

NGHIÊN CỨU KHAI THÁC CÁC GEN MÃ HÓA ENZYME OXI HÓA ĐA ĐỒNG TỪ DỮ LIỆU METAGENOME CỦA KHU HỆ VI KHUẨN QUANH NẤM MỤC TRẮNG (*Trametes versicolor*) TRONG RỪNG QUỐC GIA CÚC PHƯƠNG

Nguyễn Thị Bình^{2,3}, Đào Trọng Khoa^{1,3}, Lê Thị Thu Hồng^{1,3}, Trương Nam Hải^{1,3*}

¹ Viện Công nghệ sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

² Trường Đại học Thủ đô Hà Nội

³ Học viện Khoa học và Công nghệ sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

TÓM TẮT

Enzyme oxi hóa đa đồng có khả năng phân hủy đa dạng các loại cơ chất, vì vậy, chúng được ứng dụng nhiều trong thực tiễn. Các enzyme thuộc nhóm này rất đa dạng ở các hệ sinh thái khác nhau. Dữ liệu DNA metagenome của vi sinh vật quanh khu nấm mục trắng được giải trình tự, thu được bộ dữ liệu 51,8 Gb. Sử dụng phần mềm MGA đã dự đoán được 4.104.872 ORF. Bằng phần mềm HMM đã khai thác được 901 ORF mã hóa enzyme oxi hóa đa đồng, trong đó có 338 ORF chứa gen hoàn thiện. Trong số 338 ORF hoàn thiện có 109 ORF chứa trung tâm hoạt động là các nguyên tử đồng. Để khai thác được các ORF đó, chúng tôi sử dụng phần mềm MegaX để phân tích sự đa dạng về gen. Kết quả cho thấy các gen được phân loại theo các nhóm: laccase, ferroxidase, ascorbate oxidase, bilirubin oxidase, manganese oxidase và một số dạng enzyme oxi hóa khác. Kết quả phân tích các vùng chức năng bằng Pfam cho thấy có 80 ORF chứa 3 vùng chức năng và 29 ORF có 2 vùng chức năng. Các nhóm gen này được so sánh bằng Blastp với cơ sở dữ liệu của NCBI để dự đoán chức năng và đánh giá mức độ tương đồng. Kết quả thấy 14 ORF đại diện thuộc 6 nhóm enzyme này có độ bao phủ từ 93 - 100%, 9 ORF có độ tương đồng cao với các protein đã công bố trên ngân hàng NCBI (từ 85% trở lên), còn lại 5 ORF có độ tương đồng từ 55 - 74%. Phần mềm Phyre2 được sử dụng để xác định cấu trúc bậc hai của các gen thuộc các nhóm khác nhau, kết quả thấy cấu trúc bậc hai của các enzyme oxi hóa đa đồng với các protein đã được công bố với độ tin cậy đều là 100% và độ bao phủ từ 58 - 97%.

Từ khóa: Enzyme oxi hóa đa đồng, lignocellulose, metagenome, metagenomic, nấm mục trắng, tin sinh học.

MỞ ĐẦU

Tiền xử lý là giai đoạn quan trọng và tốn kém nhất trong quá trình chuyển hóa lignocellulose. Trong quá trình này, cấu trúc tinh thể của sợi lignocellulose được nới lỏng, tạo điều kiện để các enzyme tham gia thủy phân cellulose và hemicellulose hiệu quả hơn nguồn cơ chất này (Haghighi *et al.*, 2013; Zheng *et al.*, 2014). Trong tự nhiên, sinh khối này được phân hủy chủ yếu bởi các enzyme của các vi sinh vật trong các hệ sinh thái khác nhau. Nấm mục trắng có thể chuyển hóa tất cả các thành phần trong gỗ gồm cellulose, hemicellulose và lignin, trong đó lignin được phân hủy nhiều nhất. Khả năng này có được là do nấm tiết phức hợp enzyme không đặc hiệu ngoại bào tham gia phân hủy lignin trong suốt quá trình phân hủy lignocellulose. Tính không đặc hiệu của enzyme làm chúng chuyển hóa được rất nhiều loại hợp chất có cấu trúc tương tự lignin. Sự phân giải hiệu quả lignin của nấm thường được kết hợp cùng với các enzyme của vi khuẩn sống trong cùng hệ sinh thái. Trong số các enzyme của vi sinh vật trong khu hệ quanh khu nấm mục trắng, enzyme oxi hóa đa đồng (multicopper oxidase-MCO) là họ enzyme xúc tác cho các phản ứng oxi hóa khử các cơ chất đồng thời xảy ra sự khử phân tử oxi thành nước. Trung tâm xúc tác của enzyme oxi hóa đa đồng phù hợp cho mỗi phản ứng, gồm bốn hoặc nhiều hơn các nguyên tử đồng thuộc ba trung tâm đồng khác nhau về quang phổ là T1, T2, T3 (Aleksandra *et al.*, 2018). Enzyme thuộc họ này được mô tả nhiều nhất là laccase, ascorbate oxidase, bilirubin oxidase, ferroxidases, enzyme giống laccase (laccase-like multicopper oxidases) và một số dạng khác (Copete *et al.*, 2015; Hoegger *et al.*, 2006). Khả năng oxi hóa khử mạnh và có khả năng xúc tác đa dạng các loại phản ứng làm cho họ enzyme này ngày càng phổ biến và có tiềm năng ứng dụng cao. Việc tìm kiếm các hệ sinh thái có vi sinh vật tiết enzyme này có vai trò quan trọng. Trong đó, rừng là khu hệ sinh thái đặc biệt với các nguồn gen đa dạng, đặc biệt rừng quốc gia Cúc Phương là nơi có các nguồn gen tiến hóa tự nhiên ít bị ảnh hưởng bởi tác động của con người. Khu hệ sinh thái này là nơi rất tiềm năng để tìm kiếm các nguồn gen mới. Có nhiều phương pháp khác nhau để sàng lọc và tìm kiếm gen mới, trong đó metagenomics là phương pháp nghiên cứu về đa hệ gen vi sinh vật của các khu hệ sinh thái tự nhiên mà không cần nuôi cấy chúng trong điều kiện phòng thí nghiệm. Trong nghiên cứu này chúng tôi thông báo kết quả khai thác các gen mã hóa enzyme oxi hóa đa đồng từ dữ liệu DNA metagenome của vi sinh vật quanh nấm mục trắng (*Trametes versicolor*) bằng các công cụ tin sinh.

NGUYÊN LIỆU VÀ PHƯƠNG PHÁP

Nguyên liệu

DNA metagenome của mẫu vi sinh vật khu trú xung quanh nấm mục trắng (*Trametes versicolor*) thủy phân gỗ trong rừng nhiệt đới Cúc Phương đã được giải trình tự bằng hệ thống giải trình tự HiSeq 2500 của Illumina (Illumina, San Diego, Hoa Kỳ) tại Công ty BGI, Hồng Kông, Trung Quốc (kết quả chưa công bố).

Phương pháp nghiên cứu

Khai thác dữ liệu bằng phần mềm MGA và mô hình đại diện HMM

Bộ dữ liệu 51,8 Gb được tập hợp thành các trình tự dài "contig" bằng phần mềm idba_ud (http://www.cs.hku.hk/~alse/idba_ud), sau đó bằng phần mềm MGA (MetaGeneMark), các contig được phân tích để dự đoán các khung đọc mở (ORF) - gene mã hóa protein và sử dụng mô hình đại diện HMM (Hidden Markov Model) để tìm kiếm trình tự tương đồng trong các khung đọc mở kể trên. Để tìm kiếm những trình tự thực sự tương đồng có ý nghĩa, kết quả thô được lọc với tiêu chí giá trị E (e-value) nhỏ hơn $1e-10$, tỷ lệ chiều dài đoạn tương đồng với mô hình đại diện HMM dùng để tìm kiếm so với chiều dài mô hình đó lớn hơn 0,75 và tỷ lệ giá trị bias:score nhỏ hơn 0,1. Mô hình đại diện HMM chuyển kết quả so sánh đa trình tự thành một hệ thống điểm (score) đặc trưng cho từng vị trí, từ đó có thể sử dụng để so sánh trình tự, tìm kiếm trong cơ sở dữ liệu để tìm kiếm các trình tự tương đồng (Eddy, 1998). Mô hình đại diện HMM cho biết thông tin về mỗi vị trí trong trình tự, loại gốc amino acid hay nucleotide nào có khả năng xuất hiện cao nhất, đồng thời xác suất xuất hiện đột biến chèn hoặc mất, vì vậy cách tiếp cận này ưu việt hơn so với việc sử dụng phần mềm BLAST truyền thống trong việc tìm kiếm các trình tự tương đồng (Eddy, 2011).

Xây dựng cây phân loại của các ORF bằng phần mềm MegaX

Sử dụng MegaX (Molecular evolutionary genetics analysis) phân tích trình tự protein để tạo ra cây phát sinh từ dữ liệu. Bước đầu tiên, xác định các trình tự tương đồng dưới dạng file fasta. Để phân loại các trình tự đó, MEGA cung cấp hai công cụ khác nhau: ClustalW và MUSCLE. Sau đó tính năng mô hình của MEGA cho phép chọn mô hình thay thế phù hợp nhất. Cuối cùng, MEGA cung cấp một giao diện mạnh mẽ và linh hoạt cho bước cuối cùng là vẽ cây phân loại.

Phân tích các vùng chức năng của ORF bằng Pfam

Cơ sở dữ liệu Pfam (<http://pfam.xfam.org/>) là một tập hợp lớn các họ protein. Protein thường bao gồm một hoặc nhiều vùng chức năng, thường được gọi là miền (domain). Do đó, việc xác định các miền có trong protein có thể cung cấp hiểu biết sâu sắc về chức năng của chúng. Để dự đoán các vùng chức năng của ORF bằng Pfam, chúng tôi cung cấp các trình tự protein quan tâm, sử dụng giá trị e - value và cung cấp địa chỉ e-mail cá nhân, xác nhận gửi lên thông qua trang web của HMMer, kết quả sẽ trả về sau 2 - 3 ngày tùy thuộc chiều dài chuỗi, số lượng các miền trên chuỗi.

So sánh với dữ liệu của NCBI nhờ BLAST

BLAST (<https://blast.ncbi.nlm.nih.gov/Blast>) là tập hợp các chương trình được thiết kế để tìm kiếm các vùng tương đồng giữa các trình tự nucleotide hoặc protein với các trình tự trong cơ sở dữ liệu. Ở đây, BLASTp được sử dụng để tìm kiếm các trình tự protein tương đồng trong ngân hàng NCBI với trình tự protein đang quan tâm và trả về kết quả. Chúng tôi quan tâm hai tham số trong phân tích là độ bao phủ và mức độ tương đồng, để có thể đánh giá được mức độ tương đồng của protein đang nghiên cứu so với cơ sở dữ liệu của NCBI.

Mô phỏng cấu trúc không gian và vị trí gắn cơ chất của enzyme bằng phần mềm Phyre2

Phyre2 (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>) là một bộ công cụ có sẵn trên web để dự đoán và phân tích cấu trúc, chức năng và đột biến protein. Để dự đoán cấu trúc bậc cao của protein, người dùng gửi trình tự protein đến để xác định cấu trúc bậc hai và bậc ba của các mô hình, thành phần miền và chất lượng mô hình của protein. Kết quả dự đoán cấu trúc điển hình sẽ được trả về trong khoảng từ 30 phút đến 2 giờ sau khi gửi.

KẾT QUẢ VÀ THẢO LUẬN

Khai thác trình tự DNA mã hóa enzyme oxi hóa đa đồng

DNA metagenome của vi sinh vật xung quanh khu nấm mục trắng được tách chiết và giải trình tự toàn bộ bằng thiết bị giải trình tự thế hệ mới (NGS). Kết quả giải trình tự cho ra bộ dữ liệu lên đến 51,8 Gb và được tập hợp thành 2.611.883 contig có chiều dài trung bình là 898 bp và 4.104.872 ORF được xác định. Sử dụng mô hình đại diện HMM để tìm kiếm trình tự tương đồng trong các ORF thu được 3021 trình tự ORF, sau khi lọc thu được 901 ORF thỏa mãn các điều kiện. Trong 901 ORF gồm 275 ORF mã hóa cho ferroxidase, 6 ORF mã hóa cho laccase, 58 ORF mã hóa manganese oxidase, 23 ORF mã hóa bilirubin oxidase; 11 ORF mã hóa L-ascorbate oxidase; 528 ORF mã hóa cho các enzyme khác. Trong 901 ORF mã hóa cho enzyme oxi hóa đa đồng nói trên, có 338 (37,51%) ORF là chứa gen hoàn thiện còn lại 563 ORF là chưa hoàn thiện. Trong số 338 ORF hoàn chỉnh, có 109 ORF có trung tâm hoạt động là các nguyên tử đồng, còn lại 229 ORF không có trung tâm hoạt tính.

Trong nghiên cứu này, chúng tôi ưu tiên chọn 109 ORF hoàn chỉnh có cả đầu 5', 3' và trung tâm xúc tác để tiếp tục nghiên cứu (Bảng 1).

Bảng 1. ORF mã hóa cho enzyme oxi hóa đa đồng từ dữ liệu metagenome vi khuẩn xung quanh khu nấm mục trắng

Enzyme	Mã E.C	ORF hoàn thiện có trung tâm hoạt tính	ORF hoàn thiện	ORF thiếu đầu 3'	ORF thiếu đầu 5'	ORF thiếu 2 đầu	Tổng
Ferroxidase	1.16.3.1	48	77	87	43	68	275
Laccase	1.10.3.2	5	5	1	0	0	6
Manganase oxidase	1.16.3.3	15	9	22	9	18	58
Bilirubin oxidase	1.3.3.5	11	14	6	2	1	23
L-ascorbate oxidase	1.10.3.3	4	4	2	2	3	11
Các dạng oxi hóa khác		26	229	105	101	93	528
Tổng		109	338	223	157	183	901

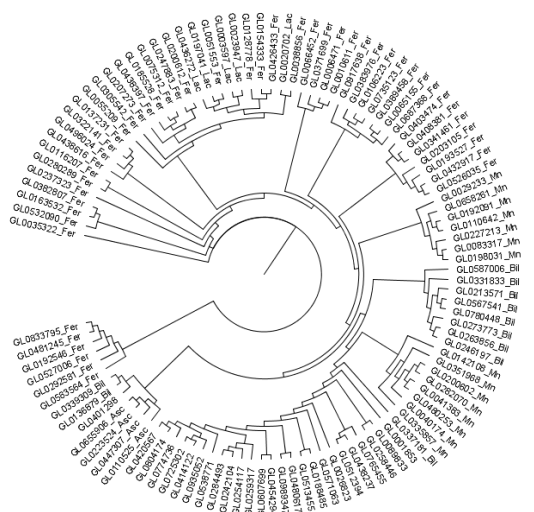
Xây dựng cây phân loại của các ORF

Kết quả dự đoán cây phân loại được thể hiện như trên Hình 1. Từ kết quả cây phát sinh chủng loài cho thấy 109 ORF được phân chia thành các nhóm khác nhau và khá tương ứng với sự phân loại các enzyme oxi hóa đa đồng theo EC. 109 ORF được chia thành 14 nhóm theo thứ tự là: 6 nhóm ferroxidase (fer), 2 nhóm laccase (lac), 2 nhóm manganase oxidase (Mn), 2 nhóm bilirubin oxidase (Bil), 1 nhóm L-ascorbate oxidase (asc), và 1 nhóm các dạng oxi hóa khác.

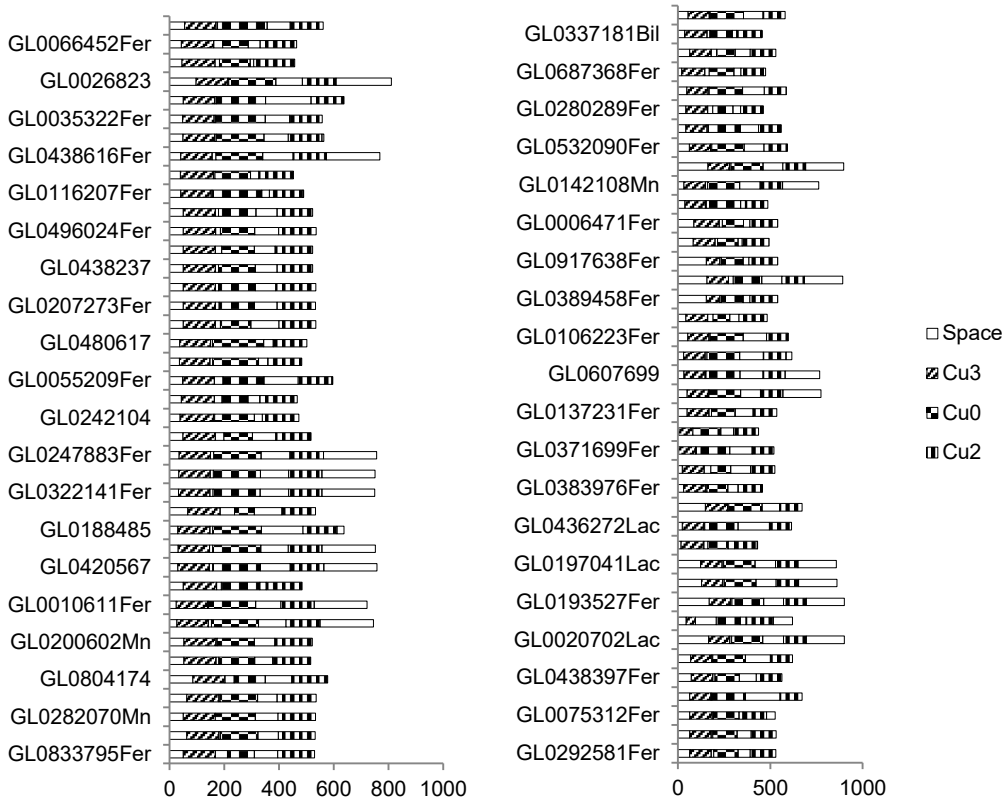
Kết quả này phù hợp với các kết quả phân loại enzyme oxi hóa đa đồng trước đó (Demet *et al.*, 2011; Aleksandra *et al.*, 2019). Theo Demet và đồng tác giả (2011), các loại enzyme oxi hóa đa đồng tìm thấy từ nhiều loài khác nhau như nấm, thực vật và trong nghiên cứu này đã tìm được 6 nhóm enzyme trong họ enzyme oxi hóa đa đồng, thể hiện sự có mặt đa dạng của loại enzyme này trong khu hệ vi khuẩn quanh nấm mục trắng.

Phân tích các vùng chức năng của ORF

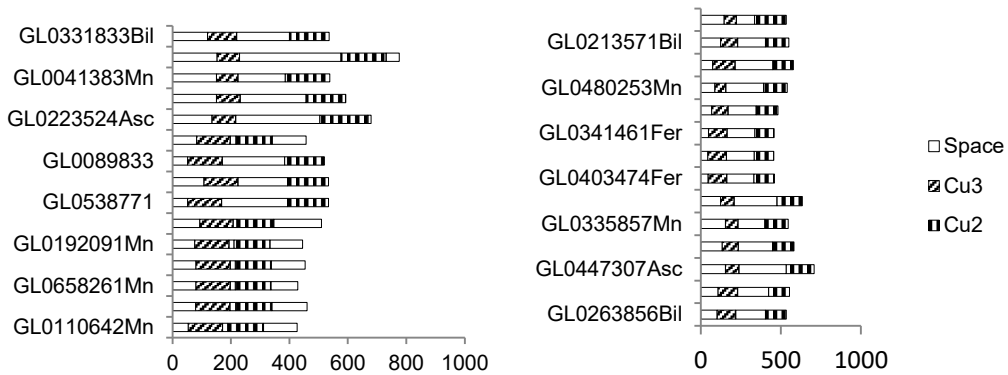
Protein thường bao gồm một hoặc nhiều vùng chức năng gọi là miền (domain). Việc xác định các miền có trong protein cung cấp hiểu biết sâu sắc về chức năng của chúng. Nhằm đánh giá về chức năng của các enzyme chúng tôi tiến hành xác định các miền của 109 ORF hoàn chỉnh. Trong số 109 ORF, có 80 ORF có 3 miền là Cu-oxidase_3, Cu-oxidase và Cu-oxidase_2, có 29 ORF có 2 miền là Cu-oxidase_3, Cu-oxidase_2. Trong số 80 ORF có 3 miền bao gồm 45 ORF ferroxidase, 5 ORF laccase, 4 ORF manganase oxidase, 3 ORF bilirubin oxidase, 23 ORF các dạng oxi hóa khác. Trong 29 ORF có 2 miền gồm 3 ORF ferroxidase, 11 ORF manganase oxidase, 8 ORF bilirubin oxidase, 4 ORF L-ascorbate oxidase, 3 ORF các dạng oxi hóa khác (Hình 2, 3). Kết quả này cũng phù hợp với kết quả nghiên cứu trước đó của Demet và đồng tác giả (2011). Theo đó, phần lớn các enzyme oxi hóa đa đồng có ba vùng chức năng, ngoại trừ một số laccase ở vi khuẩn có hai vùng chức năng và ferroxidase có sáu vùng chức năng.



Hình 1. Cây phân loại 109 ORF mã hóa MCO hoàn thiện có trung tâm hoạt tính



Hình 2. Biểu đồ thể hiện 3 miền chức năng trên 80 ORF mã hóa MCO



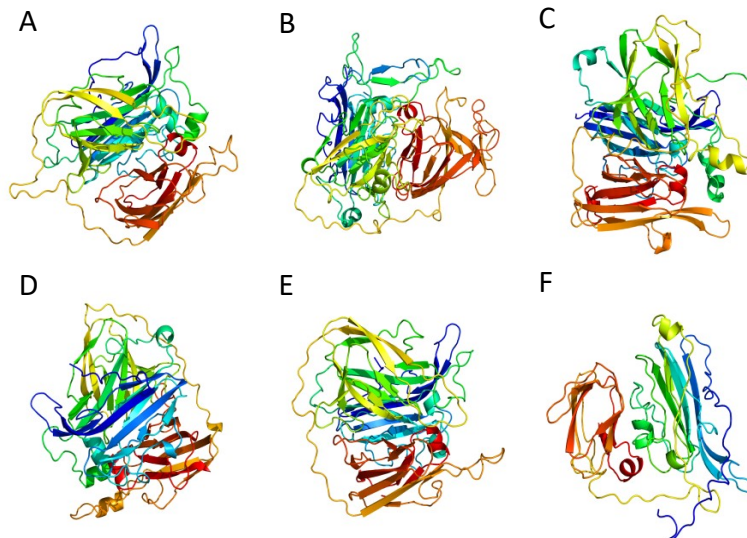
Hình 3. Biểu đồ thể hiện 2 miền chức năng trên 29 ORF mã hóa MCO

So sánh với dữ liệu của NCBI bằng BLAST

Các gen sau khi được dựng cây phân loại để phân chia thành các nhóm, sẽ được lựa chọn các đại diện để dự đoán chức năng và tính mới bằng Blastp. Trình tự axit amin của các ORF được sử dụng để tìm kiếm các trình tự protein tương đồng trên ngân hàng gen NCBI, qua đó có thể xác định chức năng của enzyme, nguồn gốc và tính mới của enzyme. Có 14 ORF được chọn tương ứng với 14 nhóm trên là: 6 ORF mã hóa ferroxidase: GL0438616, GL0128778, GL0066452, GL0383976, GL0403474, GL0833795, 2 ORF mã hóa laccase: GL0436272, GL0023947, 2 ORF mã hóa manganese oxidase: GL0029233, GL0040174, 2 ORF mã hóa bilirubin oxidase: GL0587006, GL0339309, 1 ORF mã hóa L-ascorbate oxidase: GL0110525, 1 ORF mã hóa dạng oxi hóa khác: GL0089833. Kết quả thấy các ORF có độ bao phủ từ 93 - 100% và có 9 ORF có độ tương đồng cao với các protein đã công bố trên ngân hàng NCBI (từ 85% trở lên), còn lại 5 ORF có độ tương đồng từ 55 - 74% (Bảng 2).

Bảng 2. Kết quả BLASTp đại diện 14 ORF hoàn thiện có trung tâm hoạt tính mã hóa MCO

STT	Mã gen	Enzyme	Số axit amin	Độ bao phủ (%)	Độ tương đồng (%)
1	GL0438616	Ferroxidase	592	100	93
2	GL0128778	Ferroxidase	745	100	98
3	GL0066452	Ferroxidase	490	99	63
4	GL0383976	Ferroxidase	430	99	55
5	GL0403474	Ferroxidase	460	100	98
6	GL0833795	Ferroxidase	520	100	74
7	GL0436272	Laccase	762	100	98
8	GL0023947	Laccase	758	100	92
9	GL0029233	Manganese oxidase	510	100	89
9	GL0040174	Manganese oxidase	535	97	68
11	GL0587006	Bilirubin oxidase	457	99	88
12	GL0339309	Bilirubin oxidase	527	94	87
13	GL0110525	L-ascorbate oxidase	776	93	60
14	GL0089833	Dạng oxi hóa khác	923	100	87



Hình 4. Dự đoán cấu trúc bậc hai của enzyme oxi hóa đa đồng bằng Phyre2
(A: GL0587006; B: GL0110525; GL0023947; D: GL0001853; E:GL0383976; F: GL0029233)

Mô phỏng cấu trúc không gian và vị trí gắn cơ chất của enzyme

Mỗi nhóm enzyme trong họ enzyme oxi hóa đa đồng được lựa chọn một ORF đại diện để dự đoán cấu trúc bậc hai của các ORF với sự sắp xếp của cấu trúc xoắn α và gấp nếp β khác nhau. Kết quả cho thấy cấu trúc bậc hai của các ORF mã hóa MCO từ dữ liệu DNA metagenome tương đồng với các protein có hoạt tính oxi hóa đa đồng đã được công bố với độ tin cậy 100% và độ bao phủ khá cao từ 58 – 97%. ORF GL0587006 có 2% xoắn α , 46% gấp nếp β , 4% xoắn TM và 13% không xác định; GL0110525 có 5% xoắn α , 27% gấp nếp β , 2% xoắn TM và 33% không xác định; GL0023947: có 6% xoắn α , 42% gấp nếp β , 2% xoắn TM và 18% không xác định; GL0001853: có 5% xoắn α , 40% gấp nếp β , 3% xoắn TM và 26% không xác định; GL0383976: có 2% xoắn α , 47% gấp nếp β , 4% xoắn TM và 13% không xác định; GL0029233: có 11% xoắn α , 28% gấp nếp β và 35% không xác định. Kết quả này giúp khẳng định các ORF được lựa chọn có chức năng mã hóa các enzyme oxi hóa đa đồng như dự đoán ban đầu và có thể được lựa chọn để tiến hành các thí nghiệm nghiên cứu hoạt tính sinh học của chúng. Kết quả cấu trúc không gian ba chiều của 6 nhóm thuộc họ enzyme oxi hóa đa đồng được trình bày trên Hình 4.

KẾT LUẬN

Chúng tôi đã khai thác được 901 ORF mã hóa cho enzyme oxi hóa đa đồng từ dữ liệu DNA metagenome của vi sinh vật quanh khu nấm mục trắng, trong đó có 338 ORF chứa gen hoàn chỉnh và 109 ORF chứa gen hoàn chỉnh có trung tâm hoạt tính được phân loại thành 6 nhóm mã hóa các enzyme khác nhau trong họ enzyme oxi hóa đa

đồng: ferroxidase, laccase, manganese oxidase, bilirubin oxidase, L-ascorbate oxidase và các dạng oxi hóa khác. Trong đó, có 80 ORF có 3 vùng chức năng, 29 ORF có hai vùng chức năng. Phân tích cấu trúc cho thấy các enzyme trong nhóm này có độ tương đồng cao với các protein đã được công bố.

Lời cảm ơn: Công trình được thực hiện bằng nguồn kinh phí của đề tài Nghị định thư Việt Nam - Cộng hòa Liên bang Đức giai đoạn 2018 - 2021 mã số NDT.50.GER/18 và sử dụng trang thiết bị của Phòng Thí nghiệm trọng điểm Công nghệ gen, Viện Công nghệ sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam.

TÀI LIỆU THAM KHẢO

- Aleksandra GB, Anna J, Jerzy D (2019), Characteristics and use of multicopper oxidase enzymes. *Post Mikrobiol*, 58: 7-18.
- Copete LS, Chanagá X, Barriuso J, López-Lucendo MF, Martínez MJ, Camarero S (2015). Identification and characterization of laccase-type multicopper oxidases involved in dye-decolorization by the fungus *Leptosphaerulina* sp. *BMC Biotechnol* 15:74.
- Demet S, Florian W, Lei W, Rolf DS, Jürgen P (2011). The Laccase Engineering Database: a classification and analysis system for laccases and related multicopper oxidases. *Database (Oxford)*: bar006.
- Eddy, S. R. (1998). Profile hidden Markov models. *Bioinform* 14(9): 755-763. doi: 10.1093/bioinformatics/14.9.755
- Eddy, S. R. (2011). Accelerated Profile HMM Searches. *PLoS Comput Biol* 7(10): e1002195. doi: 10.1371/journal.pcbi.1002195
- Haghighi MS, Hossein GA, Tabatabaei M, Salehi JG, Najafi GH, Gholami M (2013). Lignocellulosic biomass to bioethanol, a comprehensive review with a focus on pretreatment. *Renewable and Sustainable Energy Rev* 27: 77-93.
- Hoegger PJ, Kilaru S, James TY, Thacker JR, Kues U (2006). Phylogenetic comparison and classification of laccase and related multicopper oxidase protein sequences. *FEBS J* 273: 2308-2326.
- Zheng W, Zhao T, Feng W, et al (2014). Purification, characterization and immunomodulating activity of a polysaccharide from flowers of *Abelmoschus esculentus*. *Carbohydr Polym* 106(1): 335-342.
- http://www.cs.hku.hk/~alse/idba_ud
- <http://pfam.xfam.org>
- <https://blast.ncbi.nlm.nih.gov/Blast>
- <http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>

MINING GENES ENCODING MULTICOPPER OXIDASE ENZYME FROM METAGENOME DATA OF THE BACTERIA AROUND WHITE-ROT FUNGI ECOSYSTEM

Nguyen Thi Binh^{2,3}, Dao Trong Khoa^{2,3}, Le Thi Thu Hong^{1,3}, Truong Nam Hai^{1,3*}

¹ Institute of Biotechnology, Vietnam Academy of Science and Technology

² Hanoi Metropolitan University

³ Graduate University of Science and Technology, Vietnam Academy of Science and Technology

SUMMARY

Multicopper oxidase enzymes have the ability to decompose a variety of substrates, so they are used in many practical applications. Enzymes belonging to this group are very diverse in different ecosystems. DNA metagenome data of microorganisms around the white-rot fungi were sequenced, obtaining 51.8 Gb data set. Using the predicted MGA software, 4,104,872 ORF, using HMM representation model to search for 901 ORF coding for multi-copper oxidase enzyme. To exploit those genes, we use the MegaX software to predict the classification plants, the genes are classified into groups: laccase, ferroxidase, ascorbate oxidase, bilirubin oxidase, manganese oxidase and some other types of enzymes. These gene groups are Blastp to predict function and evaluate newness. The results show the coverage from 93-100%, the identity from 55-98%. Comparison with Pfam data to identify functional regions showed that these enzymes had 3 regions (80 ORF) and 2 functional regions (29 ORF). Phyre2 was used to determine the quadratic structure of genes belonging to different groups, resulting in the secondary structure of multi-oxidizing enzymes with published proteins with 100% confidence level. The coverage is quite high from 58 - 97%.

Keywords: Multi-copper oxidizing enzyme, lignocellulose, metagenome, metagenomic, white-rot fungi, bioinformatics.

* Author for correspondence: E-mail: tnhai@ibt.ac.vn