

XÂY DỰNG PROBE ĐỂ KHAI THÁC VÀ CHỌN GEN MÃ HOÁ BETA-GLUCOSIDASE GH3 TỪ DỮ LIỆU GIẢI TRÌNH TỰ DNA METAGENOME VI KHUẨN DẠ CỎ DÊ

NGUYỄN KHÁNH HOÀNG VIỆT ^(1, 2, 3), ĐỖ THỊ HUYỀN ^(1, 3), NGUYỄN HỮU ĐỨC ⁽⁴⁾,
VŨ THỊ LY ⁽³⁾, TRƯƠNG NAM HẢI ^(1, 3)

1. MỞ ĐẦU

Beta-glucosidase còn có tên gọi khác là beta-D-glucoside glucohydrolase (EC.3.2.1.21), xúc tác cho quá trình thủy phân các liên kết beta-glucoside ở các gốc alkyl, aryl và beta-glucoside như disaccharide và oligosaccharide chuỗi ngắn tạo ra đường glucose [3]. Beta-glucosidase rất phổ biến trong tự nhiên và có thể được tìm thấy trong vi khuẩn, nấm, thực vật, và động vật. Enzyme này làm tăng hiệu quả thủy phân chất thải rất khó phân huỷ trong tự nhiên có nguồn gốc từ thực vật như cellulose trong điều kiện hoạt động nhất định. Ngoài ra, beta-glucosidase còn được ứng dụng trong y học như làm thuốc hỗ trợ tiêu hóa, hoặc sử dụng trong các bộ kit xét nghiệm nồng độ đường [3].

Ứng dụng kỹ thuật metagenomics theo hướng phân tích dữ liệu thu được từ việc giải toàn bộ trình tự metagenome của hệ vi sinh vật dạ cỏ dê, đã khai thác được 164.644 khung đọc mở (ORF), trong đó ước đoán gồm rất nhiều trình tự mã hóa cho beta-glucosidase được chú giải chức năng dựa trên dữ liệu KEGG, CAZy [6]. Phương pháp dự đoán gen từ dữ liệu khổng lồ DNA metagenome của vi sinh vật trong dạ cỏ dê bước đầu đã được thực hiện bằng cách sử dụng các phần mềm tin sinh học, dựa trên số liệu các trình tự tương đồng với các loài đã được công bố trong ngân hàng gen [1, 6]. Trong thực tế probe đã được sử dụng trong rất nhiều các nghiên cứu như nhận diện các bản sao hoặc sản phẩm RNA của gen, các sinh vật có quan hệ gần gũi với đối tượng nghiên cứu nhằm tìm kiếm gen chức năng được bảo tồn qua tiến hoá, tìm kiếm trình tự tron vẹn của gen mã hoá cho protein trong genome,...[6]. Trong khuôn khổ bài báo này, nhóm tác giả trình bày phương pháp để xây dựng probe đặc hiệu enzyme beta-glucosidase họ GH3 và sử dụng probe cho khai thác, lựa chọn nhanh một số gen mã hóa beta-glucosidase thuộc họ GH3 từ nguồn dữ liệu rất lớn này để đưa vào biểu hiện hiệu quả trong thực nghiệm.

2. VẬT LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Vật liệu nghiên cứu

Dữ liệu metagenome DNA gồm 164.644 khung đọc mở (ORF) của vi sinh vật trong ruột dê, được giải trình tự bằng hệ thống giải trình tự HiSeq Illumina (BGI, Hồng Kông) và được chú giải chức năng dựa trên dữ liệu KEGG, CAZy [1, 3].

2.2. Phương pháp nghiên cứu

2.2.1. Xác định các họ GH có chứa beta-glucosidase theo CAZy

CAZy (Carbohydrate-Active enzymes; <http://www.cazy.org>) cung cấp số liệu trực tuyến và cập nhật liên tục các dữ liệu của GenBank về chuỗi thông tin của gần 340.000 enzyme tham gia chuyển hóa carbohydrate [4, 7] được phân vào 135 họ GH với các đặc điểm nhận biết khác nhau [2]. Dựa trên dữ liệu này, các trình tự beta-glucosidase đã được nghiên cứu tính chất thuộc họ GH3 được thu thập, tổng hợp thành bảng trong đó tích hợp trình tự với đặc điểm enzyme như khả năng chịu nhiệt, chịu kiềm/acid, cấu trúc không gian, trung tâm hoạt động của beta-glucosidase.

2.2.2. Xây dựng probe và tìm giá trị tham chiếu

Các trình tự thu thập được ở trên được so sánh mức độ tương đồng bằng ClustalW - PBIL. Kết quả so sánh sẽ cho ra một trình tự được cho là bảo tồn cao nhất (Ký hiệu *Prim.cons*), trong đó có các trình tự được đánh dấu về mức độ bảo tồn đặc thù cho họ enzyme. Dựa trên kết quả của *Prim.cons*, các vị trí giống nhau hoặc tương đối giống nhau sẽ ưu tiên lựa chọn để làm probe và các trình tự trùng được loại bỏ. Probe này sẽ được so sánh lại với các trình tự đã dùng để xây dựng nên probe bằng BLASTP. Giá trị tham chiếu được xác định là giá trị E, điểm tối đa, độ bao phủ và mức độ tương đồng thấp nhất mà probe bắt được với trình tự.

2.2.3. Lựa chọn nhanh và ước đoán cấu trúc bậc ba của các trình tự mã hóa beta-glucosidase GH3 từ dữ liệu DNA metagenome

Sử dụng BLASTP để so sánh probe với trình tự amino acid của các ORF thuộc metagenome của vi sinh vật trong dạ cỏ dê. Các trình tự có giá trị điểm tối đa, mức độ bao phủ, tương đồng cao hơn giá trị tham chiếu sẽ được lựa chọn. Sau đó các trình tự này được so sánh lại với kết quả chú giải gen dựa trên dữ liệu KEGG và CAZy. Cấu trúc bậc ba của phân tử được ước đoán dựa trên trình tự và với các protein khung đã được nghiên cứu về cấu trúc bằng phần mềm Swiss Prot.

3. KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN

3.1. Lựa chọn các trình tự beta-glucosidase GH3

Trên cơ sở số liệu của CAZy, beta-glucosidase được phân vào 8 họ GH gồm GH1, GH2, GH3, GH5, GH9, GH30, GH39, GH116 trong đó chỉ họ GH9, GH39 có cơ chế xúc tác đảo ngược [5]. Enzyme thuộc họ GH3 là enzyme phong phú, phổ biến và có ứng dụng nhiều nhất. Enzyme này có cấu trúc không gian chưa xác định, gốc cho điện tử là Asp và gốc nhận điện tử là Glu [5]. Các trình tự beta-glucosidase thuộc họ GH3 đã được nghiên cứu tính chất về pH và nhiệt độ tối ưu được thu thập, tổng hợp trong bảng 1.

Bảng 1. Tổng hợp dữ liệu về các trình tự beta-glucosidase GH3 của vi khuẩn được thu thập từ CAZy

Mã số trong GENBANK	Vi khuẩn	Số aa	pH tối ưu	Nhiệt độ tối ưu (°C)
AFS34657.1	<i>Mucilaginibacter</i> sp. QM49	781	8,0	30
AAA22082.1	<i>Agrobacterium tumefaciens</i>	818	7,2-7,4	60
AAA86753.2	<i>Prevotella bryantii</i> B14	885	7,5	-
AEX88466.1	<i>Microbacterium steraromaticum</i> KACC16318	831	7,0	37
AAB66561.1	<i>Elizabethkingia meningoseptica</i>	726	4,2-5	50
ADK91094.1	<i>Serratia</i> sp. TN49	790	-	20
AAC38196.1	<i>Cellulomonas biazotea</i>	828	7,0	30
BAA03152.1	<i>Cellvibrio gilvus</i> ATCC13127	752	6,2-6,4	35
BAJ09393.1	<i>Paenibacillus relictisesami</i>	753	6,5	37
AGS08637.1	<i>Novosphingobium</i> sp. GX9	744	-	-
CAA33665.1	<i>Ruminiclostridium thermocellum</i> DSM 1237	754	6,0	-
AGU13704.1	<i>Bacteroides uniformis</i> ZL1	755	6,0	30

Chú thích: (-) Chưa xác định.

3.2. Xây dựng probe và giá trị tham chiếu

Kết quả so sánh 12 trình tự amino acid thu thập được bằng phần mềm ClustalW - PBIL, dựa trên trình tự tương đồng đã xây dựng được một probe gồm 330 amino acid trong có 26 amino acid hoàn toàn giống nhau ở các trình tự, 37 vị trí giống nhau ở đa số các trình tự và có 21 vị trí giống nhau ở một số trình tự, còn lại là khác nhau. Kết quả này cho thấy trình tự mã hóa enzyme beta-glucosidase thuộc họ GH3 của vi khuẩn có sự bảo tồn tương đối cao (hình 1).

DIDEXIAAMTLEEKAXLLTGATTAPIERLGI¹PSXXMTDGPXGVRCF²PSA
TGLASS³WDPXLVERVGALGXEARAYGVNVLLGPGVNI⁴HSPLX⁵GRNF⁶EYY
SE⁷DP⁸LLSGEXAAAYXX⁹GV¹⁰Q¹¹SXGVGASL¹²KH¹³FAANGNQ¹⁴ET¹⁵XR¹⁶XTXDSXVDER
ALREIY¹⁷LXGFEIA¹⁸VKK¹⁹ARPWTV²⁰MSS²¹YN²²R²³LNG²⁴VYXSEN²⁵XWLLTXV²⁶L²⁷XXE²⁸WG
FXG²⁹FVMS³⁰D³¹WG³²XVXDRVAANA³³GL³⁴DLXMPGEPGQALVAAVKSGK³⁵LXEEXLD
XXVRRILXLIXRGAFKKXGAXPHALARXAAAEGM³⁶VL³⁷LKNVAV³⁸IGEXAX
DPRNXQ³⁹GGGSSQVNPTYXVSXXEGLKAAGGDA

Hình 1. Trình tự probe cho beta-glucosidase thuộc họ GH3

Chú thích: X: Gốc amino acid không xác định; chữ tô màu là gốc bảo tồn ở tất cả các trình tự (G); chữ đậm là các gốc bảo thủ ở đa số trình tự (G); chữ nghiêng là các gốc bảo thủ ở một số trình tự (G).

Để tìm giá trị tham chiếu cho việc sử dụng probe trong khai thác gen, nhóm tác giả tiến hành so sánh độ tương đồng giữa probe với từng trình tự đã sử dụng để xây dựng nên chúng. Kết quả được tổng hợp (bảng 2) cho thấy với điểm tối đa trên 122, độ bao phủ và độ tương đồng tối thiểu lần lượt là 79% và 38%, probe có thể nhận biết là beta-glucosidase GH3.

Bảng 2. So sánh tương đồng giữa probe với 12 trình tự thuộc GH3

Trình tự	Mã trình tự	Điểm tối đa	Tổng điểm	Độ bao phủ (%)	Giá trị E	Độ tương đồng (%)
11	CAA33665.1	244	244	98	2e-78	53
9	BAJ09393.1	244	244	95	2e-78	54
1	AFS34657.1	219	239	94	9e-69	50
12	AGU13704.1	207	207	98	1e-64	49
4	AEX88466.1	207	207	97	2e-64	49
2	AAA22082.1	206	225	99	5e-64	49
3	AAA86753.2	199	199	98	2e-61	43
6	ADK91094.1	184	202	85	9e-56	46
7	AAC38196.1	162	200	98	7e-48	43
5	AAB66561.1	140	140	79*	4e-40	38*
8	BAA03152.1	137	156	86	3e-39	40
10	AGS08637.1	122*	122	85	4e-34	39

Chú thích: (*) Giá trị tham chiếu.

3.3. Lựa chọn và đánh giá các trình tự từ dữ liệu DNA metagenome của vi khuẩn trong dạ cỏ dê

Kết quả BlastP probe được xây dựng với 164.644 ORF từ dữ liệu DNA metagenome của vi khuẩn trong dạ cỏ dê, dựa trên giá trị tham chiếu, probe đã nhận 59/302 ORFs mã hóa beta-glucosidase GH3 đúng như dự đoán bởi dữ liệu KEGG và CAZy (bảng 3).

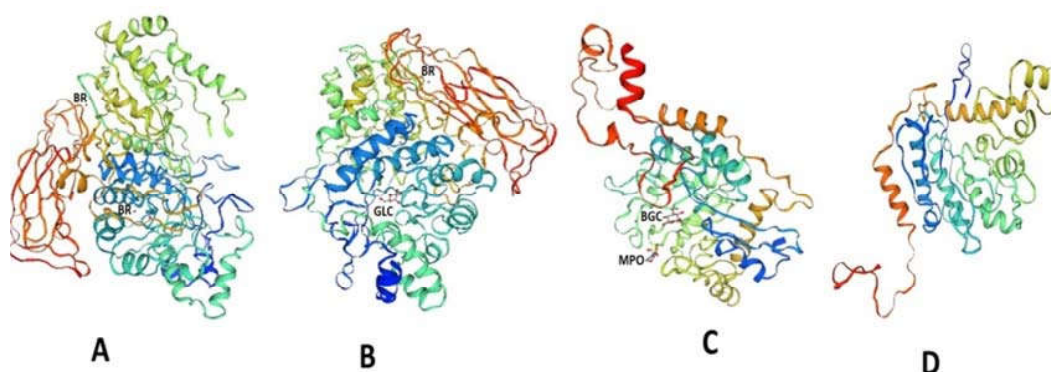
Bảng 3. Trình tự từ dữ liệu DNA metagenome của vi khuẩn dạ cỏ dê được khai thác bằng probe đặc trưng cho beta-glucosidase GH3

Mã gen	Điểm tối đa	Độ bao phủ (%)	Giá trị E	Độ tương đồng (%)	Mã gen	Điểm tối đa	Độ bao phủ (%)	Giá trị E	Độ tương đồng (%)
4506	205	98	4.00E-61	46	3894	190	94	1.00E-55	45
4363	205	98	5.00E-61	46	3914	190	86	1.00E-55	48
4221	205	98	4.00E-61	45	4178	189	98	2.00E-55	44
4563	204	98	8.00E-61	44	4103	189	94	2.00E-55	45
4566	202	98	4.00E-60	44	4378	189	94	2.00E-55	45
4750	202	90	4.00E-60	47	20457	188	95	6.00E-57	43
4252	201	98	9.00E-60	45	4125	187	94	1.00E-54	45
4509	199	98	4.00E-59	44	3979	187	89	2.00E-54	44
4580	198	98	9.00E-59	45	15453	184	90	4.00E-55	45
4222	198	98	2.00E-58	45	6729	184	90	4.00E-54	44
4456	197	98	3.00E-58	43	34030	184	89	3.00E-56	43
8686	196	98	1.00E-58	43	4807	184	84	2.00E-53	48
4081	196	94	8.00E-58	46	9590	179	93	2.00E-52	46
4148	196	94	1.00E-57	46	21627	172	96	9.00E-51	42
27003	195	98	6.00E-60	44	3999	145	89	1.00E-39	41
4273	195	98	2.00E-57	43	3920	145	83	2.00E-39	40
4240	194	98	4.00E-57	44	4317	143	89	7.00E-39	40
4532	194	98	3.00E-57	43	15912	142	89	3.00E-39	40
4309	194	98	5.00E-57	42	4279	142	89	2.00E-38	39
3916	194	98	6.00E-57	42	4885	142	87	9.00E-39	41
4151	194	94	5.00E-57	46	8510	142	87	1.00E-38	41
3960	193	97	8.00E-57	44	4005	141	89	3.00E-38	39
4170	192	94	3.00E-56	46	4040	140	89	7.00E-38	39
4133	192	94	2.00E-56	45	29414	140	87	2.00E-39	41
4109	192	94	3.00E-56	44	27578	138	79	2.00E-38	43
4110	192	84	3.00E-56	49	19474	137	88	7.00E-38	40
9646	191	94	4.00E-57	46	43583	137	83	7.00E-39	42
5940	191	90	3.00E-56	46	5059	134	79	6.00E-36	42
45892	191	82	2.00E-59	49	17700	131	79	3.00E-35	42
5436	190	98	7.00E-56	43					

Để khẳng định rõ hơn các trình tự khai thác là beta-glucosidase, cấu trúc bậc ba của 5 trình tự được lựa chọn ngẫu nhiên được ước đoán cấu trúc bằng Swiss Prot. Kết quả so sánh các trình tự này đều tương đồng với các khuôn có hoạt tính beta-glucosidase với cấu trúc bậc 3 khác nhau từ 40,92÷49,93% và độ bao phủ đều trên 85% (bảng 4, hình 2). Như vậy, kết quả thu được cho thấy probe đã xây dựng được là đặc hiệu cho beta-glucosidase GH3.

Bảng 4. Ước đoán cấu trúc bậc ba của các trình tự bằng Swiss Prot

Mã gen	Khuôn	Hoạt tính	Độ bao phủ (%)	Độ tương đồng (%)	Phương pháp	Cấu trúc	Phối tử
4221	2x40.1.A	Beta-glucosidase	90	49,22	X-ray, 2.3 Å	homo-dimer	8 x BR
4363	2x42.1.A	Beta-glucosidase	91	49,93	X-ray, 2.1 Å	monomer	1xGLC, 4xBR
45892	3abz.1.A	Beta-glucosidase I	93	40,92	X-ray, 2.1 Å	homo-tetramer	Không
29414	3abz.1.A	Beta-glucosidase I	98	53,21	X-ray, 2.1 Å	homo-tetramer	Không
27578	4i3g.1.A	Beta-glucosidase	85	41,55	X-ray, 1.4 Å	monomer	1xMPO, 1xBGC



Hình 2. Cấu trúc bậc ba của beta-glucosidase GH3 tương đồng với các trình tự được khai thác bằng probe sử dụng Swiss prot

Chú thích: A: Khuôn (template) 2x40.1.A; B: 2x42.1.A; C: 4i3g.1.A; D: 3abz.1.A thuộc họ GH3; BR: Bromide; GLC: (2S,3R,4S,5S,6R)-6-(hydroxymethyl)oxane-2,3,4,5-tetrol; BGC: (2R,3R,4S,5S,6R)-6-(hydroxymethyl)oxane-2,3,4,5-tetrol; MPO: 3-morpholin-4-ylpropane-1-sulfonic acid.

4. KẾT LUẬN

Probe đặc hiệu beta-glucosidase GH3 đã được thiết kế thành công có 330 amino acid từ 12 trình tự amino acid của beta-glucosidase GH3 của vi khuẩn. Sử dụng probe, đã lựa chọn được 59 trình tự beta-glucosidase GH3 từ dữ liệu giải trình tự DNA metagenome của vi sinh vật trong dạ cỏ dê. Các trình tự này đều được ước đoán là beta-glucosidase dựa trên dữ liệu KEGG, CAZy và SwissProt. Việc lựa chọn các trình tự acid amin của enzyme này đã được nghiên cứu chi tiết về hoạt tính từ vi khuẩn để xây dựng 01 probe thuộc họ GH3 sẽ tiết kiệm được thời gian và hỗ trợ hiệu quả cho việc lựa chọn các gen mã hóa cho beta-glucosidase từ bộ dữ liệu DNA metagenome rất lớn thu được thông qua giải trình tự đa hệ gen mẫu dạ cỏ dê tại một số vùng sinh thái của Việt Nam.

Lời cảm ơn: Công trình được thực hiện bằng nguồn kinh phí của Đề tài độc lập: “Nghiên cứu metagenome của một số hệ sinh thái mini tiềm năng nhằm khai thác các gen mới mã hóa hệ enzyme chuyển hóa hiệu quả lignocelluloses”, mã số ĐTĐLCN.15/14, trang thiết bị của Phòng Thí nghiệm trọng điểm Công nghệ gen.

TÀI LIỆU THAM KHẢO

1. Nguyễn Khánh Hoàng Việt, Lê Tùng Lâm, Phùng Thị Lan, Đỗ Thị Huyền, Trương Nam Hải, *Nghiên cứu biểu hiện GPECS1 mã hóa pectinesterase khai thác từ dữ liệu giải trình tự DNA metagenome vi khuẩn dạ cỏ dê trong tế bào Escherichia coli, sử dụng vector pET22b(+)*, Tạp chí Y học Việt Nam, 2017, **468**:197-203.
2. Cantarel B. L., Coutinho P. M., Rancurel C., Bernard T., Lombard V., Henrissat B., *The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics*, Nucleic Acids Research, 2009, **37**:233-238.
3. Do T. H., Le N. G., Dao T. K., Nguyen T. M. P., Le T. L., Luu H. L., Nguyen K. H. V., Nguyen V. L., Le L. A., Phung T. N., Straalen N. M., Roelofs D., Truong N. H., *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, accepted 2017.
4. Henrissat B., *A classification of glycosyl hydrolases based on amino acid sequence similarities*, Biochemical Journal, 1991, **280**(2):309-316.
5. Lombard V., Golaconda R. H., Drula E., Coutinho P. M., Henrissat B., *The Carbohydrate-Active EnZymes database (CAZy) in 2013*, Nucleic Acids Research, 2014, **42**:490-495.
6. Mitsuhashi M., Cooper A., Ogura M., Shinagawa T., Yano K., Hosokawa T., *Oligonucleotide probe design: a new approach*, Nature, 1994, **367**(6465):759-761.
7. Shallom D., Shoham Y., *Microbial hemicellulases*, Current Opinion in Microbiology, 2003, **6**(3):219-228.

SUMMARY

PROBE DESIGN FOR EXPLOITING GENES ENCODING BETA- GLUCOSIDASE GH3 FROM DNA METAGENOME OF BACTERIA IN GOAT RUMEN

Beta-glucosidase enzyme belongs to family GH3 is widely used in food, medicine and pharmaceutical industries. In order to quickly search for these enzyme coding sequences from DNA metagenomic data, in this study, bacterial-derived enzyme-coding sequences which was investigated empirically on the CAZy database was collected. By analogy, we have developed that a specific probe for beta-glucosidase GH3 had a length of 330 amino acids, which contains 26 conserved residues in all sequences, 37 residues similar in almost sequences, and 21 residues conserved in many sequences and homologous with all the reference sequences with the lowest coverage and identity of 79% and 38% respectively and max score of 122. Probe was used and extracted 59 coded sequences of beta-glucosidase GH3 from the metagenome DNA sequences of goat rumen bacteria. The most ORFs were annotated as beta-glucosidase GH3 by KEGG and CAZy. The sequences are estimated to have a tertiary structure similar to beta-glucosidase.

Keywords: BLASTP, beta-glucosidase GH3, DNA metagenome, ClustalW, probe.

Nhận bài ngày 25 tháng 01 năm 2018

Phản biện xong ngày 09 tháng 4 năm 2018

Hoàn thiện ngày 20 tháng 4 năm 2018

⁽¹⁾ *Viện Công nghệ sinh học, Viện Hàn lâm KH&CN Việt Nam*

⁽²⁾ *Viện Công nghệ mới, Viện Khoa học và Công nghệ quân sự*

⁽³⁾ *Học viện Khoa học và Công nghệ, Viện Hàn lâm KH&CN Việt Nam*

⁽⁴⁾ *Học viện Nông nghiệp Việt Nam*