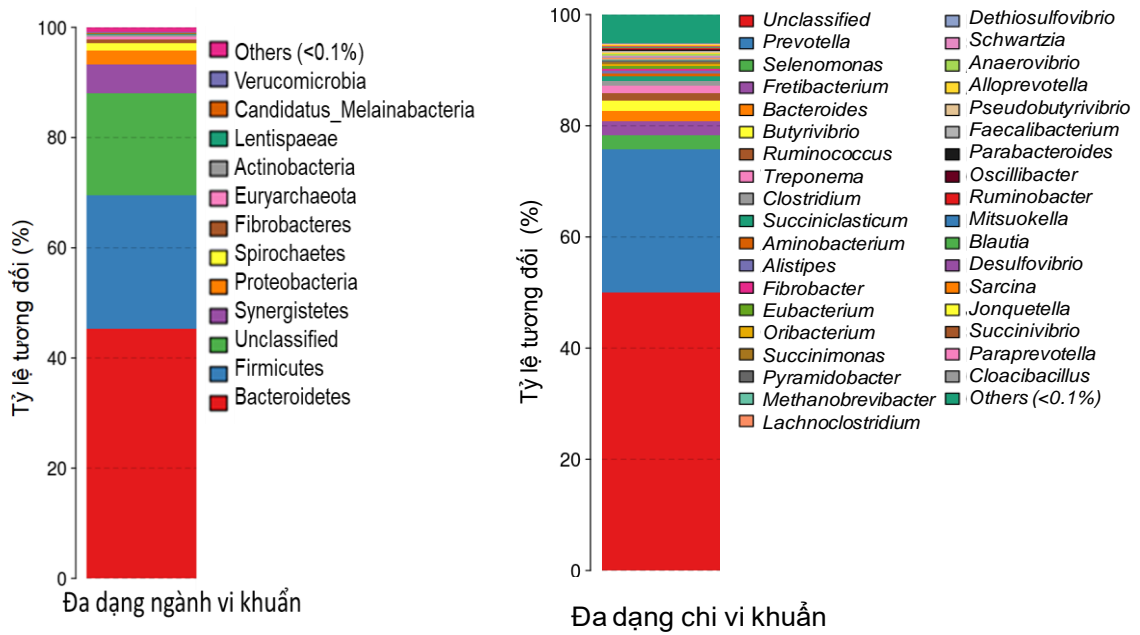


Hình 3. 3. Biểu đồ phân bố đa dạng phân loại học ở mức độ ngành và mức độ chi của vi khuẩn trong dạ cỏ dê khai thác được từ dữ liệu 8,6 Gb.

Nhìn chung, cấu trúc vi khuẩn về mặt phân loại tương tự với kết quả được mô tả trong các nghiên cứu đã được công bố về dạ cỏ động vật, đặc biệt là tỷ lệ giữa Firmicutes và Bacteroidetes. Tỷ lệ Firmicutes/ Bacteroidetes là 0,36:1, tương đương với tỷ lệ loài tương ứng trong dạ cỏ của tuần lộc Svalbard [191], cừu [192], bò Jinnan [193] và khá phù hợp với tỷ lệ 0,3:1 ở dê dạ cỏ lúc 110 ngày tuổi thích nghi với nguyên liệu thực vật [194]. Như vậy trong nghiên cứu này, tỷ lệ Firmicutes/Bacteroidetes thấp (0,36:1) cho thấy cấu trúc vi khuẩn trong dạ cỏ dê Việt Nam thích nghi tốt với chế độ ăn là các phế phụ phẩm nông nghiệp.

3.1.3.2. Đa dạng vi khuẩn trong dạ cỏ dê được đánh giá dựa trên bộ dữ liệu giải mã sâu

Kết quả giải trình tự sâu DNA metagenome vi khuẩn dạ cỏ dê thu được 48,66 Gb dữ liệu, được sắp xếp thành 3.411.867 contig với tỷ lệ bao phủ là 64,22%. Từ đó, 5.367.270 gene với tổng độ dài là 2.828.583.591 bp đã được xác định. Sử dụng phần mềm MEGAN, 4.311.093 gene (80,32% trong tổng số 5.367.270 gene) được phân vào các bậc phân loại, trong đó chỉ có một tỷ lệ nhỏ số gene thuộc về bậc phân loại không xác định (1,35%), vi khuẩn cổ (0,45%) và sinh vật nhân thật (0,14%) (Hình 3.1). So sánh với dữ liệu giải trình tự 8,6 Gb, kết quả định loại là khá tương đồng khi tỷ lệ gene được xác định có nguồn gốc vi khuẩn là 99,8%.



Hình 3. 4. Biểu đồ phân bố đa dạng phân loại học ở mức độ ngành và mức độ chi của vi khuẩn trong dạ cỏ dê khai thác được từ dữ liệu giải mã sâu

Ngành Bacteroidetes là ngành chiếm tỷ lệ lớn nhất với 45,29% tổng số gene, tiếp theo là ngành Firmicutes với 30,38%. Ngoài ra, ngành Synergistetes và Proteobacteria cũng là những ngành lớn với tỷ lệ lần lượt là 6,52% và 3,13% số gene được định loại có nguồn gốc vi khuẩn. Ở mức độ chi, 49,93% số gene vẫn chưa được phân loại. Chi phong phú nhất là *Prevotella*, chiếm 25,79% tổng số gene. Nếu tính riêng số lượng gene đã được phân loại đến cấp ngành thì chi *Prevotella* chiếm đến 32,74% số gene, tiếp theo là chi *Selenomonas* (2,62%), *Fretibacterium* (2,46%), *Bacteroides* (1,86%), *Butyrivibrio* (1,8%). Chi *Butyrivibrio* là chi vi khuẩn đóng vai trò rất quan trọng trong quá trình hydro hóa sinh học các acid béo không no C18. Các loài vi khuẩn khu trú trong dạ cỏ dê thuộc chi này có độ đa dạng rất cao, trong đó có một số lượng lớn là loài chưa được xác định. So sánh về cấu trúc vi khuẩn trong hai bộ dữ liệu MCS và MDS có thể thấy, nhờ việc tăng dung lượng và độ bao phủ của bộ dữ liệu MDS so với MCS nên tỷ lệ vi khuẩn chưa được phân loại ở bộ dữ liệu MDS thấp hơn nhiều so với bộ dữ liệu MCS. Ngược lại, có sự tương đồng về thành phần ngành vi khuẩn đa dạng nhất của cả hai bộ dữ liệu, cụ thể hai ngành Bacteroidetes và Firmicutes lần lượt là hai ngành chiếm tỷ lệ lớn nhất. Chi *Prevotella* đều là chi vi khuẩn được xác định chiếm tỷ lệ lớn nhất, gợi ý vai trò quan trọng của chi này trong quá trình chuyển hóa lignocellulose trong dạ cỏ.

3.2. Khai thác gene và thiết lập công cụ HMM cho chú giải gene, khai thác gene mã hóa protein/enzyme tham gia thủy phân lignocellulose trong dạ cỏ dê

3.2.1. Khai thác gene mã hóa enzyme thủy phân lignocellulose dựa trên cơ sở dữ

liệu KEGG**3.2.1.1. Khai thác gene từ dữ liệu giải trình tự 8,6 Gb**

Dữ liệu giải trình tự kích thước 8,6 Gb đã được phân tích, so sánh chú giải và khai thác gene dựa trên cơ sở dữ liệu KEGG (Kyoto Encyclopedia of genes and genomes - Cơ sở dữ liệu về bộ gen, con đường sinh học, bệnh tật và các chất hóa học). Kết quả khai thác gene tiền xử lý lignocellulose được thể hiện ở Bảng 3.3.

Bảng 3. 3. Kết quả khai thác gene mã hóa enzyme tiền xử lý lignocellulose từ dữ liệu giải trình tự 8,6 Gb dựa trên dữ liệu KEGG

Họ enzyme (ORFs)	COG (số ORF)	COG/KEGG/GO (số ORF)	EC
CE1 (161)	COG3693 (1)	Beta-1,4-xylanase/enterochelin esterase (1)	--
		Enterochelin esterase và enzyme liên quan(122)	--
CE1-GH10 (60)		Enterochelin esterase	
CE12 (104)	COG2755 (22)	Lysophospholipase L1 và esterases liên quan (22)	--
	COG4677 (4)	Pectin methylesterase/pectinesterase (4)	3.1.1.11
		Pectinesterase (24)	3.1.1.11
CE4 (66)	COG0726 (7)	Xylanase/chitin deacetylase (7)	--
	COG0726 (1)	Xylanase/chitin deacetylase (1)	3.5.1.-
	COG0726 (3)	Xylanase/chitin deacetylase (3)	3.5.1.41
		Polysaccharide deacetylase (12)	--
CE6 (103)	COG0656 (2)	Aldo/keto reductases, tương tự diketogulonate reductase (2)	--
	COG2272 (5)	Carboxylesterase type B (28)	3.1.1.1
		enterochelin esterase và enzyme liên quan (27)	--
CE8 (75)	COG4677 (11)	Pectin methylesterase/pectinesterase (11)	3.1.1.11
		Pectinesterase (37)	3.1.1.11
PL1 (108)	COG3866 (6)	Pectate lyase (6)	4.2.2.2
	COG3866	Pectate lyase (4)	
		Pectate lyase (32)	4.2.2.2
		Pectinesterase (1)	3.1.1.11
PL10 (36)		Pectate lyase (2)	4.2.2.2
		Pectate lyase (3)	
		Pectinesterase (9)	3.1.1.11
PL9 (8)		Pectate lyase (4)	4.2.2.2
CE1-CBM4 (7), CE13 (3), CE2 (33), CE3-GH11-CBM22 (1), CE6-GH43 (1), CE6-GH95 (1), CE7-GH26 (18), CE7 (29), PL9 (1)		Pectate lyase (6)	
Tổng số	821 ORF		

Từ dữ liệu giải trình tự DNA, 821 ORF chứa domain carbohydrate esterase (CE) và polysaccharide lyase (PL) tham gia vào quá trình tiền xử lý trong chuyển hóa lignocellulose cụ thể là lignin, 816 ORF mã hóa 11 họ glycoside hydrolase (GH) có hoạt tính cellulase, 2252 ORF mang 22 họ GH có hoạt tính hemicellulase đã được khai thác. Hầu hết các CE đều có hoạt tính pectin esterase, riêng CE6 ước đoán có hoạt tính khử của reductase và carboxylesterase có khả năng tham gia trong khử các hợp chất của lignin. Trong số này, 61 ORF được chú giải có hai chức năng

là hemicellulase- (GH10) và esterase (CE1). Thêm vào đó, có 18 ORF mã hóa cho enzyme có hai vùng chức năng hemicellulase GH26 và esterase CE7, 1 ORF có chứa hai vùng chức năng của hemicellulase GH43 và esterase CE6. Hầu hết các enzyme thuộc họ PL trong dữ liệu này đều có hoạt tính thủy phân pectin.

Các ORF được chú giải mã hóa enzyme tham gia phân giải hemicellulose bao gồm 2252 ORF. Các ORF ước đoán mã hóa cho 22 hemicellulase bao gồm endo-1,4-beta-xylanase (67 ORF), beta-xylosidase (27 ORF) và nhóm enzyme phân cắt nhánh bao gồm: alpha-galactosidase (62 ORF), alpha-galactosidases/6-phospho-beta-glucosidase (1 ORF), alpha-glucosidase (74 ORF), alpha-glucuronidase (37 ORF), alpha-L-arabinofuranosidase (93 ORF), alpha-mannosidase (1 ORF), alpha-N-acetylgalactosaminidase (1 ORF), alpha-N-arabinofuranosidase (57 ORF), arabinan endo-1,5-alpha-L-arabinosidase (36 ORF), arabinogalactan beta-galactosidase (1 ORF), arabinogalactan endo-1,4-beta-galactosidase (78 ORF), beta-galactosidase (216 ORF), beta-glucuronidase (30 ORF), beta-xylosidase/alpha-N-arabinofuranosidase (5 ORF), beta-xylosidase/arabinan endo-1,5-alpha-L-arabinosidase (5 ORF), carboxylesterase type B (1 ORF), endopolygalacturonase (35 ORF), galacturan 1,4-alpha-galacturonidase (3 ORF), mannan endo-1,4-beta-mannosidase (63 ORF), unsaturated glucuronyl hydrolase (2 ORF).

Thêm vào đó, một số ORF không được nhận biết tham gia thủy phân lignocellulose bằng CAZy nhưng lại được các cơ sở dữ liệu khác (COG/KEGG/GO) ước đoán là tham gia thủy phân lignocellulose bao gồm endo-beta-xylanase (2 ORF), endo-beta-xylanase cellulase (3 ORF), endo-beta-xylanase xyn5a (2 ORF), maltodextrin glucosidase (1 ORF), mannan endo-beta-mannosidase (2 ORF), xylan beta-xylosidase (1 ORF).

Bảng 3. 4. Danh sách enzyme thủy phân hemicellulose mã hóa từ gene trong dữ liệu giải trình tự 8,6 Gb dựa trên dữ liệu KEGG

GHs (số ORF)	COG (số ORF)	COG/KEGG/GO (số ORF)
CE1 (1)	COG3693 (1)	Endo-1,4-beta-xylanase (1)
GH10 (54), GH10-CE1 (25)	COG3693 (23)	Endo-1,4-beta-xylanase (52)
GH10-CBM6 (16), GH10-CBM0 (9), GH10-CBM22 (1), GH10-CBM9 (1)		Endo-1,4-beta-xylanase (12)
GH11 (1)		Endo-1,4-beta-xylanase (1)
GH2 (358)	COG3250 (12)	Beta-galactosidase (179)
	COG3250 (1)	Beta-glucuronidase (30)
GH2-CBM32 (4)		Beta-galactosidase (1)
GH26 (48), GH26-CBM35 (17), GH26-CE7 (42)		Mannan endo-1,4-beta-mannosidase (63)
GH27 (11), GH27-CBM35 (8)		Alpha-galactosidase (10)
		Alpha-N-acetylgalactosaminidase (1)
GH28 (194)	COG4225 (2)	Unsaturated glucuronyl hydrolase (2)
	COG5434 (35)	Endopolygalacturonase (35)
		Endo-1,4-beta-xylanase (1)
		Galacturan 1,4-alpha-galacturonidase (3)

GH30 (16)		Glucosylceramidase (7)
GH35 (64), GH35-CBM32 (11)	COG1874 (11)	Beta-galactosidase (31)
GH36 (52)	COG3345 (2)	Alpha-galactosidase (52)
GH38-CBM32 (1)		Alpha-mannosidase (1)
GH4 (2)	COG1486 (1)	6-phospho-alpha-glucosidase (1)
	COG1486 (1)	Alpha-galactosidases/6-phospho-beta-glucosidase (1)
GH42 (1)		Beta-galactosidase (1)
GH43 (492)	COG3507 (4)	Beta-xylosidase (10)
	COG3940 (2)	Beta-xylosidase/alpha-N-arabinofuranosidase (2)
	COG3507 (5)	Beta-xylosidase/arabinan endo-1,5-alpha-L-arabinosidase (5)
	COG3507 (3)	Beta-xylosidase/alpha-N-arabinofuranosidase (3)
	COG3507 (17)	Beta-xylosidase (17)
	COG2017 (35)	Galactose mutarotase/aldose 1-epimerase (35)
		Arabinan endo-1,5-alpha-L-arabinosidase (32)
		Alpha-N-arabinofuranosidase (56)
GH43-CBM13 (29)		Arabinan endo-1,5-alpha-L-arabinosidase (4)
GH51 (138)	COG3534 (62)	Alpha-L-arabinofuranosidase (93)
GH53 (76)	COG3867 (7)	Arabinogalactan endo-1,4-beta-galactosidase (75)
GH53-CBM61 (3)	COG3867 (3)	Arabinogalactan endo-1,4-beta-galactosidase (3)
GH67 (58)	COG3661 (34)	Alpha-glucuronidase (37)
GH97 (178)		Alpha-glucosidase (74)
	COG2731 (1)	Beta galactosidase (4)
		Carboxylesterase type B (1)
Arabinogalactan beta-galactosidase (1), endo-beta-xylanase (2), alpha-n-arabinofuranosidase (1), endo-beta-xylanase cellulase (3), endo-beta-xylanase xyn5a (2), maltodextrin glucosidase (1), mannan endo-beta-mannosidase (2), xylan beta-xylosidase (1)		
GH10-CBM48-CE1 (1), GH113 (1), GH115 (121), GH120 (4), GH125 (4), GH127 (62), GH2-CBM57 (9), GH26-CBM13 (2), GH43-CBM0 (1), GH43-CBM32 (17), GH43-CBM6 (100), GH43-CBM61 (1), GH64-CBM6 (1)		
Tổng số:	2252 ORF	

Tất cả endo-1,4-beta-xylanase đều thuộc họ GH10, trong đó 30,6% trình tự có chứa vùng bám carbohydrate CBM (CBM0, CBM6, CBM9, CBM22, CBM48). Tổng số 763 ORF mã hóa cho các CBM tham gia liên kết cellulose, hemicellulose và polysaccharide được xác định. Trong số các CBM này có một trình tự mã hóa CBM63 được nghiên cứu là trình tự có chức năng của expansin làm nới lỏng cấu trúc cellulose, bám cellulose giúp endoglucanase thủy phân tốt cellulose.

Từ kết quả khai thác gene từ dữ liệu giải trình tự 8,6 Gb, trình tự [denovogenes]_47103 mã hóa endoglucanase, trình tự [denovogenes]_57823 mã hóa expansin, trình tự [denovogenes]_7603 mã hóa alpha-D-glucuronidase, trình tự [denovogenes]_44301 mã hóa pectinesterase, trình tự [denovogenes]_3070 mã hóa cellobiose-phosphorylase đã được lựa chọn để thiết kế vector biểu hiện trong chủng *E. coli*. Kết quả biểu hiện, tinh chế và thử hoạt tính của trình tự [denovogenes]_57823 mã hóa expansin đã được công bố trong một bài báo nằm trong tuyển tập Báo cáo khoa học Hội nghị Công nghệ sinh học toàn quốc năm 2018 và một bằng độc quyền giải pháp hữu ích (các công bố khoa học liên quan đến luận án).

3.2.1.2. Khai thác gene từ dữ liệu giải trình tự sâu 48,6 Gb

Để khai thác gene từ dữ liệu giải trình tự sâu 48,6 Gb, các trình tự ORF được dự đoán từ kết quả giải trình tự đã được phân tích so sánh với các cơ sở dữ liệu Nr, Swissprot, KEGG, eggNOG. Kết quả tổng hợp khai thác gene mã hóa enzyme phân giải lignocellulose được thể hiện ở Bảng 3.5.

Bảng 3. 5. Bảng tổng hợp các enzyme lignocellulase khai thác từ hai bộ dữ liệu giải trình tự

Enzyme	Mã EC	Tên viết tắt	Bộ dữ liệu 8,6 Gb	Bộ dữ liệu 48,7 Gb					
				Tổng số	Thiếu đầu 3'	Thiếu đầu 5'	Thiếu hai đầu	Hoàn chỉnh	% Gene hoàn chỉnh
Cellulases			448	21029	5786	3816	9091	2336	11.1
Licheninase	3.2.1.73	LCN	1	281	49	62	68	102	36.3
Endoglucanase	3.2.1.4	EG	66	7368	1871	1406	3081	1010	13.7
Cellobiose phosphorylase	2.4.1.20	CPB	27	1439	424	268	649	98	6.8
Beta glucosidase	3.2.1.21	BGL	341	10444	3001	1780	4702	961	9.2
6-Phospho-beta-glucosidase	3.2.1.86	PBGL	13	1281	393	274	501	113	8.8
Cellobiose dehydrogenase	1.1.99.18		0	0					
Cellobiohydrolase	3.2.1.91	CPH	0	216	48	26	90	52	24.1
Hemicellulases			815	41756	12007	6658	17199	5892	14.1
Acetyl mannan esterase	3.1.1.6	AcME	0	0					
Acetylxylan esterase	3.1.1.72	AcXE	0	4	0	0	4	0	0.0
Alpha-D-xyloside xylohydrolase	3.2.1.177	AXXH	0	2833	840	460	1277	256	9.0
Alpha-glucuronidase	3.2.1.139	AGLC	37	561	155	109	224	73	13.0
Beta-D-glucuronidase	3.2.1.31	BGLC	30	888	292	127	343	126	14.2
Beta-fructofuranosidase	3.2.1.26	BFF	0	1106	368	197	427	114	10.3
Beta-mannosidase	3.2.1.25	BMAS	62	659	189	90	329	51	7.7
Endo-transglycosylase/hydrolase	2.4.1.207	XET/H	0	2	0	1	1	0	0.0
Endo-β-1,4 xylanase	3.2.1.8	XLA	67	3400	938	634	1121	707	20.8
Oligosaccharide reducing-end xylanase	3.2.1.156	OREX	0	1213	422	191	363	237	19.5
Xylan 1,4-beta-xylosidase	3.2.1.37	XBX	10	1018	262	197	337	222	21.8
Xyloglucan-active β-D-galactosidase	3.2.1.23	BGAL	290	11690	3326	1705	5490	1169	10.0
Alpha-galactosidase	3.2.1.22	AGAL	63	4020	1174	693	1638	515	12.8
Alpha-L-arabinofuranosidase	3.2.1.55	ARA	138	6229	1882	891	2282	1174	18.8
Alpha -L-fucosidase	3.2.1.51	FUC	55	6896	1890	1136	2663	1207	17.5
Beta-mannanase	3.2.1.78	BMAN	63	1237	269	227	700	41	3.3
Tiền xử lý			167	4769	1382	782	1845	760	15.9
Expansin			1	33	5	13	9	6	18.2
Feruloyl esterase	3.1.1.73	FAE	0	92	35	9	17	31	33.7
Laccases	1.10.3.2		0	9	2	2	4	1	11.1
Lignin peroxidase	1.11.1.14	LiP	0	0					

Lytic polysaccharide monooxygenase	1.14.99.56	LPMO	0	11	3	0	7	1	9.1
manganese peroxidase	1.11.1.13	MnP	0	0					
Pectinesterase	3.1.1.11	PE	86	2525	704	483	919	419	16.6
Exopolygalacturonase	3.2.1.67	EPG	38	635	216	66	223	130	20.5
Exopolygalacturonase lyase	4.2.2.9	EPGL	0	59	8	8	37	6	10.2
Endopolygalacturonase	3.2.1.15	ENPG	0	27	4	7	14	2	7.4
Endopolygalacturonase lyase	4.2.2.2	ENPGL	42	1378	405	194	615	164	11.9

Từ kết quả giải trình tự sâu DNA metagenome vi khuẩn dạ cỏ dê thu được 48,66 Gb dữ liệu, 5.367.270 gene với tổng độ dài là 2.828.583.591 bp đã được xác định. Trong số các gene trên, có 4.385.296 gene đã được ước đoán chức năng dựa trên việc so sánh trình tự protein tương ứng với các cơ sở dữ liệu Nr, Swissprot, KEGG, eggNOG. Cụ thể, với cơ sở dữ liệu KEGG, 2.809.791 gene đã được ước đoán chức năng trong đó 317.154 gene (11,3%) được xác định là liên quan đến quá trình chuyển hóa carbohydrate.

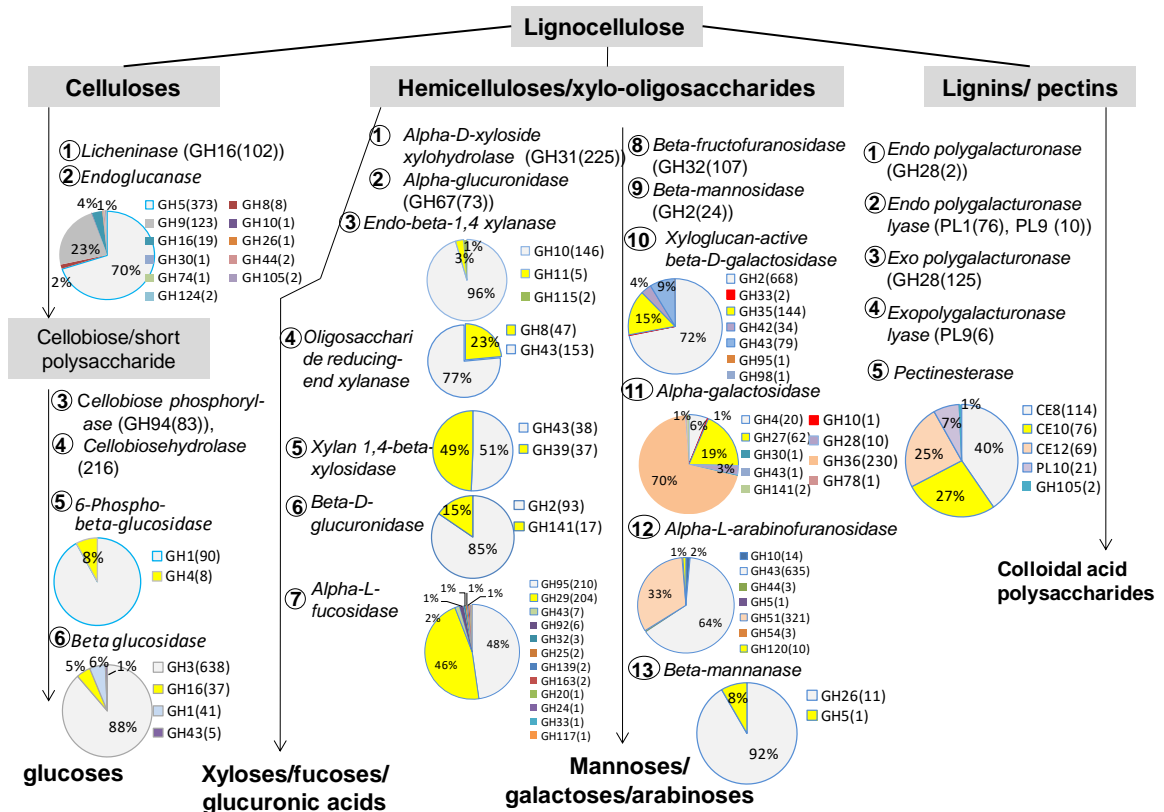
Theo lý thuyết, năm loại enzyme cellulase khác nhau đóng vai trò quan trọng trong quá trình phân giải cellulose trong hai giai đoạn: giai đoạn thứ nhất, licheninase (EC 3.2.1.73) và endo-glucanase (EC 3.2.1.4) phân cắt ngẫu nhiên mạch dài cellulose thành các đoạn ngắn oligosacchride và cellobiose, tiếp theo các enzyme cellobiose phosphorylase (EC 2.4.1.20), 6-phospho-beta glucosidase (EC 3.2.1.86) và beta-glucosidase (EC 3.2.1.21) sẽ phân giải các sản phẩm trên thành đường đơn D-glucose. Trong kết quả khai thác được, tất cả các enzyme licheninase đều thuộc về họ glycosyl hydrolase 16 (GH16), hầu hết các cellobiose phosphorylase đều thuộc họ GH94, trong khi beta-glucosidase thuộc về các họ GH3, GH16, GH1, GH4, trong đó beta-glucosidase GH3 chiếm đa số với tỷ lệ đến 88%. Enzyme endo-glucanase khai thác được rất đa dạng, thuộc về 11 họ GH, tuy nhiên họ GH5 chiếm tỷ lệ lớn nhất (70%), tiếp đó là GH9 (23%).

Quá trình chuyển hóa hemicellulose thành đường đơn trong dạ cỏ dê được xúc tác bởi ít nhất 13 loại enzyme khác nhau, trong đó alpha-D-xyloside xylohydrolase (EC 3.2.1.177), alpha-glucuronidase (EC 3.2.1.139), endo-xylanase (EC 3.2.1.8), oligosaccharide reducing-end xylanase (EC 3.2.1.156), xylan 1,4-beta xylosidase (EC 3.2.1.72), beta-D-glucuronidase (EC 3.2.1.31), alpha-L fucosidase (EC 3.2.1.51) xúc tác chuyển hóa hemicellulose và xylo-oligosaccharide thành các đơn phân xylose, fucose và glucuronic acid. Các endo-xylanase được xác định thuộc về các họ GH10, GH11 và GH115 trong đó endo-xylanase GH10 chiếm đa số với tỷ lệ

96%. Điều này khẳng định enzyme thuộc họ glycosyl hydrolase 10 đóng vai trò rất quan trọng trong việc phân giải hemicellulose thành các sản phẩm oligosaccharide trong dạ cỏ dê. Các alpha-D-xyloside xylohydrolase chỉ thuộc họ GH31 còn alpha-glucuronidase chỉ thuộc họ GH67. Oligosaccharide reducing-end xylanase được xác định thuộc họ GH8 và GH43, trong đó enzyme thuộc họ GH43 chiếm 77%; tương tự beta-D-glucuronidase thuộc hai họ GH2 và GH141 trong đó họ GH141 chiếm phần lớn với 85% số enzyme. Các xylan 1,4-beta-xylosidase phân bố tương đối đều thuộc về hai họ GH43 (51%) và GH39 (49%). Các alpha-L-fucosidase được chú giải thuộc 12 họ GH khác nhau, trong đó chiếm tỷ lệ lớn nhất là họ GH95 (48%) và GH29 (46%).

Mặt khác, các enzyme beta-fructofuranosidase (EC 3.2.1.26), beta-mannosidase (EC 3.2.1.25), xyloglucan-active beta-D-galactosidase (EC 3.2.1.23), alpha-galactosidase (EC 3.2.1.22), alpha-L-arabinofuranosidase (EC 3.2.1.55), beta-mannanase (EC 3.2.1.78) xúc tác chuyển hóa hemicellulose và xylo-oligosaccharide thành các đơn phân mannose, galactose và arabinose. Các enzyme beta-fructofuranosidase đều thuộc họ GH32 trong khi các enzyme beta-mannosidase chỉ thuộc họ GH2. Các enzyme beta-mannanase thuộc hai họ GH là GH26 và GH5, trong đó beta-mannanase GH26 chiếm 92%. Ngược lại, các enzyme xyloglucan-active beta-D-galactosidase, alpha-galactosidase, alpha-L-arabinofuranosidase đều thuộc rất nhiều họ GH khác nhau, trong đó beta-galactosidase GH2, alpha-galactosidase GH36 và alpha-L-arabinofuranosidase GH43 là những nhóm chiếm ưu thế của từng loại enzyme.

Bên cạnh cellulose và hemicellulose, lignin và pectin trong cấu trúc lignocellulose được thủy phân bởi 5 loại enzyme chính bao gồm endo-polygalacturonase (EC 3.2.1.15) với 2 trình tự thuộc họ GH28, endo-polygalacturonate lyase (EC 4.2.2.2) với 86 trình tự thuộc 2 họ polysaccharide lyase (PL) 1 và 9, exo-polygalacturonase với 125 trình tự thuộc họ GH28, exo-polygalacturonate lyase (EC 4.2.2.9) với 6 trình tự thuộc họ PL9, và 282 trình tự pectin esterase thuộc 3 họ carbohydrate esterase (CE8, CE10, CE12), một họ PL10 và một họ GH105.



Hình 3. 5. Bức tranh tổng quan về các họ GH/CE/PL liên quan đến quá trình phân giải lignocellulose của vi khuẩn trong dạ cỏ dê

(chú thích: họ GH (số lượng gene))

Lignocellulose trong chế độ ăn của dê bao gồm ba thành phần chính: cellulose, hemicellulose và lignin/pectin. Trong dạ cỏ dê, cellulose được tiêu hoá chủ yếu bằng 6 loại enzyme để giải phóng glucose. Hemicellulose bị phân hủy bởi ít nhất 13 enzyme để tạo ra xylose, fucose, glucuronic acid, mannose, galactose, arabinose. Lignin/pectin bị thủy phân bởi 5 loại enzyme chiếm ưu thế để tạo ra các polysaccharide acid keo. Trong cơ sở dữ liệu, chúng tôi tìm thấy 225 gene mã hóa alpha-D-xyloside xylohydrolase GH31 nên được ký hiệu là alpha-D-xyloside xylohydrolase (GH31(225)). Một số enzyme có GH rất đa dạng ví dụ endoglucanase.

3.2.2. Phân tích đa dạng vi khuẩn mang gene thủy phân lignocellulose

3.2.2.1. Đa dạng vi khuẩn mang gene phân giải lignocellulose khai thác được từ dữ liệu 8,6 Gb

Trong số 816 ORF mã hóa gene cellulase, 2252 ORF mã hóa gene hemicellulase và 821 ORF mã hóa gene tiền xử lý đã được xác định, chỉ có 221 gene cellulase, 544 gene hemicellulase và 226 gene tiền xử lý là có thể được định dạng phân loại, chiếm khoảng 24-27%, phần còn lại không được phân loại là rất lớn.

Nhìn chung, phần lớn các gene trên thuộc ngành Bacteroidetes, cụ thể là 854 ORF trên tổng số 991 ORF, chiếm tỷ lệ 86,2%. Ngành chiếm tỷ lệ lớn thứ hai là ngành Firmicutes với 94 ORFs (9,5%). Tỷ lệ ORF theo ngành Firmicutes/Bacteroidetes là 0,11:1 và thấp hơn tỷ lệ chung của toàn bộ quần xã vi khuẩn. Ở mức phân loại bộ, bộ Bacteroidales có số lượng ORF lớn nhất, tiếp đó là Clostridiales. Ở cấp độ chi, có 398 ORF thuộc chi *Prevotella*, 300 ORF thuộc về chi *Bacteroides*, 39 ORF được dự đoán có nguồn gốc từ *Paraprevotella*, 39 ORF có nguồn gốc từ *Ruminococcus*. Đặc biệt chi *Aerophaga* được dự đoán chỉ chứa các ORF cho hemicellulase (7 ORF), trong khi chi *Algoriphagus* được dự đoán chỉ mang ORF cho CE và PL.

Khi xem xét chi tiết hơn, có sự khác biệt về nhóm phân loại trong đa dạng nguồn gốc vi sinh vật của từng nhóm enzyme phân hủy lignocellulose. Với nhóm enzyme tiền xử lý (họ CE và PL), ngành Bacteroidetes chiếm tỷ lệ lớn nhất với 201 ORF (88,9%), tiếp đó là ngành Firmicutes với 14 ORF (6,2%), dẫn đến tỷ lệ ORF theo ngành Firmicutes/Bacteroidetes giảm rất thấp, 0,07:1. Với những ORF mã hóa enzyme esterase, tỷ lệ này có sự khác biệt đáng kể. Cụ thể, 56 ORF mã hóa CE1 có nguồn gốc từ nhiều ngành vi khuẩn như Bacteroidetes, Firmicutes, Planctomycetes, Spirochaetes, trong khi 10 ORF mã hóa CE1 có liên kết các vùng CBM (Carbohydrate-binding domain), cụ thể là CBM48 thì chỉ thuộc về ngành Bacteroidetes. Điều này cũng tương tự với CE6 và PL1, 21 ORF mã hóa CE6 có nguồn gốc từ Bacteroidetes và Verrucomicrobia, trong khi 45 ORF mã hóa PL1 có nguồn gốc từ Bacteroidetes và Firmicutes. Tuy nhiên, enzyme mang đa vùng hoạt tính CE6-GH95 và PL1 mang CBM6 hoặc vùng tương tự fibronectin-3 (FN3) chỉ có nguồn gốc từ Bacteroidetes. Ngoài ra, các enzyme mang vùng hoạt tính CE12, CE7, CE7-GH26, CE8, CE8-Ig (vùng tương tự immunoglobulin), PL9, PL10 được xác định chỉ từ Bacteroidetes, trong khi CE13 chỉ từ Firmicutes. Ở cấp độ loài, *Prevotella buccae* là loài có nhiều gene lignocellulase nhất.

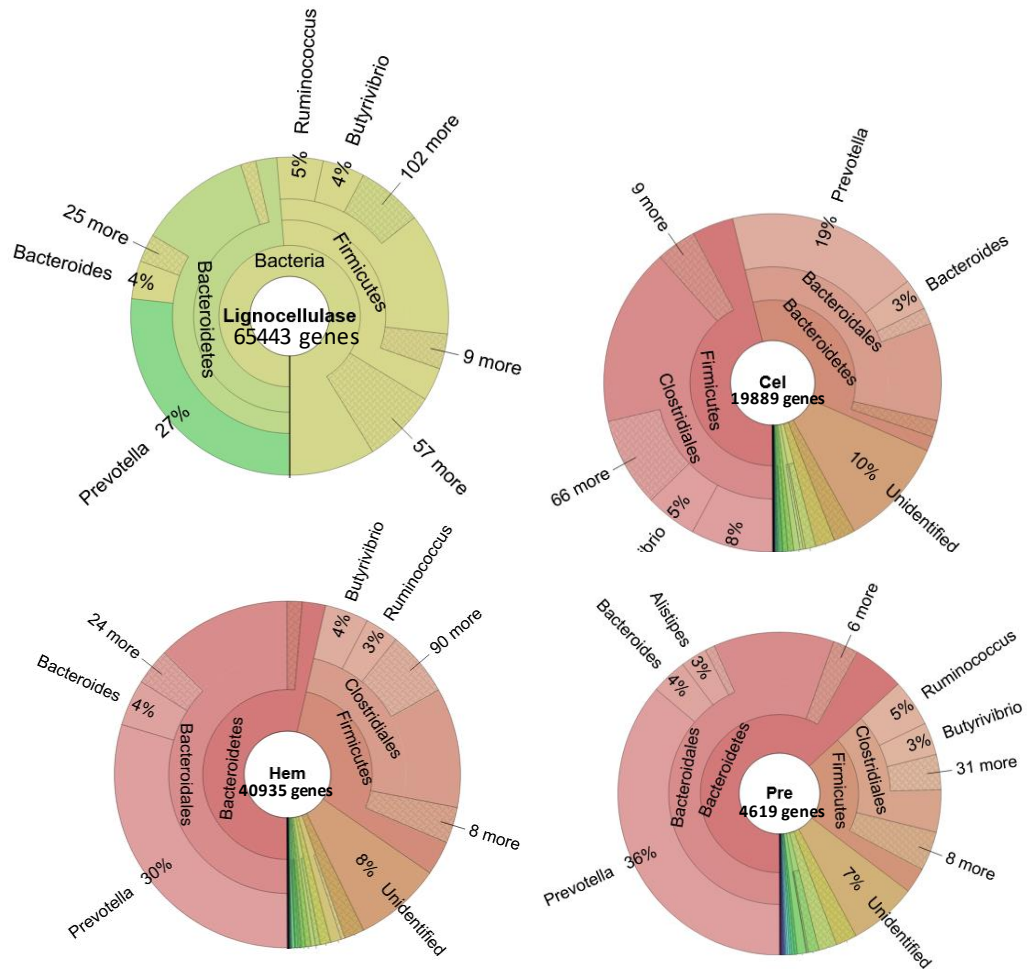
Nhóm hemicellulase là đa dạng nhất so với cellulase và gene tiền xử lý lignocellulose, với 20 nhóm GHs khác nhau. Trong nhóm này, ngành phong phú nhất tiếp tục là Bacteroidetes với 500 ORF (chiếm 91,9%), tiếp theo là Firmicutes (27 ORF), và sau đó là Proteobacteria (7 ORF). Do đó, tỷ lệ Firmicutes so với Bacteroidetes lại thấp là 0,05: 1. Ở đây, GH2 được ước đoán có nguồn gốc từ các ngành đa dạng nhất như Actinobacteria, Bacteroidetes, Firmicutes, Proteobacteria, Thermotogae. Nhóm chiếm tỷ lệ lớn tiếp theo là GH10, GH15, GH53. Firmicutes và Bacteroidetes cũng mang các gene thuộc sáu họ GH gồm GH2, GH28, GH36, GH6, GH51, GH53. Tuy nhiên, nhiều GH khác bao gồm các enzyme đa chức năng cũng chứa vùng hoạt tính CE hoặc CBM, và được dự đoán là chỉ thuộc

Bacteroidetes, trong khi GH113, GH26-CBM35 chỉ thuộc Firmicutes. Ở cấp loài, *Prevotella bergensis*, *P. bryantii*, *P. copri*, *Paraprevotella clara*, *B. graminisolvens*, *P. albensis*, *P. paludivivens*, *B. cellulosilyticus*, *P. buccae*, *B. eggerthii*, *P. ruminicola* là những loài mang gene hemicellulase phong phú nhất. Để phân hủy hemicellulose hiệu quả, các gene của *B. graminisolvens* có thể mang chín loại vùng hoạt tính khác nhau, con số này với *P. copri* và *P. paludivivens* là tám. Trong nhóm vi khuẩn tiền xử lý lignocellulose, *Prevotella* sp. MSX73 được dự đoán là phân hủy tiềm năng nhất với sáu vùng xúc tác khác nhau. Điều thú vị là có rất nhiều loài chứa gene cho rất nhiều chức năng phân hủy lignocellulose. Ví dụ, *Prevotella albensis* được dự đoán là mang cả hai loại cellulase và hemicellulases khi vi khuẩn này sở hữu cả hai vùng hoạt tính cellulase GH3-FN3, GH9-Ig và hemicellulase GH10, GH2, GH30, GH43-CBM13, GH43-CBM6, GH97. Bảy loài chứa cả GHs cho hemicellulases và CE và PLs để tiền xử lý lignocellulose. Cụ thể hơn, *B. cellulosilyticus* và *P. bryantii* có 10 gene cho các vùng xúc tác khác nhau.

3.2.2.2. Đa dạng vi khuẩn mang gene phân giải lignocellulose khai thác từ kết quả giải trình tự sâu

Dựa trên kết quả chú giải chức năng theo cơ sở dữ liệu KEGG, có tổng cộng 65554 gene mã hóa cho 30 loại enzyme/protein liên quan đến phân giải lignocellulose đã được xác định. So sánh với kết quả giải trình tự 8,6 Gb, với dữ liệu giải trình tự sâu có khối lượng dữ liệu lớn gấp 5,7 lần, tương ứng với tỷ lệ lắp ráp của các contigs so với các reads tăng 2,35 lần từ 27,3% lên 64,22%, số lượng gene liên quan đến phân giải lignocellulose khai thác được tăng 47 lần. Trong số 65.554 gene, 8988 gene được xác định là gene hoàn thiện, các gene này được so sánh với cơ sở dữ liệu CAZy để xác định các vùng chức năng protein.

Tất cả 65.554 gene mã hóa cho 30 enzyme/protein liên quan đến quá trình phân hủy lignocellulose trong dạ cỏ của dê đều được đưa vào phần mềm MEGAN. Kết quả cho thấy có 65.443 gene được xếp vào các đơn vị phân loại (99,85%). Trong đơn vị phân loại chi, chi lớn nhất là chi *Prevotella* đóng góp 27% gene liên quan đến quá trình phân giải lignocellulose, tiếp theo là *Ruminococcus* (5%) và *Bacteroides* (4%). Đáng chú ý, *Prevotella* đóng góp rất nhiều cho quá trình phân giải hemicellulose và tiền xử lý lignocellulose, với việc chi này đóng góp tới 30% số gene chuyển hóa hemicellulose và 36% gene tiền xử lý lignocellulose (Hình 3.6).



Hình 3. 6. Đa dạng phân loại vi khuẩn mang gene lignocellulase trong dạ cỏ dê Việt Nam đã được KEGG chú thích và phân loại bởi MEGAN

Lignocellulase: toàn bộ gene mã hóa enzyme phân giải lignocellulose; Cel: gene mã hóa cho cellulase; Hem: gene mã hóa cho hemicellulase; Pre: các gene mã hóa protein/enzyme liên quan đến tiền xử lý lignocellulose

3.2.2.3. Vai trò của chi *Prevotella* trong phân giải lignocellulose

Phân tích sâu hơn cho thấy rằng trong nghiên cứu này *Prevotella* không phát hiện có gene mã hóa laccase, lytic polysaccharit monooxygenase, expansin hoặc cellobiohydrolase, endo-transglycosylase/hydrolase, phospho beta glucosidase, nhưng đóng góp rất nhiều cho quá trình phân giải hemicellulose với nhiều enzyme phân nhánh đặc biệt là acetyl xylan esterase, feruloyl esterase, alpha-D-xyloside xylohydrolase (Bảng 3.6).

Bảng 3. 6. Các gene của *Prevotella* mã hóa enzyme tham gia thủy phân lignocellulose

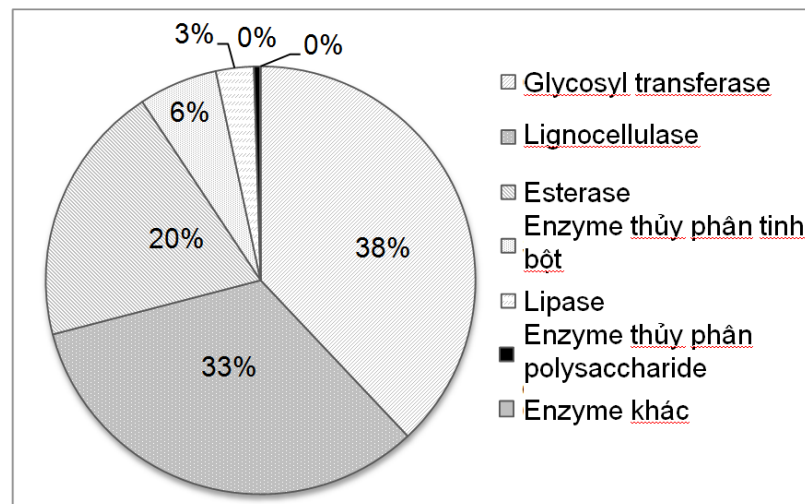
Enzyme	Ký hiệu	<i>Prevotella</i>			Nguồn gốc khác	Tổng số chung	Tỷ lệ % gene từ <i>Prevotella</i>
		Gene chú giải theo KEGG	Gene chú giải theo CAZy-HMMER	Tổng số			
Acetylxylan esterase	AcXE	0	327	327	4	331	98,8%
Feruloyl esterase	FAE	0	38	38	15	53	71,7%
Alpha-D-xyloside xylohydrolase	AXXH	1508		1508	1191	2699	55,9%
Endopolygalacturonase lyase	ENPGL	608		608	745	1353	44,9%
Exopolygalacturonase	EPG	271		271	364	635	42,7%
Alpha-glucuronidase	AGLC	227		227	317	544	41,7%
Oligosaccharide reducing-end xylanase	OREX	453		453	744	1197	37,8%
Endopolygalacturonase	ENPG	9		9	15	24	37,5%
Alpha-L-arabinofuranosidase	ARA	2258		2258	3888	6146	36,7%
Beta-D-glucuronidase	BGLC	309		309	577	886	34,9%
Licheninase	LCN	92		92	188	280	32,9%
Xyloglucan-active β -D-galactosidase	BGAL	3741		3741	7848	11589	32,3%
Pectinesterase	PE	774		774	1719	2493	31%
Cellobiose phosphorylase	CPB	416		416	1014	1430	29,1%
Alpha-L-fucosidase	FUC	1953		1953	4803	6756	28,9%
Beta glucosidase	BGL	2221		2221	7713	9934	22,4%
Beta-mannanase	BMAN	227		227	945	1172	19,4%
Endo- β -1,4 xylanase	XLA	583		583	2610	3193	18,3%
Endoglucanase	EG	948	258	1206	5592	6798	17,7%
Exopolygalacturonase lyase	EPGL	10		10	48	58	17,2%
Alpha-galactosidase	AGAL	620		620	3364	3984	15,6%
Xylan 1,4-beta-xylosidase	XBX	132		132	883	1015	13%
Beta-fructofuranosidase	BFF	97		97	1007	1104	8,8%
Beta-mannosidase	BMAS	37		37	607	644	5,7%
6-Phospho-beta-glucosidase	PBGL	1		1	1278	1279	0,1%
Laccase	Laccase	0		0	5	5	0%
Endo-transglycosylase/hydrolase	XET/H	0		0	2	2	0%
Cellobiohydrolase	CPH	0		0	168	168	0%
Lytic polysaccharide monooxygenase	LPMO	0		0	5	5	0%
Expansin	Expansin	0		0	31	31	0%
Tổng		17495	623	18118	47690	65808	27,5%

Tổng số gene bao gồm 17495 gene được chú thích chức năng bởi KEGG và 327 gene AcXE, 38 gene FAE, 258 gene EG được chú thích bởi CAZy và HMMER. EG: Endoglucanase; CPB: Cellobiose phosphorylase; BGL: Beta glucosidase; PBGL: 6-phospho-beta-glucosidase; CPH: Cellobiohydrolase; AcXE: Acetylxylan esterase; AXXH: Alpha-D-xyloside xylohydrolase; AGLC: Alpha-glucuronidase; BGLC: Beta-D-glucuronidase; BFF: Beta-fructofuranosidase; BMAS: Beta-mannosidase; XET/H: Endo-transglycosylase/hydrolase; XLA: Endo-beta-1,4 xylanase; OREX: Oligosaccharide reducing-end xylanase; XBX: Xylan 1,4-beta-xylosidase; BGAL: Xyloglucan-active beta-D-galactosidase; AGAL: Alpha-galactosidase; ARA: Alpha-L-arabinofuranosidase; FUC: Alpha-L-fucosidase; BMAN: Beta-mannanase; FAE: Feruloyl esterase; PE: Pectinesterase; EPG: Exopolygalacturonase; EPGL: Exopolygalacturonate lyase; ENPG: Endopolygalacturonase; ENPGL: Endopolygalacturonate lyase.

Đây là những enzyme rất quan trọng để tăng cường chuyển đổi lignocellulose. Xét tổng số gene, có 18.118 gene mã hóa cho các loại enzyme phân giải lignocellulose là có nguồn gốc từ *Prevotella* trong tổng số 65.808 gene, chiếm tỷ lệ 27,5%. Đáng chú ý, 98,8% gene mã hóa acetyl xylan esterase là có nguồn gốc từ *Prevotella*. Điều đó cho thấy tầm quan trọng của vi khuẩn này trong chuyển hóa chuỗi bên xylan, phân giải xylan thành đường đơn.

Vi khuẩn mang gene lignocellulase trong quần xã có thể được chia thành vi khuẩn có tiềm năng phân giải lignocellulose và vi khuẩn cơ hội. Các vi khuẩn cơ hội thường chỉ sản xuất beta glucosidase để chuyển hóa các đoạn oligosaccharide ngắn thành glucose nhằm cung cấp năng lượng cho bản thân chứ không phải cho vật chủ. Ngược lại, vi khuẩn tiềm năng được đặc trưng bởi khả năng sản xuất nhiều hơn một loại enzyme để phân hủy lignocellulose. Ở vi khuẩn có khả năng phân giải lignocellulose, các gene mã hóa cho các enzyme phân hủy lignocellulose thường nằm ở những vùng nhất định trong bộ gene.

3.2.2.4. Vai trò của chi *Prevotella* trong hỗ trợ tiêu hóa các nguồn thức ăn khác trong dạ cỏ dê



Hình 3. 7. Các gene nguồn gốc từ *Prevotella* tham gia tiêu hóa trong dạ cỏ dê

Để hiểu rõ vai trò của *Prevotella* trong quá trình tiêu hóa các chất dinh dưỡng khác trong dạ cỏ của dê, 17.495 gene có nguồn gốc từ *Prevotella* mã hóa cho các enzyme/protein liên quan đến sự phân hủy lignocellulose được KEGG chú thích về mặt chức năng trong số tất cả các gene (547.751 gene) đã được tách riêng để phân tích. Các gene còn lại bao gồm các gene chưa biết chức năng đã được phân tích bằng phần mềm HMMER và so sánh với cơ sở dữ liệu CAZy, từ đó có tất cả 18.118 gene có nguồn gốc từ *Prevotella* đã được chú thích chức năng. 510.840 gene không khớp với cơ sở dữ liệu của HMMER và CAZy, nghĩa là những gene này không liên quan đến enzyme.

HMMER và CAZy phân loại các gene thành họ GH, nhưng KEGG chú thích vào số định danh enzyme (số EC). Vì vậy, chỉ một số họ GH có thể được làm rõ thành số EC. Bằng cách này, trong số các gene được chú giải chức năng, có 327 gene mã hóa cho acetylxytan esterase, 38 gene mã hóa cho feruloyl esterase, 258 gene mã hóa endoglucanase GH5 có nguồn gốc từ *Prevotella* đã được xác định để

thêm vào các gene lignocellulase được chú thích bằng KEGG.

Bảng 3. 7. Các gene từ *Prevotella* mã hóa các enzyme tham gia phân giải các chất dinh dưỡng khác được phân tích từ dữ liệu giải trình tự sâu metagenomic của vi khuẩn trong dạ cỏ dê Việt Nam

Enzyme	Số lượng gene	Enzyme	Số lượng gene	Enzyme	Số lượng gene
1. Esterases	3725	4. Lipase	537	-Hemicellulases	
CE	1621	GDSLlike Lipase	537	GH2	153
Calcineurin-like phosphoesterase	1155	5. Phân giải tinh bột	1117	GH98	131
Thioesterase	332	Alpha amylase	926	GH2-SBD	108
Type I phosphodiesterase	251	GH57	124	GH67	41
AcylACP thioesterase	176	GH66	33	GH115-GH67	35
Esterase-like activity of phytase	96	Core2/I Branching enzyme	28	CE2	32
Glycerophosphoryl diester phosphodiesterase	33	GH63	2	GH38	30
Esterase PHB depolymerase	32	Alpha amylaseCE	1	GH35	16
Phospholipase/Carboxylesterase protein thioesterase	14	GH119	1	GH43-CE	10
LigT like Phosphoesterase	5	GH126	1	GH10CECE1	9
CE-Ig	1	GH63PE	1	GH2-SBDCE	8
2. Glycosyl transferases	7153	6. Lignocellulases	6199	GH10CE	7
GT2	3562	GH116	1	GH36	7
GT1	1925	GH118	1	GH28-CE	4
GT4	429	GH122	1	GH39	3
Transglycosylase	396	GH18	47	GH53CE	3
Lipid A disaccharide synthetase	179	- Cellulase	699	GH115-GH20	2
GT11	129	GH30	165	GH115-CE1	1
GT9 (heptosyltransferase)	114	GH124	1	GH115-CE12	1
GT	113	GH128	1	GH115-GH28	1
GT28	82	GH16	20	GH115-GH33	1
GT8	67	GH26	164	GH115-GH43	1
GT90	37	GH26AcXE	4	GH117	1
GT25 (LPS biosynthesis protein)	28	GH3	252	GH120	1
GT-CBM	27	GH3-CE	1	GH121	1
GT26	24	GH8	24	GH123	1
GT10 (fucosyltransferase)	15	GH9	67	GH125	1
GT6	13	-Hemicellulases	5344	GH127	1
GT36	10	GH43	769	GH129	1
GT21	2	GH92	647	GH130	1
Mannosyltransferase	1	GH88	576	GH2-CE	1
3. Polysaccharide degradation	87	GH28	476	GH28-PE	1
Carbohydrate phosphorylase	46	GH10	448	GH43-Lipase	1
GH76	34	GH115	432	GH43-GT1	1
GH65	4	GH53	387	GH59	1
GH85	2	GH25	301	-Tiền xử lý	156
GH70	1	GH20	274	PE	156
		GH32	244	Khác	6
		GH31	173	GH56-alpha L fucosyl transferase	6

Các gene còn lại chủ yếu mã hóa cho glycosyl transferase (7.153 gene), protein/enzyme liên quan đến phân giải lignocellulose (6.199 gene), esterase (3.725

gene), enzyme thủy phân tinh bột (1.089 gene), lipase (537 gene), enzyme phân giải polysaccharide (87 gene) và các enzyme khác (116 gene) (Bảng 3.7). Trong nhóm glycosyl transferase, GT2 chiếm ưu thế nhất với 3.562 gene, sau đó là GT1 (1925 gene). Enzyme của *Prevotella* tham gia chuyển hóa tinh bột chủ yếu là alpha amylase (927 gene), sau đó là GH57, GH119, GH126, GH63, GH66 và Core2/enzyme phân nhánh. Đa dạng nhất được tìm thấy trong các enzyme lignocellulase với 39 họ GH, esterase CE2 và pectin esterase (Bảng 3.7).

Prevotella là một trong gần 40 chi quan trọng và phổ biến trong dạ cỏ dê có chức năng hỗ trợ phân giải lignocellulose cũng như chuyển hóa carbohydrate [195]. Sự cố định của vi khuẩn *Prevotella* trong dạ cỏ dê diễn ra dần dần từ ngày thứ 14 cho đến khi trở thành vi khuẩn chiếm ưu thế trong dạ cỏ dê khi khẩu phần ăn trở nên giàu chất xơ thực vật [194, 195]. Đến 110 ngày, gene nguồn gốc *Prevotella* có thể chiếm 22,37% số gene của vi khuẩn trong dạ cỏ dê. Bằng phân tích metagenomic 16S rRNA, ở những con dê được cho ăn chế độ ăn từ sữa, *Prevotella* chiếm 8,91% quần xã vi khuẩn, nhưng với những con dê sử dụng thức ăn thực vật dạng rắn, vi khuẩn trong chi này đã tăng lên 56,02% [196]. Bekele và cộng sự đã sử dụng phương pháp điều tra sự đa dạng của vi khuẩn và chỉ ra rằng *Prevotella* chiếm hơn 50% vi khuẩn trong dạ cỏ dê [197]. Thông thường trong dạ cỏ nói chung, *Prevotella* chiếm khoảng 42 - 60% [198]. Những kết quả này phù hợp với phát hiện của chúng tôi trong dạ cỏ của dê, trong đó chế độ ăn của dê bao gồm cỏ khô và nhiều loại cỏ và lá cây trên núi cũng như tàn dư thực vật. Trong dữ liệu giải trình tự metagenome thấp, *Prevotella* đã được tìm thấy là chi phong phú nhất, chiếm 35,3% số gene phân loại. Bằng giải trình tự sâu DNA metagenome của vi khuẩn trong dạ cỏ dê Việt Nam, *Prevotella* xuất hiện nhiều nhất chiếm 32,74% gene được phân loại trong các đơn vị phân loại ngành.

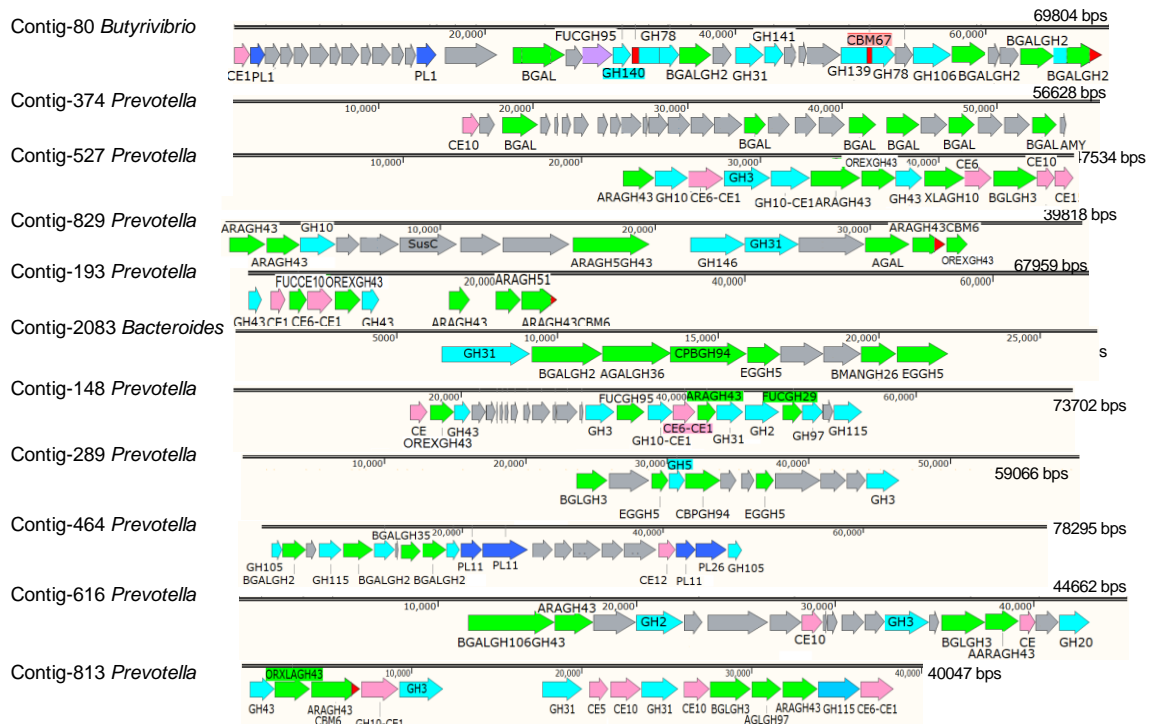
Liên quan đến quá trình phân giải lignocellulose trong dạ cỏ dê, mặc dù kích thước của dữ liệu giải trình tự sâu chỉ tăng 5,7 lần so với kích thước của dữ liệu giải trình tự của đề tài giai đoạn trước, nhưng số lượng gene mã hóa protein/enzyme liên quan đến quá trình phân hủy lignocellulose khai thác được cao hơn đáng kể, lên đến 47,24 lần, với 30 loại enzyme/protein, trong đó 8988 gene được dự đoán là hoàn chỉnh. Các enzyme thủy phân lignocellulose rất đa dạng, GH phong phú và nhiều loại enzyme có cấu trúc module chứa cả vùng xúc tác và vùng phụ trợ như vùng tương tự globulin miễn dịch (Ig), vùng giống fibronectin (FN3) hoặc vùng liên kết carbohydrate (CBM) , ... Một số GH đã được quan sát thấy trong nghiên cứu trước đây của Lim và cộng sự, phần lớn các enzyme lần đầu tiên được nhìn thấy trong dạ cỏ dê như alpha-D-xylosidexylohydrolase, acetylxylylan esterase, beta-

fructofuranosidase, endo-transglycosylase/hydrolase, oligosaccharide-reducing-end xylanase, feruloyl esterase, laccase, lytic polysaccharide monooxygenase, exopolygalacturonase lyase, endopolygalacturonase. Đáng chú ý, *Prevotella* chiếm 27% số gene phân loại theo NCBI mã hóa cho các protein/enzyme liên quan đến sự phân hủy lignocellulose. Với mỗi nhóm, *Prevotella* đóng góp 19% gene mã hóa cho cellulase, 30% gene mã hóa cho hemicellulase và 38% gene mã hóa cho enzyme/protein liên quan đến quá trình tiền xử lý lignocellulose, đặc biệt là pectinesterase. Trong nghiên cứu trước đây, với kết quả giải trình tự 8,6 Gb, chúng tôi thấy rằng *Prevotella* cung cấp tới 40% gene cho quá trình phân hủy lignocellulose. Đồng ý với kết quả này, bằng cách phân tích metagenomic của vi khuẩn mang gene mã hóa cho cellulase trong dạ cỏ dê Hàn Quốc, Lim và cộng sự cũng chỉ ra rằng các loài *Prevotella* đóng vai trò quan trọng đối với sự phân hủy cellulose [199]. Các enzyme được phát hiện có nhiều vùng chức năng khác nhau ngoài vùng xúc tác, là cơ sở để khai thác các enzyme mới có nhiều chức năng. Ví dụ, trong dữ liệu giải trình tự 8,6 Gb, chúng tôi đã tìm thấy một gene mã hóa cho endoglucanase có cấu trúc module bao gồm vùng X ở đầu N, tiếp theo là vùng FN3 rồi đến vùng hoạt tính GH5. Trong một nghiên cứu khác, FN3 đã được chứng minh là đóng vai trò tăng cường khả năng hòa tan của enzyme và giúp hình thành cấu trúc phù hợp [48]. Kết quả phân tích khẳng định rằng dữ liệu lớn về các gene mã hóa protein/enzyme tham gia vào quá trình phân hủy lignocellulose rất có giá trị để tìm kiếm các enzyme mới cho quá trình phân hủy lignocellulose hiệu quả, hỗ trợ cho sự phát triển kinh tế sinh học.

Kết quả thu được từ các phân tích sâu hơn cho thấy chi *Prevotella* có thể sản xuất nhiều loại enzyme khác nhau, chủ yếu để phân giải hemicellulose hơn là phân giải cellulose. Phát hiện này phù hợp với quan sát trong nghiên cứu khác về *P. bryantii* B14 [200]. Đáng chú ý, nhiều enzyme từ *Prevotella* là enzyme then chốt để nâng cao tốc độ phân giải lignocellulose như acetylxylyl esterase [201] và feruloyl esterase [202]. Trong tổng số 331 gene mã hóa acetylxylyl esterase thì có 327 gene thuộc *Prevotella*. Kết quả thu được khẳng định vai trò then chốt của *Prevotella* trong quá trình phân giải hemicellulose trong dạ cỏ dê.

Với mục đích tìm ra các dòng vi khuẩn tiềm năng, tất cả các contig mang gene hoàn chỉnh (8.900 gene) mã hóa protein/enzyme liên quan đến quá trình phân giải lignocellulose đã được tách riêng để phân tích. Theo kết quả thống kê, 8.900 gene nằm trong 8.364 contig, trong đó 7848 contig chỉ mang một gene trên mỗi contig. Các contig có khả năng phân hủy lignocellulose bao gồm: 4 contig chứa 6 gene/contig, 2 contig chứa 5 gene/contig, 16 contig mang 4 gene/contig, 54 contig

chứa 3 gene/contig và 440 contig mang 2 gene/contig. Đối chiếu với kết quả phân loại học, chúng tôi nhận thấy trong số 22 contig mang ít nhất 4 gene/contig thì có 18 contig thuộc chi *Prevotella*, 2 contig thuộc chi *Bacteroides*, 1 contig thuộc chi *Clostridium* và 1 contig thuộc chi *Butyrivibrio*. Cụm gene liên quan đến quá trình phân giải lignocellulose trong mỗi nhóm tiềm năng đã được xác định và vẽ bằng công cụ SnapGene DNA.



Hình 3. 8. Các locus gene phân giải celluloses/hemicelluloses trong những contig tiềm năng xây dựng từ dữ liệu giải trình tự sâu DNA metagenome vi khuẩn dạ cỏ dê.

CE: esterase; GH: glycosylase; PL: Polysaccharide lyase; EG: Endoglucanase; CBM: carbohydrate binding model; BGL: Beta glucosidase; XLA: Endo- β -1,4 xylanase; OREX: Oligosaccharide reducing-end xylanase; XBX: Xylan 1,4-beta-xylosidase; BGAL: Xyloglucan-active β -D-galactosidase; AGAL: Alpha-galactosidase; ARA: Alpha-L-arabinofuranosidase; AFUC, FUC: Alpha-L-fucosidase; BMAN: Beta-mannanase; AGL: Anpha glucosidase; AMY: amylase

Từ kết quả phân tích (Hình 3.8), có thể thấy quá trình phân hủy hemicellulose được xúc tác với những cụm gene trên hầu hết đều tương tác đặc hiệu với một số cơ chất nhất định. Ví dụ, trong contig-80, vùng hoạt tính đặc trưng của họ CE1, PL1 đã được nhìn thấy ở phần đầu của contig, tiếp đó là nhiều gene chức năng chưa biết, sau đó là vùng PL1, vùng gene BGAL (Xyloglucan-active beta-D-galactosidase) bao gồm vùng hoạt tính của họ GH43-GH143-GH142, vùng gene FUC (Alpha-L-fucosidase) liền kề GH95, kế đó là bốn gene BGALGH2 và nhiều GH khác, tất cả đều tập trung phân giải xyloglucan trong cấu trúc hemicellulose. Kết quả tương tự cũng được quan sát thấy ở contig 374 với 6 gene BGAL nằm gần nhau. Các cụm gene phân giải xyloarabinofuranose được thấy ở các contig 527, 829, 193, 148, 616,

813 với sự tham gia của arabinofuranosidase, oligosaccharide reducing-end xylanase, endoxylanase, esterase, alpha-galactosidase và đôi khi với beta-glucosidase. Một cụm gene phân giải cellulose được tìm thấy ở contig 289, trong đó, các gene mã hóa cho cellulase như endo glucanase, beta-glucosidase, cellobiose phosphorylase đã được quan sát.

Nhấn mạnh vai trò của *Prevotella* trong quá trình phân giải lignocellulose đặc biệt là hemicellulose, 22 contig chứa gene có khả năng phân hủy lignocellulose với ít nhất 4 gene/contig có 18 contig thuộc về *Prevotella*. Vùng chứa một cụm gene tham gia vào quá trình phân hủy polysaccharide được gọi chung là PUL (polysaccharide utilization loci). PULs thường có ở vi khuẩn đường tiêu hóa, tiêu biểu là *Ruminococcus* ở ruột người [203], *Bacteroides* ở dạ cỏ động vật [204, 205]. Tuy nhiên, PUL ở *Bacteroides* cũng phức tạp tương tự như ở *Prevotella*, do sự khả năng phân giải cơ chất cực đa dạng. Phân tích các cụm gene trong 11 contig, chúng tôi đã tìm thấy 9 contig liên quan đến sự phân hủy hemicellulose và tất cả đều thuộc về *Prevotella*, hai contig góp phần vào sự phân hủy cellulose thuộc về *Bacteroides* và *Prevotella*. Ngoài ra, tất cả các gene trong một cụm được sắp xếp theo cùng một hướng. Bên cạnh các enzyme chính có hoạt tính hemicellulase, nhiều gene mã hóa cho các enzyme thuộc các GH khác nhau, có thể hỗ trợ cho chức năng chính của locus và một số gene chưa rõ chức năng cũng đã được xác định. Bốn PUL phân giải arabinan được tìm thấy trong contig 527, 829, 813, 193, và ba PUL phân giải galactan nằm trong contig 80, 374, 464. Một số PUL liên quan đến quá trình phân giải arabinan, xylan, galactan nằm ở contig 148 và 616. Các PUL phân giải arabinan đã được tìm thấy trong *Prevotella copri* trong ruột người [206] nhưng các PUL phân hủy polysaccharide khác bao gồm galactan, xylan, arabinan trong *Prevotella* lần đầu tiên được phát hiện.

Nghiên cứu vai trò của *Prevotella* trong dạ cỏ dê, chúng tôi nhận thấy *Prevotella* sở hữu nhiều gene mã hóa cho các enzyme phân giải tinh bột và polysaccharide, trong đó enzyme quan trọng nhất để phân giải tinh bột là amylase có nhiều nhất (Bảng 3.7). Một nghiên cứu trước đây cũng phát hiện ra rằng chủng 25A của *Prevotella bryantiican* lên men nhanh chóng tinh bột và giảm 90% sự tích tụ lactate, đồng thời tăng lượng succinate và propionate [207]. Bằng cách này, *Prevotella* hỗ trợ cung cấp một lượng lớn năng lượng cho cơ thể vật chủ dê. Việc nghiên cứu các enzyme phân giải tinh bột và polysaccharide do *Prevotella* trong dạ cỏ còn hạn chế. Trong nghiên cứu này, chúng tôi cũng tìm thấy một số lipase từ *Prevotella*. Tuy nhiên, *Prevotella* được biết là không phải là đối tượng phân giải lipid trong dạ cỏ dê. *Prevotella* giảm khi cho dê ăn khẩu phần giàu lipid.

Tóm lại, *Prevotella* là một chi cực kỳ linh hoạt, đóng vai trò then chốt trong quá trình phân giải hemicellulose và tham gia vào quá trình phân giải tinh bột, cellulose, hemicellulose và pectin trong dạ cỏ dê.

3.2.3. Xây dựng công cụ mới khai thác hiệu quả protein/enzyme tham gia chuyển hóa và tiền xử lý lignocelullose

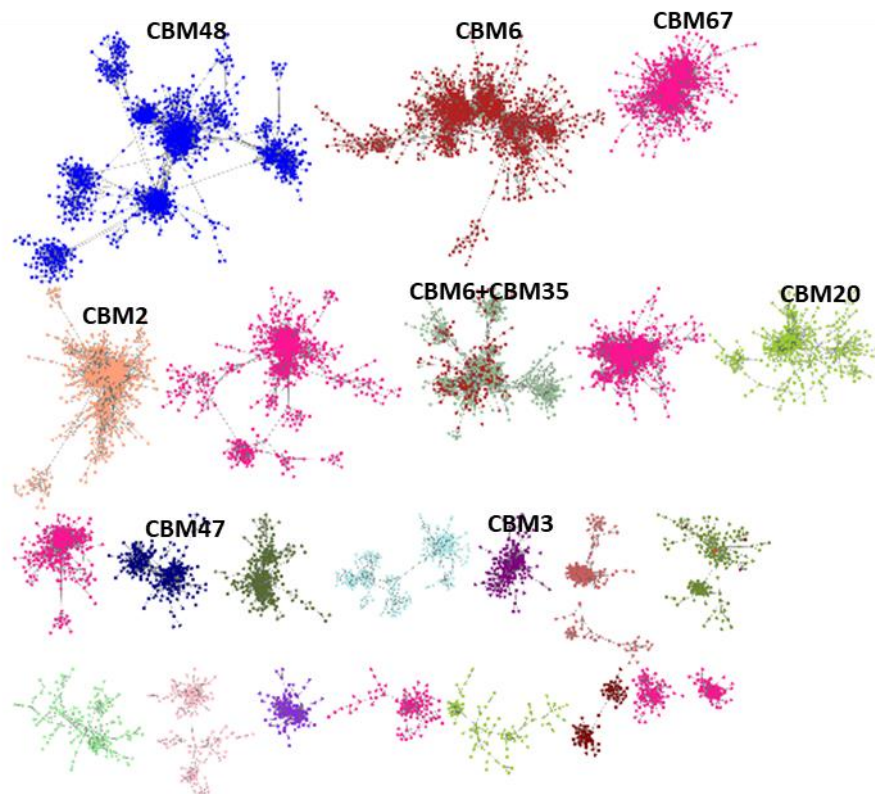
3.2.3.1. Xây dựng mô hình HMM đại diện cho các enzyme/protein cần khai thác

Theo như phân phương pháp đã trình bày, các trình tự enzyme/protein hoàn thiện đã được khẳng định về chức năng đã được thu thập từ các bộ dữ liệu và tìm ra vùng bảo thủ đặc hiệu cho từng nhóm enzyme/protein. Các trình tự bảo thủ sau đó được sử dụng để xây dựng mô hình HMM đại diện cho từng loại enzyme/protein. Mô hình HMM của một số nhóm enzyme/protein đã được xây dựng và công bố trong cơ sở dữ liệu Pfam và được trích xuất để làm tập hợp mô hình HMM cơ sở sử dụng trong việc khai thác gene phân giải lignocelullose (Bảng 3.8). Tuy nhiên vẫn còn khá nhiều enzyme chưa có trong cơ sở dữ liệu này hoặc đã có nhưng dữ liệu còn hạn chế, đặc biệt là nhóm carbohydrate-binding module (CBM). Với nhóm CBM, cơ sở dữ liệu Pfam mới công bố mô hình HMM của 23 họ, dữ liệu trình tự của 69 họ CBM đã được tổng hợp, so sánh và xây dựng mô hình HMM.

Bảng 3. 8. Thống kê các mô hình HMM được sử dụng cho khai thác nhóm protein/enzyme mới trong đề tài

STT	Tên enzyme	Nguồn gốc	STT	Tên enzyme	Nguồn gốc
1	Lignin peroxidase (LiP)	Pfam	17	Feruloyl esterase	Pfam
2	Hydrogen peroxide oxidoreductase	Pfam	18	Xyloglucanase	Pfam
3	Manganese peroxidase (MnP)	Pfam	19	Mannanase	Pfam
4	Versatile peroxidase (VP)	Pfam	20	Beta-mannosidase	Pfam
5	Laccase	Pfam	21	Arabinase	Pfam
6	Expansin	Pfam	22	Galactanase	Pfam
7	Lytic polysaccharide monooxygenase	Pfam	23	Polygalacturonase	Pfam
8	Cellobiohydrolase	Pfam	24	Beta-glucuronidase	Pfam
9	Endo-1,4-glucanase	Pfam	25	Glucuronyl esterase	Pfam
10	Beta-glucosidase	Pfam	26	CBM (1,2,3,6,9,10,11,14,15,19, 20,21,25,27,32,35,39,46, 48,49,53,60,65,77)	Pfam
11	Endo-xylanase	Pfam	27	CBM (4,5,8,12,13,16-19,22-24,26,28-31,34,36-38,40-45,47,50-52,54-59,61-64,66-76,78-84)	Tự xây dựng
12	Beta-xylosidase	Pfam	28	Fibronectin 3-like domain	Pfam
13	Alpha-L-arabinofuranosidase	Pfam	29	Dockerin	Pfam
14	Alpha-glucuronidase	Pfam	30	Immunoglobulin-like domain	Pfam
15	Licheninase	Pfam	31	Acetyl xylan esterase	Pfam
16	Acetyl xylan esterase	Pfam			

Dựa trên các dữ liệu khác nhau, các trình tự cho các nhóm CBM đã được lấy để tạo bộ dữ liệu cơ sở. Tổng số 391.140 trình tự đã được lấy để phân tích cho toàn bộ dữ liệu gồm 84 loại CBM. Các trình tự trên được phân tích so sánh tương đồng bằng phần mềm Usearch, sau đó kết quả so sánh được xây dựng mô hình mạng lưới bằng công cụ Cytoscape 3.7.1 với giới hạn cut-off của độ tương đồng các trình tự 0,55, kết quả mô hình xây dựng được cho thấy các trình tự bảo thủ của các nhóm CBM là khá tách biệt nhau (Hình 3.9). Điều này chứng tỏ các vùng trình tự bảo thủ này đảm bảo khả năng xây dựng mô hình HMM đặc hiệu cho mỗi nhóm, từ đó là công cụ tốt cho việc khai thác gen.



Hình 3. 9. Kết quả phân tích trình tự vùng CBM có cấu trúc beta-sandwich bằng Cytoscape với giới hạn cut-off của độ tương đồng trình tự là 0,55

Dựa trên kết quả phân tích, mô hình đại diện HMM của các họ CBM được xây dựng bằng phần mềm HMMER3, kết quả được tóm tắt trong Bảng 3.9.

Bảng 3. 9. Thống kê độ dài các vùng bảo thủ dùng để thiết lập mô hình HMM cho khai thác gene họ CBM

TT	Họ CBM	Độ dài (aa)	TT	Họ CBM	Độ dài (aa)	TT	Họ CBM	Độ dài (aa)
1	CBM4	153	21	CBM38	207	41	CBM64	86
2	CBM5	52	22	CBM40	196	42	CBM66	161
3	CBM8	162	23	CBM41	107	43	CBM67	186
4	CBM12	57	24	CBM42	138	44	CBM68	187
5	CBM13	158	25	CBM43	111	45	CBM69	146
6	CBM16	145	26	CBM44	194	46	CBM70	169
7	CBM17	179	27	CBM45	148	47	CBM71	188
8	CBM18	42	28	CBM47	149	48	CBM72	199
9	CBM19	69	29	CBM50	47	49	CBM73	70
10	CBM22	153	30	CBM51	166	50	CBM74	323
11	CBM23	149	31	CBM52	66	51	CBM75	298
12	CBM24	193	32	CBM54	153	52	CBM76	257
13	CBM26	89	33	CBM55	169	53	CBM78	151
14	CBM28	217	34	CBM56	85	54	CBM79	129
15	CBM29	142	35	CBM57	157	55	CBM80	94
16	CBM30	173	36	CBM58	109	56	CBM81	97
17	CBM31	107	37	CBM59	175	57	CBM82	234
18	CBM34	123	38	CBM61	159	58	CBM83	282
19	CBM36	119	39	CBM62	110	59	CBM84	165
20	CBM37	316	40	CBM63	96			

3.2.3.2. Khai thác các enzyme tham gia thủy phân lignocellulose từ dữ liệu DNA đa hệ gene vi khuẩn trong dạ cỏ dê

Dựa vào mô hình HMM đã được xây dựng để khai thác cho 29 enzyme/vùng phụ trợ (accessory domain) khác nhau có liên quan đến chuyển hóa lignocellulose, công cụ đã hỗ trợ khai thác với các nhóm gene được khai thác hiệu quả trong bộ dữ liệu là galactanase, glucuronyl esterase, hydrogen peroxide oxidoreductase (HPOXRE catalase), xyloglucanase, laccase, CBM (1-84), cellobiohydrolase, beta-glucuronidase, beta-xylosidase, beta-mannosidase GH2, lichenase, alpha-glucuronidase (GH76N) và xylanase GH44 (Bảng 3.10). Ngoài ra, các nhóm enzyme khác cũng được khai thác hiệu quả bằng công cụ mới nhưng các gene trùng với gene được phát hiện dựa trên cơ sở dữ liệu KEGG nên không được thống kê lại ở đây.

Bảng 3. 10. Khai thác một số enzyme hiệu quả từ dữ liệu DNA metagenome vi khuẩn trong dạ cỏ dê bằng mô hình HMM đã được xây dựng

Số thứ tự	Tên enzyme	Số lượng gene dựa trên HMM	Số lượng gene dựa trên KEGG
1	Galactanase	305	-
2	Glucuronyl esterase	165	-
3	HPOXRE catalase	1	-
4	Xyloglucanase	14	-
5	Laccase	350	9
6	CBM(1-84)	6875	270
7	Cellobiohydrolase	1645	216
9	Beta-glucuronidase	4171	888
10	Beta-xylosidase	3276	1018
11	Beta-mannosidase GH2	1890	659
12	Lichenase	350	281
13	Anpha-glucuronidase GH76N	615	561
14	Xylanase GH44	3534	3400
15	Axetyl xylan esterase (AXE1)	219	331
16	Anpha-L-arabinofuranosidase	2646	6896
17	Polygalacturonase	211	635
18	Arabinanase (GH43)	1894	6229
19	Endo-glucanase	2009	7368
20	Mannanase	263	1237
21	Beta-glucosidase	1582	10444
22	LPMO	1	11
23	Feruloyl esterase	7	92

3.3. Nghiên cứu lựa chọn, biểu hiện và xác định đặc điểm của endo-xylanase

Vi khuẩn trong dạ cỏ, đặc biệt là trong dạ cỏ dê rất quan tâm đến việc khai thác các enzyme mới liên quan đến sự phân hủy lignocellulose trong kỷ nguyên dữ liệu lớn. Endoxylanase là một trong những enzyme quan trọng nhất trong nhiều ngành công nghiệp như sản xuất cồn sinh học, giấy, hóa chất tốt... Endoxylanase không chỉ phân giải xylan trong cấu trúc hemicellulose để các enzyme khác chuyển hóa các sản phẩm xylanase thành xylose mà còn giúp tách cellulose để cellulase gắn vào và chuyển hóa. thành glucose để lên men thành sản phẩm cuối cùng. Các endoxylanase chủ yếu được phân loại vào các họ GH10 và GH11, một số trong họ GH115.

3.3.1. Nghiên cứu lựa chọn gene mã hóa endo-xylanase cho biểu hiện

3.3.1.1. Nghiên cứu đa dạng vi khuẩn mang enzyme endo-xylanase

Từ kết quả giải trình tự sâu 48,6 Gb, dựa theo kết quả chú giải chức năng gene với các cơ sở dữ liệu, có 3400 gene được xác định là endo-1,4-beta-xylanase. Toàn bộ số gene này được tích hợp với kết quả định danh nguồn gốc phân loại dựa vào phần mềm MEGAN

Bảng 3. 11. Đa dạng vi khuẩn mang gene mã hóa endo-xylanase

Ngành	Số ORF	Tỷ lệ	Lớp	Số ORF	Tỷ lệ	Bộ	Số ORF	Tỷ lệ
Firmicutes	1435	49%	Clostridia	1189	44%	Clostridiales	1166	44%
Bacteroidetes	1110	38%	Bacteroidia	1065	40%	Bacteroidales	1056	40%
Spirochaetes	114	4%	Spirochaetia	114	4%	Spirochaetales	110	4%
Fibrobacteres	77	3%	Bacilli	84	3%	Fibrobacterales	76	3%
Lentisphaerae	77	3%	Fibrobacteria	76	3%	Bacillales	49	2%
Khác	117	4%	Khác	163	6%	Khác	211	8%
Tổng	2930		Tổng	2691		Tổng	2668	
Họ	Số ORF	Tỷ lệ	Chi	Số ORF	Tỷ lệ	Loài	Số ORF	
Prevotellaceae	800	33%	<i>Prevotella</i>	577	30%	<i>Butyrivibrio proteoclasticus</i>	3	
Lachnospiraceae	460	19%	<i>Ruminococcus</i>	375	19%	<i>Desulfotomaculum hydrothermale</i>	5	
Ruminococcaceae	441	18%	<i>Butyrivibrio</i>	233	12%	<i>Melioribacter roseus</i>	2	
Spirochaetaceae	110	5%	<i>Treponema</i>	108	6%	<i>Petrotoxa olearia</i>	5	
Bacteroidaceae	95	4%	<i>Bacteroides</i>	81	4%	<i>Ruminococcus albus</i>	4	
Fibrobacteraceae	76	3%	<i>Fibrobacter</i>	67	3%	<i>Ruminococcus champanellensis</i>	1	
Clostridiaceae	66	3%	<i>Clostridium</i>	53	3%	<i>Treponema vincentii</i>	1	
Khác	345	14%	Khác	459	24%			
Tổng	2393		Tổng	1953		Tổng	21	

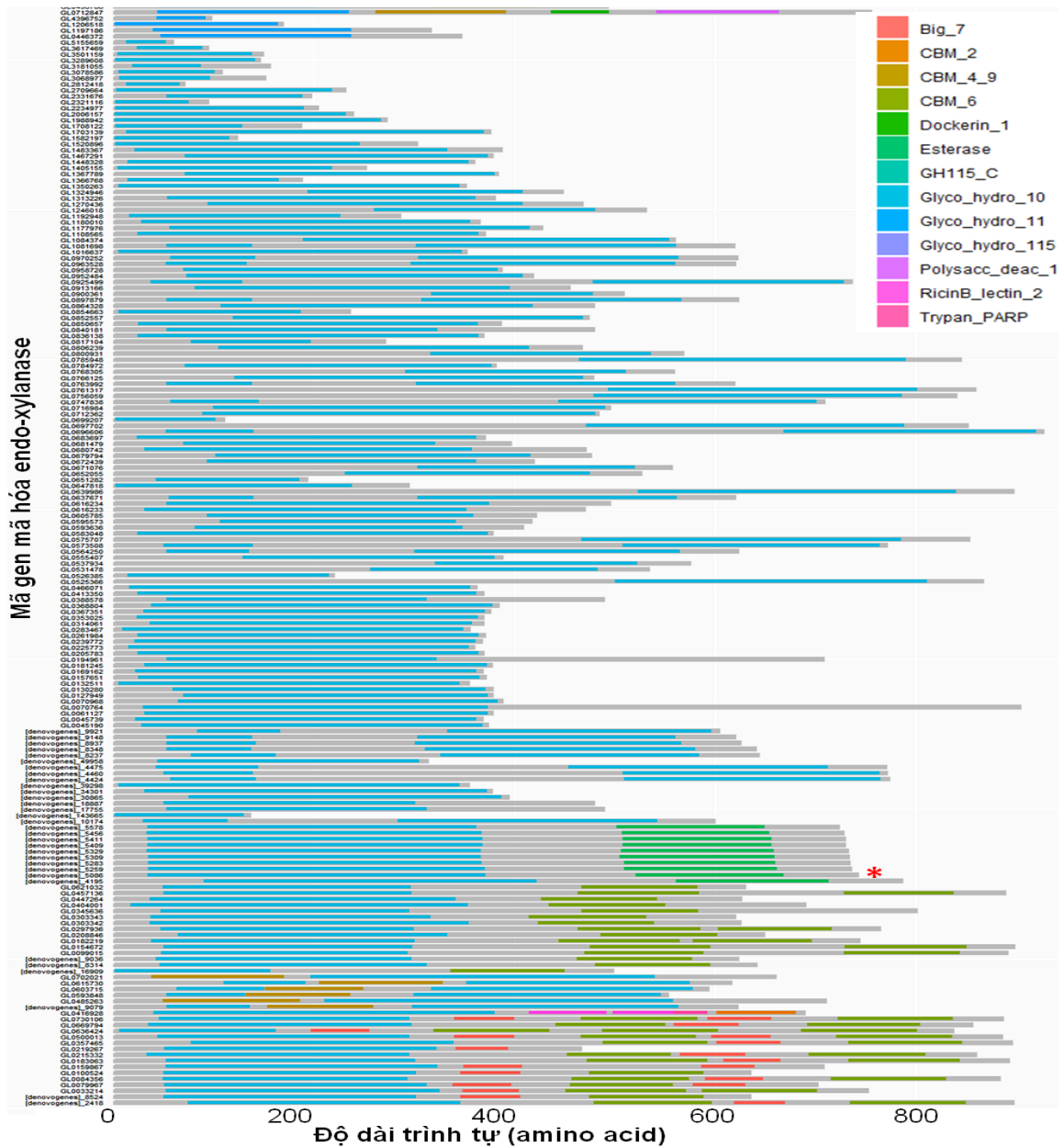
Kết quả phân tích cho thấy, có 3213 gene được phân loại đến các đơn vị phân loại thuộc 3 giới, 19 ngành, 33 lớp, 48 bộ, 67 họ, 120 chi và 9 loài, chỉ có 187 gene không xác định được đơn vị phân loại. Cụ thể, trong 3213 gene được phân loại, có 3193 gene được xác định thuộc giới vi khuẩn, chiếm 99,4%, còn lại chỉ 0,56% thuộc giới sinh vật nhân chuẩn và 0,06% thuộc giới vi khuẩn cổ. Chỉ xét riêng các gene có nguồn gốc từ vi khuẩn, 3193 gene được xác định thuộc 15 ngành, 27 lớp, 43 bộ, 62 họ, 113 chi và 7 loài (Bảng 3.11).

Ở mức độ phân loại đến chi, 30% số gene endo-xylanase có nguồn gốc từ *Prevotella*, tiếp theo là chi *Ruminococcus* (19%) và *Butyrivibrio* (12%). Kết quả này đồng nhất với kết quả chung về các gene tham gia phân giải hemicellulose và tiền xử lý sinh khối, cụ thể 30% số gene hemicellulase và 36% số gene tiền xử lý sinh khối trong dữ liệu giải trình tự là có nguồn gốc từ chi *Prevotella*.

3.3.1.2. Nghiên cứu đa dạng cấu trúc endo-xylanase

Endo-xylanase có nguồn gốc từ *Prevotella* được nhiều nhà nghiên cứu quan tâm do có hoạt tính cao và ổn định trong môi trường của một số chất tẩy rửa như tween, SDS,... Từ dữ liệu giải trình tự metagenomic thông thường 8,6 Gb và giải

trình tự sâu 48,6 Gb của vi khuẩn trong dạ cỏ dê, 739 gene hoàn chỉnh mã hóa cho endo-xylanase được chú giải bởi KEGG đã được tách riêng và phân tích cấu trúc vùng chức năng.



Hình 3. 10. Tóm lược các cấu trúc domain của enzyme nhóm endo-xylanase có mang vùng hoạt tính GH, được vẽ bằng ngôn ngữ R trên phần mềm drawProteins

Gen đánh dấu * là gene được lựa chọn cho nghiên cứu biểu hiện

Theo phân loại của cơ sở dữ liệu CAZy, vùng hoạt tính của enzyme endo-xylanase có thể thuộc về các họ GH5, 6, 8, 10, 11, 26, 30, 43, 98, trong đó phần lớn là họ GH10 và GH11. Kết quả chú giải 739 trình tự amino acid suy diễn từ các gene hoàn chỉnh bằng phần mềm Pfamscan cho thấy có 574 gene có thể xác định được các vùng chức năng, các cấu trúc vùng chức năng của nhóm này rất đa dạng tuy nhiên chỉ có 185 trình tự là mang các vùng hoạt tính glycosyl hydrolase.

Các gene endo-xylanase được xác định thuộc về 3 họ GH là GH10, GH11 và GH115, trong đó ưu thế nhất thuộc về họ GH10 (96% trong tổng số trình tự được xác định có mang ít nhất một vùng hoạt tính GH). Một lượng lớn trình tự chỉ có duy nhất vùng hoạt tính GH10. Khoảng một phần tư số trình tự endo-xylanase có mang ít nhất một vùng hoạt tính GH là các enzyme mang từ hai vùng chức năng trở lên, bên cạnh vùng hoạt tính là những vùng hỗ trợ như vùng tăng cường khả năng bám cơ chất như vùng tương tự immunoglobulin, CBM2, CBM4_9, CBM6, RicinB_lectin_2, dockerin hoặc hoạt tính phụ như vùng esterase.

Kết quả phân tích so sánh 739 gene hoàn chỉnh với BLASTP cho thấy có 180 trình tự amino acid được suy ra từ 180 gene tương đồng với endo-xylanase trong GenBank, trong đó có tất cả 108 trình tự có nguồn gốc từ *Prevotella*. Tất cả các gene mã hóa cho các trình tự tương đồng cao với *Prevotella* endo-xylanase đều có vùng hoạt tính xúc tác là GH10. Trong số này, 10 gene có trình tự cho vùng hoạt tính phụ CE1 (vùng esterase), 20 gene có trình tự thêm vùng CBM6. Các endo-xylanase được mã hóa bởi 10 gene trên có T_m cao hơn 65°C và là các enzyme có tính acid.

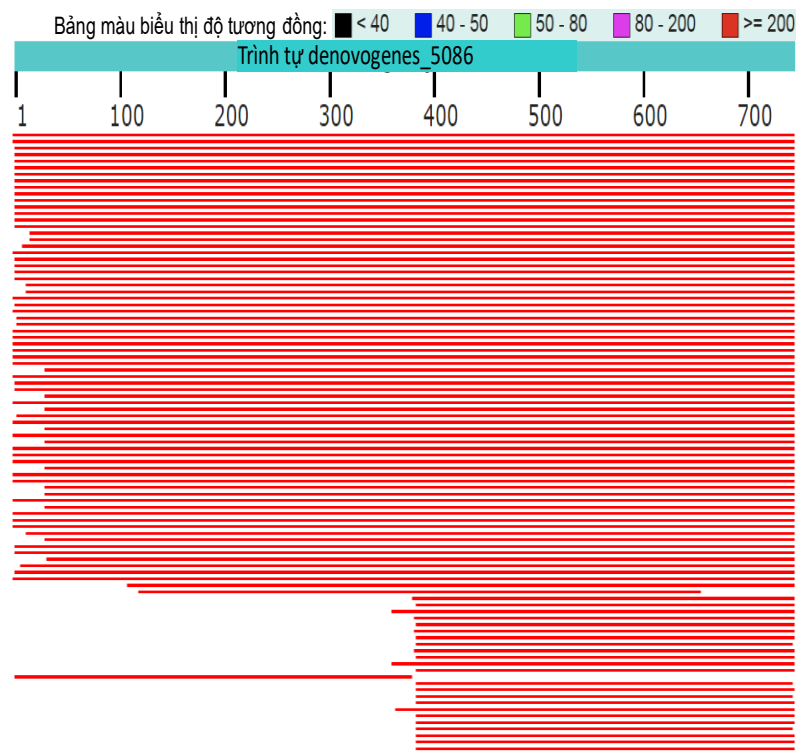
3.3.1.3. Nghiên cứu lựa chọn trình tự xylanase để biểu hiện

Kết quả tìm kiếm tương đồng với trình tự denovogenes_5086 trên cơ sở dữ liệu NCBI bằng công cụ BLASTp dựa trên dữ liệu trình tự NR (non redundant protein sequences, <https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>) được thể hiện ở Bảng 3.12.

Bảng 3. 12. Các trình tự tương đồng với trình tự denovogenes_5086

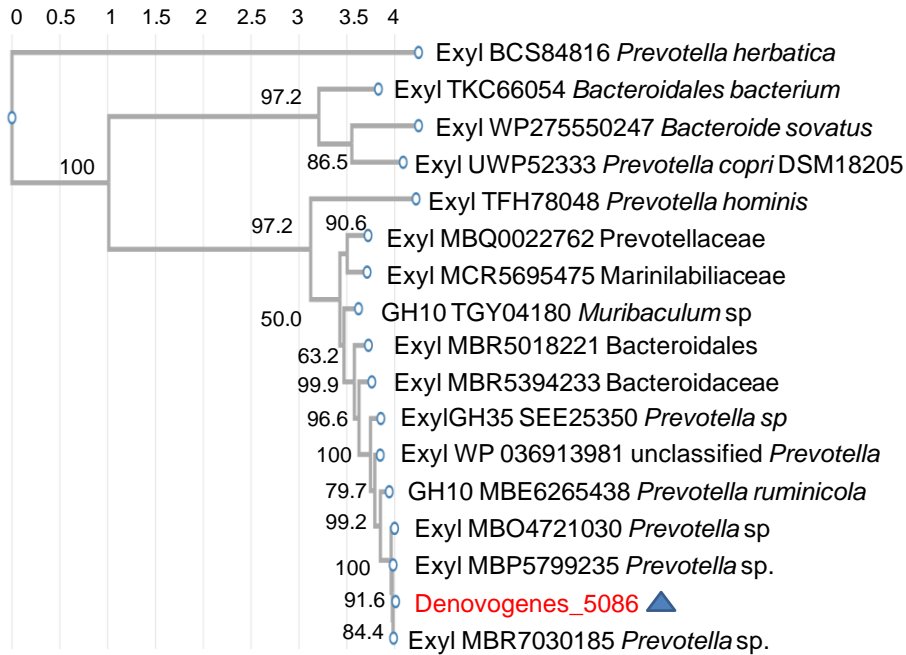
Mô tả	Điểm tối đa	Tổng điểm	Độ bao phủ	Giá trị E	Độ tương đồng
Endo-1,4-beta-xylanase [<i>Prevotella</i> sp. E2-28]	1264	1264	100%	0.0	82,3%
Endo-1,4-beta-xylanase [unclassified <i>Prevotella</i>]	1250	1250	100%	0.0	81,76%
Glycosyl hydrolase family 10 [<i>Prevotella</i> sp. MA2016]	1213	1213	99%	0.0	80%
Endo-1,4-beta- xylanase/feruloyl esterase [<i>Prevotella</i> sp. RM4]	1212	1212	99%	0.0	80%
Endo-1,4-beta- xylanase/feruloyl esterase [<i>Prevotella ruminicola</i>]	1211	1211	99%	0.0	80%
Endo-1,4-beta- xylanase/feruloyl esterase [<i>Prevotella ruminicola</i>]	1210	1210	99%	0.0	80%
Glycosyl hydrolase family 10 [<i>Candidatus Bacteroides timonensis</i>]	940	940	99%	0.0	63%
Hypothetical protein [<i>Bacteroides cellulosilyticus</i>]	939	939	100%	0.0	63%

Kết quả tìm kiếm tương đồng cho thấy trình tự protein suy diễn từ gene *denovogenes_5086* tương đồng cao nhất 96% với endo-1,4-beta-xylanase của *Prevotella* sp. (mã số MBR7030185.1) với độ bao phủ 99% (kết quả kiểm tra lại với cơ sở dữ liệu cập nhật ngày 31 tháng 3 năm 2023). Ngoài ra trình tự protein này cũng tương đồng 125 trình tự endo-1,4-beta-xylanase của *Prevotella* khác. Khi so sánh với các trình tự protein trong cơ sở dữ liệu của NCBI bằng BLASTP, trình tự suy từ gene *denovogenes_5086* có độ tương đồng cao nhất với các enzyme thủy phân xylan thuộc họ GH10 thuộc chi *Prevotella* hoặc các enzyme endo-1,4-beta-xylanase/feruloyl esterase thuộc loài *Prevotella ruminicola* (truy cập ngày 5 tháng 12 năm 2017).



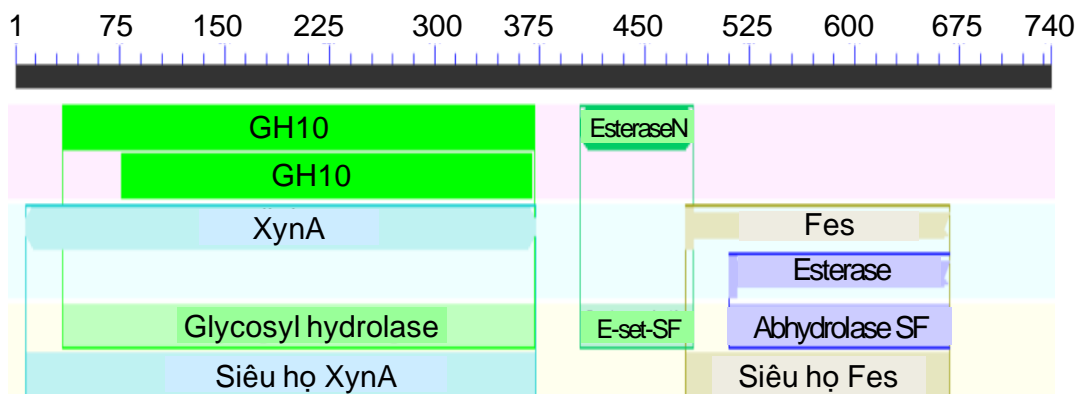
Hình 3. 11. Kết quả minh họa trình tự tương đồng với trình tự *denovogenes_5086* trên NCBI bằng BLASTP

Kết quả từ cây phân loại chủng loài cũng hoàn toàn trùng khớp với kết quả tìm kiếm tương đồng, trình tự *denovogenes_5086* mã hóa cho enzyme có nguồn gốc gần với các enzyme thủy phân GH10 và endo-1,4-beta-xylanase/feruloyl esterase thuộc chi *Prevotella*.



Hình 3. 12. Cây phát sinh chủng loài trình tự amino acid gene denovogenes_5086

Khi tìm kiếm các vùng bảo thủ trên trình tự gen, trình tự denovogenes_5086 có chứa hai vùng bảo thủ: Glyco_hydro_10 (từ 34-371) và vùng E-set_Esterase (404-485), ngoài ra có một vùng Fes được xác định là enterochelin esterase và các enzyme liên quan (479-668).

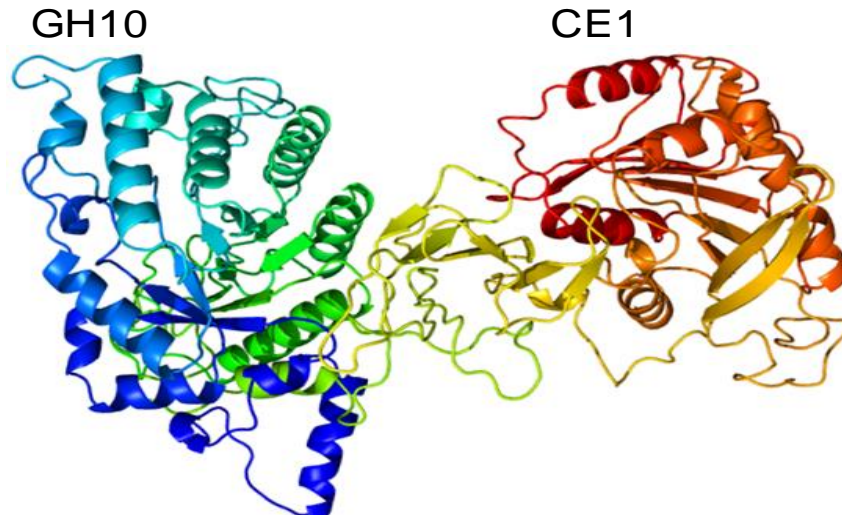


Hình 3. 13. Dự đoán các vùng bảo thủ trên trình tự denovogenes_5086

GH10: Glycosyl hydrolase family 10; XynA: Endo-1,4-beta-xylanase; E-set-SF: Siêu họ bao gồm vùng trước đầu N esterase, EsteraseN: Vùng trước của đầu N esterase; Fes: Enterochelin esterase và enzyme có liên quan; SF: siêu họ.

Bằng phần mềm Phyre2, mô hình ba chiều của enzyme đã được xây dựng. Phù hợp với kết quả so sánh BLASTP về các vùng bảo thủ, cấu trúc enzyme chứa hai vùng riêng biệt. Vùng xúc tác GH10 nằm ở đầu N và CE1 nằm ở đầu C. Trình tự của GH10 tương đồng 46% với một trình tự xylanase mang Ig từ ruột mối, với độ

tin cậy 100% và 37% với endo-xylanase kiềm chịu nhiệt từ *Bacillus* [208] cũng với độ tin cậy là 100%. Vùng CE1 tương đồng 54% với ferulic acid esterase [209] với độ tin cậy của mô hình là 100%.



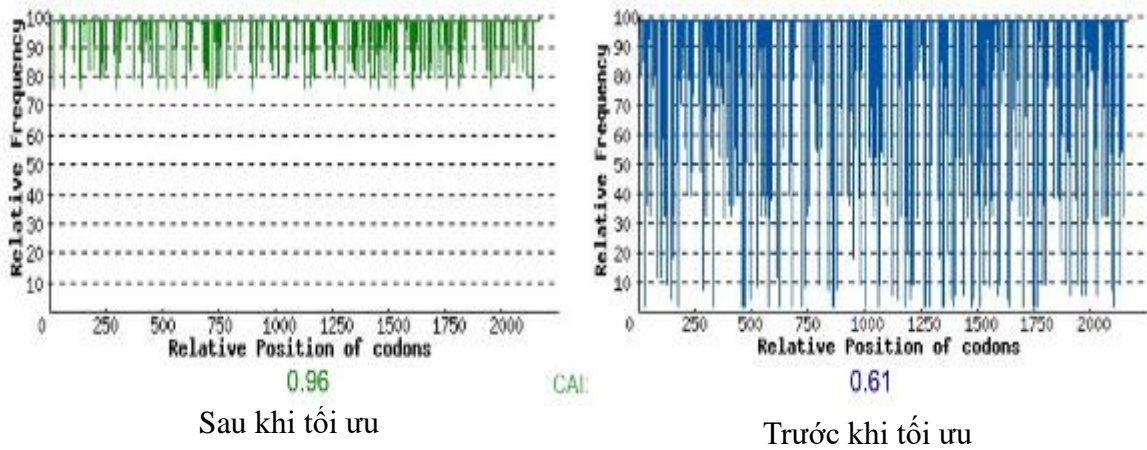
Hình 3. 14. Mô hình cấu trúc không gian của protein được dự đoán bằng Phyre2

Mã gene [denovogenes]_5086 có điểm acid thấp nhất là 0,363 trên tổng điểm là 1. Điều đó có nghĩa là enzyme do gene này mã hóa được dự đoán là có thể hoạt động ở điều kiện kiềm nhẹ. Do đó, chúng tôi đã chọn gene này để sản xuất endo-xylanase tái tổ hợp.

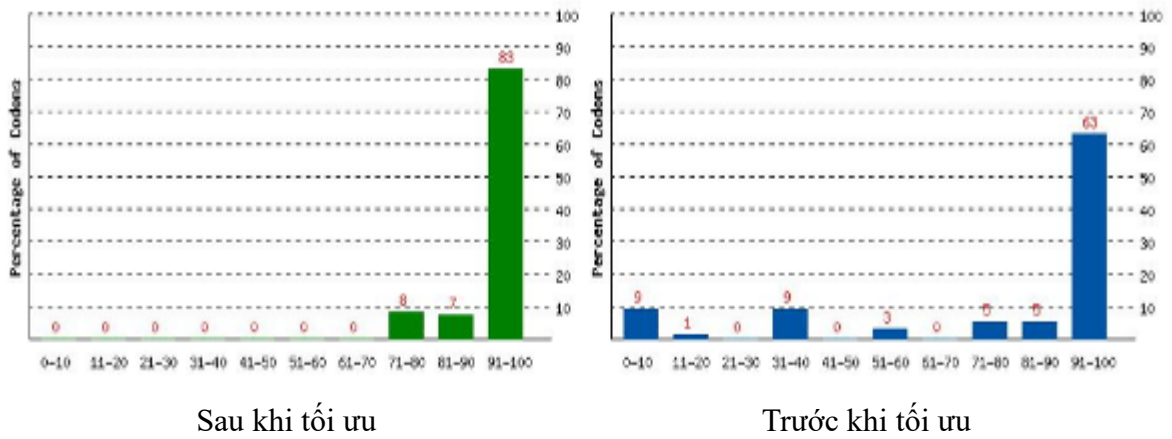
3.3.2. Nghiên cứu biểu hiện endo-xylanase

3.3.2.1. Phân tích tối ưu mã bộ ba của trình tự gene *exl*

Mã gene [denovogenes]_5086 ký hiệu *exl* mã hóa cho endo-xylanase được lựa chọn để biểu hiện trong *E. coli*. Trình tự có 2223 nucleotide với 63 nucleotide ở đầu 5' mã hóa cho vùng trình tự tín hiệu tiết được dự đoán bằng phần mềm SignalP. Trình tự tín hiệu tiết tự thân của protein thường được loại bỏ để tránh làm ảnh hưởng đến tính chất và khả năng biểu hiện cũng như hoạt tính của enzyme khi biểu hiện protein tái tổ hợp.



(A) Khả năng sử dụng các mã bộ ba trên



(B) Mức độ phù hợp mã bộ ba trên gen

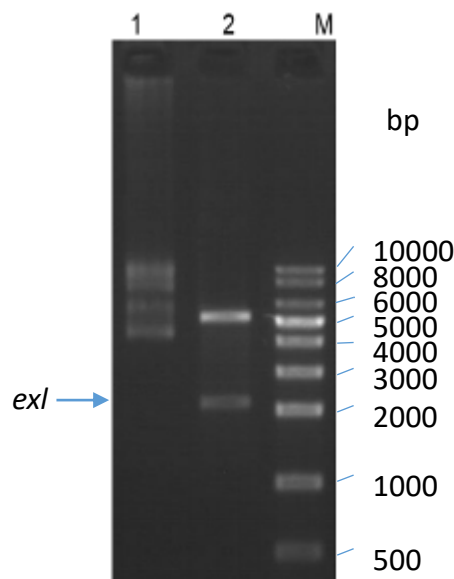
Hình 3. 15. Khả năng sử dụng các mã bộ ba trên gene (A) và mức độ phù hợp các mã bộ ba (B) của gene *exl* mã hóa được so sánh với trước và sau khi cải biến để biểu hiện trên chủng chủ *E. coli*

Trình tự gene *exl* tối ưu mã bộ ba để phù hợp cho việc biểu hiện protein tái tổ hợp trong *E. coli* trước khi được tổng hợp nhân tạo và đưa vào vector pET22b(+) bằng cặp enzyme hạn chế *NcoI/XhoI*. Kết quả sau khi được cải biến, tỷ lệ các bộ ba có khả năng sử dụng trên gene là 97% so với 61% trước khi tối ưu (Hình 3.15).

3.3.2.2. Thiết kế vector biểu hiện mang gene *exl*

Gen *exl* sau khi tối ưu mã bộ ba được tổng hợp nhân tạo bởi công ty GenScript. gene *exl* được đưa vào vector pET22b(+) bằng cặp enzyme hạn chế *NcoI/XhoI*. Vector pET22_ *exl* được biến nạp vào chủng tách dòng *E. coli* DH10B theo quy trình đã được mô tả. Khuẩn lạc mọc trên đĩa môi trường LB chứa chất kháng sinh ampicillin được nuôi cấy và tách plasmid. DNA plasmid được cắt kiểm tra bằng hai enzyme giới hạn *NcoI* và *XhoI* để kiểm tra sự có mặt của đoạn gene đã được chèn vào vector.

Theo lý thuyết, khi sản phẩm cắt từ plasmid pET22 có mang đoạn gene *exl* bằng hai enzyme *NcoI* và *XhoI* được điện di và nhuộm với chất hiện màu là ethidium bromide, hai băng DNA có kích thước khoảng 5,5 kb tương ứng với vector pET22b(+) dạng thẳng và 2,2 kb tương ứng với gene *exl* sẽ xuất hiện trên điện di đồ. Kết quả điện di thực tế được thể hiện trên Hình 3.17.



Hình 3. 17. Kết quả điện di kiểm tra vector pJET mang gene endo-xylanase

M: Thang chuẩn DNA. 1: Plasmid gốc. 2: Sản phẩm cắt

Kết quả điện di trên Hình 3.18 cho thấy, sản phẩm cắt có hai băng DNA với kích thước đúng với kích thước lý thuyết của gene endo-xylanase và vector pET22b(+) ở dạng thẳng. Để khẳng định chắc chắn hơn sự có mặt chính xác của gene *exl* trong vector tái tổ hợp, vector này được giải trình tự bằng cặp môi đặc hiệu của vector pET22. So sánh trình tự đọc được với trình tự gene endo-xylanase gửi đặt tổng hợp thấy chúng có độ tương đồng 100% với nhau, với tổng chiều dài là 2160 nucleotide. Như vậy, gene endo-xylanase đã được tổng hợp và ghép nối thành công vào vector biểu hiện pET22b(+).

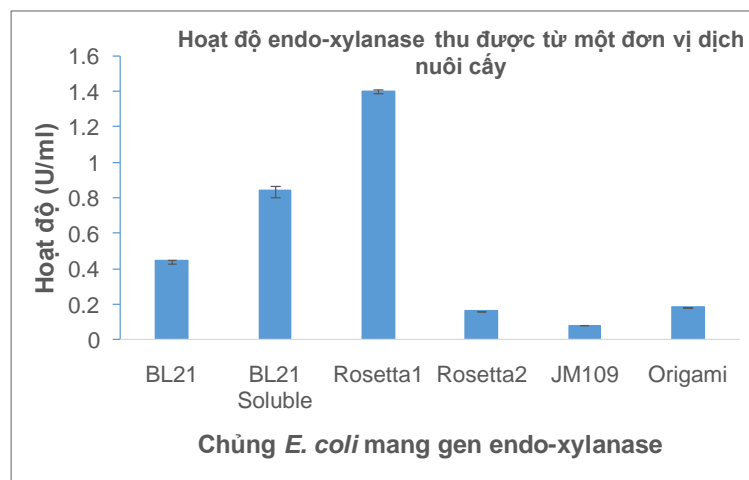
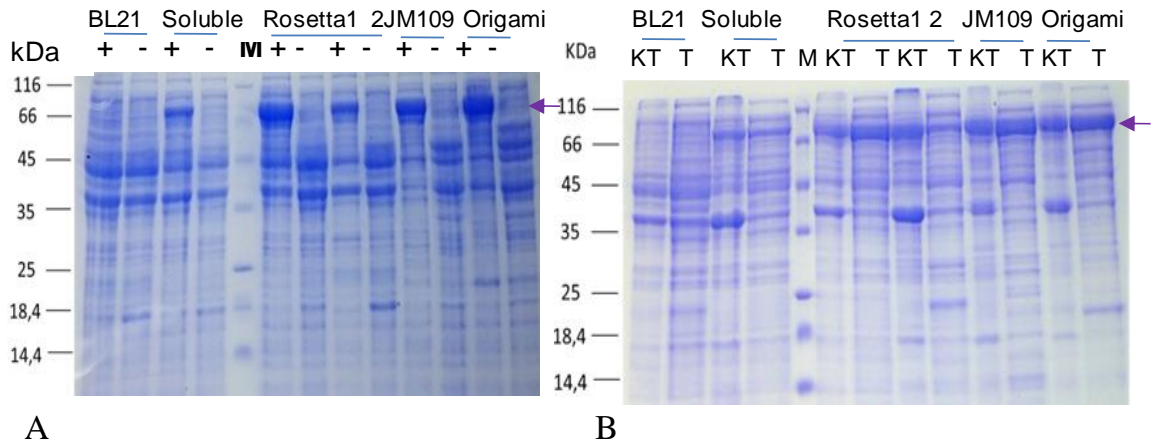
3.3.2.3. Biểu hiện gene *endo-xylanase* [*denovogenes*]₅₀₈₆

(1) Lựa chọn chủng biểu hiện *endo-xylanase*

Vector pET22b(+) hoạt động theo cơ chế của operon lac, cụ thể là chất cảm ứng IPTG (một chất đồng hình với lactose) sẽ liên kết với protein ức chế vốn luôn được biểu hiện trong tế bào, dẫn đến thay đổi cấu trúc không gian của chất ức chế, và làm mất khả năng bám của phân tử này vào trình tự operator, từ đó kích hoạt sự phiên mã và dịch mã của các gene nằm trong vùng điều khiển của operon, tổng hợp nên các protein tương ứng. gene *exl* mã hóa cho enzyme *endo xylanase* đã được đưa vào 6 chủng biểu hiện khác nhau. Các mẫu biểu hiện của các chủng được đo độ hấp thụ quang học ở bước sóng 600 nm trước khi ly tâm và hòa lại trong đệm Tris-HCl 20 mM sao cho OD đạt được là 10. Kết quả điện di protein tổng số các mẫu biểu hiện cùng một lượng dịch hòa tế bào như nhau được thể hiện ở trên Hình 3.18.

Khi nuôi cấy ở cùng một điều kiện như nhau, các chủng nuôi cấy khác nhau có khả năng sinh trưởng và tổng hợp protein khác nhau, cụ thể là độ hấp thụ quang học ở bước sóng 600 nm của các mẫu BL21(DE3), BL21 Soluble, Rosetta1, Rosetta2, JM109, Origami lần lượt là 4,8 – 6,4 – 4,23 – 1,14 – 0,572 – 1,292. Kết quả điện di trên Hình 3.19 cho thấy, ngoại trừ mẫu BL21(DE3), trên đường chạy của các chủng BL21 Soluble, Rosetta1, Rosetta2, JM109, Origami đều xuất hiện một băng đậm rõ nét có kích thước khoảng 80 kDa bằng với kích thước lý thuyết dạng đơn của protein *endo-xylanase*, trong khi mẫu đối chứng chủng không mang gene thì không xuất hiện băng này. Điều này cho thấy protein *endo-xylanase* đã được biểu hiện thành công trong cả 5 chủng BL21 Soluble, Rosetta1, Rosetta2, JM109, Origami. Trong các nghiên cứu đã được công bố, gene *xylanase* đã được biểu hiện tái tổ hợp trong nhiều chủng khác nhau như chủng nấm men *Komagataella phaffii* T07 [210], chủng *Pichia pastoris* [211], chủng *Aspergillus niger* AG11 [212], nhưng phần lớn được biểu hiện trong chủng *E. coli* BL21.

Endo-xylanase là một enzyme nên cần phải được biểu hiện ở dạng tan thì mới có khả năng thể hiện hoạt tính sinh học, vì vậy chúng tôi kiểm tra tính tan của protein biểu hiện bằng cách điện di riêng rẽ hai pha tan và không tan của dịch phá tế bào sau biểu hiện.

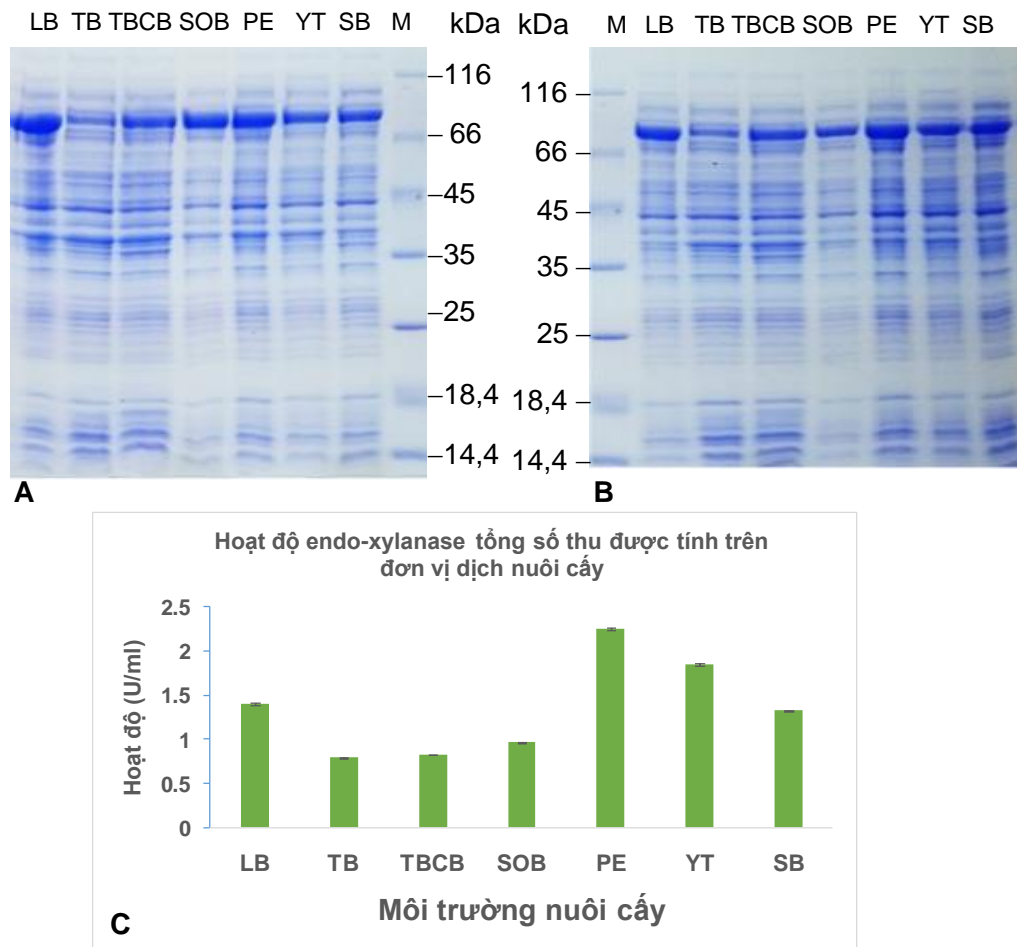


C

Hình 3. 18. Điện di đồ protein tổng số (A), pha protein tủa, pha protein tan (B) và hoạt độ endo-xylanase tổng số (C) của protein được biểu hiện trong các chủng *E. coli* mang gene *exI*

(-): chủng đối chứng mang vector pET22b(+) nhưng không có đoạn chèn gene endo-xylanase; (+): chủng mang vector pET22b(+) mang gene endo-xylanase; M: thang chuẩn protein (Thermo Scientific); mũi tên chỉ vị trí của endo-xylanase; T: Protein pha tan; KT: Protein pha không tan;

Kết quả điện di trên Hình 3.18A cho thấy, protein được biểu hiện của chủng BL21 Soluble và Rosetta2 hầu như chỉ tồn tại ở dạng không tan, còn đối với các chủng Rosetta1, JM109 và Origami, protein tồn tại ở cả dạng tan và không tan. Tuy nhiên lượng protein tan của chủng Rosetta1 là lớn nhất so với hai chủng còn lại, đồng thời khả năng sinh trưởng của chủng Rosetta1 cũng lớn hơn so với hai chủng JM109 và Origami, thể hiện ở số liệu quang phổ hấp thụ của dịch tế bào khi thu mẫu lớn nhất, đồng thời hoạt tính của dịch nuôi cấy chủng Rosetta1 là lớn nhất trong tất cả các chủng (Hình 3.18) vì vậy chúng tôi lựa chọn chủng Rosetta1 để làm chủng biểu hiện gene endo-xylanase.

(2) Lựa chọn môi trường biểu hiện endo-xylanase

Hình 3. 19. Kết quả điện di protein tổng số (A), protein pha tan (B) và hoạt endo-xylanase (C) từ các chủng *E. coli* Rosetta1 mang gene mã hóa endo-xylanase được nuôi trong các môi trường dinh dưỡng khác nhau.

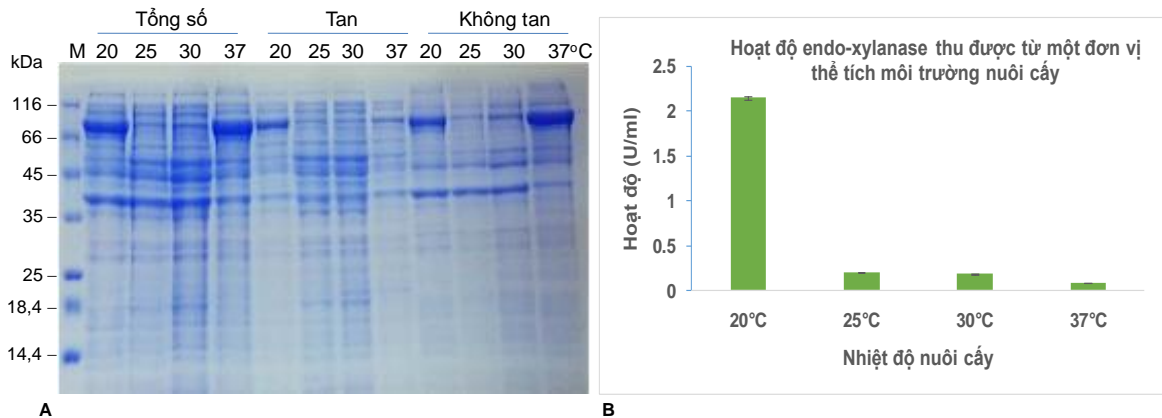
LB, TB, TBCB, SOB, PE, YT, SB: các mẫu biểu hiện trong các môi trường LB, TB, TB cải biến, SOB, PE, YT, SB; M: thang chuẩn protein (Thermo Scientific)

Chủng Rosetta1 mang vector biểu hiện pET22b(+) mang gene mã hóa endo-xylanase tiếp tục được tối ưu biểu hiện bằng cách thử nuôi cấy trong các điều kiện môi trường khác nhau nhằm tăng tối đa lượng protein biểu hiện trong tế bào đồng thời nâng cao sinh khối trong cùng một thể tích môi trường nuôi cấy. Các môi trường nuôi cấy thử nghiệm là các môi trường giàu chất dinh dưỡng nhằm cung cấp lượng lớn nguồn nitrogen và carbon cho quá trình sinh trưởng và sinh tổng hợp protein tái tổ hợp. Mật độ tế bào vi khuẩn ở trong mẫu sau khi kết thúc quá trình nuôi cấy, thể hiện bằng chỉ số OD₆₀₀, cùng với hoạt tính thô tính trên một đơn vị môi trường nuôi cấy cũng là một thông số quan trọng để đánh giá khả năng sinh trưởng và biểu hiện protein tái tổ hợp. Kết quả điện di protein tổng số của các mẫu biểu hiện trong các môi trường khác nhau được thể hiện trên Hình 3.19.

Kết quả điện di thể hiện ở hình trên cho thấy, các đường chạy mẫu protein

tổng số biểu hiện trong các môi trường LB, TB cải biến, SOB, PE, YT, SB đều xuất hiện một băng đậm rõ nét có kích thước tương ứng với kích thước lý thuyết khoảng 80 kDa của protein endo-xylanase, chỉ có ở môi trường TB là băng này không rõ. Điều này chứng tỏ ở các môi trường biểu hiện ngoại trừ môi trường TB, protein endo-xylanase đều được biểu hiện tương đối tốt. Các protein nói chung và protein tái tổ hợp nói riêng chỉ có thể thể hiện hoạt tính sinh học khi được biểu hiện đúng cấu trúc và ở dạng tan, vì vậy chúng tôi tiến hành tách riêng hai pha tan và không tan của mẫu protein tổng số và điện di kiểm tra. Kết quả điện di cho thấy, với cùng một lượng tế bào, protein endo-xylanase được biểu hiện ở dạng tan nhiều nhất trong môi trường PE, thể hiện ở việc băng protein tương ứng với kích thước lý thuyết của protein tái tổ hợp đậm nhất ở đường chạy mẫu biểu hiện trong môi trường này. Trong các nghiên cứu biểu hiện xylanase đã được công bố, protein tái tổ hợp biểu hiện trong chủng *E. coli* hầu như đều chỉ được biểu hiện trong môi trường LB bổ sung một số loại kháng sinh tương ứng với gene chọn lọc trên vector [213-215]. Hoạt tính của mẫu biểu hiện trong môi trường PE tính trên một đơn vị thể tích môi trường nuôi cấy cũng là lớn nhất so với các mẫu còn lại, vì vậy chúng tôi lựa chọn môi trường PE để biểu hiện endo-xylanase trong các thí nghiệm tiếp theo.

(3) Lựa chọn nhiệt độ nuôi cấy



Hình 3. 20. Kết quả điện di protein (A) và hoạt độ endo-xylanase tổng hợp từ chủng *E. coli* Rosetta1 mang gene mã hóa endo-xylanase được nuôi cấy trong môi trường cảm ứng ở các nhiệt độ khác nhau (B)

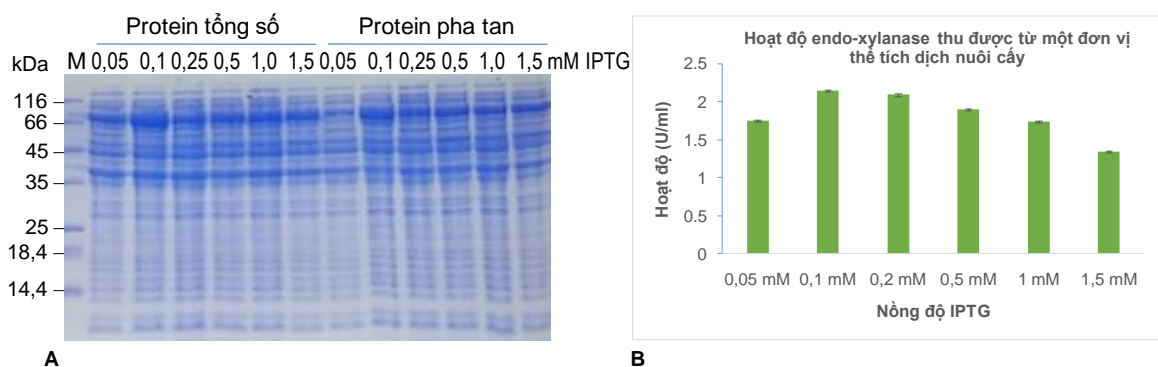
M: thang chuẩn protein (Thermo Scientific)

Nhiệt độ nuôi cấy cũng là một trong những yếu tố tác động đến quá trình tổng hợp protein của chủng tái tổ hợp. Theo lý thuyết, nhiệt độ nuôi cấy thấp sẽ làm giảm tốc độ sinh trưởng của chủng, tuy nhiên lại tạo điều kiện để protein tái tổ hợp được biểu hiện ở dạng tan trong một số trường hợp. Chủng Rosetta1 mang gene mã hóa endo-xylanase được thử nghiệm biểu hiện ở các nhiệt độ 20°C - 25°C - 30°C - 37°C, mẫu sau khi biểu hiện được ly tâm loại bỏ môi trường, sinh khối tế bào được

hòa vào đệm Tris-HCl trước khi xử lý để điện di kiểm tra. Kết quả điện di protein của các mẫu biểu hiện trong các nhiệt độ cảm ứng khác nhau được thể hiện trên Hình 3.20.

Kết quả điện di cho thấy, các đường chạy tương ứng với mẫu protein tổng số của mẫu biểu hiện ở nhiệt độ 20°C và 37°C đều xuất hiện một băng đậm rõ nét có kích thước tương ứng với kích thước lý thuyết của protein endo-xylanase, còn ở đường chạy tương ứng với mẫu protein tổng số biểu hiện ở 25°C và 30°C, băng protein tương ứng rất nhỏ. Dịch protein tổng số của mẫu biểu hiện ở cả bốn điều kiện nhiệt độ đều thể hiện hoạt tính endo-xylanase với độ mạnh yếu khác nhau, điều này cho thấy protein tái tổ hợp đều được biểu hiện dù ở nhiệt độ nào. Tuy nhiên băng protein tương ứng với kích thước ký thuyết của xylanase rất đậm ở mẫu pha tan của mẫu biểu hiện ở 20°C và pha không tan của mẫu biểu hiện ở 37°C. Điều này chứng tỏ protein được biểu hiện ở nhiệt độ 20°C tồn tại chủ yếu ở dạng tan, ngược lại protein tái tổ hợp biểu hiện ở nhiệt độ 37°C tồn tại chủ yếu ở dạng không tan, còn ở hai nhiệt độ trung gian, mẫu biểu hiện kém. Kết hợp với hoạt độ của enzyme xylanase tính trên một đơn vị thể tích môi trường nuôi cấy của mẫu biểu hiện ở nhiệt độ 20°C là lớn nhất vì vậy 20°C được chọn là nhiệt độ biểu hiện endo-xylanase. Trong các nghiên cứu liên quan đến biểu hiện xylanase đã được công bố, thông số hoạt tính của enzyme thô cũng như hoạt tính tính trên một đơn vị thể tích môi trường nuôi cấy không được nghiên cứu. Phần lớn xylanase được biểu hiện ở nhiệt độ 37°C [214, 215] nhưng cũng có nghiên cứu biểu hiện protein ở nhiệt độ thấp 16°C [213].

(4) Lựa chọn nồng độ chất cảm ứng IPTG



Hình 3. 21. Kết quả điện di protein các mẫu biểu hiện endo-xylanase ở các nồng độ chất cảm ứng

M: thang chuẩn protein (Thermo Scientific)

Nồng độ chất cảm ứng IPTG là một trong những yếu tố lớn ảnh hưởng đến số lượng và tính chất của protein biểu hiện tái tổ hợp bằng mô hình operon *lac*. Để lựa

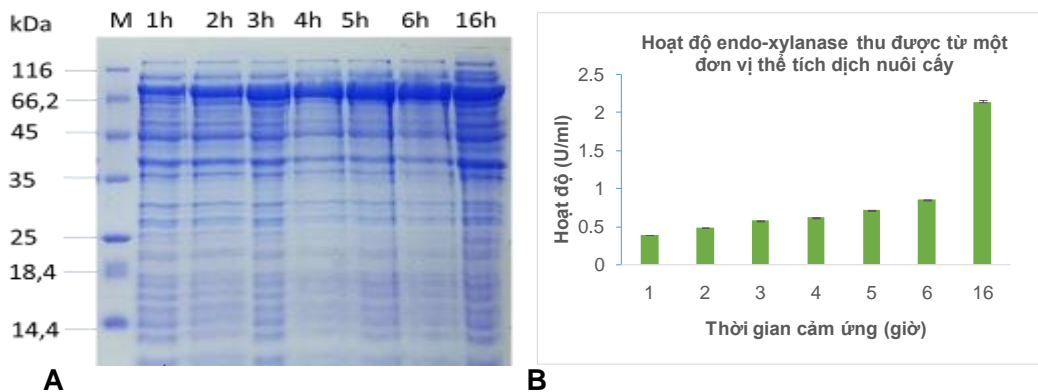
chọn nồng độ chất cảm ứng thích hợp cho biểu hiện endo-xylanase trong chủng *E. coli* Rosetta1, các mẫu biểu hiện sau khi được nuôi cấy đến OD thích hợp cho cảm ứng trước khi bổ sung chất cảm ứng vào các mẫu với nồng độ lần lượt là 0,05 mM, 0,1 mM, 0,25 mM, 0,5 mM, 1 mM, 1,5 mM. Các mẫu sau khi biểu hiện được xử lý thu hồi và điện di SDS-PAGE trên gel 12,6%.

Kết quả điện di cho thấy, protein endo-xylanase biểu hiện ngay từ khi mẫu nuôi cấy được cảm ứng ở nồng độ IPTG thấp là 0,05 mM, và lượng protein biểu hiện là gần như tương đương nhau ở các mẫu từ 0,1 mM đến 1,5 mM nhưng mẫu cảm ứng ở nồng độ 0,1 mM có băng endo-xylanase là tương đối đậm hơn các mẫu còn lại. Enzyme xylanase tái tổ hợp trong các nghiên cứu đã được công bố cũng được biểu hiện với nồng độ chất cảm ứng trong khoảng 0,05 mM đến 1 mM [213-216]. Khi so sánh hoạt tính dịch nuôi cấy của mẫu biểu hiện với các nồng độ chất cảm ứng khác nhau có xu hướng giảm khi nồng độ chất cảm ứng tăng lên, vì vậy chúng tôi lựa chọn nồng độ chất cảm ứng là 0,1 mM để biểu hiện endo-xylanase.

(5) Lựa chọn thời điểm thu mẫu biểu hiện endo-xylanase

Theo lý thuyết, chủng vi khuẩn tái tổ hợp mang vector điều hòa bởi hệ thống operon lac sẽ bắt đầu biểu hiện protein tái tổ hợp khi trong môi trường xuất hiện chất cảm ứng. Thời điểm thu mẫu tế bào sau khi bổ sung chất cảm ứng sẽ ảnh hưởng đến số lượng và tính chất protein tái tổ hợp cần biểu hiện, tùy thuộc vào bản thân các chủng và các gene cần biểu hiện.

Chủng Rosetta1 mang gene endo-xylanase được nghiên cứu so sánh thời điểm thu mẫu sau cảm ứng với các mốc là 1 giờ, 2 giờ, 3 giờ, 4 giờ, 5 giờ, 6 giờ và 16 giờ (qua đêm). Sản phẩm của quá trình biểu hiện trên được xử lý điện di và đo hoạt độ tính trên một đơn vị môi trường. Kết quả được trình bày ở Hình 3.22.



Hình 3. 22. Kết quả điện di protein pha tan và kiểm tra hoạt tính của các mẫu biểu hiện endo-xylanase ở các thời điểm thu mẫu sau cảm ứng

Kết quả điện di cho thấy, băng protein tương ứng với kích thước khoảng 80

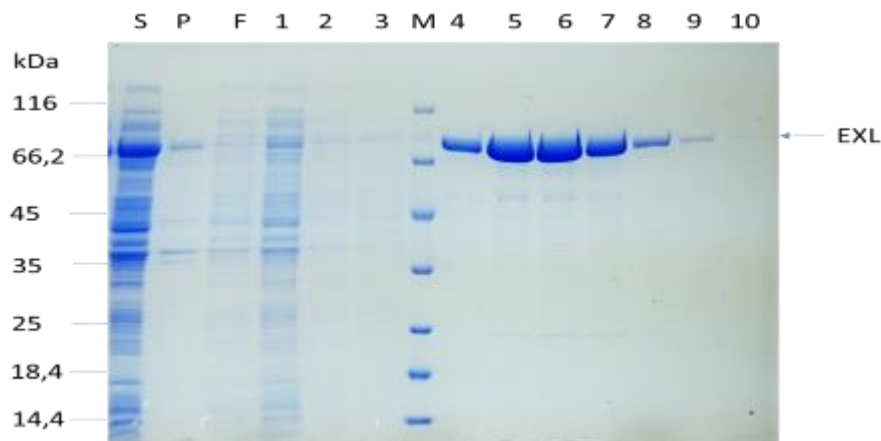
kDa của protein endo-xylanase được biểu hiện ngay sau khi bắt đầu kích hoạt quá trình biểu hiện bằng cách thêm IPTG khoảng 1 giờ. Với cùng một lượng sinh khối tế bào, lượng protein tái tổ hợp được biểu hiện tăng dần đến khoảng 3 giờ sau cảm ứng, sau đó lượng protein được duy trì tương đối đều. Trong một số nghiên cứu biểu hiện xylanase tái tổ hợp đã được công bố, thời gian thu mẫu sau khi cảm ứng là từ 3-5 giờ sau khi cảm ứng đến 20 giờ sau khi cảm ứng. Cụ thể, Zhang và cộng sự biểu hiện xylanase trong 12 giờ sau khi cảm ứng [213], thông số này là 16 giờ ở nghiên cứu của Kiribayeva và cộng sự [215], và 5 giờ trong nghiên cứu của Al-Darkazali [214].

Hoạt tính thô tính trên một đơn vị thể tích môi trường nuôi cấy của chủng biểu hiện tăng dần sau khi cảm ứng, và đạt giá trị cao nhất ở mẫu biểu hiện sau 16 giờ so với các mẫu được thu hồi ở thời gian ngắn, tăng khoảng 3 lần so với thời điểm 6 giờ sau khi cảm ứng. Vì vậy, thời điểm được lựa chọn để thu mẫu tế bào biểu hiện endo-xylanase là 16 giờ sau khi cảm ứng.

3.3.3. Tinh chế protein endo-xylanase tái tổ hợp

3.3.3.1. Tinh chế endo-xylanase bằng sắc ký ái lực His-tag

Theo nguyên lý của sắc ký, trình tự 6 His ở đầu C của protein xylanase tái tổ hợp sẽ liên kết mạnh với các ion kim loại hóa trị II đã được cố định trong giá thể của cột, còn các protein khác có các His nằm riêng rẽ và rải rác trong protein nên có ái lực kém với những ion trên và sẽ dễ dàng bị rửa trôi khỏi giá thể.

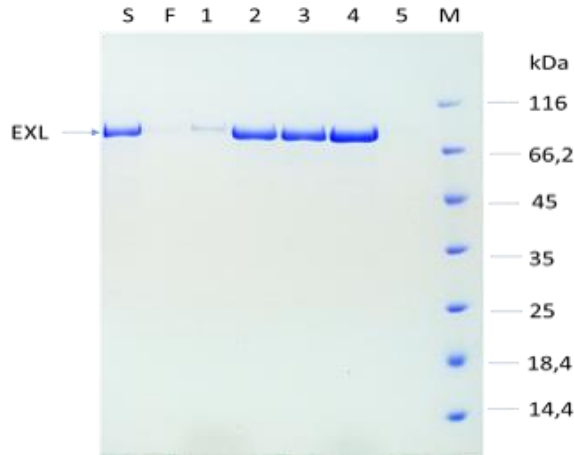


Hình 3.23. Điện di đồ kiểm tra sản phẩm trong các phân đoạn tinh chế lần 1.

S, P: protein pha tan và pha không tan của mẫu biểu hiện endo-xylanase; F: mẫu protein sau khi đưa qua giá thể; 1: dịch rửa bằng đệm chứa 50 mM imidazol; 2-3: dịch rửa bằng đệm chứa 100 mM imidazol; 4-10: các phân đoạn mẫu khi rửa bằng đệm chứa 250 mM imidazol; M: thang protein chuẩn (Thermo Scientific)

Mẫu protein sau khi biểu hiện được hòa loãng trong đệm cân bằng với pH

thích hợp cho việc bám của protein lên giá thể. Kết quả điện di trên Hình 3.24 cho thấy protein tái tổ hợp đã bám hoàn toàn vào chất giá sau khi được đưa qua cột, thể hiện ở việc hầu như không xuất hiện băng protein tương ứng với xylanase ở trong mẫu sau khi đưa qua giá thể.



Hình 3. 24. Điện di đồ kiểm tra sản phẩm trong các phân đoạn tinh chế lần 2

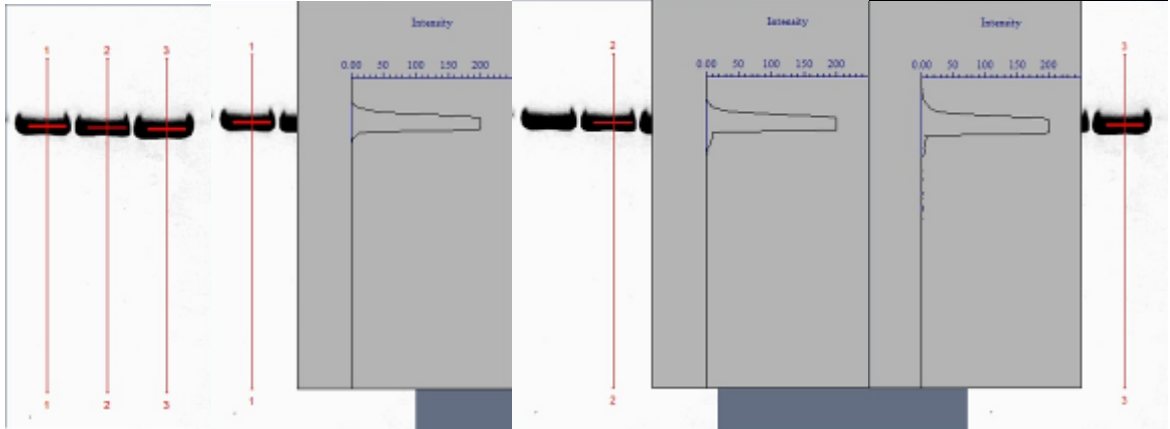
S; protein tinh chế lần 1 đã được pha loãng; F: dịch sau khi đưa lên cột lần 2; 1: dịch rửa mẫu protein bằng đệm chứa 100 mM imidazol; 2-5: các phân đoạn thu hồi EXL bằng đệm chứa 250 mM imidazol; M: thang protein chuẩn (Thermo Scientific)

Với nồng độ thích hợp của chất cạnh tranh imidazole, các protein nội bào bám không đặc hiệu với giá thể đã bị rửa trôi bằng đệm rửa, cho phép thu lại protein tái tổ hợp xylanase tương đối sạch trong các phân đoạn thu lại bằng đệm đầy. Tuy nhiên, vẫn còn có thể quan sát được những protein tạp trong các phân đoạn trên. Vì vậy để thu được protein có độ tinh sạch cao, sản phẩm của lần tinh chế lần 1 được sử dụng để tinh chế lại lần 2 cũng bằng cột sắc ký ái lực His-tag. Mẫu protein sau khi tinh chế lần 1 được pha loãng bằng đệm cân bằng để giảm nồng độ Imidazole từ 250 mM xuống còn 50 mM, là nồng độ thấp hơn nồng độ của Imidazole trong đệm rửa nhằm đảm bảo khả năng bám của protein tái tổ hợp vào giá thể trên cột sắc ký. Kết quả điện di cho thấy, đường chạy tương ứng với các phân đoạn của xylanase thu lại sau khi tinh chế lần 2 chỉ xuất hiện một băng duy nhất mà không quan sát thấy có băng nào khác, kích thước của băng protein này so với thang chuẩn protein là tương ứng với kích thước lý thuyết của protein xylanase dạng đơn. Để đảm bảo protein xylanase sau khi tinh chế đạt được độ sạch cần thiết, chúng tôi sử dụng phần mềm Quantity One để phân tích và đánh giá.

* Kiểm tra độ sạch của sản phẩm sau khi tinh chế bằng phần mềm Quantity One

Phần mềm Quantity One của BioRad là phần mềm thông dụng và được sử dụng nhiều trên thế giới để xác định độ đậm nhạt của các băng protein trên điện di

đồ. Chúng tôi sử dụng phần mềm này để xác định độ tinh khiết của protein sản phẩm được điện di trên gel acrylamide bằng kỹ thuật SDS-PAGE.

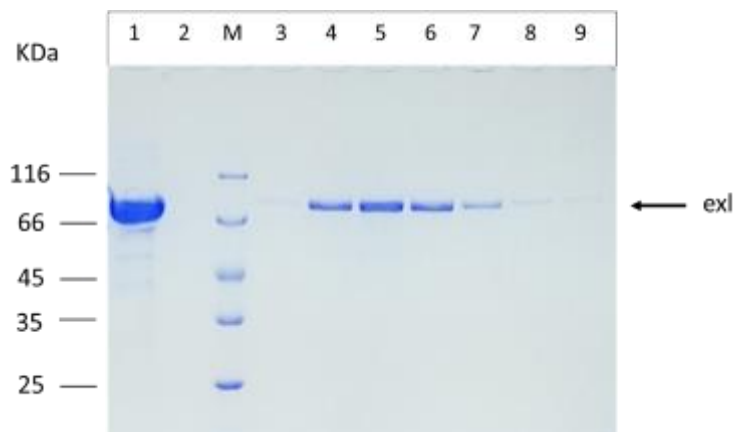


Hình 3. 25. Xác định độ tinh khiết của protein bằng phần mềm Quantity One

Kết quả phân tích bằng phần mềm dựa trên phương pháp so sánh tỷ lệ băng protein cần phân tích với toàn bộ đường chạy cho thấy, các mẫu protein sau khi tinh sạch chỉ có một băng đậm, các băng protein xylanase đều có độ sạch rất cao, lần lượt là 99,53% - 95,87% - 95,56%, đạt tiêu chuẩn của protein tinh sạch có độ tinh khiết trên 95%.

3.3.3.2. Loại muối protein *endo-xylanase* sau khi tinh sạch

Mẫu sau khi tinh sạch cần phải được rửa lại bằng ammonium sulfate bão hòa và hòa lại với nước deion với thể tích tối thiểu. Để loại bỏ hoàn toàn muối và các hóa chất khác, mẫu sau khi rửa được tiến hành loại muối bằng cột loại muối PD-10. Kết quả loại muối được thể hiện ở hình sau.



Hình 3. 26. Điện di đồ kiểm tra sản phẩm trong các phân đoạn loại muối.

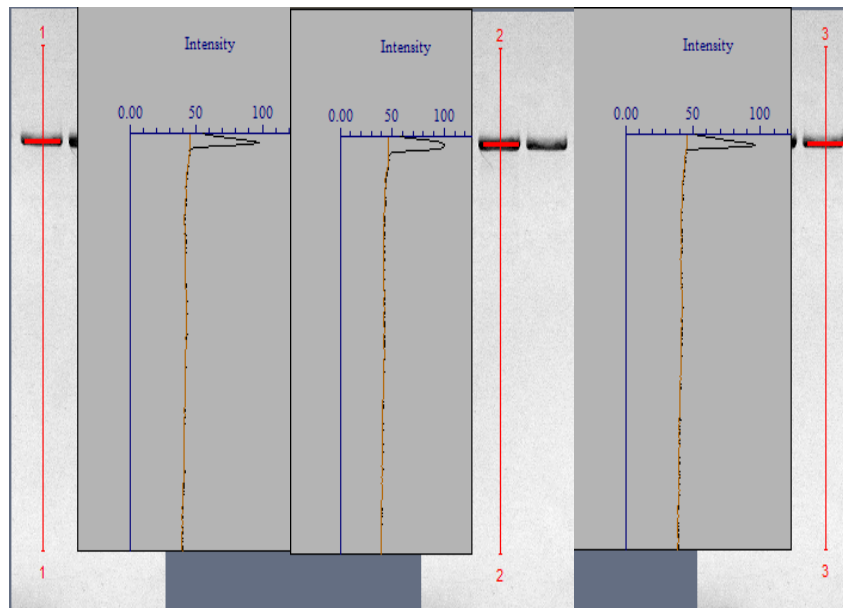
Đường chạy 1: protein tổng số trước khi loại muối; Đường chạy 2: dịch thu lại sau khi đưa mẫu protein qua cột; Đường chạy 3-9: các phân đoạn thu protein xylanase sau khi loại muối; Đường chạy M: thang protein chuẩn (Thermo Scientific)

Mẫu protein sau khi tinh sạch bằng sắc ký thường có nồng độ tương đối thấp

và thể tích lớn, không thuận tiện cho việc thu hồi và bảo quản cũng như thử nghiệm các tính chất của protein. Đồng thời, imidazole và muối có trong mẫu sau khi tinh sạch thường gây tác động xấu, khiến cho enzyme hoạt động kém hiệu quả hoặc không thể hoạt động xúc tác. Kết quả điện di sản phẩm loại muối của protein sau khi tinh sạch thể hiện trên Hình 3.26 cho thấy lượng mẫu protein tái tổ hợp đưa lên cột là nằm trong giới hạn của cột, thể hiện ở việc hầu như không xuất hiện băng protein tương ứng với xylanase ở trong mẫu sau khi đưa qua giá thể. Đường chạy tương ứng với các phân đoạn của xylanase thu lại sau khi loại muối chỉ xuất hiện một băng duy nhất khoảng 80 kDa của protein xylanase dạng đơn mà không quan sát thấy có băng nào khác. Các protein sau khi loại muối tập trung vào khoảng 3 phân đoạn, điều này giúp đảm bảo mẫu protein sau khi loại muối có nồng độ tương đối cao. Để kiểm tra độ tinh sạch của protein xylanase, chúng tôi sử dụng phần mềm Quantity One để phân tích và đánh giá.

* Kiểm tra độ sạch và nồng độ của protein sau khi loại muối

3 phân đoạn tương ứng với mẫu trên trên đường chạy 4,5,6 của Hình 3.26 được kiểm tra độ sạch bằng phần mềm Quantity One.



Hình 3. 27. Kết quả kiểm tra độ sạch của mẫu bằng phần mềm Quantity One

Độ sạch của mẫu protein được tính bằng phần trăm tỷ lệ protein đã được tinh sạch trên tổng số protein có trong mẫu. Bằng phần mềm Quantity One, độ sạch của 3 phân đoạn trên Hình 3.27 được tính toán đạt các giá trị lần lượt là 97% - 95% - 98%, hoàn toàn thỏa mãn yêu cầu mẫu protein sau khi tinh sạch có độ tinh sạch từ trên 95%. Bằng phương pháp Bradford, phân đoạn 3 của mẫu protein sau khi loại muối tương ứng với đường chạy số 5 trên Hình 3.26 có nồng độ protein đạt 300

$\mu\text{g/ml}$. Mẫu protein sau khi loại muối sẽ được kiểm tra hoạt tính sinh học bằng phương pháp DNS.

* Các thông số của quá trình tinh sạch và đo hiệu suất thu hồi

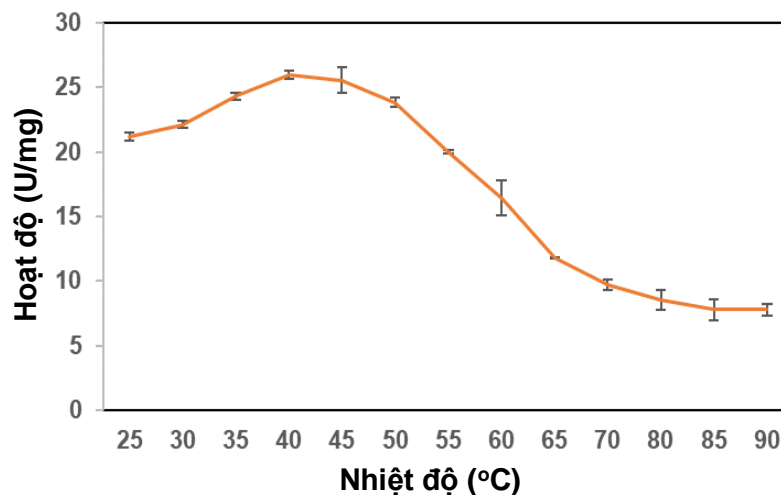
Bảng 3. 13. Các thông số các lần tinh chế enzyme endo-xylanase

Các thông số tinh chế enzyme	Lần 1	Lần 2	Lần 3
Hàm lượng protein tổng số (mg/ml)	8,35	8,2	8,5
Hoạt tính enzyme tổng số (IU/ml)	3,344	3,319	3,39
Lượng mẫu load lên cột tinh chế (ml)	20	20	20
Hoạt tính tổng của mẫu load lên cột (IU)	66,89	66,39	67,81
Thể tích mẫu thu từ mẻ tinh chế (ml)	5	5	5
Nồng độ enzyme của mẫu tinh chế (mg/ml)	0,3	0,295	0,31
Hoạt tính của mẫu tinh chế (IU/mg)	34,88	35,37	32,57
Hiệu suất thu hồi theo đơn vị hoạt tính (IU)	25,02%	25,15%	23,82%

Protein xylanase tái tổ hợp đã được tinh sạch và loại muối lặp lại 3 lần với độ lặp lại cao. Hoạt tính của protein sau khi tinh sạch được thu hồi với tỷ lệ khoảng 25%, cho thấy để đảm bảo độ tinh sạch cao của protein sau khi tinh sạch, có một lượng tương đối lớn protein đã bị thất thoát.

3.3.4. Nghiên cứu đặc tính enzyme xylanase và các thông số động học enzyme

3.3.4.1. Nghiên cứu xác định nhiệt độ tối ưu cho enzyme hoạt động

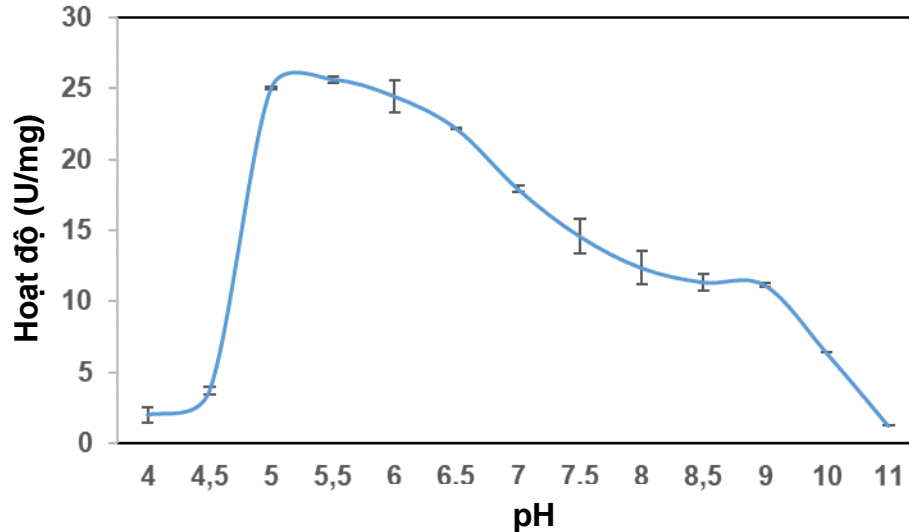


Hình 3.28. Biểu đồ thể hiện sự ảnh hưởng của nhiệt độ đến hoạt tính xylanase

Kết quả nghiên cứu cho thấy, enzyme xylanase thể hiện hoạt tính cao nhất khi được ủ ở nhiệt độ 40°C, và thể hiện hoạt tính tương đối mạnh khi nhiệt độ phản ứng nằm trong khoảng 35°C - 50°C (đạt trên 90% hoạt tính tối đa so với nhiệt độ 40°C).

Mặc dù protein xylanase được dự đoán là có chỉ số nóng chảy T_m trên 65°C nhưng xylanase xúc tác phản ứng phân giải xylan yếu hơn khi được ủ ở nhiệt độ 30°C trở xuống và 55°C trở lên, và hoạt tính giảm mạnh khi được ủ ở nhiệt độ trên 65°C . Như vậy khoảng nhiệt độ thích hợp nhất cho hoạt động của enzyme là từ 40°C đến 45°C . Theo kết quả nghiên cứu của Taberero và cộng sự, xylanase có nguồn gốc từ *Bacillus* sp. được biểu hiện tái tổ hợp trong *E. coli*, enzyme tái tổ hợp có nhiệt độ hoạt động tối ưu là 40°C [217]. Đây cũng là nhiệt độ hoạt động tối ưu của xylanase có nguồn gốc từ *Clostridium cellulovorans* [218], *Paenibacillus xylanilyticus* [216], đây đều là những trực khuẩn Gram (+), bên cạnh xylanase có nguồn gốc từ *Fibrobacter succinogenes* [219], một vi khuẩn cellulolytic và thường xuyên được tìm thấy trong dạ cỏ của các loài gia súc, hay là xylanase phân lập từ DNA metagenome vi sinh vật của mẫu đất [220], trong mẫu phân của người [221]. Tuy nhiên, dựa trên nghiên cứu của Alvarez và cộng sự năm 2013, nhiệt độ này thấp hơn nhiệt độ tối ưu của endo-xylanase có nguồn gốc từ một loài vi khuẩn chưa được định danh trong mẫu DNA metagenome của đất [222]. Nhiệt độ tối ưu của một số xylanase đã được nghiên cứu được liệt kê ở bảng S1 của phụ lục.

3.3.4.2. Nghiên cứu xác định pH tối ưu cho enzyme hoạt động



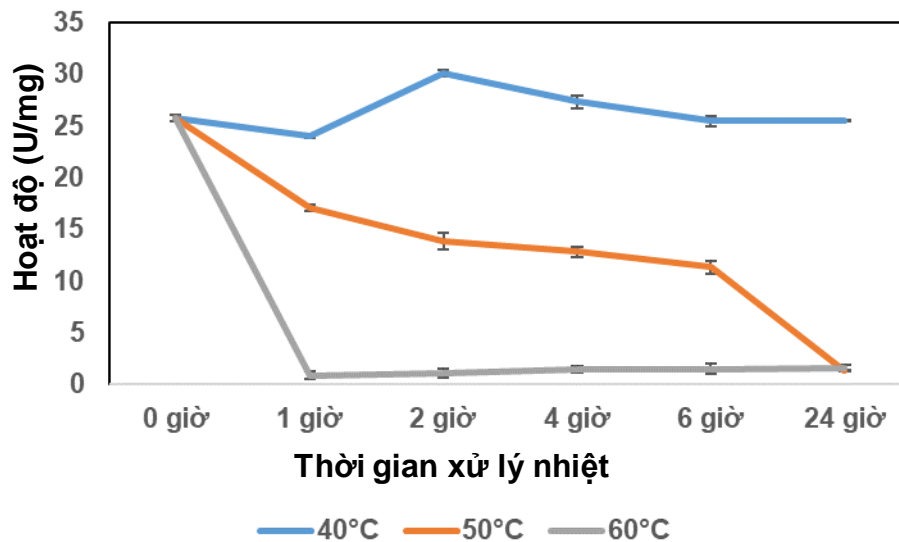
Hình 3. 29. Biểu đồ thể hiện sự ảnh hưởng của pH đến hoạt tính xylanase

Về ảnh hưởng của giá trị pH, trong nghiên cứu này, enzyme xylanase được thử hoạt tính trong dải pH từ 4 đến 11 để tìm điều kiện tối ưu. Mặc dù được dự đoán là enzyme hoạt động ở môi trường kiềm nhẹ nhưng kết quả nghiên cứu cho thấy, protein xylanase thể hiện hoạt tính cao nhất khi được ủ ở pH 5,5 đạt 25,6 U/mg, và thể hiện hoạt tính tương đối mạnh khi pH phản ứng nằm trong khoảng 5-6 (đạt trên 90% hoạt tính tối đa so với pH 5,5). Xylanase xúc tác phản ứng phân giải xylan yếu

hơn khi được ủ ở pH 6,5 trở lên, và gần như mất khả năng xúc tác khi được ủ ở pH dưới 4,5 hoặc trên 10. Kết luận được xylanase tái tổ hợp EXL là enzyme hoạt động ở môi trường pH trung tính và hơi acid. Theo các kết quả nghiên cứu của Mannarelli và cộng sự năm 1990, Meng và cộng sự năm 2009, Qiu và cộng sự năm 2010, Kim và cộng sự năm 2014, xylanase có nguồn gốc từ *Butyrivibrio fibrisolvens* [223], *Demequina* sp. JK4 [224], *Streptomyces megasporus* DSM 41476 [225], *Streptomyces mexicanus* HY-14 [226] cũng có pH hoạt động tối ưu ở 5,5. Phần lớn các xylanase đã được biểu hiện và nghiên cứu sâu về tính chất cũng đều hoạt động tốt nhất ở khoảng pH trung tính và hơi acid. pH tối ưu của một số xylanase đã được nghiên cứu được liệt kê ở bảng S2 của phụ lục.

3.3.4.3. Nghiên cứu tính bền nhiệt của enzyme

Để kiểm tra độ bền nhiệt của enzyme, thành phần và điều kiện của phản ứng kiểm tra hoạt tính xylanase được giữ nguyên, tuy nhiên xylanase tái tổ hợp được xử lý ở 3 nhiệt độ 40°C, 50°C, 60°C trong các khoảng thời gian khác nhau 1, 2, 4, 6 và 24 giờ trước khi được trộn với cơ chất. Các thành phần của đối chứng âm cũng được xử lý ở cùng điều kiện để đảm bảo loại bỏ yếu tố ảnh hưởng. Kết quả đo hoạt tính của xylanase sau khi xử lý nhiệt được thể hiện ở hình sau.



Hình 3. 30. Biểu đồ nghiên cứu sự bền nhiệt của xylanase

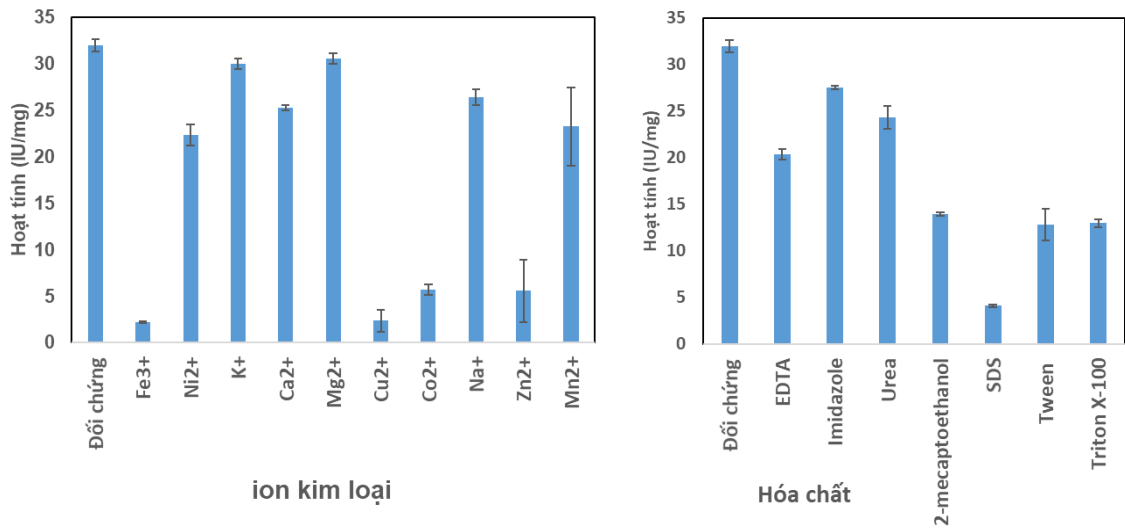
Kết quả nghiên cứu cho thấy, enzyme xylanase bền tốt ở nhiệt độ tối ưu cho hoạt tính là 40°C, thể hiện ở việc hoạt tính enzyme được giữ nguyên sau khi được ủ ở nhiệt độ này đến 24 giờ. Ở nhiệt độ 50°C, hoạt tính bắt đầu giảm chỉ sau 1 giờ xử lý (còn khoảng 66%) và giảm dần theo thời gian, chỉ còn khoảng 44% sau khi ủ 6 giờ và hoàn toàn mất hoạt tính sau khi ủ trong 24 giờ. Ở nhiệt độ cao (60°C), enzyme mất hoàn toàn hoạt tính chỉ sau 1 giờ ủ. So sánh với kết quả nghiên cứu của

Amaya-Delgado và cộng sự, enzyme endo-xylanase Xyn11A nguồn gốc từ *Cellulomonas flavigena* được biểu hiện tái tổ hợp trong *E. coli* có hoạt tính tối ưu ở nhiệt độ 55°C nhưng chỉ giữ được 50% hoạt tính tối đa sau khi ủ ở nhiệt độ tối ưu trong một giờ [227]. Tương tự, enzyme xylanase tái tổ hợp XylK1 nguồn gốc từ *Cellulosimicrobium* sp. HY-13 biểu hiện trong *E. coli* cũng có hoạt tính tối ưu ở nhiệt độ 55°C, tuy nhiên enzyme mất 50% hoạt tính sau khi ủ 10 phút ở nhiệt độ tối ưu [228]. Một enzyme xylanase khác, xynAS9 có nguồn gốc từ *Streptomyces* sp. S9 biểu hiện tái tổ hợp trong *E. coli* có hoạt tính tối ưu ở nhiệt độ 60°C và giữ được 80% hoạt tính sau khi được xử lý ở nhiệt độ này trong một giờ [229]. Vì vậy có thể kết luận enzyme EXL từ luận án không phải là enzyme chịu nhiệt nhưng có khả năng bền ở nhiệt độ tối ưu tốt hơn một số enzyme xylanase tái tổ hợp đã được công bố.

3.3.4.4. Nghiên cứu ảnh hưởng của một số ion kim loại và hóa chất lên hoạt tính xylanase

Thông thường, các ion kim loại và các loại hóa chất có thể làm tăng cường hoặc hạn chế sự hoạt động xúc tác của enzyme đối với cơ chất, đặc biệt là với những enzyme có phối tử kim loại. Theo nghiên cứu của Bai và cộng sự năm 2012, xylanase có nguồn gốc từ *Bacillus* sp. SN5 được biểu hiện tái tổ hợp trong *E. coli* thể hiện hoạt tính cao nhất tăng 34% trong điều kiện môi trường đệm chứa ion Na^+ 0,5 M và vẫn giữ được 90% hoạt tính trong điều kiện môi trường chứa đến 2,5 M NaCl [230]. Xylanase có nguồn gốc từ *Geobacillus* sp. 71 và biểu hiện tái tổ hợp trong *E. coli* bị mất hoạt tính hoàn toàn khi bổ sung Hg^{2+} và bị giảm hoạt tính một phần khi có sự có mặt của Al^{3+} và Cu^{2+} [231]. Trong một nghiên cứu khác, xylanase tái tổ hợp nguồn gốc từ *Geobacillus* sp. MT-1 biểu hiện trong *E. coli* thể hiện hoạt tính ổn định trong 1 mM các chất ức chế (PMSF, EDTA, 2-ME hoặc DTT) và trong 0,1% chất tẩy rửa (tween 20 hoặc triton X-100), trong khi đó, enzyme bị ức chế mạnh bởi 1 mM SDS, về mặt ion kim loại, enzyme bị ức chế mạnh bởi Ni^{2+} , Mn^{2+} , Co^{2+} , Cu^{2+} , Zn^{2+} , Cd^{2+} , Hg^{2+} và Al^{3+} nồng độ 1 hoặc 0,1 mM [232]. Tương tự, Cu^{2+} , Zn^{2+} , Pb^{2+} và Hg^{2+} cũng làm ức chế hoạt động của xylanase nguồn gốc từ *Kocuria* sp. Mn22 biểu hiện tái tổ hợp trong *E. coli* [233].

Xylanase tái tổ hợp EXL được đo hoạt tính trong điều kiện phản ứng bổ sung một số loại ion kim loại cũng như một số loại hóa chất ở nồng độ cao. Cụ thể khi so sánh hoạt độ enzyme khi thủy phân xylan bằng enzyme xylanase tái tổ hợp có bổ sung các loại hóa chất với khi thủy phân trong môi trường đệm không bổ sung thì kết quả thể hiện trên Hình 3.31.



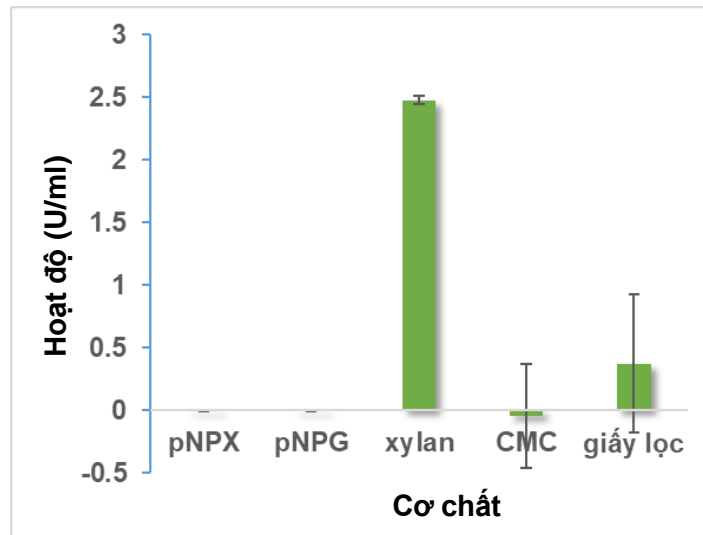
Hình 3. 31. Biểu đồ ảnh hưởng của ion kim loại và hóa chất đến hoạt độ xylanase

Kết quả thử hoạt tính xylanase chứng minh khi có sự có mặt của các ion kim loại nặng Fe³⁺, Cu²⁺, Co²⁺, Zn²⁺ và chất biến tính SDS thì hoạt tính của enzyme xylanase giảm mạnh, đặc biệt hai ion Fe³⁺, Co²⁺ gần như làm cho enzyme mất hoạt tính (hoạt tính chỉ đạt 7% so với mẫu đối chứng không bổ sung ion). Khi có sự có mặt của ion K⁺ và Mg²⁺ thì hoạt tính của enzyme giảm không đáng kể, còn các ion còn lại như Ni²⁺, Ca²⁺, Na⁺, Mn²⁺ đều làm giảm rõ rệt hoạt tính của enzyme với các mức độ khác nhau. Tuy nhiên đặc biệt là khi có sự có mặt của EDTA là một chất có ái lực cao với các ion kim loại tự do, enzyme vẫn giữ được 64% hoạt tính so với mẫu đối chứng, điều này chứng tỏ enzyme xylanase tái tổ hợp có khả năng hoạt động mà ít phụ thuộc sự có mặt của ion kim loại. Các hóa chất 2-mecaptoethanol, Tween-20 và Triton X100 cũng ức chế hoạt tính của endo-xylanase EXL, chỉ còn khoảng 40% so với mẫu không bổ sung các chất này, tuy nhiên imidazole nồng độ cao chỉ làm hoạt tính của EXL giảm tương đối nhỏ, điều này giúp hạn chế ảnh hưởng của quá trình tinh sạch enzyme đến quá trình xúc tác của enzyme.

3.3.4.5. Nghiên cứu tính đặc hiệu cơ chất của enzyme

Enzyme xylanase được xác định có hai vùng hoạt tính là vùng xúc tác glycosyl hydrolase 10 nằm ở đầu N và carbohydrate esterase 1 nằm ở đầu C. Theo phân loại của CAZy, enzyme thuộc họ GH10 có thể thể hiện hoạt tính endo-1,4-β-xylanase (EC 3.2.1.8), endo-1,3-β-xylanase (EC 3.2.1.32) và cả endo-β-1,4-glucanase (EC 3.2.1.4). Trên thực tế, có những enzyme đã được công bố là mang cả hai hoạt tính xylanase và glucanase. Năm 2011, Chang và cộng sự đã tách dòng gene *RuCelA* từ thư viện metagenomic khai thác từ hệ vi sinh vật dạ cỏ bò và biểu

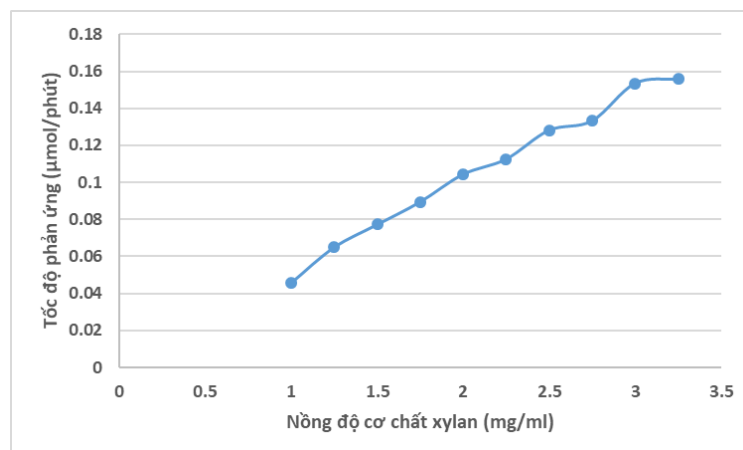
hiện thành enzyme endo-glucanase RuCelA, enzyme này xác định có hoạt tính của xylanase [234]. Kết quả tương tự cũng được phát hiện trong nghiên cứu của Ghatge và cộng sự (2014) cho thấy endo-glucanase Cel5 từ chủng *Hahella chejuensis* KCTC 2396 là một enzyme đa chức năng có cả hoạt tính của endo-glucanase và xylanase [235]. Để kiểm tra khả năng phân giải cơ chất của EXL tái tổ hợp, enzyme sau khi được tinh sạch được ủ với các loại cơ chất bao gồm xylan, pNPX, pNPG, CMC, giấy lọc.



Hình 3. 32. Biểu đồ thể hiện tính đặc hiệu cơ chất của xylanase

Dựa trên biểu đồ cho thấy, xylanase thể hiện hoạt tính mạnh với cơ chất đặc hiệu là xylan, ngược lại gần như không thể hiện hoạt tính với các loại cơ chất khác là CMC, giấy lọc, pNPG và pNPX, cụ thể là hoạt tính xúc tác đối với các loại cơ chất này là không đáng kể. Vì vậy có thể kết luận enzyme xylanase có tính đặc hiệu với cơ chất xylan.

3.3.4.6. Nghiên cứu thông số động học của enzyme endo xylanase tái tổ hợp



Hình 3. 33. Đồ thị tương quan giữa nồng độ cơ chất và tốc độ phản ứng của enzyme endo xylanase tái tổ hợp.

Dựa trên kết quả thực nghiệm, đường chuẩn độ tương quan giữa nồng độ cơ chất và tốc độ phản ứng của enzyme endo-xylanase đã được xây dựng. Để thuận tiện cho việc xác định các thông số động học, đồ thị tương quan giữa nghịch đảo giá trị nồng độ cơ chất $1/[S]$ và tốc độ phản ứng $1/V$ được xây dựng (phụ lục hình S1) là $y=16,978x + 1,1658$ với hệ số tương quan R^2 lên đến 0,9971 hay độ tin cậy lên đến hơn 99%. Từ phương trình trên dựa trên công thức của phương trình Lineweaver-Berk, các thông số động học là K_m và V_{max} của enzyme xylanase được xác định tương ứng là 14,56 mg/ml và 0,86 $\mu\text{mol/phút}$ với hoạt tính riêng đạt 171,56 IU/mg đối với cơ chất xylan.

Protein xylanase được khai thác từ cơ sở dữ liệu DNA metagenome vi sinh vật ruột dê bằng cách sử dụng trình tự tương đồng khi so sánh các trình tự xylanase thuộc họ GH10 trên ngân hàng gen, với điều kiện là các trình tự này đã được biểu hiện và nghiên cứu tính chất. Khi so sánh với các trình tự xylanase đã được nghiên cứu tính chất này, protein xylanase tái tổ hợp có giá trị K_m cao nhất, cụ thể là gấp hơn hai lần so với các trình tự có nguồn gốc từ *Paenibacillus* sp. HPL-001 [236], *Kocuria* sp. [233] và gấp khoảng 10 lần so với các trình tự có nguồn gốc từ *Bacteroides xylanisolvens* [237] và *Geobacillus* sp. MT-1 [232] so với cùng cơ chất xylan. Tuy nhiên nếu so sánh với giá trị K_m của các protein xylanase khác đã được nghiên cứu tính chất có nguồn gốc từ *Nesterenkonia xinjiangensis* [238] và *Paenibacillus barcinonensis* [239] lần lượt là 16,08 mg/ml và 12,98 mg/ml thì giá trị K_m của xylanase tái tổ hợp EXL không chênh lệch quá lớn. Giá trị K_m thể hiện tính ái lực của enzyme với cơ chất, nên giá trị này cho thấy enzyme xylanase tái tổ hợp có ái lực trung bình với cơ chất xylan.

Mặt khác, giá trị V_{max} thể hiện tốc độ phản ứng tối đa mà enzyme có thể đạt được, và giá trị này càng lớn thể hiện enzyme càng có hoạt tính mạnh. So sánh với các trình tự xylanase đã được nghiên cứu và công bố, hoạt tính riêng của enzyme xylanase tái tổ hợp lớn gấp khoảng 2-3 lần so với các xylanase có nguồn gốc từ *Streptomyces* sp. TN119 [240] và *Paenibacillus barcinonensis* [241]. Enzyme xylanase tái tổ hợp được biểu hiện vì vậy có tiềm năng cao trong ứng dụng xử lý phân hủy chất thải lignocellulose bảo vệ môi trường.

KẾT LUẬN VÀ KIẾN NGHỊ

KẾT LUẬN

1. DNA đa hệ gene của vi khuẩn trong dạ cỏ dê đã được giải mã với dung lượng thông thường là 8,6 Gb và dung lượng lớn (giải mã sâu) là 48,66 Gb. Từ bộ dữ liệu giải mã sâu, 3.411.867 contig đã được lắp ghép với tỷ lệ bao phủ là 64,22%, 5.367.270 gene đã được chú giải trong đó 4.311.093 gene (80,32%) đã được phân loại. Cả hai bộ dữ liệu đều dự đoán các gene có nguồn gốc vi khuẩn chiếm 99,8%. Ngành Bacteroidetes là ngành chiếm tỷ lệ lớn nhất với 45,29% tổng số gen, tiếp theo là ngành Firmicutes với 30,38%. Ở mức độ chi, 49,93% số gene vẫn chưa được phân loại. Chi chiếm tỷ lệ lớn nhất là *Prevotella*, chiếm 25,79% tổng số gen.

2. Từ bộ dữ liệu 48,66 Gb, 65.443 gene mã hóa cho 30 enzyme/protein liên quan đến quá trình phân hủy lignocellulose đã được xếp vào các đơn vị phân loại (chiếm 99,85% tổng số gene lignocellulase) trong đó có 21029 gene cellulase, 41756 gene hemicellulase và 4769 gene tiền xử lý lignocellulose. Chi *Prevotella* có vai trò quan trọng, cung cấp tới 27% gene tham gia phân giải lignocellulose. Các gene này được xác định trên các cấu trúc PUL để tăng cường thủy phân lignocellulose. Công cụ mô hình đại diện HMM đã được thiết kế thành công cho khai thác hiệu quả các gene galactanase, glucuronyl esterase, hydrogen peroxide oxidoreductase, xyloglucanase, laccase, CBM (1-84), cellobiohydrolase, beta-glucuronidase, beta-xylosidase, beta-mannosidase GH2, lichenase, alpha-glucuronidase (GH76N) và xylanase GH44.

3. Trong số 739 gene hoàn chỉnh mã hóa cho endo-xylanase có 108 trình tự được xác định có nguồn gốc từ *Prevotella* đều thuộc họ GH10 và 10 enzyme có thêm vùng phụ trợ CE1 có T_m cao hơn 65°C và là các enzyme có tính acid. Trình tự denovogenes_5086 mã hóa cho endo-xylanase GH10-CE1 đã được lựa chọn, tối ưu mã bộ ba và biểu hiện thành công trong *E. coli*. Enzyme được tinh chế bằng cột sắc ký ái lực His-tag với độ tinh sạch trên 95%, hiệu suất thu hồi 25%. Enzyme tinh sạch có hoạt tính tối ưu ở 40°C, pH 5,5. Enzyme ổn định và bền ở nhiệt độ tối ưu đến 24 giờ nhưng mất hoạt tính ngay khi ủ ở 60°C trong 1 giờ. Enzyme hoạt động trên cơ chất đặc hiệu là xylan. Hoạt độ của enzyme giảm khi bổ sung các ion Fe^{3+} , Cu^{2+} , Co^{2+} , Zn^{2+} , Ni^{2+} , Ca^{2+} , Na^+ , Mn^{2+} (1 mM) và các loại hóa chất thường dùng là SDS (1%), urea (1 μ M), 2-mercaptoethanol (1 μ M), EDTA (1 μ M), tween 80 (1mM), triton X-100 (1 μ M). Các thông số động học là K_m và V_{max} của enzyme xylanase được xác định tương ứng là 14,56 mg/ml và 0,86 μ mol/phút với hoạt tính

riêng đạt 171,56 IU/mg.

KIẾN NGHỊ

Tiếp tục nghiên cứu phương pháp và điều kiện thích hợp để xác định hoạt tính esterase của enzyme EXL, tiến tới sản xuất hỗn hợp enzyme ứng dụng trong xử lý môi trường và sản xuất nguyên liệu - nhiên liệu sinh học.

DANH MỤC CÔNG TRÌNH CÔNG BỐ LIÊN QUAN ĐẾN LUẬN ÁN

1. **Trong-Khoa Dao**, Thi-Huyen Do, Ngoc-Giang Le, Hong-Duong Nguyen, Thi-Quy Nguyen, Thi-Thu-Hong Le, Nam-Hai Truong, 2021, Understanding the role of *Prevotella* genus in the digestion of lignocellulose and other substrates in Vietnamese native goats' rumen by metagenomic deep sequencing, *Animals* 11(11), 3257.
2. Thi-Huyen Do, **Trong-Khoa Dao**, Khanh-Hoang-Viet Nguyen, Ngoc-Giang Le, Thi-Mai-Phuong Nguyen, Tung-Lam Le, Thu-Nguyet Phung, Nico M. van Straalen, Dick Roelofs, Nam-Hai Truong, 2018, Metagenomic analysis of bacterial community structure and diversity of lignocellulolytic bacteria in Vietnamese native goat rumen. *Asian-Australasian journal of animal sciences*, 31(5), 738–747.
3. Thi Huyen Do, Ngoc Giang Le, **Trong Khoa Dao**, Thi Mai Phuong Nguyen, Tung Lam Le, Han Ly Luu, Khanh Hoang Viet Nguyen, Van Lam Nguyen, Lan Anh Le, Thu Nguyet Phung, Nico M. van Straalen, Dick Roelofs, Nam Hai Truong, 2018, Metagenomic insights into lignocellulose-degrading genes through Illumina-based de novo sequencing of the microbiome in Vietnamese native goats rumen. *Journal of General and Applied Microbiology*. 64(3):108-116.
4. Khanh Hoang Viet Nguyen, **Trong Khoa Dao**, Hong Duong Nguyen, Khanh Hai Nguyen, Thi Quy Nguyen, Thuy Tien Nguyen, Thi Mai Phuong Nguyen, Nam Hai Truong, Thi Huyen Do, 2021, Some characters of bacterial cellulases in goats' rumen elucidated by metagenomic DNA analysis and the role of fibronectin 3 module for endoglucanase function. *Animal Bioscience* 34(5): 867-879.
5. **Đào Trọng Khoa**, Đỗ Thị Huyền, Trương Nam Hải, 2018, Nghiên cứu biểu hiện expansin tái tổ hợp trong *Escherichia coli*, *Báo cáo khoa học Hội nghị Công nghệ sinh học toàn quốc 2018*, 138-142.
6. **Đào Trọng Khoa**, Đỗ Thị Huyền, Trương Nam Hải, 2020, Khai thác gene mã hóa endo-1,4-beta-xylanase từ dữ liệu DNA metagenome vi khuẩn trong dạ cỏ dê bằng mẫu dò, *Tạp chí Công nghệ Sinh học* 19(3): 519-528.
7. Trương Nam Hải, Đỗ Thị Huyền, **Đào Trọng Khoa**, 2021, Trình tự gene mã hóa expansin có nguồn gốc từ vi khuẩn trong dạ cỏ dê và expansin tái tổ hợp có khả năng làm tăng chuyển hóa xenluloza tinh thể của xenlulaza. Bằng độc quyền giải pháp hữu ích số 2701, số đơn 2-2017-00312, Cục sở hữu trí tuệ.

8. Nguyen Hai Dang, Do Thi Huyen, Nguyen Thi Kien, Ha Thi Thuy Hoa, Le Quynh Giang, **Dao Trong Khoa**, Truong Nam Hai, 2021, Expression of gene coding endoglucanase GH5-4 derived from meagenomic DNA data of bacteria in goats rumen in *Escherichia coli*, *Academia Journal of Biology*, 43(2): 17–26.

DANH MỤC TÀI LIỆU THAM KHẢO

1. Surriya O., Saleem S., Waqar K., Gul Kazi A., Öztürk M., 2015, Bio-fuels: A Blessing in Disguise, in *Phytoremediation for Green Energy*, M. Öztürk, M. Ashraf, A. Aksoy and M. Ahmad, Editors. Springer, Dordrecht. p. 11-54, DOI: doi.org/10.1007/978-94-007-7887-0_2.
2. Chen S., Wojcieszak R., Dumeignil F., Marceau E., Royer S., 2018, How Catalysts and Experimental Conditions Determine the Selective Hydroconversion of Furfural and 5-Hydroxymethylfurfural. *Chemical Reviews*, **118**(22): p. 11023-11117, DOI: 10.1021/acs.chemrev.8b00134.
3. Shrinivas N.S., Padmaja P.S., Krishnaraj P.U., 2020, Soil Metagenomics: Concepts and Applications, in *Metagenomics - Basics, Methods and Applications*. IntechOpen, DOI: 10.5772/intechopen.88958.
4. Xie F., Jin W., Si H., Yuan Y., Tao Y., Liu J., . . . Mao S., 2021, An integrated gene catalog and over 10,000 metagenome-assembled genomes from the gastrointestinal microbiome of ruminants. *Microbiome*, **9**(1): p. 137, DOI: 10.1186/s40168-021-01078-x.
5. Bodor A., Bounedjoum N., Vincze G.E., Erdeiné Kis Á., Laczi K., Bende G., . . . Rákhely G., 2020, Challenges of unculturable bacteria: environmental perspectives. *Reviews in Environmental Science and Bio/Technology*, **19**(1): p. 1-22, DOI: 10.1007/s11157-020-09522-4.
6. Doi R.H., Tamaru Y., 2001, The *Clostridium cellulovorans* cellulosome: an enzyme complex with plant cell wall degrading activity. *Chemical Record*, **1**(1): p. 24-32, DOI: 10.1002/1528-0691(2001)1:1<24::aid-tcr5>3.0.co;2-w.
7. Handelsman J., 2004, Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiology and Molecular Biology Reviews*, **68**(4): p. 669-685, DOI: 10.1128/mnbr.68.4.669-685.2004.
8. Kowalchuk G.A., Speksnijder A.G., Zhang K., Goodman R.M., van Veen J.A., 2007, Finding the needles in the metagenome haystack. *Microbial ecology*, **53**(3): p. 475-85, DOI: 10.1007/s00248-006-9201-2.
9. Saxena R., Dhakan D.B., Mittal P., Waiker P., Chowdhury A., Ghatak A., Sharma V.K., 2017, Metagenomic Analysis of Hot Springs in Central India Reveals Hydrocarbon Degrading Thermophiles and Pathways Essential for Survival in Extreme Environments. *Frontiers in Microbiology*, **7**(2123), DOI: 10.3389/fmicb.2016.02123.
10. Kennedy J., Flemer B., Jackson S.A., Lejon D.P.H., Morrissey J.P., O’Gara F., Dobson A.D.W., 2010, Marine Metagenomics: New Tools for the Study and Exploitation of Marine Microbial Metabolism. *Marine Drugs*, **8**(3): p. 608-628.
11. Xu Z., Hansen M.A., Hansen L.H., Jacquiod S., Sørensen S.J., 2014, Bioinformatic Approaches Reveal Metagenomic Characterization of Soil Microbial Community. *PLOS ONE*, **9**(4): p. e93445, DOI: 10.1371/journal.pone.0093445.
12. Martín R., Miquel S., Langella P., Bermúdez-Humarán L.G., 2014, The role of metagenomics in understanding the human microbiome in health and disease. *Virulence*, **5**(3): p. 413-423, DOI: 10.4161/viru.27864.

13. Stewart R.D., Auffret M.D., Warr A., Walker A.W., Roehe R., Watson M., 2019, Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nature Biotechnology*, **37**(8): p. 953-961, DOI: 10.1038/s41587-019-0202-3.
14. Kittelmann S., Janssen P.H., 2011, Characterization of rumen ciliate community composition in domestic sheep, deer, and cattle, feeding on varying diets, by means of PCR-DGGE and clone libraries. *FEMS microbiology ecology*, **75**(3): p. 468-81, DOI: 10.1111/j.1574-6941.2010.01022.x.
15. Lam K.N., Charles T.C., 2015, Strong spurious transcription likely contributes to DNA insert bias in typical metagenomic clone libraries. *Microbiome*, **3**(1): p. 22, DOI: 10.1186/s40168-015-0086-5.
16. Rondon M.R., August P.R., Bettermann A.D., Brady S.F., Grossman T.H., Liles M.R., . . . Goodman R.M., 2000, Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Applied and environmental microbiology*, **66**(6): p. 2541-7.
17. Henne A., Schmitz R.A., Bomeke M., Gottschalk G., Daniel R., 2000, Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Applied and environmental microbiology*, **66**(7): p. 3113-6.
18. Pareek C.S., Smoczynski R., Tretyn A., 2011, Sequencing technologies and genome sequencing. *Journal of applied genetics*, **52**(4): p. 413-35, DOI: 10.1007/s13353-011-0057-x.
19. Teeling H., Glockner F.O., 2012, Current opportunities and challenges in microbial metagenome analysis--a bioinformatic perspective. *Briefings in bioinformatics*, **13**(6): p. 728-42, DOI: 10.1093/bib/bbs039.
20. Kanehisa M., Araki M., Goto S., Hattori M., Hirakawa M., Itoh M., . . . Yamanishi Y., 2008, KEGG for linking genomes to life and the environment. *Nucleic acids research*, **36**(Database issue): p. D480-4, DOI: 10.1093/nar/gkm882.
21. Powell S., Szklarczyk D., Trachana K., Roth A., Kuhn M., Muller J., . . . Bork P., 2012, eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic acids research*, **40**(Database issue): p. D284-9, DOI: 10.1093/nar/gkr1060.
22. Tatusov R.L., Fedorova N.D., Jackson J.D., Jacobs A.R., Kiryutin B., Koonin E.V., . . . Natale D.A., 2003, The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*, **4**: p. 41, DOI: 10.1186/1471-2105-4-41.
23. El-Gebali S., Mistry J., Bateman A., Eddy S.R., Luciani A., Potter S.C., . . . Finn R.D., 2018, The Pfam protein families database in 2019. *Nucleic Acids Research*, **47**(D1): p. D427-D432, DOI: 10.1093/nar/gky995.
24. Haft D.H., Selengut J.D., Richter R.A., Harkins D., Basu M.K., Beck E., 2012, TIGRFAMs and Genome Properties in 2013. *Nucleic Acids Research*, **41**(D1): p. D387-D395, DOI: 10.1093/nar/gks1234.
25. Zhang L., Chen F., Zeng Z., Xu M., Sun F., Yang L., . . . Xie Y., 2021, Advances in Metagenomics and Its Application in Environmental

- Microorganisms. *Frontiers in microbiology*, **12**: p. 766364, DOI: 10.3389/fmicb.2021.766364.
26. Wu F., Zhao S., Yu B., Chen Y.M., Wang W., Song Z.G., . . . Zhang Y.Z., 2020, Author Correction: A new coronavirus associated with human respiratory disease in China. *Nature*, **580**(7803): p. E7, DOI: 10.1038/s41586-020-2202-3.
 27. Aron-Wisnewsky J., Vigliotti C., Witjes J., Le P., Holleboom A.G., Verheij J., . . . Clément K., 2020, Gut microbiota and human NAFLD: disentangling microbial signatures from metabolic disorders. *Nature reviews. Gastroenterology & hepatology*, **17**(5): p. 279-297, DOI: 10.1038/s41575-020-0269-9.
 28. Svoboda E., 2021, Gut feeling yields evidence of microbial involvement in autoimmunity. *Nature*, **595**(7867): p. S54–S55, DOI: 10.1038/d41586-021-01837-8
 29. Andrews M.C., Duong C.P.M., Gopalakrishnan V., Iebba V., Chen W.S., Derosa L., . . . Wargo J.A., 2021, Gut microbiota signatures are associated with toxicity to combined CTLA-4 and PD-1 blockade. *Nature medicine*, **27**(8): p. 1432-1441, DOI: 10.1038/s41591-021-01406-6.
 30. Nnadozie C.F., Odume O.N., 2019, Freshwater environments as reservoirs of antibiotic resistant bacteria and their role in the dissemination of antibiotic resistance genes. *Environmental pollution*, **254**(Pt B): p. 113067, DOI: 10.1016/j.envpol.2019.113067.
 31. Arango-Argoty G., Garner E., Pruden A., Heath L.S., Vikesland P., Zhang L., 2018, DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome*, **6**(1): p. 23, DOI: 10.1186/s40168-018-0401-z.
 32. Miller S., Naccache S.N., Samayoa E., Messacar K., Arevalo S., Federman S., . . . Chiu C.Y., 2019, Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. *Genome research*, **29**(5): p. 831-842, DOI: 10.1101/gr.238170.118.
 33. Di L., Jean-Luc L., Pangzhen Z., Deli C., Kate H., 2021, Diversity and dynamics of fungi during spontaneous fermentations and association with unique aroma profiles in wine. *International Journal of Food Microbiology*, **338**: p. 108983, DOI: <https://doi.org/10.1016/j.ijfoodmicro.2020.108983>.
 34. Sharma T.R., Devanna B.N., Kiran K., Singh P.K., Arora K., Jain P., . . . Mondal T.K., 2018, Status and Prospects of Next Generation Sequencing Technologies in Crop Plants. *Current issues in molecular biology*, **27**: p. 1-36, DOI: 10.21775/cimb.027.001.
 35. Gidlund H., Vaga M., Ahvenjärvi S., Rinne M., Ramin M., Huhtanen P., 2018, Predicting omasal flow of nonammonia N and milk protein yield from in vitro-determined utilizable crude protein at the duodenum. *Journal of dairy science*, **101**(2): p. 1164-1176, DOI: 10.3168/jds.2017-13422.
 36. Seshadri R., Leahy S.C., Attwood G.T., Teh K.H., Lambie S.C., Cookson A.L., . . . Kelly W.J., 2018, Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nature Biotechnology*, **36**(4): p. 359-367, DOI: 10.1038/nbt.4110.

37. Abdelmegeid M.K., Elolimy A.A., Zhou Z., Lopreiato V., McCann J.C., Loor J.J., 2018, Rumen-protected methionine during the periparturient period in dairy cows and its effects on abundance of major species of ruminal bacteria. *Journal of Animal Science and Biotechnology*, **9**(1): p. 17, DOI: 10.1186/s40104-018-0230-8.
38. Wang L., Liu K., Wang Z., Bai X., Peng Q., Jin L., 2019, Bacterial Community Diversity Associated With Different Utilization Efficiencies of Nitrogen in the Gastrointestinal Tract of Goats. *Frontiers in microbiology*, **10**: p. 239, DOI: 10.3389/fmicb.2019.00239.
39. Niu Q., Li P., Hao S., Zhang Y., Kim S.W., Li H., . . . Huang R., 2015, Dynamic Distribution of the Gut Microbiota and the Relationship with Apparent Crude Fiber Digestibility and Growth Stages in Pigs. *Scientific Reports*, **5**(1): p. 9938, DOI: 10.1038/srep09938.
40. Thuy T.T., Lan N.N., Tung N.V., Man T.D., Hoang N.H., Thoa N.K., 2022, Analysis of microbial communities in Binh Chau hot spring through metagenomic DNA sequencing. *Vietnam Journal of Biotechnology*, **20**(1): p. 151-161, DOI: 10.15625/1811-4989/16073.
41. Kagaya Y., Minei R., Duong H.T.T., Le B.T.N., Dang L.T., Tran T.T.H., . . . Kim O.T.P., 2020, Metagenome Sequences from the Environment of Diseased Otter Clams, *Lutraria rhynchaena*, from a Farm in Vietnam. *Microbiology resource announcements*, **9**(2), DOI: 10.1128/mra.01068-19.
42. Kim O.T.P., Kagaya Y., Tran H.S., Minei R., Tran T.T.H., Duong H.T.T., . . . Yura K., 2020, A novel circular ssDNA virus of the phylum Cressnaviricota discovered in metagenomic data from otter clams (*Lutraria rhynchaena*). *Archives of virology*, **165**(12): p. 2921-2926, DOI: 10.1007/s00705-020-04819-9.
43. Lan-Anh L., Thien-Phuc N.-H., Van-Phuc H., Tan-Huy N., Thuy-Vy N., Thuy-Duong H.-H., 2020, Microbiome dataset analysis from a shrimp pond in Ninh Thuan, Vietnam using shotgun metagenomics. *Data in Brief*, **31**: p. 105731, DOI: <https://doi.org/10.1016/j.dib.2020.105731>.
44. Dinh Minh T., To Uyen H., Thi Huyen N., Tu Oanh D., Quang Vinh N., Anh Dzung N., 2022, Metagenomic next-generation sequencing of the microbiome dataset from the surface water sample collected from Serepok River in Yok Don National Park, Vietnam. *Data in Brief*, **45**: p. 108614, DOI: <https://doi.org/10.1016/j.dib.2022.108614>.
45. Hoang H.T., Le D.H., Le T.T.H., Nguyen T.T.N., Chu H.H., Nguyen N.T., 2021, Metagenomic 16S rDNA amplicon data of microbial diversity of guts in Vietnamese humans with type 2 diabetes and nondiabetic adults. *Data in Brief*, **34**: p. 106690, DOI: 10.1016/j.dib.2020.106690.
46. Hoang Duc L., Trung Nam N., Thi Hoang Oanh D., Georg S., Trung Thanh T., Huyen Trang V., . . . Hoang Ha C., 2018, Characterization of Bacterial Community in the Gut of *Penaeus monodon* and Its Culture Water in Shrimp Ponds. *Turkish Journal of Fisheries and Aquatic Sciences*, **19**(11): p. 977-986, DOI: 10.4194/1303-2712-v19_11_09.
47. Trung Thanh T., Nathan J. B., Dai Nhan L., Trung Nam N., Thi Hoang Oanh D., Hoang Duc L., . . . Hoang Ha C., 2019, The Role of *Pseudomonas* in Heterotrophic Nitrification: A Case Study on Shrimp Ponds (*Litopenaeus*

- vannamei) in Soc Trang Province. *Microorganisms*, **7**(6), DOI: 10.3390/microorganisms7060155.
48. Nguyen K.H.V., Dao T.K., Nguyen H.D., Nguyen K.H., Nguyen T.Q., Nguyen T.T., . . . Do T.H., 2021, Some characters of bacterial cellulases in goats' rumen elucidated by metagenomic DNA analysis and the role of fibronectin 3 module for endoglucanase function. *Animal bioscience*, **34**(5): p. 867-879, DOI: 10.5713/ajas.20.0115.
 49. Cantalapiedra C.P., Hernández-Plaza A., Letunic I., Bork P., Huerta-Cepas J., 2021, eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular biology and evolution*, **38**(12): p. 5825-5829, DOI: 10.1093/molbev/msab293.
 50. Törönen P., Holm L., 2022, PANNZER—A practical tool for protein function prediction. *Protein Science*, **31**(1): p. 118-128, DOI: <https://doi.org/10.1002/pro.4193>.
 51. Kanehisa M., Sato Y., Morishima K., 2016, BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*, **428**(4): p. 726-731, DOI: <https://doi.org/10.1016/j.jmb.2015.11.006>.
 52. Chen T.-W., Gan R.-C., Fang Y.-K., Chien K.-Y., Liao W.-C., Chen C.-C., . . . Tang P., 2017, FunctionAnnotator, a versatile and efficient web tool for non-model organism annotation. *Scientific Reports*, **7**(1): p. 10430, DOI: 10.1038/s41598-017-10952-4.
 53. Zheng J., Ge Q., Yan Y., Zhang X., Huang L., Yin Y., 2023, dbCAN3: automated carbohydrate-active enzyme and substrate annotation. *Nucleic Acids Research*, **51**(W1): p. W115-W121, DOI: 10.1093/nar/gkad328.
 54. Kelley L.A., Mezulis S., Yates C.M., Wass M.N., Sternberg M.J., 2015, The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*, **10**(6): p. 845-58, DOI: 10.1038/nprot.2015.053.
 55. Paysan-Lafosse T., Blum M., Chuguransky S., Grego T., Pinto B.L., Salazar Gustavo A., . . . Bateman A., 2023, InterPro in 2022. *Nucleic Acids Research*, **51**(D1): p. D418-D427, DOI: 10.1093/nar/gkac993.
 56. Zhao B., Hu S., Li X., Zhang F., Tian Q., Ni W., 2016, An efficient method for protein function annotation based on multilayer protein networks. *Human Genomics*, **10**(1): p. 33, DOI: 10.1186/s40246-016-0087-x.
 57. Anand S., Kuntal B.K., Mohapatra A., Bhatt V., Mande S.S., 2020, FunGeCo: a web-based tool for estimation of functional potential of bacterial genomes and microbiomes using gene context information. *Bioinformatics*, **36**(8): p. 2575-2577, DOI: 10.1093/bioinformatics/btz957.
 58. Martinez-Guerrero C.E., Ciria R., Abreu-Goodger C., Moreno-Hagelsieb G., Merino E., 2008, GeConT 2: gene context analysis for orthologous proteins, conserved domains and metabolic pathways. *Nucleic acids research*, **36**(Web Server issue): p. W176-80, DOI: 10.1093/nar/gkn330.
 59. Higdon R., Louie B., Kolker E., 2010, Modeling sequence and function similarity between proteins for protein functional annotation. *Proceedings of the International Symposium on High Performance Distributed Computing*, **2010**: p. 499-502, DOI: 10.1145/1851476.1851548.

60. Schwartz A., Hannum G., Dwiel Z., Smoot M., Grant A., Knight J., . . . Richardson T., 2018, Deep Semantic Protein Representation for Annotation, Discovery, and Engineering. DOI: 10.1101/365965.
61. Madera M., Gough J., 2002, A comparison of profile hidden Markov model procedures for remote homology detection. *Nucleic Acids Research*, **30**(19): p. 4321-4328, DOI: 10.1093/nar/gkf544.
62. Yoon B.J., 2009, Hidden Markov Models and their Applications in Biological Sequence Analysis. *Current genomics*, **10**(6): p. 402-15, DOI: 10.2174/138920209789177575.
63. Gough J., Chothia C., 2002, SUPERFAMILY: HMMs representing all proteins of known structure. SCOP sequence searches, alignments and genome assignments. *Nucleic Acids Research*, **30**(1): p. 268-272, DOI: 10.1093/nar/30.1.268.
64. Sillitoe I., Lewis T.E., Cuff A., Das S., Ashford P., Dawson N.L., . . . Orengo C.A., 2015, CATH: comprehensive structural and functional annotations for genome sequences. *Nucleic acids research*, **43**(Database issue): p. D376-81, DOI: 10.1093/nar/gku947.
65. Deng X., Cheng J., 2014, Enhancing HMM-based protein profile-profile alignment with structural features and evolutionary coupling information. *BMC Bioinformatics*, **15**(1): p. 252, DOI: 10.1186/1471-2105-15-252.
66. *RENEWABLES 2020 GLOBAL STATUS REPORT* (2020). REN21. REN21 Secretariat: Paris, ISBN 978-3-948393-00-7.
67. Voloshin A.R., Rodionova V.M., Zharmukhamedov K.S., Hou J.M.H., Shen J.-R., Allakhverdiev I.S., 2016, Components of natural photosynthetic apparatus in solar cells, in *Applied photosynthesis - New progress*, M.M. Najafpour, Editor. InTech d.o.o: Rijeka, Croatia. p. 161-188, DOI: DOI: 10.5772/62238.
68. Demirbas A., 2008, Biofuels sources, biofuel policy, biofuel economy and global biofuel projections. *Energy Conversion and Management*, **49**(8): p. 2106-2116, DOI: <https://doi.org/10.1016/j.enconman.2008.02.020>.
69. McKendry P., 2002, Energy production from biomass (Part 1): Overview of biomass. *Bioresource technology*, **83**(1): p. 37-46, DOI: 10.1016/s0960-8524(01)00118-3.
70. Sims R.E., Mabee W., Saddler J.N., Taylor M., 2010, An overview of second generation biofuel technologies. *Bioresource technology*, **101**(6): p. 1570-80, DOI: 10.1016/j.biortech.2009.11.046.
71. Nigam P., Singh A., 2011, Production of liquid biofuels from renewable resources. *Progress in Energy and Combustion Science*, **37**: p. 52-68, DOI: 10.1016/j.pecs.2010.01.003.
72. Dionisio K.L., Phillips K., Price P.S., Grulke C.M., Williams A., Biryol D., . . . Isaacs K.K., 2018, The Chemical and Products Database, a resource for exposure-relevant data on chemicals in consumer products. *Scientific data*, **5**: p. 180125, DOI: 10.1038/sdata.2018.125.
73. Scott L.N., Fiume M., Bergfeld W.F., Belsito D.V., Hill R.A., Klaassen C.D., . . . Heldreth B., 2024, Safety Assessment of Alkane Diols as Used in Cosmetics. *International journal of toxicology*, **43**(2_suppl): p. 70s-131s, DOI: 10.1177/10915818231224234.

74. Hugouvieux-Cotte-Pattat N., Condemine G., Shevchik V.E., 2014, Bacterial pectate lyases, structural and functional diversity. *Environmental microbiology reports*, **6**(5): p. 427-40, DOI: 10.1111/1758-2229.12166.
75. Yahia E.M., Carrillo-López A., Bello-Perez L.A., 2019, Chapter 9 - Carbohydrates, in *Postharvest Physiology and Biochemistry of Fruits and Vegetables*, E.M. Yahia, Editor. Woodhead Publishing. p. 175-205, DOI: <https://doi.org/10.1016/B978-0-12-813278-4.00009-9>.
76. Lemaire A., Duran Garzon C., Perrin A., Habrylo O., Trezel P., Bassard S., . . . Pelloux J., 2020, Three novel rhamnogalacturonan I- pectins degrading enzymes from *Aspergillus aculeatinus*: Biochemical characterization and application potential. *Carbohydrate Polymers*, **248**: p. 116752, DOI: <https://doi.org/10.1016/j.carbpol.2020.116752>.
77. Wu H.C., Bulgakov V.P., Jinn T.L., 2018, Pectin Methylsterases: Cell Wall Remodeling Proteins Are Required for Plant Response to Heat Stress. *Frontiers in plant science*, **9**: p. 1612, DOI: 10.3389/fpls.2018.01612.
78. Philippe F., Pelloux J., Rayon C., 2017, Plant pectin acetylsterase structure and function: new insights from bioinformatic analysis. *BMC Genomics*, **18**(1): p. 456, DOI: 10.1186/s12864-017-3833-0.
79. Hsiao Y.M., Zheng M.H., Hu R.M., Yang T.C., Tseng Y.H., 2008, Regulation of the *pehA* gene encoding the major polygalacturonase of *Xanthomonas campestris* by Clp and RpfF. *Microbiology*, **154**(Pt 3): p. 705-713, DOI: 10.1099/mic.0.2007/012930-0.
80. Sweeney M., Xu F., 2012, Biomass Converting Enzymes as Industrial Biocatalysts for Fuels and Chemicals: Recent Developments. *Catalysts*, **2**: p. 244-263, DOI: 10.3390/catal2020244.
81. Cosgrove D.J., 2015, Plant expansins: diversity and interactions with plant cell walls. *Current opinion in plant biology*, **25**: p. 162-72, DOI: 10.1016/j.pbi.2015.05.014.
82. Quiroz Castañeda R., Martínez-Anaya C., Cuervo-Soto L., Segovia L., Folch-Mallol J., 2011, Loosenin, a novel protein with cellulose-disrupting activity from *Bjerkandera adusta*. *Microbial cell factories*, **10**: p. 8, DOI: 10.1186/1475-2859-10-8.
83. Suzuki H., Vuong T.V., Gong Y., Chan K., Ho C.Y., Master E.R., Kondo A., 2014, Sequence diversity and gene expression analyses of expansin-related proteins in the white-rot basidiomycete, *Phanerochaete carnosa*. *Fungal genetics and biology*, **72**: p. 115-123, DOI: 10.1016/j.fgb.2014.05.008.
84. Santos C.A., Ferreira-Filho J.A., O'Donovan A., Gupta V.K., Tuohy M.G., Souza A.P., 2017, Production of a recombinant swollenin from *Trichoderma harzianum* in *Escherichia coli* and its potential synergistic role in biomass degradation. *Microbial Cell Factories*, **16**(1): p. 83, DOI: 10.1186/s12934-017-0697-6.
85. Falade A.O., Nwodo U.U., Iweriebor B.C., Green E., Mabinya L.V., Okoh A.I., 2017, Lignin peroxidase functionalities and prospective applications. *Microbiology Open*, **6**(1), DOI: 10.1002/mbo3.394.
86. Shradha, Shekher R., Sehgal S., Kamthania M., Kumar A., 2011, Laccase: microbial sources, production, purification, and potential biotechnological

- applications. *Enzyme research*, **2011**: p. 217861, DOI: 10.4061/2011/217861.
87. Chauhan P.S., Goradia B., Saxena A., 2017, Bacterial laccase: recent update on production, properties and industrial applications. *3 Biotech*, **7**(5): p. 323, DOI: 10.1007/s13205-017-0955-7.
 88. Uzuner S., Cekmecelioglu D., 2019, Chapter 3 - Enzymes in the Beverage Industry, in *Enzymes in Food Biotechnology*, M. Kuddus, Editor. Academic Press. p. 29-43, DOI: <https://doi.org/10.1016/B978-0-12-813280-7.00003-7>.
 89. Seiboth B., Ivanova C., Seidl-Seiboth V., 2010, *Trichoderma reesei*: A Fungal Enzyme Producer for Cellulosic Biofuels, in *Biofuel Production - Recent Developments and Prospects*, M.A.d.S. Bernardes, Editor. IntechOpen, DOI: DOI: 10.5772/16848.
 90. Meena M., Zehra A., Dubey M.K., Aamir M., Upadhyay R.S., 2018, Chapter 9 - *Penicillium* Enzymes for the Food Industries, in *New and Future Developments in Microbial Biotechnology and Bioengineering*, V.K. Gupta and S. Rodriguez-Couto, Editors. Elsevier: Amsterdam. p. 167-186, DOI: <https://doi.org/10.1016/B978-0-444-63501-3.00009-0>.
 91. Mamo G., 2020, Alkaline Active Hemicellulases, in *Alkaliphiles in Biotechnology*, G. Mamo and B. Mattiasson, Editors. Springer International Publishing: Cham. p. 245-291, DOI: 10.1007/10_2019_101.
 92. Krause D.O., Denman S.E., Mackie R.I., Morrison M., Rae A.L., Attwood G.T., McSweeney C.S., 2003, Opportunities to improve fiber degradation in the rumen: microbiology, ecology, and genomics. *FEMS Microbiology Reviews*, **27**(5): p. 663-693, DOI: 10.1016/s0168-6445(03)00072-x.
 93. Choudhury P., Sirohi S., Puniya A., Chaudhary P.P., 2012, Harnessing the Diversity of Rumen Microbes using Molecular Approaches, in *Livestock Green House Gases: emission and options for mitigation*, S.K. Sirohi, T.K. Walli, B. Singh and N. Singh, Editors. Satish Serial Publishing House: Delhi, India. p. 65–82.
 94. Castillo-González A.R., Burrola-Barraza M.E., Viveros J., Chavez-Martinez A., 2013, Rumen microorganisms and fermentation. *Archivos de medicina veterinaria*, **46**: p. 349-361, DOI: 10.4067/S0301-732X2014000300003.
 95. O'Hara E., Neves A.L.A., Song Y., Guan L.L., 2020, The Role of the Gut Microbiome in Cattle Production and Health: Driver or Passenger? *Annual review of animal biosciences*, **8**: p. 199-220, DOI: 10.1146/annurev-animal-021419-083952.
 96. Henderson G., Cox F., Ganesh S., Jonker A., Young W., Abecia L., . . . Global Rumen Census C., 2015, Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Scientific Reports*, **5**(1): p. 14567, DOI: 10.1038/srep14567.
 97. Tong F., Wang T., Gao N.L., Liu Z., Cui K., Duan Y., . . . Liu Q., 2022, The microbiome of the buffalo digestive tract. *Nature Communications*, **13**(1): p. 823, DOI: 10.1038/s41467-022-28402-9.
 98. Kim J.N., Méndez–García C., Geier R.R., Iakiviak M., Chang J., Cann I., Mackie R.I., 2017, Metabolic networks for nitrogen utilization in *Prevotella*

- ruminicola* 23. *Scientific Reports*, **7**(1): p. 7851, DOI: 10.1038/s41598-017-08463-3.
99. Accetto T., Avguštin G., 2019, The diverse and extensive plant polysaccharide degradative apparatuses of the rumen and hindgut *Prevotella* species: A factor in their ubiquity? *Systematic and applied microbiology*, **42**(2): p. 107-116, DOI: 10.1016/j.syapm.2018.10.001.
 100. Portincasa P., Bonfrate L., Vacca M., De Angelis M., Farella I., Lanza E., . . . Di Ciaula A., 2022, Gut Microbiota and Short Chain Fatty Acids: Implications in Glucose Homeostasis. *International journal of molecular sciences*, **23**(3), DOI: 10.3390/ijms23031105.
 101. Tett A., Pasolli E., Masetti G., Ercolini D., Segata N., 2021, *Prevotella* diversity, niches and interactions with the human host. *Nature reviews Microbiology*, **19**(9): p. 585-599, DOI: 10.1038/s41579-021-00559-y.
 102. Aguilar-Marin S.B., Betancur-Murillo C.L., Isaza G.A., Mesa H., Jovel J., 2020, Lower methane emissions were associated with higher abundance of ruminal *Prevotella* in a cohort of Colombian buffalos. *BMC Microbiology*, **20**(1): p. 364, DOI: 10.1186/s12866-020-02037-6.
 103. Zhang D., Wang Y., Zheng D., Guo P., Cheng W., Cui Z., 2016, New combination of xylanolytic bacteria isolated from the lignocellulose degradation microbial consortium XDC-2 with enhanced xylanase activity. *Bioresource technology*, **221**: p. 686-690, DOI: 10.1016/j.biortech.2016.09.087.
 104. Driss D., Bhiri F., Elleuch L., Bouly N., Stals I., Miled N., . . . Chaabouni S.E., 2011, Purification and properties of an extracellular acidophilic endo-1,4- β -xylanase, naturally deleted in the “thumb”, from *Penicillium occitanis* Pol6. *Process Biochemistry*, **46**(6): p. 1299-1306, DOI: <https://doi.org/10.1016/j.procbio.2011.02.022>.
 105. Hunt C.J., Tanksale A., Haritos V.S., 2016, Biochemical characterization of a halotolerant feruloyl esterase from *Actinomyces* spp.: refolding and activity following thermal deactivation. *Applied Microbiology and Biotechnology*, **100**(4): p. 1777-1787, DOI: 10.1007/s00253-015-7044-9.
 106. Dou T.-Y., Chen J., Liu C., 2020, Isolation and subunit compositions of the xylanosome complexes produced by *Cellulosimicrobium* species. *Enzyme and Microbial Technology*, **133**: p. 109445, DOI: <https://doi.org/10.1016/j.enzmictec.2019.109445>.
 107. Pohlschroder M., Leschine S.B., Canale-Parola E., 1994, Multicomplex cellulase-xylanase system of *Clostridium papyrosolvans* C7. *Journal of bacteriology*, **176**(1): p. 70-6, DOI: 10.1128/jb.176.1.70-76.1994.
 108. Verma D., Satyanarayana T., 2012, Molecular approaches for ameliorating microbial xylanases. *Bioresource Technology*, **117**: p. 360-367, DOI: <https://doi.org/10.1016/j.biortech.2012.04.034>.
 109. Collins T., Gerday C., Feller G., 2005, Xylanases, xylanase families and extremophilic xylanases. *FEMS Microbiology Reviews*, **29**(1): p. 3-23, DOI: 10.1016/j.femsre.2004.06.005.
 110. Henrissat B., Coutinho P.M., 2001, Classification of glycoside hydrolases and glycosyltransferases from hyperthermophiles. *Methods Enzymol*, **330**: p. 183-201, DOI: 10.1016/s0076-6879(01)30375-0.

111. Lombard V., Golaconda Ramulu H., Drula E., Coutinho P.M., Henrissat B., 2013, The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, **42**(D1): p. D490-D495, DOI: 10.1093/nar/gkt1178.
112. Ardèvol A., Rovira C., 2015, Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from ab Initio Quantum Mechanics/Molecular Mechanics Dynamic Simulations. *Journal of the American Chemical Society*, **137**(24): p. 7528-7547, DOI: 10.1021/jacs.5b01156.
113. Motta F.L., Andrade C.C.P., Santana M.H.A., 2013, A Review of Xylanase Production by the Fermentation of Xylan: Classification, Characterization and Applications, in *Sustainable Degradation of Lignocellulosic Biomass - Techniques, Applications and Commercialization*, K.A. Chandel and S.S. da Silva, Editors. IntechOpen, DOI: 10.5772/53544.
114. Zhang Y., An J., Yang G., Zhang X., Xie Y., Chen L., Feng Y., 2016, Structure features of GH10 xylanase from *Caldicellulosiruptor bescii*: implication for its thermophilic adaption and substrate binding preference. *Acta Biochimica et Biophysica Sinica*, **48**(10): p. 948-957, DOI: 10.1093/abbs/gmw086.
115. Faulds C.B., Mandalari G., Lo Curto R.B., Bisignano G., Christakopoulos P., Waldron K.W., 2006, Synergy between xylanases from glycoside hydrolase family 10 and family 11 and a feruloyl esterase in the release of phenolic acids from cereal arabinoxylan. *Applied microbiology and biotechnology*, **71**(5): p. 622-9, DOI: 10.1007/s00253-005-0184-6.
116. Dodd D., Cann I.K., 2009, Enzymatic deconstruction of xylan for biofuel production. *Global change biology Bioenergy*, **1**(1): p. 2-17, DOI: 10.1111/j.1757-1707.2009.01004.x.
117. Chakdar H., Kumar M., Pandiyan K., Singh A., Nanjappan K., Kashyap P.L., Srivastava A.K., 2016, Bacterial xylanases: biology to biotechnology. *3 Biotech*, **6**(2): p. 150, DOI: 10.1007/s13205-016-0457-z.
118. Chadha B.S., Kaur B., Basotra N., Tsang A., Pandey A., 2019, Thermostable xylanases from thermophilic fungi and bacteria: Current perspective. *Bioresource Technology*, **277**: p. 195-203, DOI: <https://doi.org/10.1016/j.biortech.2019.01.044>.
119. Singh S., Sidhu G.K., Kumar V., Dhanjal D.S., Datta S., Singh J., 2019, Fungal Xylanases: Sources, Types, and Biotechnological Applications, in *Recent Advancement in White Biotechnology Through Fungi: Volume 1: Diversity and Enzymes Perspectives*, A.N. Yadav, S. Mishra, S. Singh and A. Gupta, Editors. Springer International Publishing: Cham. p. 405-428, DOI: 10.1007/978-3-030-10480-1_12.
120. Chen N.J., Paull R.E., 2003, Endoxylanase expressed during papaya fruit ripening: purification, cloning and characterization. *Functional Plant Biology*, **30**(4): p. 433-441.
121. Béra-Maillet C., Devillard E., Cezette M., Jouany J.P., Forano E., 2005, Xylanases and carboxymethylcellulases of the rumen protozoa *Polyplastron multivesiculatum*, *Eudiplodinium maggii* and *Entodinium* sp. *FEMS microbiology letters*, **244**(1): p. 149-56, DOI: 10.1016/j.femsle.2005.01.035.

122. Suzuki M., Yoshida K., Ashida K., 1991, Purification and Characterization of Xylanase from the Mid-gut Gland of the Apple Snail (*Pomacea canaliculata*). *Agricultural and Biological Chemistry*, **55**(3): p. 693-700, DOI: 10.1080/00021369.1991.10870640.
123. Brennan Y., Callen W.N., Christoffersen L., Dupree P., Goubet F., Healey S., . . . Steer B.A., 2004, Unusual microbial xylanases from insect guts. *Applied and environmental microbiology*, **70**(6): p. 3609-17, DOI: 10.1128/aem.70.6.3609-3617.2004.
124. Jensen J.K., Busse-Wicher M., Poulsen C.P., Fangel J.U., Smith P.J., Yang J.Y., . . . Harholt J., 2018, Identification of an algal xylan synthase indicates that there is functional orthology between algal and plant cell wall biosynthesis. *The New phytologist*, **218**(3): p. 1049-1060, DOI: 10.1111/nph.15050.
125. Gupta V., Garg S., Capalash N., Gupta N., Sharma P., 2015, Production of thermo-alkali-stable laccase and xylanase by co-culturing of *Bacillus* sp. and *B. halodurans* for biobleaching of kraft pulp and deinking of waste paper. *Bioprocess and biosystems engineering*, **38**(5): p. 947-56, DOI: 10.1007/s00449-014-1340-0.
126. Thomas L., Ushasree M.V., Pandey A., 2014, An alkali-thermostable xylanase from *Bacillus pumilus* functionally expressed in *Kluyveromyces lactis* and evaluation of its deinking efficiency. *Bioresource technology*, **165**: p. 309-13, DOI: 10.1016/j.biortech.2014.03.037.
127. Banka A.L., Guralp S.A., Gulari E., 2014, Secretory expression and characterization of two hemicellulases, xylanase, and β -xylosidase, isolated from *Bacillus subtilis* M015. *Applied biochemistry and biotechnology*, **174**(8): p. 2702-10, DOI: 10.1007/s12010-014-1219-1.
128. Kumar V., Satyanarayana T., 2014, Production of endoxylanase with enhanced thermostability by a novel polyextremophilic *Bacillus halodurans* TSEV1 and its applicability in waste paper deinking. *Process Biochemistry*, **49**(3): p. 386-394, DOI: <https://doi.org/10.1016/j.procbio.2013.12.005>.
129. FERNANDES A.C., FONTES C.M.G.A., GILBERT H.J., HAZLEWOOD G.P., FERNANDES T.H., FERREIRA L.M.A., 1999, Homologous xylanases from *Clostridium thermocellum*: evidence for bi-functional activity, synergism between xylanase catalytic modules and the presence of xylan-binding domains in enzyme complexes. *Biochemical Journal*, **342**(1): p. 105-110, DOI: 10.1042/bj3420105.
130. Karlsson E.N., Hachem M.A., Ramchuran S., Costa H., Holst O., Svenningsen Å.F., Hreggvidsson G.O., 2004, The modular xylanase Xyn10A from *Rhodothermus marinus* is cell-attached, and its C-terminal domain has several putative homologues among cell-attached proteins within the phylum Bacteroidetes. *FEMS Microbiology Letters*, **241**(2): p. 233-242, DOI: 10.1016/j.femsle.2004.10.026.
131. Sukhumsirichart W., Deesukon W., Kawakami T., Matsumoto S., Seesom W., Sakamoto T., 2014, Expression and characterization of recombinant GH11 xylanase from thermotolerant *Streptomyces* sp. SWU10. *Applied biochemistry and biotechnology*, **172**(1): p. 436-46, DOI: 10.1007/s12010-013-0508-4.

132. Raj A., Kumar S., Singh S.K., 2013, A Highly Thermostable Xylanase from *Stenotrophomonas maltophilia*: Purification and Partial Characterization. *Enzyme Research*, **2013**: p. 429305, DOI: 10.1155/2013/429305.
133. Shi H., Zhang Y., Li X., Huang Y., Wang L., Wang Y., . . . Wang F., 2013, A novel highly thermostable xylanase stimulated by Ca²⁺ from *Thermotoga thermarum*: cloning, expression and characterization. *Biotechnology for Biofuels*, **6**(1): p. 26, DOI: 10.1186/1754-6834-6-26.
134. Akila G., Chandra T.S., 2003, A novel cold-tolerant *Clostridium* strain PXYL1 isolated from a psychrophilic cattle manure digester that secretes thermolabile xylanase and cellulase. *FEMS Microbiology Letters*, **219**(1): p. 63-7, DOI: 10.1016/s0378-1097(02)01196-5.
135. Dornez E., Verjans P., Arnaut F., Delcour J.A., Courtin C.M., 2011, Use of psychrophilic xylanases provides insight into the xylanase functionality in bread making. *Journal of agricultural and food chemistry*, **59**(17): p. 9553-62, DOI: 10.1021/jf201752g.
136. Collins T., Meuwis M.A., Stals I., Claeysens M., Feller G., Gerday C., 2002, A novel family 8 xylanase, functional and physicochemical characterization. *The Journal of biological chemistry*, **277**(38): p. 35133-9, DOI: 10.1074/jbc.M204517200.
137. Humphry D.R., George A., Black G.W., Cummings S.P., 2001, *Flavobacterium frigidarium* sp. nov., an aerobic, psychrophilic, xylanolytic and laminarinolytic bacterium from Antarctica. *International journal of systematic and evolutionary microbiology*, **51**(Pt 4): p. 1235-1243, DOI: 10.1099/00207713-51-4-1235.
138. Digvijay V., Satyanarayana T., 2012, Cloning, expression and applicability of thermo-alkali-stable xylanase of *Geobacillus thermoleovorans* in generating xylooligosaccharides from agro-residues. *Bioresource Technology*, **107**: p. 333-338, DOI: <https://doi.org/10.1016/j.biortech.2011.12.055>.
139. Luo L., Cai J., Wang C., Lin J., Du X., Zhou A., Mengxiong X., 2015, Purification and characterization of an alkaliphilic endo-xylanase from *Streptomyces althioticus* LMZM and utilization in pulp industry. *Journal of Chemical Technology and Biotechnology*, **91**, DOI: 10.1002/jctb.4690.
140. Twomey L.N., Pluske J.R., Rowe J.B., Choct M., Brown W., McConnell M.F., 2003, The effects of increasing levels of soluble non-starch polysaccharides and inclusion of feed enzymes in dog diets on faecal quality and digestibility. *Animal Feed Science and Technology*, **108**(1-4): p. 71-82.
141. Whiting I.M., Rose S.P., Mackenzie A.M., Amerah A.M., Pirgozliev V.R., 2019, Effect of wheat distillers dried grains with solubles and exogenous xylanase on laying hen performance and egg quality. *Poultry science*, **98**(9): p. 3756-3762, DOI: 10.3382/ps/pez063.
142. Van Dorn R., Shanahan D., Ciofalo V., 2018, Safety evaluation of xylanase 50316 enzyme preparation (also known as VR007), expressed in *Pseudomonas fluorescens*, intended for use in animal feed. *Regulatory toxicology and pharmacology*, **97**: p. 48-56, DOI: 10.1016/j.yrtph.2018.05.016.

143. Courtin C.M., Delcour J.A., 2002, Arabinoxylans and Endoxylanases in Wheat Flour Bread-making. *Journal of Cereal Science*, **35**(3): p. 225-243, DOI: <https://doi.org/10.1006/jcrs.2001.0433>.
144. Camacho N.A., Aguilar O G., 2003, Production, purification, and characterization of a low-molecular-mass xylanase from *Aspergillus* sp. and its application in baking. *Applied Biochemistry and Biotechnology*, **104**(3): p. 159-171, DOI: 10.1385/ABAB:104:3:159.
145. Polizeli M.L., Rizzatti A.C., Monti R., Terenzi H.F., Jorge J.A., Amorim D.S., 2005, Xylanases from fungi: properties and industrial applications. *Applied microbiology and biotechnology*, **67**(5): p. 577-91, DOI: 10.1007/s00253-005-1904-7.
146. Danalache F., Mata P., Alves V.D., Moldão-Martins M., 2018, Chapter 10 - Enzyme-Assisted Extraction of Fruit Juices, in *Fruit Juices*, G. Rajauria and B.K. Tiwari, Editors. Academic Press: San Diego. p. 183-200, DOI: <https://doi.org/10.1016/B978-0-12-802230-6.00010-2>.
147. Rosmine E., Sainjan N.C., Silvester R., Alikkunju A., Varghese S.A., 2017, Statistical optimisation of xylanase production by estuarine *Streptomyces* sp. and its application in clarification of fruit juice. *Journal of Genetic Engineering and Biotechnology*, **15**(2): p. 393-401, DOI: <https://doi.org/10.1016/j.jgeb.2017.06.001>.
148. Kumar L., Nagar S., Mittal A., Garg N., Gupta V.K., 2014, Immobilization of xylanase purified from *Bacillus pumilus* VLK-1 and its application in enrichment of orange and grape juices. *Journal of food science and technology*, **51**(9): p. 1737-49, DOI: 10.1007/s13197-014-1268-z.
149. Adiguzel G., Faiz O., Sisecioglu M., Sari B., Baltaci O., Akbulut S., . . . Adiguzel A., 2019, A novel endo- β -1,4-xylanase from *Pediococcus acidilactici* GC25; purification, characterization and application in clarification of fruit juices. *International journal of biological macromolecules*, **129**: p. 571-578, DOI: 10.1016/j.ijbiomac.2019.02.054.
150. Nagar S., Gupta V.K., 2021, Hyper Production and Eco-Friendly Bleaching of Kraft Pulp by Xylanase From *Bacillus pumilus* SV-205 Using Agro Waste Material. *Waste and Biomass Valorization*, **12**(7): p. 4019-4031, DOI: 10.1007/s12649-020-01258-0.
151. Azeri C., Tamer A., Oskay M., 2010, Thermoactive cellulase-free xylanase production from alkaliphilic *Bacillus* strains using various agro-residues and their potential in bioleaching of kraft pulp. *African Journal of Biotechnology*, **9**: p. 063-072.
152. Joshi C., Khare S.K., 2012, Induction of xylanase in thermophilic fungi *Scytalidium thermophilum* and *Sporotrichum thermophile*. *Brazilian Archives of Biology and Technology*, **55**: p. 21-27.
153. Maity C., Ghosh K., Halder S.K., Jana A., Adak A., Das Mohapatra P.K., . . . Mondal K.C., 2012, Xylanase isozymes from the newly isolated *Bacillus* sp. CKBx1D and optimization of its deinking potentiality. *Applied biochemistry and biotechnology*, **167**(5): p. 1208-19, DOI: 10.1007/s12010-012-9556-4.
154. Dhiman S.S., Garg G., Sharma J., Kalia V.C., Kang Y.C., Lee J.-K., 2014, Reduction in Acute Ecotoxicity of Paper Mill Effluent by Sequential

- Application of Xylanase and Laccase. *PLOS ONE*, **9**(7): p. e102581, DOI: 10.1371/journal.pone.0102581.
155. Dhiman S.S., Sharma J., Battan B., 2008, Pretreatment processing of fabrics by alkalothermophilic xylanase from *Bacillus stearothermophilus* SDX. *Enzyme and Microbial Technology*, **43**(3): p. 262-269, DOI: <https://doi.org/10.1016/j.enzmictec.2008.03.016>.
 156. Garg G., Dhiman S.S., Gautam R., Mahajan R., Patra A.K., Sharma J., 2013, Bioscouring of jute fabric by cellulase-free alkalo-thermostable xylanase from *Bacillus pumilus* ASH. *Journal of Molecular Catalysis B: Enzymatic*, **85-86**: p. 43-48, DOI: <https://doi.org/10.1016/j.molcatb.2012.08.002>.
 157. Battan B., Dhiman S.S., Ahlawat S., Mahajan R., Sharma J., 2012, Application of Thermostable Xylanase of *Bacillus pumilus* in Textile Processing. *Indian journal of microbiology*, **52**(2): p. 222-9, DOI: 10.1007/s12088-011-0118-1.
 158. Abd El Aty A.A., Saleh S.A.A., Eid B.M., Ibrahim N.A., Mostafa F.A., 2018, Thermodynamics characterization and potential textile applications of *Trichoderma longibrachiatum* KT693225 xylanase. *Biocatalysis and Agricultural Biotechnology*, **14**: p. 129-137, DOI: <https://doi.org/10.1016/j.bcab.2018.02.011>.
 159. Vázquez M.J., Alonso J.L., Domínguez H., Parajó J.C., 2000, Xylooligosaccharides: manufacture and applications. *Trends in Food Science & Technology*, **11**(11): p. 387-393, DOI: [https://doi.org/10.1016/S0924-2244\(01\)00031-0](https://doi.org/10.1016/S0924-2244(01)00031-0).
 160. Chen H.H., Chen Y.K., Chang H.C., Lin S.Y., 2012, Immunomodulatory Effects of Xylooligosaccharides. *Food Science and Technology Research*, **18**(2): p. 195-199, DOI: 10.3136/fstr.18.195.
 161. Gupta P.K., Agrawal P., Hedge P., Akhtar M.S., 2018, Xylooligosaccharides and Their Anticancer Potential: An Update, in *Anticancer Plants: Natural Products and Biotechnological Implements: Volume 2*, M.S. Akhtar and M.K. Swamy, Editors. Springer Singapore: Singapore. p. 255-271, DOI: 10.1007/978-981-10-8064-7_11.
 162. Kallel F., Driss D., Chaabouni S.E., Ghorbel R., 2015, Biological Activities of Xylooligosaccharides Generated from Garlic Straw Xylan by Purified Xylanase from *Bacillus mojavensis* UEB-FK. *Applied Biochemistry and Biotechnology*, **175**(2): p. 950-964, DOI: 10.1007/s12010-014-1308-1.
 163. Aachary A., Siddalingaiya P., 2010, Xylooligosaccharides (XOS) as an Emerging Prebiotic: Microbial Synthesis, Utilization, Structural Characterization, Bioactive Properties, and Applications. *Comprehensive Reviews in Food Science and Food Safety*, **10**: p. 2-16, DOI: 10.1111/j.1541-4337.2010.00135.x.
 164. Long L., Zhang Y., Ren H., Sun H., Sun F.F., Qin W., 2020, Recombinant expression of *Aspergillus niger* GH10 endo-xylanase in *Pichia pastoris* by constructing a double-plasmid co-expression system. *Journal of Chemical Technology and Biotechnology*, **95**(3): p. 535-543, DOI: <https://doi.org/10.1002/jctb.6250>.
 165. Tuyen D.T., Cuong N.T., le Thanh N.S., Thao N.T., Hoang L.T., Trang N.T.H., . . . Anh D.T.M., 2021, Cloning, Expression, and Characterization of

- Xylanase G2 from *Aspergillus oryzae* VTCC-F187 in *Aspergillus niger* VTCC-F017. *BioMed Research International*, **2021**: p. 8840038, DOI: 10.1155/2021/8840038.
166. Boonyapakron K., Pootanakit K., Chantasingh D., Kirtikara K., Eurwilaichitr L., 2005, Cloning and expression of xylanase 10 from *Cryptovalsa mangrovei* (BCC7197) in *Pichia pastoris*. *DNA sequence : the journal of DNA sequencing and mapping*, **16**(5): p. 372-8, DOI: 10.1080/10425170500186629.
 167. Miao T., Basit A., Liu J., Zheng F., Rahim K., Lou H., Jiang W., 2021, Improved Production of Xylanase in *Pichia pastoris* and Its Application in Xylose Production From Xylan. *Frontiers in bioengineering and biotechnology*, **9**: p. 690702, DOI: 10.3389/fbioe.2021.690702.
 168. Pauchet Y., Kirsch R., Giraud S., Vogel H., Heckel D.G., 2014, Identification and characterization of plant cell wall degrading enzymes from three glycoside hydrolase families in the cerambycid beetle *Apriona japonica*. *Insect biochemistry and molecular biology*, **49**: p. 1-13, DOI: 10.1016/j.ibmb.2014.03.004.
 169. Padilla-Hurtado B., Flórez-Ramos C., Aguilera-Gálvez C., Medina-Olaya J., Ramírez-Sanjuan A., Rubio-Gómez J., Acuña-Zornosa R., 2012, Cloning and expression of an endo-1,4- β -xylanase from the coffee berry borer, *Hypothenemus hampei*. *BMC Reseach Notes*, **5**: p. 23, DOI: 10.1186/1756-0500-5-23.
 170. Ding M., Teng Y., Yin Q., Zhao J., Zhao F., 2008, The N-terminal cellulose-binding domain of EGXA increases thermal stability of xylanase and changes its specific activities on different substrates. *Acta biochimica et biophysica Sinica* (Shanghai), **40**(11): p. 949-54, DOI: 10.1111/j.1745-7270.2008.00481.x.
 171. Caspers M.P., Lok F., Sinjorgo K.M., van Zeijl M.J., Nielsen K.A., Cameron-Mills V., 2001, Synthesis, processing and export of cytoplasmic endo-beta-1,4-xylanase from barley aleurone during germination. *The Plant journal: for cell and molecular biology*, **26**(2): p. 191-204, DOI: 10.1046/j.0960-7412.2001.01019.x.
 172. Devillard E., Bera-Maillet C., Flint H.J., Scott K.P., Newbold C.J., Wallace R.J., . . . Forano E., 2003, Characterization of XYN10B, a modular xylanase from the ruminal protozoan *Polyplastron multivesiculatum*, with a family 22 carbohydrate-binding module that binds to cellulose. *The Biochemical journal*, **373**(Pt 2): p. 495-503, DOI: 10.1042/bj20021784.
 173. Thi Kim Anh D., Thi Huong T., Thanh Thuy N., Tat Thanh D., Nguyen Thanh V., Lisbeth Cecilia O., 2016, Expression of Synthetic Gene Encoding Acidic, Thermostable Endo-1,4- β -Xylanase for Application in Feed Industry *Vietnam Journal of Agricultural Sciences*, **14**(3): p. 400-408.
 174. Liên Hà T., Thị Ngọc Ngà N., 2017, Tách dòng và biểu hiện gene mã hóa enzyme xylanase từ nấm mốc *Aspergillus niger* C1 vào *E. coli* BL21. *Vietnam Journal of Science and Technology*, **51**(1): p. 73, DOI: 10.15625/0866-708X/51/1/9563.
 175. Sambrook J., Russell D.W., 2001, Molecular cloning: a laboratory manual. 3rd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

176. Hu T., Chitnis N., Monos D., Dinh A., 2021, Next-generation sequencing technologies: An overview. *Human immunology*, **82**(11): p. 801-811, DOI: 10.1016/j.humimm.2021.02.012.
177. Sambrook J., Russell D.W., 2006, Directional cloning into plasmid vectors. *CSH Protocols*, **2006**(1), DOI: 10.1101/pdb.prot3919.
178. Froger A., Hall J.E., 2007, Transformation of plasmid DNA into *E. coli* using the heat shock method. *Journal of visualized experiments*, (6): p. 253, DOI: 10.3791/253.
179. Sambrook J., Russell D.W., 2001, Preparation of Plasmid DNA by Alkaline Lysis with SDS: Miniprep, in *Molecular cloning: a laboratory manual* Cold Spring Harbor Laboratory: Cold Spring Harbor, NY. p. 4.
180. Green M.R., Sambrook J., 2019, Agarose Gel Electrophoresis. *Cold Spring Harbor Protocols*, **2019**(1), DOI: 10.1101/pdb.prot100404.
181. Sambrook J., Russell D.W., 2001, Expression of Cloned Genes in *E. coli* Using IPTG-inducible Promoters, in *Molecular cloning: a laboratory manual* Cold Spring Harbor Laboratory: Cold Spring Harbor, NY. p. 310-311.
182. Laemmli U.K., 1970, Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, **227**(5259): p. 680-685, DOI: 10.1038/227680a0.
183. Bradford M.M., 1976, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry*, **72**: p. 248-54, DOI: 10.1006/abio.1976.9999.
184. Miller G.L., 1959, Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Analytical Chemistry*, **31**(3): p. 426-428, DOI: 10.1021/ac60147a030.
185. Dashtban M., Maki M., Leung K.T., Mao C., Qin W., 2010, Cellulase activities in biomass conversion: measurement methods and comparison. *Critical reviews in biotechnology*, **30**(4): p. 302-9, DOI: 10.3109/07388551.2010.490938.
186. Marli C., 2012, Cellulase Determination: Modifications to Make the Filter Paper Assay Easy, Fast, Practical and Efficient. *Journal of analytical and bioanalytical techniques*, **1**, DOI: 10.4172/scientificreports.125.
187. Lineweaver H., Burk D., 1934, The Determination of Enzyme Dissociation Constants. *Journal of the American Chemical Society*, **56**(3): p. 658-666, DOI: 10.1021/ja01318a036.
188. Lin H., Chen W., Ding H., 2013, AcalPred: A Sequence-Based Tool for Discriminating between Acidic and Alkaline Enzymes. *PLOS ONE*, **8**(10): p. e75726, DOI: 10.1371/journal.pone.0075726.
189. Ebrahimi M., Lakizadeh A., Agha-Golzadeh P., Ebrahimie E., Ebrahimi M., 2011, Prediction of thermostability from amino acid attributes by combination of clustering with attribute weighting: a new vista in engineering enzymes. *PLoS One*, **6**(8): p. e23146, DOI: 10.1371/journal.pone.0023146.

190. Huson D.H., Auch A.F., Qi J., Schuster S.C., 2007, MEGAN analysis of metagenomic data. *Genome Research*, **17**(3): p. 377-86, DOI: 10.1101/gr.5969107.
191. Pope P.B., Mackenzie A.K., Gregor I., Smith W., Sundset M.A., McHardy A.C., . . . Eijsink V.G., 2012, Metagenomics of the Svalbard reindeer rumen microbiome reveals abundance of polysaccharide utilization loci. *PLoS One*, **7**(6): p. e38571, DOI: 10.1371/journal.pone.0038571.
192. Henderson G., Cox F., Kittelmann S., Miri V.H., Zethof M., Noel S.J., . . . Janssen P.H., 2013, Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities. *PLoS One*, **8**(9): p. e74787, DOI: 10.1371/journal.pone.0074787.
193. An D., Dong X., Dong Z., 2005, Prokaryote diversity in the rumen of yak (*Bos grunniens*) and Jinnan cattle (*Bos taurus*) estimated by 16S rDNA homology analyses. *Anaerobe*, **11**(4): p. 207-15, DOI: 10.1016/j.anaerobe.2005.02.001.
194. Han X., Yang Y., Yan H., Wang X., Qu L., Chen Y., 2015, Rumen bacterial diversity of 80 to 110-day-old goats using 16S rRNA sequencing. *PLoS One*, **10**(2): p. e0117811, DOI: 10.1371/journal.pone.0117811.
195. Zhang K., Li B., Guo M., Liu G., Yang Y., Wang X., . . . Zhang E., 2019, Maturation of the Goat Rumen Microbiota Involves Three Stages of Microbial Colonization. *Animals (Basel)*, **9**(12), DOI: 10.3390/ani9121028.
196. Lv X., Chai J., Diao Q., Huang W., Zhuang Y., Zhang N., 2019, The Signature Microbiota Drive Rumen Function Shifts in Goat Kids Introduced to Solid Diet Regimes. *Microorganisms*, **7**, DOI: 10.3390/microorganisms7110516.
197. Bekele A.Z., Koike S., Kobayashi Y., 2010, Genetic diversity and diet specificity of ruminal *Prevotella* revealed by 16S rRNA gene-based analysis. *FEMS Microbiology Letters*, **305**(1): p. 49-57, DOI: 10.1111/j.1574-6968.2010.01911.x.
198. Stevenson D.M., Weimer P.J., 2007, Dominance of *Prevotella* and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. *Applied microbiology and biotechnology*, **75**(1): p. 165-74, DOI: 10.1007/s00253-006-0802-y.
199. Lim S., Seo J., Choi H., Yoon D., Nam J., Kim H., . . . Chang J., 2013, Metagenome Analysis of Protein Domain Collocation within Cellulase Genes of Goat Rumen Microbes. *Asian-Australasian journal of animal sciences*, **26**(8): p. 1144-51, DOI: 10.5713/ajas.2013.13219.
200. Dodd D., Moon Y.H., Swaminathan K., Mackie R.I., Cann I.K., 2010, Transcriptomic analyses of xylan degradation by *Prevotella bryantii* and insights into energy acquisition by xylanolytic bacteroidetes. *The Journal of biological chemistry*, **285**(39): p. 30261-73, DOI: 10.1074/jbc.M110.141788.
201. Xu J., Zhao X., Yao Q., Zong W., Dai S., Deng Z., . . . Li H., 2021, Cloning, characterization of a novel acetyl xylan esterase, and its potential application on wheat straw utilization. *All Life*, **14**(1): p. 622-635, DOI: 10.1080/26895293.2021.1947393.

202. Underlin E.N., Frommhagen M., Dilokpimol A., van Erven G., de Vries R.P., Kabel M.A., 2020, Feruloyl Esterases for Biorefineries: Subfamily Classified Specificity for Natural Substrates. *Frontiers in bioengineering and biotechnology*, **8**: p. 332, DOI: 10.3389/fbioe.2020.00332.
203. P O.S., Martin J.C., Lawley T.D., Browne H.P., Harris H.M.B., Bernalier-Donadille A., . . . H J.F., 2016, Polysaccharide utilization loci and nutritional specialization in a dominant group of butyrate-producing human colonic Firmicutes. *Microbial Genomics*, **2**(2): p. e000043, DOI: 10.1099/mgen.0.000043.
204. Mackenzie A.K., Naas A.E., Kracun S.K., Schüchel J., Fangel J.U., Agger J.W., . . . Pope P.B., 2015, A polysaccharide utilization locus from an uncultured bacteroidetes phylotype suggests ecological adaptation and substrate versatility. *Applied and environmental microbiology*, **81**(1): p. 187-95, DOI: 10.1128/aem.02858-14.
205. Ilmberger N., Güllert S., Dannenberg J., Rabausch U., Torres J., Wemheuer B., . . . Streit W.R., 2014, A Comparative Metagenome Survey of the Fecal Microbiota of a Breast- and a Plant-Fed Asian Elephant Reveals an Unexpectedly High Diversity of Glycoside Hydrolase Family Enzymes. *PLOS ONE*, **9**(9): p. e106707, DOI: 10.1371/journal.pone.0106707.
206. Jing L., Eric J.C.G., Lena A., Éva A., Aida I., Till R.L., . . . Till S., 2021, A versatile genetic toolbox for *Prevotella copri* enables studying polysaccharide utilization systems. *bioRxiv*: p. 2021.03.19.436125, DOI: 10.1101/2021.03.19.436125.
207. Rodríguez F., 2003, *Control of lactate accumulation in ruminants using Prevotella bryantii*. Iowa State University: Iowa, USA, DOI: <https://doi.org/10.31274/rtd-180813-57>.
208. Manikandan K., Bhardwaj A., Gupta N., Lokanath N.K., Ghosh A., Reddy V.S., Ramakumar S., 2006, Crystal structures of native and xylosaccharide-bound alkali thermostable xylanase from an alkalophilic *Bacillus* sp. NG-27: Structural insights into alkalophilicity and implications for adaptation to polyextreme conditions. *Protein Science*, **15**(8): p. 1951-1960, DOI: <https://doi.org/10.1110/ps.062220206>.
209. Holck J., Fredslund F., Møller M.S., Brask J., Krogh K., Lange L., . . . Wilkens C., 2019, A carbohydrate-binding family 48 module enables feruloyl esterase action on polymeric arabinoxylan. *The Journal of biological chemistry*, **294**(46): p. 17339-17353, DOI: 10.1074/jbc.RA119.009523.
210. Zadorozhny A.V., Ushakov V.S., Rozanov A.S., Bogacheva N.V., Shlyakhtun V.N., Voskoboev M.E., . . . Peltek S.E., 2022, Heterologous Expression of Xylanase xAor from *Aspergillus oryzae* in *Komagataella phaffii* T07. *International journal of molecular sciences*, **23**(15), DOI: 10.3390/ijms23158741.
211. Hu H., Dai S., Wen A., Bai X., 2019, Efficient Expression of Xylanase by Codon Optimization and Its Effects on the Growth Performance and Carcass Characteristics of Broiler. *Animals (Basel)*, **9**(2), DOI: 10.3390/ani9020065.
212. Li Y., Li C., Muhammad Aqeel S., Wang Y., Zhang Q., Ma J., . . . Liu S., 2023, Enhanced expression of xylanase in *Aspergillus niger* enabling a two-step enzymatic pathway for extracting β -glucan from oat bran. *Bioresource*

- Technology*, **377**: p. 128962, DOI: <https://doi.org/10.1016/j.biortech.2023.128962>.
213. Zhang J., Qin Y., Wang Q., Liu S., Zhou J., He B., . . . Wu J., 2023, Gene cloning, expression, and characterization of two endo-xylanases from *Bacillus velezensis* and *Streptomyces rochei*, and their application in xylooligosaccharide production. *Frontiers in microbiology*, **14**: p. 1292726, DOI: 10.3389/fmicb.2023.1292726.
 214. Al-Darkazali H., Meevootisom V., Isarangkul D., Wiyakrutta S., 2017, Gene Expression and Molecular Characterization of a Xylanase from Chicken Cecum Metagenome. *International journal of microbiology*, **2017**: p. 4018398, DOI: 10.1155/2017/4018398.
 215. Kiribayeva A., Mukanov B., Silayev D., Akishev Z., Ramankulov Y., Khassenov B., 2022, Cloning, expression, and characterization of a recombinant xylanase from *Bacillus sonorensis* T6. *PLoS One*, **17**(3): p. e0265647, DOI: 10.1371/journal.pone.0265647.
 216. Park D.J., Lee Y.S., Chang J., Fang S.J., Choi Y.L., 2013, An β -1,4-xylanase with exo-enzyme activity produced by *Paenibacillus xylanilyticus* KJ-03 and its cloning and characterization. *Journal of microbiology and biotechnology*, **23**(3): p. 397-404, DOI: 10.4014/jmb.1212.12017.
 217. Taberner C., Sánchez-Torres J., Pérez P., Santamaría R.I., 1995, Cloning and DNA sequencing of xyaA, a gene encoding an endo-beta-1,4-xylanase from an alkalophilic *Bacillus* strain (N137). *Applied and environmental microbiology*, **61**(6): p. 2420-2424, DOI: 10.1128/aem.61.6.2420-2424.1995.
 218. Han S.O., Yukawa H., Inui M., Doi R.H., 2004, Isolation and expression of the xynB gene and its product, XynB, a consistent component of the *Clostridium cellulovorans* cellulosome. *Journal of bacteriology*, **186**(24): p. 8347-8355, DOI: 10.1128/jb.186.24.8347-8355.2004.
 219. Jun H.S., Ha J.K., Malburg L.M., Jr., Verrinder G.A., Forsberg C.W., 2003, Characteristics of a cluster of xylanase genes in *Fibrobacter succinogenes* S85. *Canadian journal of microbiology*, **49**(3): p. 171-180, DOI: 10.1139/w03-024.
 220. Jeong Y.S., Na H.B., Kim S.K., Kim Y.H., Kwon E.J., Kim J., . . . Kim H., 2012, Characterization of xyn10J, a novel family 10 xylanase from a compost metagenomic library. *Applied biochemistry and biotechnology*, **166**(5): p. 1328-1339, DOI: 10.1007/s12010-011-9520-8.
 221. Hayashi H., Abe T., Sakamoto M., Ohara H., Ikemura T., Sakka K., Benno Y., 2005, Direct cloning of genes encoding novel xylanases from the human gut. *Canadian journal of microbiology*, **51**(3): p. 251-259, DOI: 10.1139/w04-136.
 222. Alvarez T.M., Goldbeck R., dos Santos C.R., Paixão D.A., Gonçalves T.A., Franco Cairo J.P., . . . Squina F.M., 2013, Development and biotechnological application of a novel endoxylanase family GH10 identified from sugarcane soil metagenome. *PLoS One*, **8**(7): p. e70014, DOI: 10.1371/journal.pone.0070014.
 223. Mannarelli B.M., Evans S., Lee D., 1990, Cloning, sequencing, and expression of a xylanase gene from the anaerobic ruminal bacterium

- Butyrivibrio fibrisolvens*. *Journal of bacteriology*, **172**(8): p. 4247-4254, DOI: 10.1128/jb.172.8.4247-4254.1990.
224. Meng X., Shao Z., Hong Y., Lin L., Li C., Liu Z., 2009, A novel pH-stable, bifunctional xylanase isolated from a deep-sea microorganism, *Demequina* sp. JK4. *Journal of microbiology and biotechnology*, **19**(10): p. 1077-1084.
225. Qiu Z., Shi P., Luo H., Bai Y., Yuan T., Yang P., . . . Yao B., 2010, A xylanase with broad pH and temperature adaptability from *Streptomyces megasporus* DSM 41476, and its potential application in brewing industry. *Enzyme and microbial technology*, **46**(6): p. 506-512, DOI: 10.1016/j.enzmictec.2010.02.003.
226. Kim D.Y., Shin D.H., Jung S., Lee J.S., Cho H.Y., Bae K.S., . . . Park H.Y., 2014, Biocatalytic properties and substrate-binding ability of a modular GH10 β -1,4-xylanase from an insect-symbiotic bacterium, *Streptomyces mexicanus* HY-14. *Journal of microbiology (Seoul, Korea)*, **52**(10): p. 863-870, DOI: 10.1007/s12275-014-4390-8.
227. Amaya-Delgado L., Mejía-Castillo T., Santiago-Hernández A., Vega-Estrada J., Amelia F.G., Xoconostle-Cázares B., . . . Hidalgo-Lara M.E., 2010, Cloning and expression of a novel, moderately thermostable xylanase-encoding gene (Cflxyn11A) from *Cellulomonas flavigena*. *Bioresource technology*, **101**(14): p. 5539-45, DOI: 10.1016/j.biortech.2010.02.057.
228. Kim D.Y., Han M.K., Park D.S., Lee J.S., Oh H.W., Shin D.H., . . . Park H.Y., 2009, Novel GH10 xylanase, with a fibronectin type 3 domain, from *Cellulosimicrobium* sp. strain HY-13, a bacterium in the gut of *Eisenia fetida*. *Applied and environmental microbiology*, **75**(22): p. 7275-7279, DOI: 10.1128/aem.01075-09.
229. Li N., Meng K., Wang Y., Shi P., Luo H., Bai Y., . . . Yao B., 2008, Cloning, expression, and characterization of a new xylanase with broad temperature adaptability from *Streptomyces* sp. S9. *Applied microbiology and biotechnology*, **80**(2): p. 231-240, DOI: 10.1007/s00253-008-1533-z.
230. Bai W., Xue Y., Zhou C., Ma Y., 2012, Cloning, expression and characterization of a novel salt-tolerant xylanase from *Bacillus* sp. SN5. *Biotechnology letters*, **34**(11): p. 2093-2099, DOI: 10.1007/s10529-012-1011-7.
231. Canakçı S., Cevher Z., Inan K., Tokgoz M., Bahar F., Kacagan M., . . . Belduz A.O., 2012, Cloning, purification and characterization of an alkali-stable endoxylanase from thermophilic *Geobacillus* sp. 71. *World journal of microbiology and biotechnology*, **28**(5): p. 1981-1988, DOI: 10.1007/s11274-011-1000-3.
232. Wu S., Liu B., Zhang X., 2006, Characterization of a recombinant thermostable xylanase from deep-sea thermophilic *Geobacillus* sp. MT-1 in East Pacific. *Applied microbiology and biotechnology*, **72**(6): p. 1210-1216, DOI: 10.1007/s00253-006-0416-4.
233. Chanjuan L., Hong Y., Shao Z., Lin L., Huang X., Liu P., . . . Liu Z., 2009, Novel alkali-stable, cellulase-free xylanase from deep-sea *Kocuria* sp. Mn22. *Journal of microbiology and biotechnology*, **19**(9): p. 873-880, DOI: 10.4014/jmb.0812.689.

234. Chang L., Ding M., Bao L., Chen Y., Zhou J., Lu H., 2011, Characterization of a bifunctional xylanase/endoglucanase from yak rumen microorganisms. *Applied microbiology and biotechnology*, **90**(6): p. 1933-42, DOI: 10.1007/s00253-011-3182-x.
235. Ghatge S.S., Telke A.A., Kang S.H., Arulalapperumal V., Lee K.W., Govindwar S.P., . . . Kim S.W., 2014, Characterization of modular bifunctional processive endoglucanase Cel5 from *Hahella chejuensis* KCTC 2396. *Applied microbiology and biotechnology*, **98**(10): p. 4421-35, DOI: 10.1007/s00253-013-5446-0.
236. Hwang I.T., Lim H.K., Song H.Y., Cho S.J., Chang J.S., Park N.J., 2010, Cloning and characterization of a xylanase, KRICT PX1 from the strain *Paenibacillus* sp. HPL-001. *Biotechnology advances*, **28**(5): p. 594-601, DOI: 10.1016/j.biotechadv.2010.05.007.
237. Mirande C., Mosoni P., Béra-Maillet C., Bernalier-Donadille A., Forano E., 2010, Characterization of Xyn10A, a highly active xylanase from the human gut bacterium *Bacteroides xylanisolvens* XB1A. *Applied microbiology and biotechnology*, **87**(6): p. 2097-2105, DOI: 10.1007/s00253-010-2694-0.
238. Kui H., Luo H., Shi P., Bai Y., Yuan T., Wang Y., . . . Yao B., 2010, Gene cloning, expression, and characterization of a thermostable xylanase from *Nesterenkonia xinjiangensis* CCTCC AA001025. *Applied biochemistry and biotechnology*, **162**(4): p. 953-965, DOI: 10.1007/s12010-009-8815-5.
239. Valenzuela S.V., Diaz P., Pastor F.I., 2014, Xyn11E from *Paenibacillus barcinonensis* BP-23: a LppX-chaperone-dependent xylanase with potential for upgrading paper pulps. *Applied microbiology and biotechnology*, **98**(13): p. 5949-5957, DOI: 10.1007/s00253-014-5565-2.
240. Zhou J., Huang H., Meng K., Shi P., Wang Y., Luo H., . . . Yao B., 2010, Cloning of a new xylanase gene from *Streptomyces* sp. TN119 using a modified thermal asymmetric interlaced-PCR specific for GC-rich genes and biochemical characterization. *Applied biochemistry and biotechnology*, **160**(5): p. 1277-1292, DOI: 10.1007/s12010-009-8642-8.
241. Valenzuela S.V., Díaz P., Javier Pastor F.I., 2010, Recombinant expression of an alkali stable GH10 xylanase from *Paenibacillus barcinonensis*. *Journal of agricultural and food chemistry*, **58**(8): p. 4814-4818, DOI: 10.1021/jf9045792.