

**BỘ GIÁO DỤC VÀ
ĐÀO TẠO**

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

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ



Nguyễn Thị Bình

**NGHIÊN CỨU ĐA DẠNG KHU HỆ VI
KHUẨN QUANH NẤM MỤC TRẮNG THỦY
PHÂN LIGNOCELLULOSE VÀ KHAI THÁC
GEN MÃ HÓA CELLULASE BẰNG KỸ THUẬT
METAGENOMICS**

**TÓM TẮT LUẬN ÁN TIẾN SĨ
CÔNG NGHỆ SINH HỌC**

Mã số: 9.42.02.01

Hà Nội – 2023

Công trình được hoàn thành tại: Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

Người hướng dẫn khoa học:

1. Người hướng dẫn khoa học 1: GS.TS. Trương Nam Hải - Viện Công nghệ Sinh học
2. Người hướng dẫn khoa học 2: TS. Lê Thị Thu Hồng – Viện Công nghệ Sinh học

Phản biện 1: GS. TS. Phạm Xuân Hội – Viện Di truyền Nông nghiệp

Phản biện 2: PGS. TS. Đỗ Hữu Nghị - Viện Hóa học các hợp chất thiên nhiên

Phản biện 3: PGS. TS. Nguyễn Quang Huy – Đại học Khoa học Tự nhiên - ĐHQGHN

Luận án được bảo vệ trước Hội đồng đánh giá luận án tiến sĩ cấp Học viện họp tại Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam vào hồi giờ, ngày.... tháng..... năm.....

Có thể tìm hiểu luận án tại:

1. Thư viện Học viện Khoa học và Công nghệ
2. Thư viện Quốc gia Việt Nam

MỞ ĐẦU

1. Tính cấp thiết của luận án

Lignocellulose là nguồn sinh khối tự nhiên, rẻ tiền có trữ lượng lớn và được tái tạo liên tục được xác định là nguồn sinh khối tự nhiên để thúc đẩy nền công nghiệp sinh học. Tuy nhiên, việc chuyển hóa sinh khối này bằng enzyme sinh học gặp nhiều khó khăn, đặc biệt là quá trình đường phân. Vì vậy, giá thành các sản phẩm sinh học còn khá cao so với các sản phẩm sản xuất bằng nhiên liệu hóa thạch. Để hạ giá thành cũng như thúc đẩy phát triển kinh tế sinh học thì việc tìm ra enzyme có thể tham gia hiệu quả vào quá trình thủy phân cellulose có vai trò quan trọng vì cellulose chứa hàm lượng cao trong các sinh khối lignocellulose. Các hệ sinh thái mà sự phân giải lignocellulose diễn ra mạnh mẽ sẽ là nguồn tiềm năng để tìm kiếm và khai thác các enzyme có giá trị trong công nghiệp. Nấm mục trắng và đất xung quanh nấm mục trắng cũng là hệ sinh thái mà sự phân hủy lignocellulose diễn ra mạnh mẽ. Trong đó, nấm mục trắng là sinh vật nhân chuẩn có khả năng phân giải tốt các loại cơ chất khác nhau. Tuy nhiên, sự thủy phân này thường được kết hợp cùng với các enzyme của vi khuẩn sống trong cùng hệ sinh thái. Vi khuẩn tồn tại trong những điều kiện này phải có những đặc điểm đặc biệt về thành phần loại và các enzyme chuyển hóa các chất. Mặc dù có nhiều nghiên cứu về nấm mục trắng và vi khuẩn đất xung quanh khu nấm mục trắng nhưng cơ chế đằng sau sự tương tác này vẫn chưa được sáng tỏ, các đặc tính chức năng của chúng vẫn cần được xác định lại bằng thực nghiệm. Ở Việt Nam, cho đến nay vẫn chưa có nghiên cứu về đa dạng các loài vi khuẩn ở rừng Quốc Gia Cúc Phương nói chung và đa dạng loài các vi khuẩn đất xung quanh khu nấm mục trắng nói riêng cũng như khai thác các enzyme phân giải cellulose của vi khuẩn trong hệ

sinh thái này. Để sàng lọc enzyme mong muốn từ khu hệ vi sinh vật không thông qua nuôi cấy, kỹ thuật metagenomics có nhiều ưu thế. Nhằm đánh giá đa dạng thành phần loài vi sinh vật trong đất xung quanh khu nấm mục trắng cũng như sàng lọc enzyme mã hóa cellulase có nhiều đặc tính mới không thông qua nuôi cấy, chúng tôi đã sử dụng kỹ thuật metagenomic để tiến hành đề tài: ***“Nghiên cứu đa dạng khu hệ vi khuẩn quanh nấm mục trắng thủy phân lignocellulose và khai thác gen mã hóa cellulase bằng kỹ thuật Metagenomics”***.

2. Mục tiêu nghiên cứu của luận án

Đánh giá được đa dạng của khu hệ vi sinh vật đất mùn xung quanh khu nấm mục trắng phân hủy lignocellulose và xác định được đa dạng enzyme tham gia vào quá trình phân giải lignocellulose, khai thác và lựa chọn được enzyme phân giải cellulase có tiềm năng ứng dụng trong thực tiễn sản xuất từ khu hệ vi khuẩn đất xung quanh khu nấm mục trắng ở rừng Quốc gia Cúc Phương bằng kỹ thuật Metagenomics.

3. Các nội dung nghiên cứu chính của luận án

- Phân tích và đánh giá mức độ đa dạng loài của khu hệ vi khuẩn trong đất xung quanh khu nấm mục trắng thủy phân lignocellulose bằng kỹ thuật Metagenomics;
- Phân tích và đánh giá mức độ đa dạng của các enzyme tham gia phân giải lignocellulose của khu hệ vi khuẩn đất xung quanh khu nấm mục trắng thủy phân lignocellulose bằng kỹ thuật Metagenomics;
- Tìm kiếm và lựa chọn các trình tự gen mới mã hóa cellulase có tiềm năng ứng dụng bằng các công cụ tin sinh học;
- Nghiên cứu biểu hiện tái tổ hợp của một gen đã lựa chọn, tinh chế và đánh giá tính chất của enzyme β -glucosidase.

CHƯƠNG 1. TỔNG QUAN TÀI LIỆU

1.1. Khái quát chung về lignocellulose

Lignocellulose được cấu tạo từ ba thành phần chính là cellulose, hemicellulose và lignin. Cellulose và hemicelluloses liên kết chặt chẽ với lignin. Cellulose là thành phần chính trong cấu trúc của thành tế bào thực vật thường chiếm tỷ lệ 38 – 50%. Cellulose được cấu tạo từ đơn phân D-glucose và chịu trách nhiệm về độ bền cơ học của thành tế bào thực vật. Tiếp đến là hemicellulose chiếm tỷ lệ 17 – 32% có cấu trúc không đồng nhất, có sự phân nhánh cao thường được cấu tạo từ các đường đơn pentose và hexose. Các hemicellulose tạo ra các liên kết chéo giữa các cellulose. Lignin chiếm 15 – 30% bao gồm các polyphenol thơm, được sinh tổng hợp và tạo thành cấu trúc bao bọc xung quanh hai thành phần cellulose và hemicelluloses, cung cấp thêm độ bền cơ học cho thành tế bào, chống lại côn trùng hoặc điều kiện ẩm ướt.

1.2. Cellulase

Enzym cellulase là một trong 3 nhóm enzym quan trọng tham gia vào phân hủy sinh khối lignocellulose. Cellulase được xếp vào nhóm enzyme glycoside hydrolase (GH) (EC 3.2.1.-), có khả năng cắt mỗi liên kết β -1,4-glycoside trong phân tử cellulose tạo thành các sản phẩm cello-oligosaccharide, cellobiose và glucose. Trong tự nhiên, quá trình thủy phân hoàn toàn cellulose được thực hiện nhờ sự hoạt động phối hợp của ba loại cellulase chính: (1) endoglucanase (EC 3.2.1.4) thủy phân các liên kết β -1,4-glucoside nội phân tử của chuỗi cellulose một cách ngẫu nhiên để tạo ra các đầu chuỗi mới, (2) exoglucanase (EC 3.2.1.74, EC 3.2.1.91) phân cắt chuỗi cellulose ở đầu khử và không khử để giải phóng các phân tử cellobiose hoặc glucose hòa tan và (3) β -glucosidase (EC 3.2.1.21) thủy phân các

cellobiose thành glucose. Trong những năm gần đây, trên thế giới đã có nhiều nghiên cứu khai thác gen mã hóa cellulase sử dụng kỹ thuật metagenomic. Ở Việt Nam, từ năm 2012, nhóm nghiên cứu của Trương Nam Hải đã bắt đầu sử dụng kỹ thuật metagenomic trong khai thác gen mã hóa enzyme thủy phân lignocellulose từ khu hệ vi sinh vật ruột mối Việt Nam, vi sinh vật trong dạ cỏ dê bản địa ở Việt Nam, vi sinh vật suối nước nóng Bình Châu...

1.3. Nấm mục trắng và khu hệ vi sinh vật xung quanh khu nấm mục trắng thủy phân lignocellulose

Nấm mục trắng là nhóm sinh vật duy nhất có khả năng phân giải tất cả các thành phần của lignocellulose. Để phân giải hiệu quả lignocellulose thì không chỉ có sự tham gia của nấm mà còn có cả cộng đồng vi sinh vật trong đất xung quanh khu nấm mục trắng. Sự tương tác giữa nấm và vi khuẩn sống trong cùng khu vực có thể là quan hệ hỗ trợ và/hoặc cạnh tranh. Trong quá trình nấm phân giải gỗ, điều kiện môi trường trở lên rất chọn lọc đối với vi khuẩn. Vi khuẩn tồn tại trong những điều kiện này phải có những đặc tính đặc biệt và mới mẻ. Vì vậy, vi sinh vật trong đất xung quanh khu nấm mục trắng là nguồn quan trọng để tìm kiếm các gen phân giải lignocellulase.

1.4. Metagenomic và một số công cụ tin sinh, cơ sở dữ liệu được sử dụng trong khai thác DNA đa hệ gen

Metagenomic là phương pháp nghiên cứu đa hệ gen (metagenome) của tất cả các vi sinh vật thu nhận trực tiếp từ mẫu môi trường tự nhiên mà không thông qua nuôi cấy. Metagenomics tiếp cận đa hệ gen theo hai phương pháp chính là: (1) Phân lập gen dựa trên việc thiết lập thư viện DNA đa hệ gen và (2) khai thác trình tự DNA và phân lập gen dựa trên dữ liệu giải trình tự trực tiếp DNA đa hệ gen. Trong nghiên cứu này, chúng tôi khai thác gen mã hóa

cellulase bằng kỹ thuật metagenomics dựa trên dữ liệu giải trình tự trực tiếp DNA đa hệ gen. Có nhiều công cụ tin sinh được sử dụng để khai thác gen từ dữ liệu DNA đa hệ gen như: BLAST, Periscope, Phyre2, AcalPred... và một số CSDL được sử dụng là: NCBI, KEGG, Pfam...

CHƯƠNG 2. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Vật liệu, hóa chất

- Các mẫu đất mùn xung quanh nấm mục trắng trong Vườn Quốc gia Cúc Phương được lấy vào mùa mưa (tháng 5 – 6 trong năm)

- Địa điểm nghiên cứu: Phòng Kỹ thuật di truyền, Viện Công nghệ sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

- Các chủng vi sinh vật của hãng Invitrogen (Mỹ) và phòng thí nghiệm Hóa sinh, Đại học Tổng hợp Saarland (CHLB Đức) được sử dụng làm chủng tách dòng và chủng biểu hiện gen. Vector pET22b(+) của hãng Novagene (Mỹ) được sử dụng làm vector biểu hiện gen gh3s2 trong *E. coli* Rosetta 1.

- Cặp môi khuếch đại gen 16S rDNA của vi khuẩn:

27F: 5'-GAGTTTGATCCTGGCTCAG-3'

1527R: 5'-AGAAAGGAGGTGATCCAGCC-3'

- Hóa chất, thiết bị sử dụng trong các thí nghiệm đều có nguồn gốc, xuất xứ rõ ràng.

2.2. Phương pháp nghiên cứu

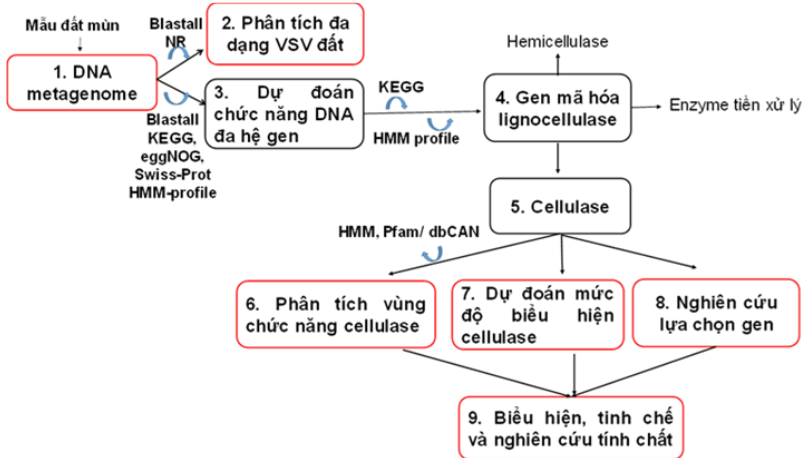
Có 3 nhóm phương pháp nghiên cứu được sử dụng trong luận án gồm:

2.2.1. Các phương pháp vi sinh và sinh học phân tử

2.2.2. Các phương pháp hóa sinh protein

2.2.3. Các phương pháp tin sinh học

Quy trình nghiên cứu của luận án được thực hiện như hình 2.2



Hình 2.1. Sơ đồ quy trình nghiên cứu trong luận án

CHƯƠNG 3: KẾT QUẢ VÀ THẢO LUẬN

3.1. Nghiên cứu đa dạng khu hệ vi khuẩn đất quanh khu nấm mục trắng

3.1.1. Tách chiết, tinh sạch DNA đa hệ gen vi sinh vật đất

Các mẫu đất mùn được thu thập, trộn lại với nhau để đảm bảo tính đa dạng của vi sinh vật dùng cho tách chiết DNA đa hệ gen. Kết quả tách chiết thành công DNA đa hệ gen với nồng độ DNA thu được từ 112 cho đến 145 ng/ μ l, A260/A280 đều nằm trong khoảng 1,8 đến 2, khi PCR bằng môi 16S cho kết quả tốt. Như vậy, DNA đa hệ gen đã được tách chiết và tinh sạch, đủ điều kiện đi giải trình tự.

3.1.2. Kết quả giải trình tự DNA đa hệ gen vi sinh vật đất

Từ khoảng 100 μ g DNA đa hệ gen được đưa đi giải trình tự, thu được tổng kích thước khoảng 51,82 Gb (Bảng 3.2).

Bảng 3.2. Kết quả giải trình tự DNA đa hệ gen bằng hệ thống giải trình tự thế hệ mới HiSeq Illuminar

| Loại phân tích | Kết quả thu được | Đơn vị |
|----------------|------------------|--------|
|----------------|------------------|--------|

| | | | |
|---------------|--------------------------|----------------|----------|
| Read | Số lượng | 345.471.086 | read |
| | Tổng kích thước các read | 51.820.662.900 | cặp base |
| Contig | Số lượng | 2.611.883 | contig |
| | Kích thước trung bình | 898 | cặp base |
| | Kích thước N50 | 1117 | cặp base |
| | Kích thước lớn nhất | 611.845 | cặp base |
| Gen | Số lượng | 4.104.872 | gen |
| | Kích thước trung bình | 505 | cặp base |
| | Kích thước N50 | 615 | cặp base |
| | Kích thước lớn nhất | 20.541 | cặp base |

3.1.3. Phân tích đa dạng vi sinh vật đất

Bảng 3.3. Kết quả phân tích đa dạng từ dữ liệu DNA đa hệ gen vi sinh vật đất bằng phần mềm MEGAN (version 6) dựa trên CSDL NR

| | Số gen | Tỉ lệ (%) | Ngành | Lớp | Bộ | Họ | Chi | Loài |
|----------------------|-----------|-----------|-------|-----|-----|-----|------|------|
| Vi khuẩn | 3.884.879 | 99,69 | 111 | 83 | 170 | 406 | 1971 | 738 |
| Vi khuẩn cổ | 293 | 0,01 | 9 | 12 | 18 | 23 | 50 | 8 |
| SV nhân chuẩn | 1144 | 0,03 | 7 | 26 | 46 | 79 | 113 | 86 |
| Virus | 10565 | 0,27 | 0 | 0 | 2 | 14 | 101 | 84 |
| Tổng | 3.896.881 | 100 | 131 | 118 | 237 | 523 | 2240 | 916 |

Từ dữ liệu 51,82 Gb DNA đa hệ gen của vi sinh vật đất quanh khu nấm mục trắng ở vườn Quốc gia Cúc Phương, có 4.104.872 gen mã hóa protein đã được xác định. Trong đó, có 3.923.046 gen (khoảng 95,57%) được chú giải trong cơ sở dữ liệu NR. Bằng phần mềm MEGAN các gen này đã được xác định phân loại, có 3.896.881 gen được xếp vào các giới vi khuẩn, sinh vật nhân chuẩn (Eukaryote), vi khuẩn cổ (Archaea) và virus. Trong đó, số gen được xếp vào giới vi khuẩn là chiếm ưu thế tuyệt đối với 3.884.879 gen (chiếm khoảng 99,69% tổng số gen), các giới còn lại là vi khuẩn cổ với 293 gen

(0,01%), sinh vật nhân thực với 1144 gen (0,03%) và virus là 10.565 gen (0,27%). Như vậy, vi khuẩn có số lượng gen nhiều nhất và các gen của vi khuẩn được xếp vào 111 ngành, 83 lớp, 170 bộ, 406 họ, 1971 chi và chỉ có 738 loài được xác định. Trong giới vi khuẩn này, có 93,26% của tổng số gen được xác định ở bậc phân loại ngành. Trong 111 ngành vi khuẩn được xác định, có 5 ngành phổ biến chiếm 92,59% tổng số còn lại là các ngành khác. Trong số đó, Proteobacteria là ngành phổ biến nhất với 3.106.400 gen chiếm khoảng 75,68%. Các ngành tiếp theo là Bacteroidetes chiếm 13,11%, Actinobacteria 1,6%, Firmicutes 1,4%, Acidobacteria 0,8%. Như vậy, Proteobacteria là ngành chiếm ưu thế lớn có số lượng gen cao gấp 5,77 lần ngành Bacteroidetes phổ biến thứ hai.

3.2. Nghiên cứu khai thác gen mã hóa lignocellulase

3.2.1. Dự đoán chức năng DNA đa hệ gen của vi khuẩn đất

Các gen đã được chú giải chức năng dựa trên các CSDL Swiss-Prot, KEGG, eggNOG, Nr và HMM-profile. Có 3.925.740 gen (95,64%) được chú giải chức năng dựa trên ít nhất một trong các CSDL trên. Dựa trên dữ liệu KEGG, có 2.809.791 gen (tương ứng khoảng 68,45% tổng số gen) được xác định chức năng mã hóa protein tham gia vào chuyển hóa các chất trong tế bào và cơ thể. Các protein này tham gia vào 5 nhóm chuyển hóa bao gồm: các quá trình trong tế bào, xử lý thông tin môi trường, xử lý thông tin di truyền, bệnh ở người, sự trao đổi chất. Trong quá trình trao đổi các chất khác nhau, trao đổi carbohydrate có 297.103 gen mã hóa protein tham gia (chiếm khoảng 13,56% trong tổng số các gen tham gia trao đổi chất)

3.2.2. Khai thác gen dựa trên kết quả chú giải bởi KEGG

Từ 297.103 gen được xác định chức năng là tham gia vào quá trình chuyển hóa carbohydrate trên CSDL KEGG, có 22.226 gen được

ước đoán là các gen mã hóa các enzyme có tham gia vào quá trình phân giải sinh khối lignocellulose gồm: (1) 907 gen mã hóa các enzyme tham gia vào tiền xử lý sinh khối chia thành 4 nhóm là pectinesterase 89,96%, feruloyl esterase 8,27%, laccase 1,10% và expansin 0,67%; (2) 8301 gen mã hóa cellulase (gồm 5 nhóm sắp xếp theo thứ tự giảm dần là β -glucosidas, endoglucanase, 6-phospho-beta-glucosidase, cellobiohydrolase, cellobiose phosphorylase); (3) 13.018 gen mã hóa hemicellulase được xếp vào 20 nhóm, trong đó nhóm xyloglucan-active β -D-galactosidase là nhóm phổ biến nhất 25,26% (3288 ORF).

3.2.3. Khai thác gen dựa trên mô hình đại diện HMM

Sử dụng mô hình HMM mà bản chất là dựa trên sự tương đồng về motif, có 13 enzyme tham gia thủy phân lignocellulose đã được khai thác hiệu quả hơn so với việc khai thác gen dựa trên sự tương đồng về trình tự trong KEGG. Đó là CBM (1-84), arabinanase (GH43), galactanase, glucuronyl esterase, HPOXRE catalase, hydrogen peroxide oxidoreductase, LPMO, laccase, axetylxylanesterase, beta-glucuronidase, cellobiohydrolase, lichenase, beta-xylosidase. Trong đó hydrogen peroxide oxidoreductase (thuộc nhóm hemicellulase) và LPMO (enzyme tiền xử lý) là chưa được tìm thấy dựa trên dữ liệu KEGG, CAZy.

3.2.4. Nghiên cứu đa dạng vi sinh vật mang gen mã hóa lignocellulase

Trong số 22.226 gen mã hóa enzyme tham gia phân hủy lignocellulose có 22.092 gen (chiếm 99,39%) là thuộc về vi khuẩn được xếp vào 28 ngành, trội nhất là ngành Proteobacteria (11.288 gen, chiếm 50,79%), tiếp theo là Bacteroidetes (8.164 gen, 36,73%). Tỷ lệ

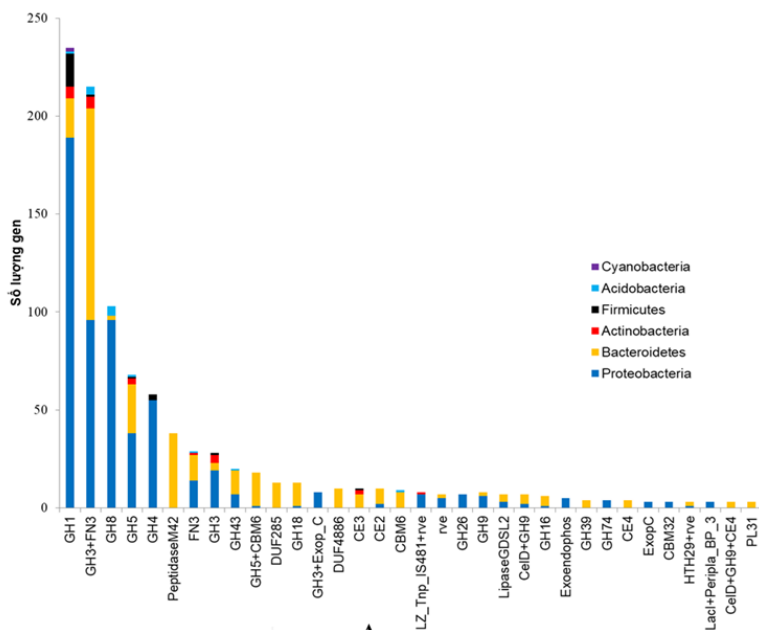
Bacteroidetes/Proteobacteria (0,72; 1) trong gen mã hóa enzyme tham gia phân hủy lignocellulose cao hơn nhiều so với tỷ lệ này trong tổng số cấu trúc vi khuẩn của mùn xung quanh khu nấm mục trắng (0,17: 1). Điều này cho thấy Bacteroidetes đóng vai trò quan trọng trong quá trình thủy phân lignocellulose. Ở cấp độ bộ, phân tích cũng cho thấy Enterobacterales là bộ nổi bật nhất chiếm 20,06%, tiếp theo là Flavobacters 15,14%, Sphingobacteria 11,62%. Phân tích sâu hơn với nhóm enzyme tiền xử lý cho thấy rằng ngành Bacteroidetes là ngành phong phú nhất (427 gen, chiếm 47,08%), cao hơn một chút so với Proteobacteria (45,31%). Trong khi đó, đối với nhóm hemicellulase, Proteobacteria (44,20%) cao hơn so với Bacteroidetes (43,52%). Đối với cellulase, tỷ lệ giữa Proteobacteria và Bacteroidetes là 2,4 lần tương ứng với Proteobacteria 61,72% và Bacteroidetes 24,96%. Như vậy, tỷ lệ Proteobacteria/Bacteroidetes trong DNA đa hệ gen vi sinh vật xung quanh nấm mục trắng là 5,77, trong khi đối với cellulase thì tỷ lệ Proteobacteria/Bacteroidetes là 2,4. Do đó, Bacteroidetes dường như đóng một vai trò quan trọng hơn trong quá trình thủy phân lignocellulose.

3.3. Nghiên cứu khai thác và lựa chọn gen tiềm năng mã hóa cellulase

3.3.1. Phân tích các vùng chức năng của cellulase

Dựa trên CSDL KEGG, có 8301 ORF mã hóa cho enzyme cellulase được chú giải trong đó có 1279 ORF hoàn chỉnh, 1058 ORF hoàn chỉnh có 81 loại domain bao gồm: 367 ORF mã hóa endoglucanase, 6 ORF mã hóa cellobiohydrolase, 475 ORF mã hóa β -glucosidase, 210 ORF mã hóa 6-phospho β -glucosidase. Trong đó, domain phổ biến nhất thuộc họ GH (chiếm trên 80% ORF hoàn chỉnh

có domain). Đại diện là GH1 có 245 ORF trong đó 189 ORF (tương ứng 77,14%) thuộc ngành Proteobacteria, 20 ORF (8,16%) thuộc ngành Bacteroidetes còn lại là thuộc các ngành khác. Tiếp theo là domain GH3+FN3 (220 ORF) trong đó 96 ORF (43,67%) thuộc ngành Proteobacteria, 108 ORF (49,09%) thuộc ngành Bacteroidetes. Sau đó là các họ GH khác như họ GH8 (105 ORF), GH5 (72 ORF), GH4 (58 ORF) trong đó tỉ lệ các ORF thuộc ngành Proteobacteria lần lượt là 91,43%; 52,78%, 94,83%.



Hình 3.5. Các ngành vi khuẩn mang ORF hoàn chỉnh có domain mã hóa cellulase

Phân tích theo từng nhóm enzyme cho thấy, thuộc nhóm endoglucase có 367 ORF với 47 loại domain. Trong đó, domain GH8 là phổ biến nhất với 105 ORF, 96 ORF trong số này (91,43%) thuộc ngành Proteobacteria, chỉ có 2 ORF (1,90%) thuộc ngành

Bacteroidetes. Loại domain phổ biến thứ hai trong nhóm enzyme này là GH5 với 72 ORF. Các ORF chứa domain GH5 hầu hết thuộc hai ngành Proteobacteria (52.78%) và Bacteroidetes (34.72%). Trong nhóm enzyme exoglucanase chỉ có 6 ORF với 6 loại domain khác nhau. Trong số các domain thuộc ORF mã hóa exoglucanase có 3 domain Alginate_lyase, Amidase 3, GH128+Laminin G3 thuộc ngành Bacteroidetes, 2 domain CBM2 và Znribbon8 thuộc ngành Acidobacteria, domain CBP_BcsO thuộc ngành Proteobacteria. Nhóm enzyme β -glucosidase là nhóm enzyme có số lượng ORF nhiều nhất 475 ORF (44,90%) với 27 loại domain. Trong nhóm enzyme này, domain có số lượng nhiều nhất là domain GH3+FN3 với 220 ORF, trong đó 96 ORF (43,64%) thuộc ngành Proteobacteria và 108 ORF (49,09%) thuộc ngành Bacteroidetes còn lại 16 ORF thuộc các ngành khác. Tiếp đó là domain GH1 (93 ORF) trong đó 60 ORF (64,52%) thuộc ngành Proteobacteria, 20 ORF (21,51%) thuộc ngành Bacteroidetes và 13 ORF còn lại thuộc một số ngành khác. Trong nhóm 6-phospho- β -glucosidase, domain phổ biến nhất là GH1 với 152 ORF trong đó có 129 ORF (84,87%) thuộc ngành Proteobacteria, 15 ORF thuộc ngành Firmicutes còn lại chưa được phân loại ngành. Domain còn lại trong nhóm enzyme này là GH4 với 58 ORF, trong đó 55 ORF (94,83%) thuộc ngành Proteobacteria và 3 ORF thuộc ngành Firmicutes.

3.3.2. Dự đoán mức độ biểu hiện của các gen cellulase

Kết quả xác định mức độ biểu hiện ở *E. coli* cho thấy trong 1058 gen hoàn chỉnh có domain mã hóa cellulase, các gen thuộc nhóm endoglucanase và β -glucosidase được xác định có khả năng biểu hiện cao hơn nhóm exoglucanase. Trong nhóm endoglucase, các gen chứa domain GH8 được dự đoán có mức biểu hiện cao nhất với các mã

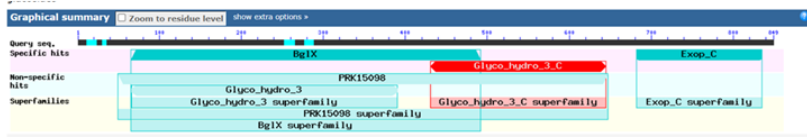
GL0183420, GL1155166, GL0051672, GL0127466, GL0176868 đều có khả năng biểu hiện trên 3000 mg/l. Các gen này thuộc ngành Proteobacteria và Acidobacteria. Tiếp theo là các gen chứa domain GH5 thuộc ngành Proteobacteria và Bacteroidetes có mức biểu hiện trên 2000 mg/l. Ngoài ra một số gen chứa domain PeptidaseM42, GH5-CBM6, DUF285 thuộc ngành Bacteroidetes cũng biểu hiện tốt trong hệ biểu hiện *E. coli*. Trong nhóm β -glucosidase, các gen có domain GH3 thuộc ngành Proteobacteria đều mức độ biểu hiện tốt: nhiều đại diện cấu trúc domain GH3+FN3 có mức độ biểu hiện cao trên 4000 mg/l, gen có domain GH3 có mức biểu hiện trên 1800 mg/l, mã gen GL0050362 với cấu trúc domain GH3+Exop_C có mức độ biểu hiện cao nhất 4626 mg/l. Bên cạnh đó, các gen β -glucosidase chứa domain GH4, GH43, GH1 thuộc ngành Proteobacteria, gen β -glucosidase chứa domain GH16 thuộc ngành Bacteroidetes cũng biểu hiện tốt. Nhóm 6-phospho β -glucosidase, gen có domain GH1 có mức độ biểu hiện cao nhất 4714 mg/l, một số gen chứa domain GH4 có mức độ biểu hiện trên 1000 mg/l.

3.3.3. Nghiên cứu lựa chọn gen mã hóa cellulase

3.3.3.1. Dự đoán vùng bảo thủ của gen bằng BLASTp

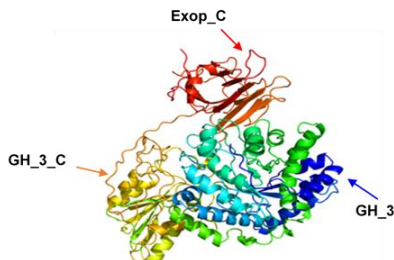
Kết quả xác định cấu trúc protein do gen GL0050362 mã hóa bằng BLASTp cho thấy protein này có ba vùng đặc hiệu (specific hit) gồm: (1) là vùng BglX (thuộc siêu họ BglX) tương ứng với hai vùng không đặc hiệu (non-specific hit) PRK15098 và GH3-N (theo số liệu được liên kết với pfam00933) mã hóa β -glucosidase và các glycosidase tham gia vào quá trình trao đổi carbohydrate (theo cơ sở dữ liệu COG1472); (2) là vùng GH3-C (thuộc siêu họ GH3-C) tham gia vào quá trình xúc tác và có thể liên kết beta-glucan (theo số liệu được liên kết với pfam01915); (3) Exop_C (thuộc siêu họ Exop_C)

giống vùng liên kết với Galactose, đây là vùng đầu C được tìm thấy trong ExoP (exo-1,3/1,4-beta-glucanase) từ *Pseudoalteromonas*. Vùng này chứa một nếp gấp β thường gặp trong glycosyl hydrolase (GH7, 11, 12 và 16) và trong một số vùng/cấu trúc liên kết carbohydrate. Vùng này được cho rằng không chỉ có vai trò định hướng liên kết với cơ chất mà còn giúp làm ổn định cấu trúc cần thiết cho hoạt động của ExoP.



Hình 3.6. Kết quả dự đoán chức năng gen GL0050362 bằng BLASTp.

3.3.3.2. Dự đoán cấu trúc không gian và ước đoán vị trí gắn cơ chất của protein gh3s2



Hình 3.7. Mô hình cấu trúc không gian của gen ứng viên sử dụng Pyre2 dựa trên khuôn c3f93D

Trong mô hình cấu trúc không gian ba chiều, protein được xác định dựa trên khuôn enzyme β -glucosidase từ *Pseudoalteromonas* sp. bb1 (c3f93D) có độ bao phủ 93% và độ tin cậy 100%. Cấu trúc không gian bậc 3 của gen này có 47% tương đồng với β -glucosidase của khuôn c3f93D với độ tin cậy 100%, có ba vùng đặc hiệu GH-3, GH-3-C và Exop_C, ngoài ra gen này có vùng bảo tồn cao [HIS]249 giống

nhau giữa protein ứng viên và khuôn c3f93D, mặt khác enzyme ứng viên còn có vị trí xúc tác [GLY]848 liên quan đến hoạt tính β -glucosidase theo ước đoán của Phyer2.

3.3.3.3. Dự đoán một số tính chất của enzyme ứng viên

Phần mềm AcalPred xác định khả năng chịu axit/kiềm của GH3S2 lần lượt là 0,507957 và 0,49204, enzyme này có pH trung tính, hơi ngả axit. Kết quả khảo sát khả năng chịu nhiệt của protein GH3S2 có Tm là 0,6606, như vậy nhiệt độ tối ưu cho hoạt động của enzyme được dự đoán từ 55°C-65°C.

3.4. Biểu hiện, tinh chế và nghiên cứu tính chất GH3S2

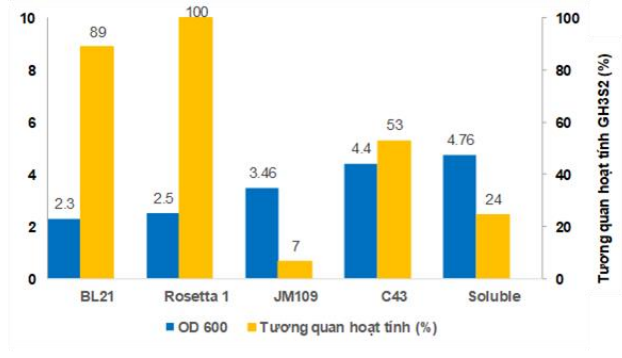
3.4.1. Nghiên cứu biểu hiện gen *gh3s2*

3.4.1.1. Thiết kế vector tái tổ hợp pET22b(+) mang *gh3s2*

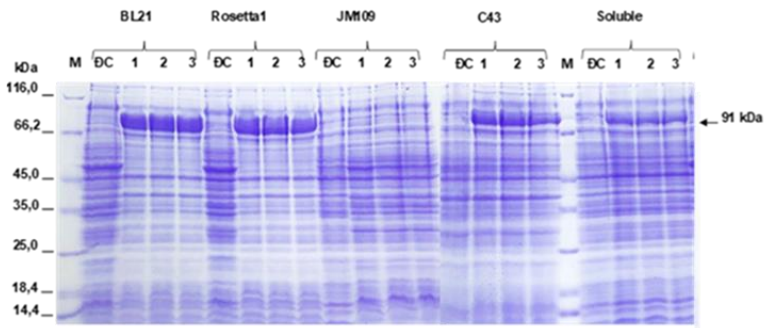
Gen *gh3s2* sau khi được lựa chọn và tối ưu mã sẽ được đặt tổng hợp và gắn vào vector biểu hiện pET22b(+). DNA plasmid tái tổ hợp pET22b(+)*gh3s2* sẽ được biến nạp vào tế bào tách dòng *E.coli* DH10b bằng phương pháp sốc nhiệt và thực hiện tách chiết plasmid.

3.4.1.2. Nghiên cứu lựa chọn chủng biểu hiện GH3S2

Chúng tôi nghiên cứu biểu hiện GH3S2 trong 5 chủng *E. coli* gồm: BL21, Rosetta 1, JM109, C43, Soluble. Kết quả cho thấy GH3S2 biểu hiện tốt trong các chủng BL21, Rosetta 1 (Hình 3.9 B). Trong 2 chủng này, ở chủng Rosetta có mật độ tế bào khi thu mẫu cao hơn, hoạt tính tương đối của enzyme tổng số khi biểu hiện ở chủng Rosetta cao hơn (Hình 3.9 A). Kết hợp mức độ biểu hiện, mật độ tế bào và hoạt tính enzyme thu được, chúng tôi lựa chọn Rosetta 1 làm chủng biểu hiện GH3S2. Sau khi biểu hiện trong chủng *E. coli* Rosetta 1, hoạt tính β -glucosidase của enzyme GH3S2 được kiểm tra với cơ chất esculin. Kết quả cho thấy kích thước vòng màu nâu đối với các mẫu khác nhau có mức độ hoạt động β -glucosidase khác nhau.

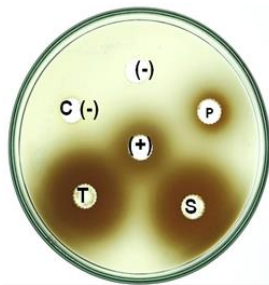


A



B

Hình 3.9. (A). Mật độ tế bào và hoạt tính của enzyme thu được khi biểu hiện trong các chủng biểu hiện; (B). Điện di đồ GH3S2 tổng số, ĐC: protein tổng số của đối chứng vector không mang gen; 1, 2, 3: protein tổng số các dòng khác nhau 1, 2, 3 mang gen *gh3s2* cảm ứng IPTG; M: protein chuẩn (Fermentas)



Hình 3.10. Kiểm tra hoạt tính của GH3S2 trên đĩa thạch LB sử dụng cơ chất esculin.

Điều này chứng tỏ rằng protein GH3S2 đã được biểu hiện thành công ở dạng hòa tan và cho thấy có hoạt tính rõ ràng (Hình 3.10).

3.4.1.3. Nghiên cứu ảnh hưởng của nhiệt độ nuôi cấy đến sự biểu hiện của GH3S2 trong E.coli Rosetta 1

Chủng *E. coli* Rosetta 1 mang DNA tái tổ hợp được nuôi cấy ở các điều kiện nhiệt độ: 18°C, 20°C, 25°C, 30°C và 37°C. Để nâng cao hiệu quả biểu hiện, giảm chi phí năng lượng chúng tôi lựa chọn 25°C là nhiệt độ biểu hiện GH3S2 trong các nghiên cứu tiếp theo.

3.4.1.4. Nghiên cứu ảnh hưởng của thành phần môi trường nuôi cấy đến sự biểu hiện của GH3S2 trong E.coli Rosetta 1

Năm môi trường nuôi cấy đã được kiểm tra bao gồm: LB, TB, TB cải biến, SB và PE. Môi trường được chọn để biểu hiện protein tái tổ hợp GH3S2 là môi trường TB cải biến.

3.4.1.5. Nghiên cứu ảnh hưởng của nồng độ chất cảm ứng IPTG đến sự biểu hiện của GH3S2 trong E. coli Rosetta 1

Các nồng độ IPTG khác nhau từ 0,05 đến 2 mM được kiểm tra để biểu hiện GH3S2 tái tổ hợp. Kết quả cho thấy nồng độ IPTG thích hợp được lựa chọn là 0,3 mM để cho các nghiên cứu tiếp theo.

3.4.1.6. Nghiên cứu ảnh hưởng của mật độ tế bào lúc cảm ứng đến sự biểu hiện của GH3S2 trong E. coli Rosetta 1

Để biết mật độ tế bào tối ưu cho sự cảm ứng tổng hợp protein GH3S2, chúng tôi khảo sát OD khi cảm ứng từ 0,4 đến 4. Kết quả cho thấy mật độ tế bào khi cảm ứng là 1 thì protein GH3S2 có mức độ biểu hiện tốt nhất và hoạt tính cao.

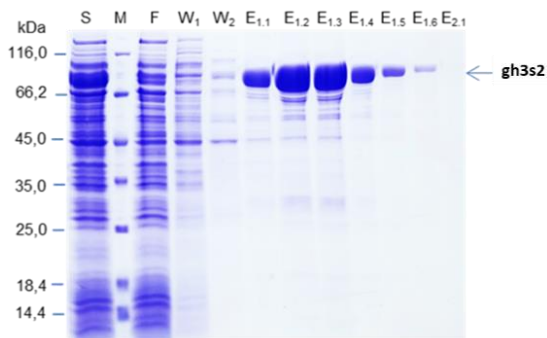
3.4.1.7. Nghiên cứu xác định thời gian thu mẫu GH3S2 tối ưu sau cảm ứng

Để xác định khoảng thời gian sau cảm ứng phù hợp cho sự biểu hiện của protein GH3S2 tái tổ hợp, mẫu lên men được thu trong

khoảng thời gian từ 1 - 6 giờ và 22 giờ sau khi cảm ứng. Kết quả cho thấy lượng protein GH3S2 tái tổ hợp có thể đạt được tốt nhất sau 4 giờ kể từ khi cảm ứng.

3.4.2. Tinh chế protein GH3S2 bằng cột sắc ký ái lực

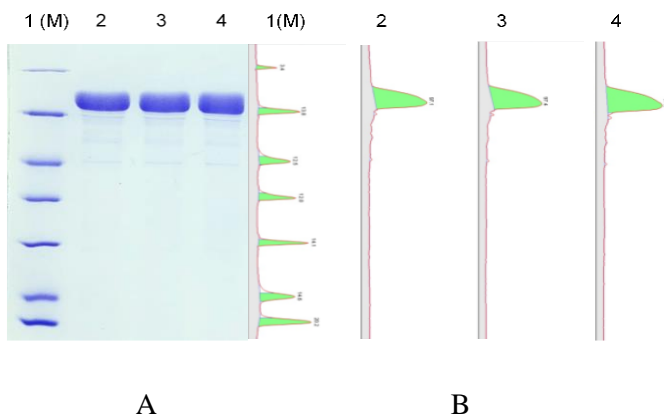
Khi tinh chế protein GH3S2 bám cột rất tốt và được thổi ra khỏi giá thể ở các phân đoạn thu protein đích ở nồng độ 300 mM imidazol. Kết quả thu được hàm lượng protein GH3S2 trong mẫu tinh sạch là 1,54 mg/ml. Như vậy, trong 1 lít dịch nuôi cấy thu được lượng protein tổng số 735,68 mg trong đó lượng GH3S2 tinh sạch thu được là 41,80 mg.



Hình 3.1. Điện di đồ kiểm tra sản phẩm trong các phân đoạn tinh chế enzyme GH3S2 trên gel polyacrylamide 12,5%

Để xác định độ sạch của protein GH3S2, chúng tôi sử dụng điện di SDS-PAGE với lượng mẫu xác định và phân tích kết quả bằng phần mềm Image Lab để đánh giá độ sạch tương đối của protein GH3S2. Kết quả thu được độ sạch của GH3S2 sau tinh chế là 97,3%. Hoạt tính β -glucosidase của mẫu tinh sạch và mẫu protein tổng số được xác định trong đó hoạt tính của protein tổng số là $0,156 \pm 0,01$ U/mg, hoạt tính của GH3S2 tinh chế là $1,10 \pm 0,02$ U/mg. Như vậy

protein GH3S2 được tinh chế 7,05 lần, hiệu suất tinh chế là 40,06% (Bảng 3.10).



Hình 3.17. Kết quả kiểm tra độ sạch GH3S2 sau tinh chế (A). Điện di đồ GH3S2 sau tinh chế (2 μ g); (B). Kết quả đo độ sạch bằng phần mềm Image Lab

Bảng 3.10. Bảng tổng kết hiệu suất tinh chế GH3S2 tái tổ hợp

(*: tính trên 1 lit dịch nuôi cấy)

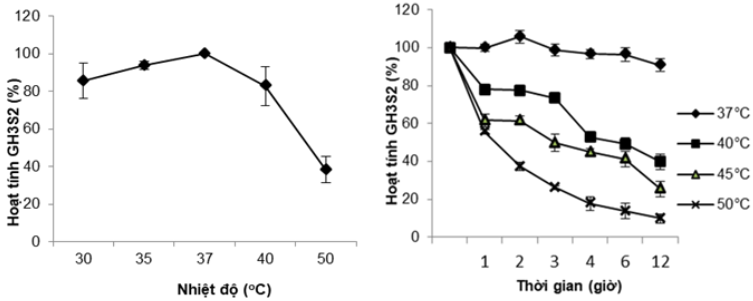
| | Tổng lượng protein (mg)* | Tổng hoạt tính (U)* | Hoạt tính riêng (U/mg) | Độ sạch (lần) | Hiệu suất thu hồi (%) |
|------------------------|--------------------------|---------------------|------------------------|---------------|-----------------------|
| Protein tổng số | 735,68 \pm 0,6 | 114,77 \pm 1,2 | 0,156 \pm 0,01 | 1 | 100 |
| GH3S2 tinh chế | 41,80 \pm 0,3 | 45,98 \pm 0,8 | 1,10 \pm 0,02 | 7,05 | 40,06 |

3.4.3. Nghiên cứu tính chất của protein tái tổ hợp GH3S2

3.4.3.1. Ảnh hưởng của nhiệt độ đến hoạt tính và độ bền nhiệt của enzyme GH3S2

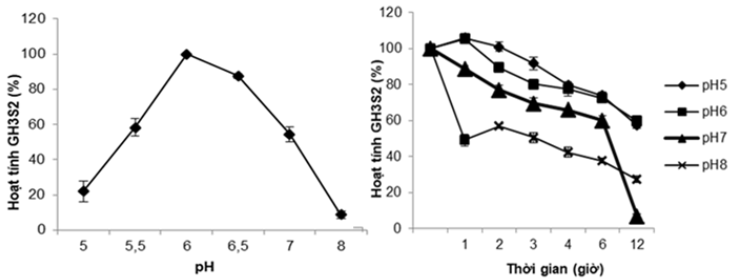
Để đánh giá ảnh hưởng của nhiệt độ đến hoạt tính của GH3S2, thì các nhiệt độ 30°C, 35°C, 37°C, 40°C, 50°C đã được kiểm tra. Kết quả cho thấy ở 37°C thì enzyme GH3S2 thể hiện hoạt tính cao nhất. Để kiểm tra độ bền nhiệt, GH3S2 được xử lý ở các nhiệt độ 37°C,

40°C, 45°C, 50°C trong các khoảng thời gian khác nhau 1, 2, 3, 4, 6 và 12 giờ trước khi ủ với cơ chất. Kết quả cho thấy, ở nhiệt độ 37°C hoạt tính enzyme của GH3S2 gần như ổn định, sau 12 giờ xử lý hoạt tính của enzyme vẫn đạt 90,78%.



Hình 3.18. Ảnh hưởng của nhiệt độ đến hoạt tính và độ bền nhiệt của enzyme GH3S2 theo thời gian

3.4.3.2. Ảnh hưởng của pH đến hoạt tính và độ bền pH của enzyme GH3S2



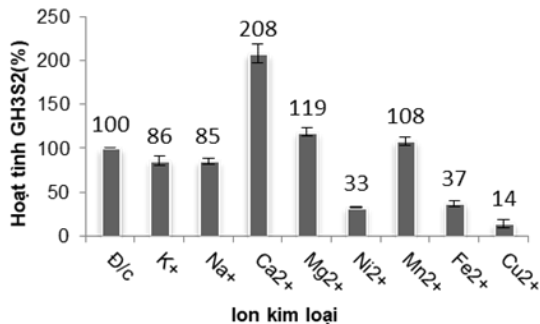
Hình 3.19. Ảnh hưởng của pH đến hoạt tính và độ bền pH của enzyme GH3S2 theo thời gian

Trong nghiên cứu này, chúng tôi lựa chọn dải pH gồm 6 giá trị khác nhau từ 5,0 đến 8,0 để tìm pH tối ưu cho ghs2 hoạt động. Kết quả cho thấy, ở pH 6,0 enzyme GH3S2 có hoạt tính cao nhất. Khi kiểm tra độ bền của enzyme với các điều kiện pH khác nhau, ở pH 6,0

enzyme được duy trì sự ổn định khoảng 70% hoạt tính sau 6 giờ, ở pH 7,0 hoạt tính enzyme được duy trì 70% sau 4 giờ sau đó giảm xuống và pH càng cao hoạt tính của enzyme càng giảm.

3.4.3.3. Ảnh hưởng của một số ion kim loại đến hoạt tính của enzyme GH3S2

Để đánh giá ảnh hưởng của các ion kim loại đến hoạt tính enzyme của GH3S2, trong nghiên cứu này chúng tôi sử dụng 8 ion kim loại K^+ , Na^+ , Ca^{2+} , Mg^{2+} , Ni^{2+} , Mn^{2+} , Fe^{2+} , Cu^{2+} với nồng độ cuối cùng là 1 mM và phản ứng xác định hoạt tính được thực hiện điều kiện tối ưu là 37°C, pH 6,0 trong thời gian 15 phút. Kết quả cho thấy khi thêm Ca^{2+} , Mg^{2+} , Mn^{2+} hoạt tính của enzyme lần lượt là 208%, 119%, 108%. Trong khi đó, các ion Fe^{2+} , Ni^{2+} , Cu^{2+} làm giảm mạnh hoạt tính của enzyme xuống còn lần lượt là 37%, 33%, 14%. Các ion K^+ , Na^+ ảnh hưởng không đáng kể đến hoạt tính GH3S2.

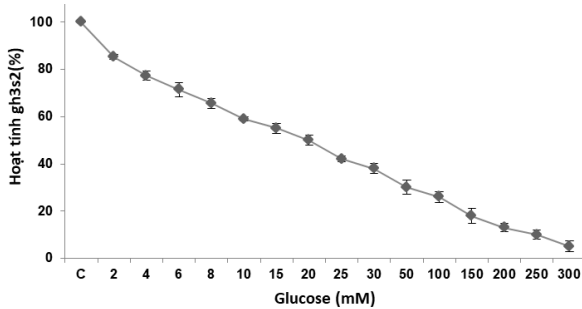


Hình 3.20. Ảnh hưởng của các ion kim loại đến hoạt tính GH3S2

3.4.3.4. Nghiên cứu ảnh hưởng của glucose đến hoạt tính của enzyme GH3S2

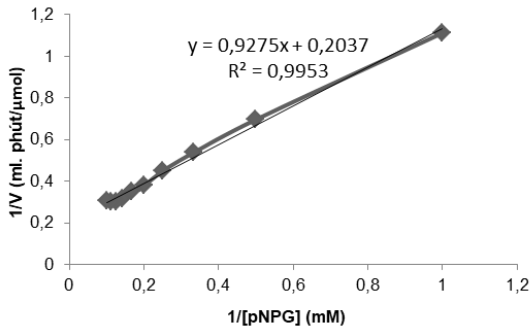
Ảnh hưởng của glucose lên hoạt tính của GH3S2 được thực hiện ở các nồng độ glucose khác nhau từ 2-300 mM. Kết quả cho thấy, glucose đến nồng độ 6 mM, hoạt tính của enzyme chỉ bị ảnh hưởng ít

và duy trì được khoảng 70% sau đó giảm xuống chỉ còn 6% khi nồng độ glucose tăng lên 300 mM.



Hình 3.21. Ảnh hưởng của glucose đến hoạt tính GH3S2

3.4.3.5. Đặc điểm động học của enzyme GH3S2



Hình 3.22. Sự phụ thuộc tốc độ phản ứng của gh3s2 vào nồng độ cơ chất pNPG theo Lineweaver – Burk

Hằng số động học của GH3S2 (K_m , V_{max}) được xác định bởi dải nồng độ cơ chất pNPG từ 1-10 mM trong đệm PBS 50 mM, pH 6,0, với lượng enzyme là 1 μ g trong 1 phản ứng. Kết quả cho thấy sự phụ thuộc tốc độ phản ứng vào nồng độ pNPG tuân theo phương trình $y = 0,9275x + 0,2037$ với độ tin cậy $R^2 = 0,9953$ (Hình 3.22). Dựa trên phương trình $1/v = K_m.1/V_{max}.1/[S] + 1/V_{max}$, các giá trị K_m , V_{max} của GH3S2 được tính tương ứng là 4,55 mM và 4,91 μ mol/phút.

Ở điều kiện này, enzyme GH3S2 có hoạt tính riêng là 2,23 U/mg với cơ chất pNPG.

KẾT LUẬN VÀ KIẾN NGHỊ

KẾT LUẬN

Từ những kết quả thu được của luận án, chúng tôi rút ra được các kết luận sau:

1. Đã xây dựng bộ dữ liệu DNA đa hệ gen của quần xã vi sinh vật đất xung quanh khu nấm mục trắng ở vườn Quốc gia Cúc Phương với dung lượng 51,82 Gb và phân tích được 3.896.881 ORF thuộc 131 ngành, 118 lớp, 237 bộ, 523 họ, 2240 chi và 916 loài, trong đó có 3.884.879 ORF thuộc giới vi khuẩn được thuộc 111 ngành, 83 lớp, 170 bộ, 406 họ, 1971 chi và 738 loài. Proteobacteria là ngành phổ biến nhất với 3.106.400 gen (75,68%) và ngành Bacteroidetes lớn thứ hai (13,11%) trong số các ngành được phân tích;
2. Dựa trên CSDL KEGG đã chú giải được chức năng của 22.226 gen mã hóa enzyme tham gia thủy phân lignocellulose trong đó 907 gen mã hóa enzyme và protein tham gia tiền xử lý, 8301 gen mã hóa enzyme cellulase và 13.018 gen mã hóa enzyme hemicellulase. Có 22.092 gen được phân loại thuộc 28 ngành của vi khuẩn, trong đó, trội nhất là ngành Proteobacteria (50,79%) và ngành Bacteroidetes (36,73%). Đã khai thác được 13 họ enzyme tham gia thủy phân lignocellulose bằng mô hình đại diện HMM;
3. Trong số 8301 gen mã hóa cellulase được chú giải bằng CSDL KEGG có 1058 gen hoàn chỉnh được phân tích các vùng/cấu trúc chức năng bao gồm các nhóm: (1) endoglucanase với 47 loại domain (367 gen), trong đó domain GH8 là phổ biến nhất; (2) exoglucanase với 6 domain (6 gen); (3) β -glucosidase với 27 loại domain (475 gen), trong

đó domain phổ biến là GH3 với vùng/cấu trúc FN3, Exop_C), GH1;
(4) 6-phospho- β -glucosidase với 2 domain GH1 và GH4 (210 gen);

4. Đã biểu hiện thành công protein GH3S2 có kích thước khoảng 91 kDa trong chủng *E. coli* Rosetta1 ở 25°C, môi trường TB cải biến, 0,3 mM IPTG, cảm ứng ở OD₆₀₀ là 1 và thu mẫu 4 giờ sau khi cảm ứng. Đã tinh chế được enzyme tái tổ hợp GH3S2 từ vi khuẩn *E. coli* Rosetta có độ sạch là 97,3 %, hàm lượng đạt 41,8 mg/lít dịch lên men; enzyme GH3S2 có Km = 4,55 mM và Vmax = 4,91 U/mg. Ion Ca²⁺ và Mg²⁺ làm tăng hoạt tính enzyme, trong khi đó, ion Ni²⁺ và Cu²⁺ làm giảm hoạt tính. Glucose ở nồng độ 6 mM ảnh hưởng nhẹ đến hoạt tính của GH3S2.

KIẾN NGHỊ

Nghiên cứu khả năng phối hợp các enzyme β -glucosidase GH3S2 với enzyme endoglucanase và exoglucanase để đánh giá hiệu quả thủy phân nguồn cơ chất cellulose.

NHỮNG ĐÓNG GÓP MỚI CỦA LUẬN ÁN

- Đây là nghiên cứu đầu tiên về đa dạng vi khuẩn xung quanh khu nấm mục trắng phân giải lignocellulose ở rừng Quốc gia Cúc Phương bằng kỹ thuật Metagenomics.

- Đã nghiên cứu được đa dạng enzyme tham gia phân giải lignocellulose ở khu hệ vi khuẩn xung quanh nấm mục trắng ở rừng Quốc gia Cúc Phương, lựa chọn và đánh giá được tính chất của enzyme β -glucosidase GH3S2 từ DNA đa hệ gen của vi khuẩn đất mùn xung quanh nấm mục trắng.

DANH MỤC CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ

1. **Nguyễn Thị Bình**, Đào Trọng Khoa, Lê Thị Thu Hồng, Trương Nam Hải, Nghiên cứu khai thác các gen mã hóa enzyme oxy hóa đa đồng tử từ dữ liệu metagenome của khu hệ vi khuẩn quanh nấm mục trắng (*Trametes versicolor*) trong rừng Quốc gia Cúc Phương, Hội nghị Công nghệ sinh học toàn quốc, 2020, tr 187-192.

2. **Nguyễn Thị Bình**, Nguyễn Hồng Dương, Nguyễn Thị Quý, Lê Thị Thu Hồng, Trương Nam Hải, Nghiên cứu khai thác và biểu hiện gen mã hóa enzyme β -glucosidase từ dữ liệu metagenome của khu hệ vi khuẩn quanh nấm mục trắng (*Trametes versicolor*), Hội nghị Công nghệ sinh học toàn quốc, 2021, tr 16 -22.

3. Thi-Thu-Hong Le, **Thi-Binh Nguyen**, Hong-Duong Nguyen, Hai-Dang Nguyen, Ngoc-Giang Le, Trong-Khoa Dao, Thi-Quy Nguyen, Thi-Huyen Do, Nam-Hai Truong, De Novo metagenomic analysis of microbial community contributing in lignocellulose degradation in humus samples harvested from Cuc Phuong tropical forest in Vietnam, Diversity, 2022, 14(3), 220; <https://doi.org/10.3390/d14030220>

4. **Nguyen Thi Binh**, Nguyen Thi Quy, Do Thi Huyen, Le Thi Thu Hong, Trương Nam Hải, Selection of optimal culture conditions for expression of recombinant beta-glucosidase in *Escherichia Coli*, Tạp chí Công nghệ sinh học, 2022, 20(3): 425-433.

5. **Nguyen Thi Binh**, Le Thi Thu Hong, Trương Nam Hải, Using some bioinformatic tools to mining genes coding cellobiohydrolase from metagenome data of the bacteria surrounding white-rot fungi (*Trametes versicolor*) in Cuc Phuong National Park, Tạp chí Khoa học Đại học Thủ đô Hà Nội tập 62/2022: 119-126.

6. **Nguyen Thi Binh**, Nguyen Thi Quy, Le Thi Thu Hong, Trương Nam Hải, Purification and characterization of a recombinant beta-glucosidase in *Escherichia Coli*, Tạp chí Công nghệ sinh học, 2022, 20(4): 1-9.