

**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 DẠ CỎ ĐÊ VÀ
KHAI THÁC, NGHIÊN CỨU TÍNH CHẤT CỦA
ENDO-XYLANASE**

TÓM TẮT LUẬN ÁN TIẾN SĨ SINH HỌC

Ngành: Hóa sinh học

Mã số: 9 42 01 16

Hà Nội – 2024

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:

Người hướng dẫn 1: **GS.TS. Trương Nam Hải**

Viện Công nghệ sinh học

Người hướng dẫn 2: **PGS.TS. Đỗ Thị Huyền**

Viện Công nghệ sinh học

Phản biện 1: PGS. TS. Phạm Thế Hải

Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội

Phản biện 2: GS. TS. Lê Mai Hương

Viện Hóa học các hợp chất thiên nhiên, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

Phản biện 3: PGS. TS. Trương Quốc Phong

Trường Hóa và Khoa học Sự sống, Đại học Bách Khoa Hà Nội

Luận án sẽ 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 202....

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

1.1. Tính cấp thiết của luận án

Lignocellulose, một trong những nguồn năng lượng tái tạo dồi dào trên Trái Đất phần lớn bị đem đi đốt, gây lãng phí và ảnh hưởng nghiêm trọng đến chất lượng môi trường sống cũng như sức khỏe của người dân. Vì vậy, việc tận dụng nguồn nguyên liệu dư thừa này để chuyển hóa chúng thành nhiên liệu sinh học không những làm giảm thiểu ô nhiễm môi trường mà còn góp phần giải quyết nhu cầu năng lượng quốc gia. Tuy nhiên, trên thực tế, lignocellulose là sinh khối rắn chắc khó chuyển hóa và đường hóa. Hướng xử lý phân giải lignocellulose bằng phương pháp sinh học theo hướng thân thiện với môi trường ngày càng được xem trọng và ứng dụng rộng rãi. Việc tìm kiếm nguồn lignocellulase có hoạt tính mạnh đã và đang là một trong những hướng nghiên cứu trọng tâm của nhiều nhà khoa học trên thế giới. Vi khuẩn khu trú trong các khu hệ giàu lignocellulose 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% 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. 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 và giải mã sâu 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 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”.

1.2. 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.

1.3. 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 gene, 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.

CHƯƠNG 1. TỔNG QUAN

1.1. Tổng quan về lignocellulose

Lignocellulose là một thành phần quan trọng và chiếm tỷ lệ lớn nhất của sinh khối thực vật, chủ yếu cấu thành nên thành tế bào thực vật. Lignocellulose được cấu thành từ ba thành phần chính đều là các polymer phân tử lớn: cellulose, hemicellulose, lignin. Sinh khối lignocellulose là một trong ba nguồn sinh khối chính có thể được sử dụng để sản xuất nhiên liệu sinh học, là nguồn năng lượng mới, khắc phục những nhược điểm của nguồn năng lượng hóa thạch. Thành phần của lignocellulose khi được phân giải ngoài việc cung cấp năng lượng còn có ứng dụng trong nhiều ngành kinh tế - xã hội khác như công nghiệp thực phẩm, y dược, miễn dịch...

1.2. Xylanase

Xylanase là một trong những enzyme phân giải xylan quan trọng nhất, với vai trò phân cắt mạch chính của xylan, tạo điều kiện cho các enzyme khác hoạt động. Những họ GH quan trọng nhất có hoạt tính xylanase là GH 5, 7, 8, 10, 11 và 43, theo cơ sở dữ liệu CAZy. Xylanase rất phổ biến trong tự nhiên, có nguồn gốc từ rất nhiều lớp sinh vật, trong đó xylanase từ vi khuẩn và nấm đã và đang được nghiên cứu và ứng dụng rộng rãi trong nhiều ngành công nghiệp.

1.3. Kỹ thuật metagenomics nhằm khai thác hiệu quả gene tiềm năng

Kỹ thuật metagenomics là kỹ thuật nghiên cứu trực tiếp DNA đa hệ gene mà không thông qua nuôi cấy, trong đó hướng đi mới nhất là bằng phương pháp giải trình tự toàn bộ nhờ những tiến bộ của kỹ thuật giải trình tự. Thông tin trình tự được phân tích xử lý bằng phần mềm để định dạng về phân loại và chức năng. Rất nhiều phương pháp mới đã và đang được phát

triển để hỗ trợ cho việc phân tích và chú giải chức năng gene một cách hiệu quả, trong đó phương pháp sử dụng mô hình đại diện HMM là một phương pháp có độ nhạy và độ chính xác cao nhất trong việc đại diện cho các trình tự tương đồng trong họ trình tự.

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

✓ **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 phẩm nông nghiệp khác nhau, không cho ăn cám.

✓ **Các chủng vi sinh vật:** Chủng vi khuẩn *E. coli* DH10B của hãng Invitrogen (Mỹ) được sử dụng làm thể nhận trong thí nghiệm tách dòng gene; chủng *E. coli* BL21(DE3), Rosetta1, JM109, SoluBL21 (BL21 Soluble), Origami 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 hóa chất được mua của Merck, Sigma, các bộ kit mua của Qiagen, Fermentas, Amersham.

2.1.3. Máy móc và thiết bị

Các máy móc và thiết bị đều có nguồn gốc từ những hãng uy tín như Shimadzu (Nhật Bản), BioTek (Mỹ), Bio-rad (Mỹ). Sorvall (Mỹ), Amersham Pharmacia (Mỹ), Applied Biosystems (Mỹ), GE-Healthcare (Thụy Điển), Implen (Đức), Precisa (Thụy Sĩ).

2.2. Phương pháp nghiên cứu

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

- Tách chiết, tinh chế DNA metagenome
- Tổng hợp gene, thiết kế vector biểu hiện mang gene *exl*
- Biến nạp DNA plasmid vào vi khuẩn *E. coli*
- Tách chiết DNA plasmid từ tế bào vi khuẩn *E. coli*
- Điện di trên gel agarose
- Tinh chế DNA từ gel agarose

2.2.2. Các phương pháp hóa sinh protein

- Biểu hiện protein tái tổ hợp trong *E. coli*
- Điện di protein trên gel polyacrylamide
- Tinh chế protein bằng sắc kí ái lực His-tag
- Xác định độ sạch của enzyme tái tổ hợp bằng phần mềm Quantity One
- Định lượng protein bằng phương pháp Bradford
- Xác định hoạt tính xylanase
- Xác định ảnh hưởng của nhiệt độ, pH, các ion kim loại và một số hóa chất lên hoạt tính enzyme
- Xác định độ bền nhiệt của enzyme
- Xác định thông số động học của enzyme

2.2.3. Các phương pháp tin sinh học

- Lắp ráp DNA đa hệ gene, chú giải các gene chức năng
- Phương pháp nghiên cứu Pfam của các trình tự
- Nghiên cứu vùng bảo thủ và dự đoán cấu trúc bậc ba của các trình tự
- Dự đoán khả năng chịu kiềm/acid
- Dự đoán khả năng chịu nhiệt của enzyme
- Định loại loài các trình tự ORF
- 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.

- Phương pháp xử lý số liệu.

CHƯƠNG 3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Nghiên cứu giải mã sâu, 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

Dựa trên kết quả tách chiết, 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. 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. Cả 10 mẫu DNA metagenome từ vi khuẩn dạ cỏ dê được sử dụng làm khuôn cho khuếch đại đoạn gene 16S của vi khuẩn, kết quả đảm bảo trong mẫu DNA không có chất ức chế quá trình tổng hợp DNA *in vitro*, vì vậy mẫu DNA đã sẵn sàng để giải trình tự.

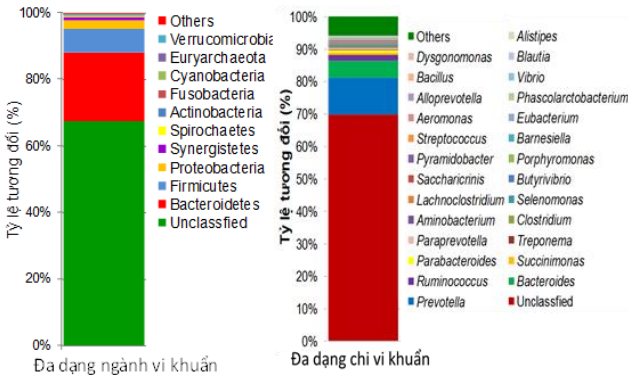
3.1.2. Giải trình tự, đánh giá chất lượng hai bộ dữ liệu và chú giải gene

Kết quả cho thấy cả hai bộ dữ liệu đều có chất lượng tốt, Q30 đều đạt trên 90%. Tổng dung lượng dữ liệu giải trình tự metagenome của vi khuẩn trong dạ cỏ dê là 392,63 triệu read. Sau khi lọc bỏ đã thu được 324 triệu read dữ liệu tinh tương đương với 48,66 Gb. Các read có chất lượng Q30 chiếm 94,59% và tỷ lệ read sạch đạt 82,61%. Sau khi lắp ghép thành contig, số lượng các contig của hai bộ dữ liệu là khá lớn. Bộ dữ liệu DNA metagenome 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.

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. Trong số đó, 39.579 ORF đã được xác định và định loại, trong đó 99,8% thuộc về vi khuẩn. 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%). Ở cấp độ chi, *Prevotella* (35,3%) và *Bacteroides* (16%) thuộc về bộ Bacteroidales là phổ biến nhất.

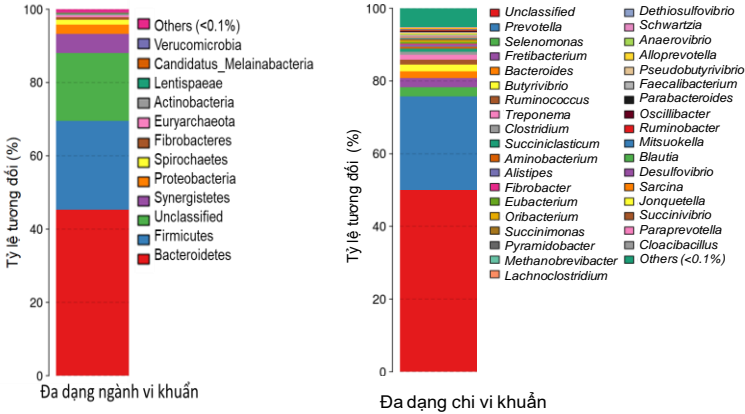


Hình 3. 3. Biểu đồ phân bố đa dạng phân loại học ở mức độ ngành và mức độ chi của vi khuẩn trong dạ cỏ dê khai thác được từ dữ liệu 8,6 Gb

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, 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%. Ngành Bacteroidetes là ngành chiếm tỷ lệ lớn nhất với

45,29% tổng số gene, tiếp theo là ngành Firmicutes với 30,38%. 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.



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

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

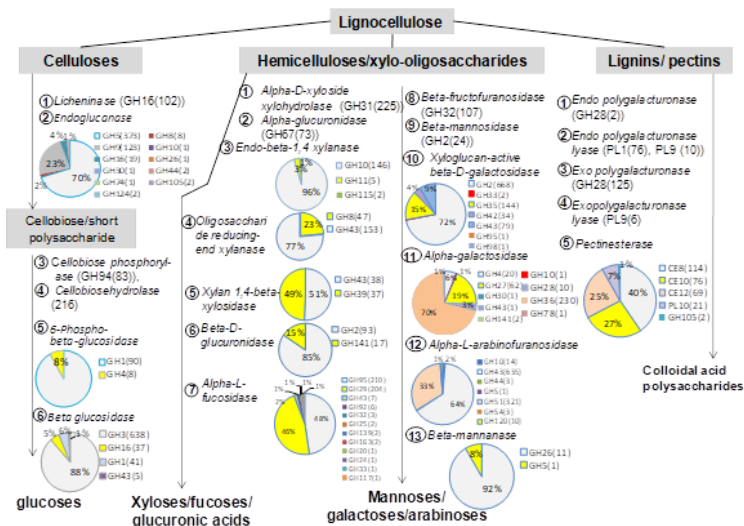
3.2.1. Khai thác gene mã hóa enzyme thủy phân lignocellulose dựa trên cơ sở dữ liệu KEGG

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

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.

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

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.



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

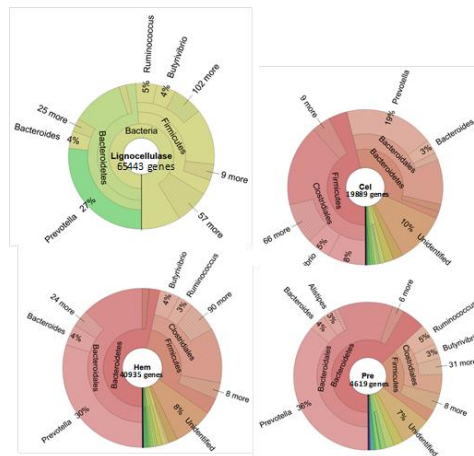
3.2.2. Phân tích đa dạng vi khuẩn mang gene thủy phân lignocellulose

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%).

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

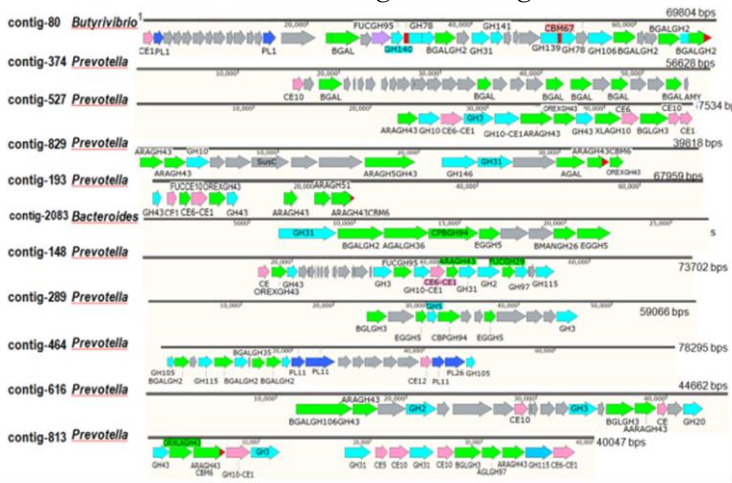


Hình 3. 6. Đa dạng phân loại vi khuẩn mang gene lignocellulase trong dạ cỏ dê Việt Nam đã được KEGG chú thích và phân loại bởi MEGAN

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.

3.2.2.3. Vai trò của chi *Prevotella* trong tiêu hóa lignocellulose



Hình 3.8. Các locus gene phân giải celluloses/hemicellulose trong những contig tiềm năng xây dựng từ dữ liệu giải trình tự sâu DNA metagenome vi khuẩn dạ cỏ dê.

8.900 gene lignocellulase hoàn chỉnh nằm trong 8.364 contig, trong đó 7848 contig chỉ mang một gene trên mỗi contig. Trong số 22 contig mang ít nhất 4 gen/contig thì có 18 contig thuộc chi *Prevotella*, 2 contig thuộc chi *Bacteroides*, 1 contig thuộc chi *Clostridium* và 1 contig thuộc chi *Butyrivibrio*. Hầu hết các cụm gene liên quan đến quá trình phân hủy hemicellulose và hoạt động đặc hiệu trên một số cơ chất nhất định. 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.

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

. 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.

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

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. Trong đó 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. 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.

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%).

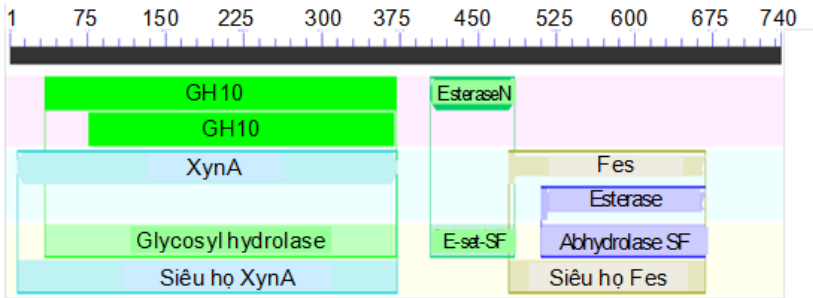
3.3.1.2. Nghiên cứu đa dạng cấu trúc endo-xylanase



Hình 3. 10. Tóm lược các cấu trúc domain của enzyme nhóm endo-xylanase có mang vùng hoạt tính GH

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, có 185 trình tự là mang các vùng hoạt tính glycosyl hydrolase. 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*. Trong số này, 10 gene có trình tự cho vùng hoạt tính phụ CE1, 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



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

Trình tự protein 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%. 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). Mã gene [denovogenes]_5086 có điểm acid thấp nhất là 0,363 trên tổng điểm là 1.

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 [desnovogenes]_5086 ký hiệu *exl* mã hóa cho endo-xylanase được lựa chọn để biểu hiện trong *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*.

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

Sản phẩm cắt kiểm tra vector có hai băng DNA có kích thước đúng với kích thước lý thuyết của gene endo-xylanase và vector pET22b(+) ở dạng thẳng. So sánh trình tự *exl* trong vector khi kiểm tra lại bằng giải trình tự với trình tự gene endo-xylanase gửi đặt tổng hợp thấy chúng có độ tương đồng 100% với nhau. 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

Protein endo-xylanase đã được biểu hiện thành công trong cả 5 chủng BL21 Soluble, Rosetta1, Rosetta2, JM109, Origami. Lượng protein tan cùng với hoạt độ thu được từ một đơn vị dịch nuôi cấy của chủng Rosetta1 là lớn nhất 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

Protein EXL đã được biểu hiện trong các loại môi trường LB, TB cải biến, SOB, PE, YT, SB. Môi trường PE được lựa chọn để biểu hiện endo-xylanase trong các thí nghiệm tiếp theo.

(3) Lựa chọn nhiệt độ nuôi cấy

Endo-xylanase được nghiên cứu biểu hiện ở các nhiệt độ 20°C, 25°C, 30°C và 37°C. Protein được biểu hiện tốt ở dạng tan và có hoạt độ trên một đơn vị môi trường nuôi cấy ở nhiệt độ 20°C so với các nhiệt độ khác biểu hiện ở dạng không tan hoặc biểu hiện kém. Vì vậy chúng tôi chọn nhiệt độ biểu hiện endo-xylanase là 20°C.

(4) Lựa chọn nồng độ chất cảm ứng IPTG

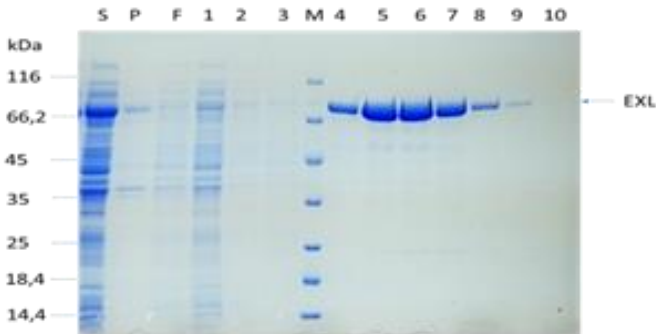
Các nồng độ IPTG khác nhau từ 0,05 đến 1,5 mM được kiểm tra để biểu hiện EXL tái tổ hợp. Kết quả cho thấy nồng độ IPTG thích hợp được lựa chọn là 0,1 mM để cho các nghiên cứu tiếp theo.

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

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). 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



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 EXL tổng số; F: mẫu sau khi đưa qua cột; 1: mẫu rửa bằng đệm 50 mM imidazol; 2-3: mẫu rửa bằng đệm 100 mM imidazol; 4-10: các phân đoạn mẫu khi rửa bằng đệm 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. 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.

Kết quả phân tích bằng phần mềm dựa trên phương pháp so sánh độ đậm của 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

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 sau khi loại muối 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%. 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%.

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

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).

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

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

đổi mạnh khi pH phản ứng nằm trong khoảng 5-6 (đạt trên 90% hoạt tính tối đa so với pH 5,5).

3.3.4.3. Nghiên cứu tính bền nhiệt của enzyme

Kết quả nghiên cứu cho thấy, enzyme xylanase bền tốt ở nhiệt độ tối ưu cho hoạt tính là 40°C, ở nhiệt độ 50°C, hoạt tính bắt đầu giảm chỉ sau 1 giờ xử lý 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. Vì vậy có thể kết luận enzyme xylanase không bền nhiệt.

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

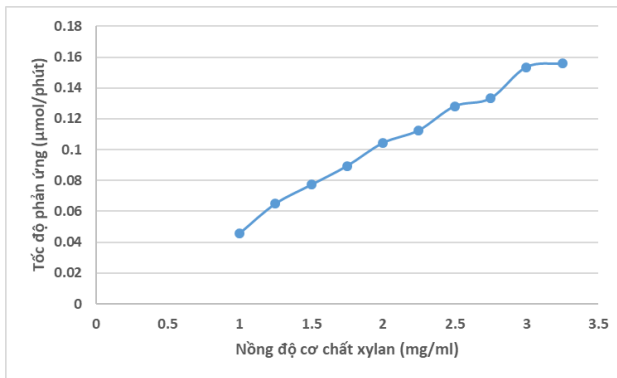
Kết quả thử hoạt tính xylanase chứng minh khi có sự có mặt của các ion Fe^{3+} , Cu^{2+} , Co^{2+} , Zn^{2+} và chất biến tính SDS thì hoạt tính của enzyme xylanase giảm mạnh, đặc biệt hai ion Fe^{3+} , Cu^{2+} gần như làm cho enzyme mất hoạt tính (hoạt tính chỉ đạt 7% so với mẫu đối chứng không bổ sung ion). Trong môi trường có sự có mặt của các hóa chất 2-mecaptoethanol, Tween-20 và Triton X100, hoạt tính của endo-xylanase cũng bị ức chế chỉ còn khoảng 40%.

3.3.4.5. Nghiên cứu tính đặc hiệu cơ chất của enzyme

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. Bên cạnh đó, enzyme EXL cũng đã được thử hoạt tính với một số cơ chất esterase như gelatin, pectin, tributyrin hay sữa tách bơ ở nồng độ 1% tuy nhiên enzyme không thể hiện hoạt tính.

3.3.4.6. Nghiên cứu thông số động học của enzyme

Dựa trên kết quả thực nghiệm, đường chuẩn độ tương quan giữa nghịch đảo giá trị nồng độ cơ chất $1/[S]$ và tốc độ phản ứng $1/V$ được xây dựng 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 $\mu\text{mol}/\text{phút}$ với hoạt tính riêng đạt 171,56 IU/mg.



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

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 μM), 2-mercaptoethanol (1 μM), EDTA (1 μM), tween 80 (1mM), triton X-100 (1 μM). 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 $\mu\text{mol}/\text{phút}$ 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.

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 trong dạ cỏ dê đã được biểu hiện, tinh sạch thành công với hoạt tính cao.

**DANH MỤC CÁC BÀI BÁO ĐÃ XUẤT BẢN 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, Trọng 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, **Đào Trọng Khoa**, Trương Nam Hải, 2021, Expression of gene coding endoglucanase GH5-4 derived from metagenomic DNA data of bacteria in goats rumen in *Escherichia coli*, *Academia Journal of Biology*, **43**(2): 17–26.