

## KHAI THÁC GEN MÃ HÓA ENDO-1,4-BETA-XYLANASE TỪ DỮ LIỆU DNA METAGENOME VI KHUẨN TRONG DẠ CỎ ĐÉ BẰNG MẪU DÒ

Đào Trọng Khoa<sup>1,2</sup>, Đỗ Thị Huyền<sup>1,2</sup>, Trương Nam Hải<sup>1,2,✉</sup>

<sup>1</sup>Viện Công nghệ sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

<sup>2</sup>Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

✉Người chịu trách nhiệm liên lạc. E-mail: tnhai@ibt.ac.vn

Ngày nhận bài: 27.5.2020

Ngày nhận đăng: 30.7.2020

### TÓM TẮT

Endo-1,4-beta-xylanase được phân loại vào 9 họ glycoside hydrolase là GH5, 8, 10, 11, 30, 43, 51, 98, 141 dựa trên cơ sở dữ liệu CAZy. Trình tự mẫu dò (probe) đại diện cho enzyme xylanase được xây dựng từ các trình tự đã được công bố mang hoạt tính phân hủy xylan từ các nghiên cứu thực nghiệm. Cụ thể, có một trình tự thuộc họ GH5, 6 trình tự thuộc họ GH8 và 5 trình tự thuộc họ GH30 đã được thu thập, từ đó xây dựng nên các mẫu dò đặc hiệu cho endo-xylanase họ GH8 và GH30. Mẫu dò được xây dựng từ các trình tự thuộc họ GH8 và GH30 có độ dài tương ứng là 351 và 425 amino acid. Giá trị tham chiếu cho mẫu dò của họ GH8 được xác định là các trình tự có điểm tối đa lớn hơn 168, độ bao phủ thấp nhất là 84%, độ tương đồng thấp nhất là 36%; giá trị tương ứng với mẫu dò của họ GH30 là điểm tối đa lớn hơn 316, độ bao phủ lớn hơn 98%, độ tương đồng lớn hơn 41%. Sử dụng các mẫu dò bảo thủ cho các họ GH có hoạt tính endo-1,4-beta-xylanase đã được xây dựng cùng với mẫu dò của hai họ GH10 và GH11 đã được công bố, chúng tôi đã khai thác được 41 trình tự mã hóa xylanase từ dữ liệu giải trình tự DNA metagenome của vi sinh vật trong dạ cỏ đê bản địa Việt Nam. Trong số 41 trình tự khai thác được, 19 trình tự trùng với kết quả chú giải của công ty BGI dựa trên dữ liệu KEGG, ngược lại có 16 trình tự không được chú giải dựa trên dữ liệu KEGG. 28 trong số 41 trình tự khai thác được là khung đọc mở hoàn chỉnh, có cấu trúc bậc ba ước đoán tương đồng cao với các cấu trúc của xylanase đã được công bố.

**Từ khóa:** endo-1,4-beta-xylanase, DNA metagenome, glycoside hydrolase (GH), mẫu dò, vi khuẩn dạ cỏ đê

### MỞ ĐẦU

Xylan là hemicellulose chủ yếu ở trong rơm rạ và các loại gỗ cứng của thực vật hạt kín (chiếm 15-30%) và chỉ chiếm tỷ lệ nhỏ trong gỗ mềm của thực vật hạt trần (chiếm 7-12%). Mạch chính của xylan được cấu thành từ các đơn phân là các gốc đường 5 carbon D-xylose ( $\beta$ -D-xylopyranose) liên kết với nhau bằng liên kết  $\beta$ -1,4-glycoside (Souza 2013). Ngoài ra mạch nhánh của xylan có thể có các nhóm bên khác như D-galactose, L-arabinose, glucuronic acid, acetyl-, feruloyl-, p-coumaroyl (de Vries *et al.*, 2001). Do tính phức tạp trong cấu trúc này, việc

thủy phân hoàn toàn phân tử hemicellulose cần tới một tổ hợp nhiều loại enzyme thủy phân khác nhau. Các enzyme chính tham gia thủy phân có thể kể đến là endo-1,4- $\beta$ -xylanase,  $\beta$ -xylosidase,  $\alpha$ -L-arabinofuranosidase,  $\alpha$ -glucuronidase, acetyl xylan esterase, phenolic acid esterase.

Trong số những enzyme thủy phân xylan, xylanase (endo-1,4-beta-xylanase) đóng vai trò quan trọng bậc nhất vì xúc tác phân cắt mạch chính của phân tử xylan, tạo thành các xylooligosaccharide kích thước ngắn (Verma *et al.*, 2012), từ đó tạo điều kiện cho các enzyme khác tiếp tục thủy phân để giải phóng các đơn

phân. Vì vậy, xylanase có tiềm năng rất lớn để ứng dụng trong rất nhiều ngành công nghiệp như công nghiệp thực phẩm (Polizeli *et al.*, 2005), công nghiệp giấy (Viikari *et al.*, 1994) và trong thức ăn chăn nuôi (Twomey *et al.*, 2003).

Hệ tiêu hóa của động vật nhai lại (trâu, bò, dê...) là nơi khu trú lý tưởng của vô số vi khuẩn có khả năng thủy phân lignocellulose mạnh, vì vậy đây là đối tượng tiềm năng cho việc nghiên cứu khai thác các gen mới mã hóa cho các enzyme tham gia chuyển hóa lignocellulose (An *et al.*, 2005; Kittelmann *et al.*, 2011). Sử dụng kỹ thuật giải trình tự thế hệ mới, một hướng trong ứng dụng công cụ Metagenomics nhằm nghiên cứu trực tiếp dữ liệu DNA của khu hệ vi sinh vật mà không thông qua nuôi cấy, dữ liệu DNA metagenome của hệ vi sinh vật khu trú trong dạ cỏ của dê bản địa Việt Nam đã được khai thác với 164.644 khung đọc mở (ORF) (Do *et al.* 2018). Các trình tự này được dự đoán chức năng sinh học dựa vào mức độ tương đồng của trình tự với nhiều cơ sở dữ liệu tham khảo có sẵn như KEGG (Kyoto Encyclopedia of Genes and Genomes, là cơ sở dữ liệu trực tuyến liên quan đến hệ gen, các con đường enzyme và các sản phẩm sinh học), eggNOG (evolutionary genealogy of genes: Non-supervised Orthologous Groups, là cơ sở dữ liệu chứa các nhóm orthologous), COG (COG - Clusters of Orthologous Group, là cơ sở dữ liệu protein của sinh vật nhân sơ, nhân chuẩn đơn bào), PFAM/TIGRFAM -protein families (là cơ sở dữ liệu các họ protein). Việc khai thác và lựa chọn gen mã hóa endo-1,4-beta-xylanase từ nguồn dữ liệu rất lớn này được chú trọng, từ đó có thể đưa vào biểu hiện trong thực nghiệm một cách hiệu quả. Mẫu dò (probe) đại diện cho enzyme xylanase của hai họ glycoside hydrolase số 10 và 11 (GH10 và GH11) đã được xây dựng trên việc so sánh tương đồng của nhiều trình tự đã được nghiên cứu kỹ tính chất trong điều kiện thực nghiệm (Nguyễn Minh Giang *et al.* 2018), đây là cơ sở để lựa chọn gen xylanase có thể biểu hiện trong điều kiện thực nghiệm với khả năng thể hiện hoạt tính thành công, tiết kiệm thời gian khai thác gen. Tuy nhiên, theo phân loại của CAZy, enzyme xylanase (EC 3.2.1.8) có thể thuộc về các họ GH

khác, nghiên cứu này sẽ tiếp tục xây dựng mẫu dò cho các họ GH mới để khai thác gen mã hóa xylanase.

## VẬT LIỆU VÀ PHƯƠNG PHÁP

### Vật liệu

Dữ liệu DNA metagenome gồm 164.644 khung đọc mở (ORF) của vi khuẩn khu trú trong dạ cỏ của dê bản địa Việt Nam, được giải trình tự bằng hệ thống giải trình tự HiSeq Illumina (BGI, Hồng Kông) và được chú giải chức năng dựa trên các cơ sở dữ liệu CAZy, KEGG, PFAM, COG, GO (Do *et al.*, 2018).

### Phương pháp

#### *Tìm kiếm các trình tự amino acid của enzyme endo-1,4-beta-xylanase đã được nghiên cứu đặc tính*

Sử dụng nguồn dữ liệu từ cơ sở dữ liệu CAZy và NCBI, dựa trên các thông tin phân loại của endo-1,4-beta-xylanase trong các họ GH từ CAZy, chúng tôi tiến hành tìm kiếm các nghiên cứu trên thế giới đã công bố về trình tự amino acid và các đặc tính như khả năng chịu nhiệt, chịu kiềm/axit, nhiệt độ tối ưu, thông số động học của enzyme. Các trình tự này được tập hợp thành bộ dữ liệu con dùng để xây dựng mẫu dò đặc hiệu cho từng họ GH có hoạt tính endo-1,4-beta-xylanase.

#### *Xác định mức độ tương đồng của các trình tự và tạo trình tự bảo tồn*

Để tạo các mẫu dò đặc hiệu cho từng họ GH, các trình tự thu được ở trên được so sánh giống hàng bằng công cụ Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) sau đó tìm trình tự bảo tồn bằng công cụ EMBOSS Cons ([https://www.ebi.ac.uk/Tools/msa/emboss\\_cons/](https://www.ebi.ac.uk/Tools/msa/emboss_cons/)). Clustal Omega là phần mềm cho phép so sánh nhiều trình tự nucleotide hoặc amino acid và đưa ra bức tranh tổng thể và tính toán các điểm tương đồng giữa các trình tự khác nhau. Kết quả so sánh của phần mềm này chỉ ra các vị trí mà amino acid giống nhau hoàn toàn hoặc một phần giữa các trình tự để người dùng dễ dàng quan sát và phân tích. Sau khi nhập dữ liệu Clustal Omega

sẽ tính toán và trả kết quả sau vài phút tùy vào số lượng trình tự cần so sánh. Tất cả trình tự amino acid thu thập được của enzyme endo-1,4-beta-xylanases sẽ so sánh với nhau cho từng họ GH. Sử dụng file kết quả so sánh đa trình tự, EMBOSS cons sẽ trả về trình tự bảo tồn được cho là đại diện cho tất cả các trình tự được sử dụng để so sánh.

#### **Xây dựng mẫu dò và giá trị tham chiếu**

Phân tích kết quả so sánh đa trình tự, trình tự bảo tồn sẽ ưu tiên lựa chọn bao gồm các vị trí giống nhau hoặc tương đối giống nhau, loại bỏ các vị trí khác nhau quá nhiều. Phần mềm BLASTP được sử dụng để so sánh trình tự bảo tồn với các trình tự ban đầu, xác định các giá trị tham chiếu (mức độ bao phủ và độ tương đồng) để khai thác gen. Giá trị tham chiếu của mẫu dò được sử dụng như ngưỡng để lựa chọn các gen mã hóa cho endo-1,4-beta-xylanase từ số liệu DNA metagenome vi khuẩn trong dạ cỏ dê.

#### **Khai thác trình tự mã hóa endo-1,4-beta-xylanase từ dữ liệu DNA metagenome vi khuẩn trong dạ cỏ dê**

Sau khi có các mẫu dò mã hóa cho endo-1,4-beta-xylanase thuộc các họ GH khác nhau, chúng tôi sử dụng BLASTP để so sánh mẫu dò với trình tự amino acid của các ORF khai thác được từ dữ liệu DNA metagenome vi khuẩn trong dạ cỏ dê. Kết quả của việc sử dụng mẫu dò tìm kiếm các gen mã hóa cho enzyme sẽ được so sánh với kết quả dự đoán ban đầu của BGI. Sau đó các trình tự có chỉ số điểm tối đa (max score) giá trị tương đồng và bao phủ lớn hơn hoặc bằng giá trị tham chiếu sẽ được lựa chọn.

#### **Ước đoán cấu trúc bậc ba và của trình tự mã hóa endo-1,4-beta-xylanase được khai thác**

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 Model (<https://swissmodel.expasy.org>).

#### **KẾT QUẢ VÀ THẢO LUẬN**

#### **Xây dựng mẫu dò cho endo-1,4-beta-xylanase thuộc họ GH8 và GH30**

Theo kết quả cập nhật của cơ sở dữ liệu CAZy, enzyme endo-1,4-beta-xylanase đã được phát hiện và xếp vào 9 họ glycoside hydrolase là GH5, 8, 10, 11, 30, 43, 51, 98, 141, trong đó số lượng trình tự nhiều nhất thuộc về hai họ GH10 và GH11. Mẫu dò của enzyme endo-1,4-beta-xylanase của hai họ này đã được xây dựng trong công bố năm 2018 của Nguyễn Minh Giang và đồng tác giả (2018), tuy nhiên hiện tại đã có những nghiên cứu thực nghiệm mới khẳng định khả năng biểu hiện và hoạt tính của enzyme xylanase thuộc về các họ GH khác như GH5, GH8 và GH30 (Bảng 1).

Với yêu cầu là các trình tự mã hóa endo-1,4-beta-xylanase đã được nghiên cứu hoạt tính sinh học bằng thực nghiệm, chúng tôi đã thu thập được 6 trình tự thuộc họ GH8, 5 trình tự thuộc họ GH30 và chỉ có một trình tự thuộc họ GH5 (Bảng 1). Không có trình tự nào thuộc các họ GH43, 51, 98, 141 đã được nghiên cứu hoạt tính sinh học được tìm thấy. Vì vậy, chỉ 6 trình tự thuộc họ GH8 và 5 trình tự thuộc họ GH30 được sử dụng để xây dựng trình tự mẫu dò cho hai họ này.

**Bảng 1.** Tổng hợp dữ liệu các endo 1-4 beta xylanase thuộc các họ GH5, GH8 và GH30 đã được nghiên cứu tính chất.

STT	Mã số trong GENBANK	Vi khuẩn	Số amino acid	pH tối ưu	Nhiệt độ tối ưu (°C)	Nguồn thu thập số liệu
<b>GH5</b>						
1	ADT62000.1	<i>Prevotella bryantii</i> B14	464	6	40	<a href="https://pubmed.ncbi.nlm.nih.gov/20622018/">https://pubmed.ncbi.nlm.nih.gov/20622018/</a>
<b>GH8</b>						
2	AAC27700.1	<i>Bacillus</i> sp. KK-1	434		55	<a href="https://pubmed.ncbi.nlm.nih.gov/9678255">https://pubmed.ncbi.nlm.nih.gov/9678255</a>

3	CBY88881.1	<i>Pseudoalteromonas arctica</i>	426		50	<a href="https://pubmed.ncbi.nlm.nih.gov/21331632">https://pubmed.ncbi.nlm.nih.gov/21331632</a>
4	CAD20872.1	<i>Pseudoalteromonas haloplanktis</i>	426	4	25	<a href="https://pubmed.ncbi.nlm.nih.gov/12089151">https://pubmed.ncbi.nlm.nih.gov/12089151</a>
5	AEC33258.1	<i>Paraglaciecola mesophila</i>	423	7	30	<a href="https://pubmed.ncbi.nlm.nih.gov/19506861">https://pubmed.ncbi.nlm.nih.gov/19506861</a>
6	AAS85781.1	uncultured bacterium	1395			<a href="https://pubmed.ncbi.nlm.nih.gov/15184164">https://pubmed.ncbi.nlm.nih.gov/15184164</a>
7	ABB71891.1	uncultured bacterium	399		20	<a href="https://pubmed.ncbi.nlm.nih.gov/16532363">https://pubmed.ncbi.nlm.nih.gov/16532363</a>
<b>GH30</b>						
8	ADM15019.1	<i>Bacillus</i> sp. BP7	423	6	55	<a href="http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2937475/">http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2937475/</a>
9	AAB63573.1	<i>Aeromonas caviae</i>	564			<a href="https://www.sciencedirect.com/science/article/abs/pii/S0922338X97827924">https://www.sciencedirect.com/science/article/abs/pii/S0922338X97827924</a>
10	AAB53151.1	<i>Dickeya chrysanthemi</i> D1	413			<a href="https://pubmed.ncbi.nlm.nih.gov/8810080/">https://pubmed.ncbi.nlm.nih.gov/8810080/</a>
11	AHA38215.1	<i>Paenibacillus favisporus</i>	430	6	65	<a href="https://pubmed.ncbi.nlm.nih.gov/26419457/">https://pubmed.ncbi.nlm.nih.gov/26419457/</a>
12	BAA13641.1	<i>Paenibacillus</i> sp. W-61	528	6	40	<a href="https://www.researchgate.net/publication/45725629">https://www.researchgate.net/publication/45725629</a>

trinhtu1	NGGYYSVKTEGMSYGMMITLQMNDEYKFKQLWDFVRKYMHRHSDRN--DSLGYHSHWMKKT
trinhtu2	-DEGDDIRTEGQSWGMTAAVMLNKQEEFDNLWRFKAYQKNPDNHPDAKKQGVYAWKLLK
trinhtu3	-DEGDDIRTEGQSWGMTAAVMLNKQEEFDNLWRFKAYQKNPDNHPDAKKQGVYAWKLLK
trinhtu4	-IDSNDIRSEGMSYGMMSVMMDDQETFNKLWRFKAKMQNTS----GNSKDFFAWRLS-
trinhtu5	-FDSNDVRSEGMSYGMMSVMMDDQETFNKLWRFKAKMQNTS----GNSKDFFAWRLS-
trinhtu6	-VNSGDVRSEGMSYGMMSVMMDDQETFNKLWRFKAKMQNTS----GNSKDFFAWRLS-
	.:::* *:* * : : : : * : * * : : : : : . : * .
trinhtu1	NGSD-VQTIQNVASDGEVWFAAALMMASGRWGDKKYPYD---YKARAQDMLDALAGDGE
trinhtu2	NQNGFVYKVDGEPAPDGEEYFALFNASARWGNSSGE-FN---YYNDIAIAMLNTIKNK--
trinhtu3	NQNGFVYKVDGEPAPDGEEYFALFNASARWGNSSGE-FN---YYNDIAIAMLNTIKNK--
trinhtu4	-GNAPYNAIDTNPAPDGEEYFAMALFFANNRWGSADGIFD---YQREANDILHDMIFTKS
trinhtu5	-TSG--SKMDKGPAPDGEEYFITALLFAHARWGSASGTTNINNYAQQARQIIYDLTRRKP
trinhtu6	-TEG--RHKDQMPAPDGEEYFAMALFFANNRWGSADGIFD---YQREANDILHDMIFTKS
	. * * * * : * * * * * : * * * * : : : *
trinhtu1	YANTGKESRVFIKNSKDQRYAMVRFPGYV---NWTDPYHVPAFFELFAKSA-----
trinhtu2	-----LMENQIIRFSPYID--NLTDPSYHIPAFYDYFANNVT-----
trinhtu3	-----LMENQIIRFSPYID--NLTDPSYHIPAFYDYFANNVT-----
trinhtu4	DNS-----STRLLMHPVYQVQVEFVTTTINVASFSDPSYHLPAFYEMWALWA-----
trinhtu5	GNGD----PYGEPSPFNVDNYMVRFATLGN SATFTDPSYHLPAFYDVWALELQADYDNSK
trinhtu6	ITGQVGEQIQSAGNIFSLYKMRIFTPDIANSEHTDPSYHLPAFYELWALWG-----
	: . * : * * * * : * * * * : * * * * : * * * * : * * * * : * * * * : *
trinhtu1	-----KSSQQYFWKDAANKSRTYLSSETTFKSVLNGSTVTNAATGLFPDEAGFDG
trinhtu2	-----NQADKTYWRQVATKSRLLKNHFTKV-----SGSPHWNLPFLSRLDG
trinhtu3	-----NQADKTYWRQVATKSRLLKNHFTKV-----SGSPHWNLPFLSRLDG
trinhtu4	-----DENNDYWHETAISRQYLAKSAH-----PVTGLFSDYANHEG
trinhtu5	LYGIWADKADLKKDIDYFKQAATTSRFFAKTTN-----GTTGLGPDYAGFDG
trinhtu6	-----PEKSDDFWKEAAKVSRRDFFFI SAN-----PKTGLTPDYANFDG
	. : : : : * * * : . * : : *
trinhtu1	VSDAAH---SSTETDRNFSYDAWRTVSHIAMDHTLWSSADNAYRASEQKAVNKFLTFMKR
trinhtu2	SPVIGYIFNGQANPGQWYEFDAWRVIMNVGLDAHLMGAQAWHKSA--VNKALGFLSYAKT
trinhtu3	SPVIGYIFNGQANPGQWYEFDAWRVIMNVGLDAHLMGAQAWHKSA--VNKALGFLSYAKT
trinhtu4	EPQTTS-FN---PDSHKSAYDSFRVMGNMAMDYHWSQSPVLQSL--VEQQVTFVAG-EV
trinhtu5	TPKN---E---GDHKYFEYDAWRIAMNIGMDYAWFAKDSWQKTF--ADRIQAFFVSKGV
trinhtu6	TPWVCP-WN---PNAANFQYDAWRTVMNWSVDWAWWEKDERQREL--SDRLQAFFESRGI
	: * * * : : * * * : . : * *

Hình 1. Kết quả so sánh đa trình tự các xylanase thuộc họ GH8.

trinhtu3	-----MNGNVSLWVRHCLHAALFVSATAGSF-SVYADTVKIDANVNYYQIIQGFGGM
trinhtu2	--MKVVKFKFKKTINVLL-ACLTAALPLMLS---PTQVSAASDANINLSSEKQLIKGFGGI
trinhtu5	-----MNLKKSVMHLL-ACLTAALPLMLT---PTHVSAASDANINLSSEKQLIKGFGGI
trinhtu1	----Missvkkpicvll-vcftmlsvmlagpgatevlasdvtinlsaeqvirgfggm
trinhtu4	MKMHATWSRLKKSVCGL-AGVTALSIVLAVPGPAQVSASSNATINLSAEKQVIRGFGGI
	.: : : . : : : . . .*: . : *:*:****:
trinhtu3	SGVGWINDLTTEQINTAYGSGVGQIGLSIMRVRIDPDSKWNILPQSARQAVSLGAKIMA
trinhtu2	NLPAWIGDLTPAQRETAFGNDQNQLGFSILRIYVDPDSNNWYREVATAKRAIEKGAIVFA
trinhtu5	NHPAWIGDLTAAQRETAFGNGPNQLGFSILRIYVDENRNNWYREVATAKRAIEQGGALVFA
trinhtu1	nhpawigdltaaqrtafgngqngqlgfsilrihvdenrnnwyrevetaknaikhgaivfa
trinhtu4	NHPVWIGDLTEAQRETAFGNGDNQLGFSILRIHVDEDRNWSKEVETAKSAIAHGAIVFA
	. **.*** * :*:*.. .*:**:* : * : .* : :*: * : ** :**
trinhtu3	TPWSPPAYMKS---NSLINGRLLPANYSAITSHLLDFSKYMQTNGAPLYAISIQNEPD
trinhtu2	SPWNPPSSMVETFNRRGDTNAKRLKYDKYAAYAQLHNDFTYMKNNGVLDYAI SVQNEPD
trinhtu5	SPWNPPSDMVETFNRRGDTNAKRLRYDKYAAYAQLHNDFTYMKDNGVNLVYAI SVQNEPD
trinhtu1	spwnppsdmvetfnrngdtsakrlrydkyaayaqlndfvtymkngvnlvyaivsqnepd
trinhtu4	SPWNPPSDMTETFNRRGDTSAKRLRYDKYADYAQYLHNDFTFMKNNGVLDYAI SVQNEPD
	:**.*: * .. * .. ** :*: :*: * ** :*: ** . *****:****
trinhtu3	WKPDIYESCEWSGDEFKSYLKSQGSKFGSLKVIVAESLGFNPALTDVPLKDSASKYVSI
trinhtu2	YAHWT--WWTPQEILRFMKENAGSIQGTQVMAPEFQYLNKINSDPILNDPQALANMDIL
trinhtu5	YAHWT--WWTPQEMLRFMKENAGSIQNTQVMAPEFQYLNKINSDPILNDPQALANMDIL
trinhtu1	yahwt--wwtpqeilrfmrenagsin-arviapesfgyfknisdpilndpqualrnm dil
trinhtu4	YAHWT--WWTPPEMLRFMKENAGSIN-CRVIAPESFQYLNKINSDPILNDPQALANMDIL
	: : * : * : : : : : : : * : . ** : : : : * : * : * : *
trinhtu3	GGHLYGTTPKPY--PL--AQNAGQLWMTEHYVDS--KQSANNWTS AIEVGTELNASMV-
trinhtu2	GAHTYGTQISNFAYPLFKQKAGKELWMTEVYVYVNSDNNADRWEALVSYHMHNAMVE
trinhtu5	GAHTYGTQFKDFAYPLFKQKAGKELWMTEVYVYVNSDNNADRWEALDVSYHMHNAMVE
trinhtu1	gthlygtqvsqfpyplfkqkagkelwmtevyvypnsdnnsadrwealgvsehishsmve
trinhtu4	GTHLYGTQISDFAYPLFKQKAGKELWMTEVYVYVNSDANSADRWEALDVSYHINAMVE
	* * ** . : ** : .***:***** * . :*:*. * : * . : : *
trinhtu3	SNYSAYVWWYIRRSYGLLEDGKVS KRGYVMSQYARFVRPGALRIQATENPQSNVHLTAY
trinhtu2	GDFQAYVWWYIRRQYGP MKEDGTISKRGYNMAHFSKFVRPGYLRVDATKNPDTNTVFSAY
trinhtu5	GDFQAYVWWYIRRQYGMNEDGSI SKRGYNMAHFSKFVRPGYRVDATKNPDTNTVFSAY
trinhtu1	gdfqsyvwwyirrsygpmedgtiskrgynmahfskfvrpgyvrdatknpnanvysay
trinhtu4	GDFEAYVWWYIRRQYGP MKEDGTISKRGYNMAHFSKFVRPGYVVDATKNPDTNVTYSAY
	. : . : *****.* : .***:***** * : : : :***** * : * : * : * : *
trinhtu3	KNTDGKMMIVAVNTNDSQMLSLNISNANVTKFEKYSTSASLNVEYGGSSQVDSSGKATV
trinhtu2	KGD-NKVVVVA INRGTSAVSQRFVLQNGNASTVSSVYVTDSSRNLASLAPINV-SNGAFTA
trinhtu5	KGD-NKAVIVAINRGTSAVSQKRFVLQNGNASTVSSVYVTDSSRNLAGAQITV-SGGAFTA
trinhtu1	kgd-nkvvivainksntgvnqnfvlqngsasqvsrwitsgssnlqpgtnlv-tgnhfwa
trinhtu4	KGN-NKVVIVAINKGTSAVSQTFVLQNGKTSKVTWPVTDASRNMTAGSSIRV-KGGSFTA
	* . * * :*:* : : : : : : : : * : . * * * : * . . . .

Hình 2. Kết quả so sánh đa trình tự các xylanase thuộc họ GH30.

Theo kết quả so sánh, 6 trình tự protein thuộc họ GH8 thu thập được có 52 vị trí amino acid giống nhau hoàn toàn, 51 vị trí giống nhau ở đa số các trình tự, 31 vị trí giống nhau một phần, còn lại là khác nhau (Hình 1). Tương tự đối với họ GH30, khi so sánh 5 trình tự thu thập được có 137 vị trí giống nhau hoàn toàn, 91 vị trí giống nhau ở đa số và 57 vị trí giống nhau một phần ở các trình tự (Hình 2). Như vậy có thể nói, so với họ GH8, các trình tự xylanase thuộc họ GH30 có tính bảo thủ cao hơn.

Dựa trên kết quả so sánh các trình tự amino acid, trình tự mẫu dò của các endo-1,4-beta-xylanase thuộc họ GH8 và GH30 đã được xác định (Hình 3).

Để tìm giá trị tham chiếu cho việc sử dụng mẫu dò trong khai thác gen, chúng tôi đã so sánh tương đồng giữa mẫu dò với từng trình tự được sử dụng để xây dựng nên chúng. Kết quả cho thấy các trình tự phải có điểm tối đa trên 168, độ bao phủ và độ tương đồng tối thiểu 84% và 36% so với mẫu dò để có thể dự đoán là trình tự mã hóa

endo-1,4-beta-xylanase GH8. Tương tự đối với endo-1,4-beta-xylanase GH30, các trình tự khai thác phải có điểm tối đa lớn hơn 300, độ bao phủ lớn hơn 90% và độ tương đồng lớn hơn 40%.

A:

MVFIITLISYTSLAAPSGASSGYRNLFQEGKTNITQKVNSAFDNMFNQSYYPYENDMAYIKADDSNDIRTEGMSYGMIVM  
MNDQEEFDKLRFAKAYMKNNDNHANSQGYAWKLNNGVKVDEGPAPDGEEYFAALLFASARWGNSTGFNYKEAMLNDLKNK  
NMNMEYQMIRFSPYIDANFTDPSYHIPAFYDLWALNQKDKDYWKQAATKSRTFLASNKSPGLPDYAFDGTVPYFNDQWFE  
YDAWRVVMNVGMDYHWWKAWKSAVDKVFVFSYKVNNGYVYVTLDTGKQKSEGLVAANAVAALASTNAQANEFINEFWLSLM  
PTGYRYDGLYMLAMLHVSGNFKYPTN

B:

LKKSXVXLXACLALXLMALXGPTQVSAASDAXINLSAEKQIIKGFGINHPAWIGDLTXAQRETAFGNGXNQLGFSIL  
RIHVDEDRNNWYREVXTAKRAIEXGAIIVFASPWNPPSDMVEFNRNGDTNAKRLRYDKYAAQAHLNDFVTYMKNNGVDLY  
AISVQNEPDYAHDWTTXWWTPEMLRFMKENAGSIXXKVIAPESFYLNKMSDPILNDPQALANMDILGXHLYGTQISDF  
AYPLFKQKAGKELWMTVEVYYPNSDNNSADRWPEALDVSYHMHNAMVEGDFQAYVWWYIRRQYGPMKEDGTISKRGYNMAH  
FSKFVRPGYLRVDATKNPDTNVVVSAYKGDKNKVVIVAINRGTSAVSQSFLQNGNASKVSVXWVTDASRNLSXGASIXVXS  
GGAFTAQLPAQSVTTFVAEL

**Hình 3.** A: Trình tự mẫu dò cho endo-1,4-beta-xylanase thuộc họ GH8. B: Trình tự mẫu dò cho endo-1,4-beta-xylanase thuộc họ GH30. X: gốc amino acid không xác định.

**Bảng 2.** So sánh tương đồng giữa mẫu dò với các trình tự ban đầu.

Trình tự	Điểm tối đa	Độ bao phủ	Giá trị E	Độ tương đồng
<b>GH8</b>				
trinhtu2	360	99%	2e-126	60%
trinhtu3	356	99%	6e-125	59%
trinhtu5	316	93%	2e-101	48%
trinhtu6	279	92%	5e-95	46%
trinhtu4	234	84%	3e-73	46%
trinhtu1	168	98%	6e-52	36%
<b>GH30</b>				
trinhtu5	765	100%	0.0	88%
trinhtu2	749	98%	0.0	88%
trinhtu4	738	100%	0.0	85%
trinhtu1	701	99%	0.0	80%
trinhtu3	316	99%	2e-108	41%

**Khai thác trình tự mã hóa endo-1,4-beta-xylanase bằng mẫu dò từ dữ liệu DNA metagenome vi khuẩn dạ cỏ dê**

Các mẫu dò của họ GH10, GH11 đã được xây dựng trước đây (Nguyễn Minh Giang *et al.* 2018) cùng với mẫu dò của họ GH8 và GH30 mới được xây dựng ở trên được sử dụng để tìm kiếm các trình tự đích trong dữ liệu DNA metagenome vi khuẩn dạ cỏ dê. Kết quả khai thác đã tìm thấy 21 trình tự tương đồng với mẫu dò của họ GH8, 20 trình tự tương đồng với mẫu dò

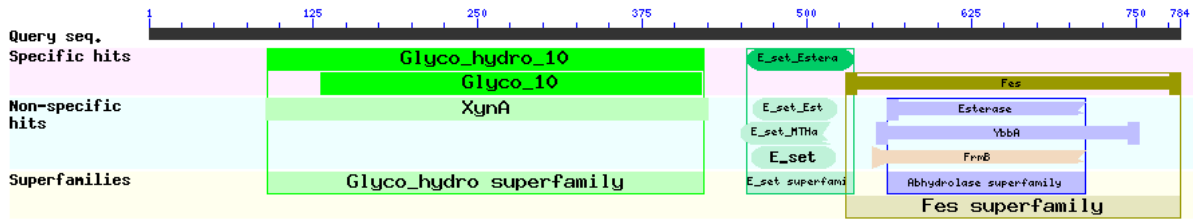
của họ GH10, và không tìm thấy trình tự nào tương đồng với mẫu dò của họ GH11 và GH30. Kết quả cụ thể được trình bày trong Bảng 3.

Kết quả chú giải của BGI dựa trên dữ liệu KEGG tìm thấy 67 trình tự có thể có hoạt tính endo-1,4-beta-xylanase, trong đó có 19 trình tự trùng với kết quả tìm kiếm bằng mẫu dò của họ GH10. Tuy nhiên phần lớn các trình tự tìm được bằng mẫu dò của họ GH8 là các trình tự không được chú giải khi so sánh với dữ liệu KEGG. Điều này chứng tỏ việc sử dụng mẫu

dò cho phép phát hiện nhiều hơn các trình tự đích mã hóa cho endo-1,4-beta-xylanase trong dữ liệu DNA metagenome. Trong 41 trình tự đã được khai thác bằng mẫu dò, có 28 trình tự là ORF hoàn chỉnh, 8 trình tự thiếu đầu 3', 1 trình tự thiếu đầu 3' và 4 trình tự thiếu cả 2 đầu.

**Bảng 3.** Kết quả khai thác endo-1,4-beta-xylanase bằng mẫu dò từ dữ liệu DNA metagenome vi khuẩn dạ cỏ dê.

ORF được khai thác	Điểm tối đa	Độ bao phủ	Giá trị E	Độ tương đồng	Dự đoán dựa trên CSDL KEGG
<b>GH8</b>					
[denovogenes]_24919	200	98%	7E-61	35%	-
[denovogenes]_22225	196	93%	4E-59	34%	endoglucanase
[denovogenes]_21763	194	99%	2E-58	33%	-
[denovogenes]_19998	193	93%	7E-58	34%	endoglucanase
[denovogenes]_22379	192	99%	1E-57	33%	-
[denovogenes]_23675	191	98%	2E-57	32%	-
[denovogenes]_25522	191	98%	2E-57	34%	-
[denovogenes]_26578	189	98%	7E-57	34%	-
[denovogenes]_28222	189	98%	7E-57	35%	Endoglucanase
[denovogenes]_21275	188	98%	4E-56	32%	-
[denovogenes]_22196	187	98%	7E-56	32%	-
[denovogenes]_27221	186	98%	1E-55	34%	-
[denovogenes]_35308	182	86%	6E-55	35%	-
[denovogenes]_24635	184	98%	6E-55	33%	-
[denovogenes]_26598	183	92%	1E-54	34%	Endoglucanase
[denovogenes]_28004	181	98%	7E-54	34%	-
[denovogenes]_22112	180	99%	3E-53	30%	-
[denovogenes]_23600	177	84%	5E-52	36%	-
[denovogenes]_21959	177	93%	6E-52	32%	-
[denovogenes]_20668	175	98%	4E-51	30%	-
[denovogenes]_39514	171	82%	6E-51	34%	Endoglucanase
<b>GH10</b>					
[denovogenes]_4195	245	91%	7E-78	43%	endo-1,4-beta-xylanase
[denovogenes]_4472	244	96%	2E-77	43%	endo-1,4-beta-xylanase
[denovogenes]_34159	236	94%	2E-78	41%	endo-1,4-beta-xylanase
[denovogenes]_5309	233	91%	6E-74	41%	endo-1,4-beta-xylanase
[denovogenes]_9650	231	97%	4E-74	39%	endo-1,4-beta-xylanase
[denovogenes]_5456	230	92%	1E-72	41%	endo-1,4-beta-xylanase
[denovogenes]_5409	228	91%	7E-72	41%	endo-1,4-beta-xylanase
[denovogenes]_5086	227	98%	1E-71	39%	endo-1,4-beta-xylanase
[denovogenes]_17825	224	92%	2E-72	40%	endo-1,4-beta-xylanase
[denovogenes]_5578	222	92%	5E-70	39%	endo-1,4-beta-xylanase
[denovogenes]_5411	221	95%	1E-69	40%	endo-1,4-beta-xylanase
[denovogenes]_5913	221	95%	2E-69	39%	endo-1,4-beta-xylanase
[denovogenes]_5329	218	95%	2E-68	39%	endo-1,4-beta-xylanase
[denovogenes]_5283	218	94%	2E-68	39%	endo-1,4-beta-xylanase
[denovogenes]_32356	216	95%	2E-70	40%	endo-1,4-beta-xylanase
[denovogenes]_5259	215	96%	3E-67	38%	endo-1,4-beta-xylanase
[denovogenes]_17041	215	92%	5E-69	40%	endo-1,4-beta-xylanase
[denovogenes]_21241	210	88%	1E-67	40%	endo-1,4-beta-xylanase
[denovogenes]_7833	209	79%	2E-65	44%	enterochelin esterase and related enzymes
[denovogenes]_43594	206	78%	2E-67	43%	endo-1,4-beta-xylanase



**Hình 4.** Kết quả dự đoán tương đồng đặc hiệu trình tự bằng BLASTP với các trình tự khai thác bằng mẫu dò.

Khi so sánh trình tự của các ORF hoàn chỉnh với cơ sở dữ liệu nr (non-redundant protein) của NCBI bằng phần mềm BLASTP, các trình tự được khai thác bằng mẫu dò của họ GH10 đều chứa vùng đặc thù của enzyme endo-1,4-beta-xylanase (XynA) bên cạnh một vùng đặc thù cho enzyme esterase (Fes) (Hình 4).

Cấu trúc không gian của phân tử cho phép ước đoán cụ thể hơn về trung tâm hoạt động, cấu

hình phân tử và các phối tử liên quan đến hoạt tính sinh học của enzyme. Kết quả ước đoán cấu trúc không gian bằng phần mềm SwissModel cho thấy tất cả các trình tự đều có cấu trúc tương đồng cao với xylanase, với độ bao phủ ít nhất là 43% so với trình tự khuôn xylanase và phối tử là các dạng oligosaccharide của beta-D-xylopyranose, tạo cơ sở để khẳng định đây là các trình tự mã hóa endo-xylanase.

**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
[denovogenes]_21959	3a3v.1.A	Xylanase	0,82	42,01	X-ray, 1,4Å	Monomer	1 x <u>NI</u>
[denovogenes]_23600	1wu6.1.A	Xylanase	0,84	33,88	X-ray, 1,4Å	Monomer	1 x <u>XYP-XYP</u> , 1 x <u>NI</u>
[denovogenes]_22112	5yxt.1.A	Reducing end xylose-releasing exo-oligoxylanase	0,83	43,94	X-ray, 1,9Å	homo-tetramer	Không
[denovogenes]_28004	5yxt.1.A	Reducing end xylose-releasing exo-oligoxylanase	0,89	34,62	X-ray, 1,9Å	homo-tetramer	Không
[denovogenes]_26598	3a3v.1.A	Xylanase	0,89	42,9	X-ray, 1,4Å	Monomer	1 x <u>NI</u>
[denovogenes]_27221	5yxt.1.A	Reducing end xylose-releasing exo-oligoxylanase	0,9	35,71	X-ray, 1,9Å	homo-tetramer	Không
[denovogenes]_22196	3a3v.1.A	Xylanase	0,82	46,47	X-ray, 1,4Å	Monomer	1 x <u>NI</u>
[denovogenes]_21275	5yxt.1.A	Reducing end xylose-releasing exo-oligoxylanase	0,81	43,36	X-ray, 1,9Å	homo-tetramer	Không
[denovogenes]_28222	1wu5.1.A	Xylanase	0,89	33,7	X-ray, 2,2Å	Monomer	1 x <u>XYP</u> , 1 x <u>NI</u>
[denovogenes]_26578	5yxt.1.A	Reducing end xylose-releasing exo-oligoxylanase	0,86	35,75	X-ray, 1,9Å	homo-tetramer	Không
[denovogenes]_25522	5yxt.1.A	Reducing end xylose-releasing exo-oligoxylanase	0,85	35,26	X-ray, 1,9Å	homo-tetramer	Không
[denovogenes]_23675	3a3v.1.A	Xylanase	0,82	45,0	X-ray, 1,4Å	Monomer	1 x <u>NI</u>
[denovogenes]_22379	3a3v.1.A	Xylanase	0,82	46,03	X-ray, 1,4Å	Monomer	1 x <u>NI</u>
[denovogenes]_19998	3a3v.1.A	Xylanase	0,78	46,85	X-ray, 1,4Å	Monomer	1 x <u>NI</u>
[denovogenes]_21763	5yxt.1.A	Reducing end xylose-releasing exo-oligoxylanase	0,83	41,6	X-ray, 1,9Å	homo-tetramer	Không



[denovogenes]_22225	3a3v.1. A	Xylanase	0,81	46,85	X-ray, 1,4Å	Monomer	1 x <u>NI</u>
[denovogenes]_24919	3a3v.1. A	Xylanase	0,84	44,32	X-ray, 1,4Å	Monomer	1 x <u>NI</u>
[denovogenes]_5259	4pud.1. A	Endo-1,4-beta-xylanase	0,44	38,65	X-ray, 2,0Å	Monomer	1 x <u>XYP-XYP-XYP-XYP-XYP</u> , 9 x <u>ZN</u>
[denovogenes]_5283	5ay7.1. A	xylanase	0,46	57,14	X-ray, 2,1Å	Monomer	Không
[denovogenes]_5329	5ay7.1. A	xylanase	0,46	54,73	X-ray, 2,1Å	Monomer	Không
[denovogenes]_5411	5ay7.1. A	xylanase	0,46	56,25	X-ray, 2,1Å	Monomer	Không
[denovogenes]_5578	5ay7.1. A	xylanase	0,46	55,82	X-ray, 2,1Å	Monomer	Không
[denovogenes]_5086	5ay7.1. A	xylanase	0,46	56,68	X-ray, 2,1Å	Monomer	Không
[denovogenes]_5409	5ay7.1. A	xylanase	0,46	56,46	X-ray, 2,1Å	Monomer	Không
[denovogenes]_5456	5ay7.1. A	xylanase	0,46	56,97	X-ray, 2,1Å	Monomer	Không
[denovogenes]_5309	5ay7.1. A	xylanase	0,46	56,55	X-ray, 2,1Å	Monomer	Không
[denovogenes]_4472	5ay7.1. A	xylanase	0,43	63,3	X-ray, 2,1Å	Monomer	Không
[denovogenes]_4195	1us2.1. A	Endo-1,4-beta-xylanase	0,50	30,2	X-ray, 1,8Å	Monomer	7 x <u>XYP</u>

Chú thích: NI: Ni<sup>2+</sup>, ZN: Zn<sup>2+</sup>, XYP: beta-D-xylopyranose.

## KẾT LUẬN

Phương pháp xây dựng mẫu dò dựa trên các trình tự mã hóa enzyme endo-1,4-beta xylanase là một hướng tiềm năng ứng dụng trong việc tìm kiếm trình tự đích từ dữ liệu DNA metagenome. Việc lựa chọn các trình tự amino acid của enzyme này đã được nghiên cứu chi tiết về hoạt tính từ vi khuẩn để xây dựng các mẫu dò thuộc các họ GH mà enzyme này được phân loại. Mẫu dò được xây dựng sẽ hỗ trợ hiệu quả cho việc lựa chọn các gen mã hóa cho endo-1,4-beta-xylanase từ dữ liệu khổng lồ DNA metagenome.

**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 ĐTDLCN.15/14, kinh phí từ đề tài Nghị định thư Việt Đức "Nghiên cứu sản xuất một số enzyme phân hủy lignocellulose trên cơ sở khai thác dữ liệu metagenome" mã số NDT.50.GER/18 và trang thiết bị của Phòng thí nghiệm Trọng điểm Công nghệ gen.

## TÀI LIỆU THAM KHẢO

An D, X Dong, Z Dong (2005) Prokaryote diversity

in the rumen of yak (*Bos grunniens*) and Jinnan cattle (*Bos taurus*) estimated by 16S rDNA homology analyses. *Anaerobe* 11: 207-215.

de Vries RP, J Visser (2001) *Aspergillus* enzymes involved in degradation of plant cell wall polysaccharides. *Microbiol Mol Biol Rev* 65: 497-522.

Do TH, NG Le, TK Dao, TMP Nguyen, TL Le, HL Luu, KHV Nguyen, VL Nguyen, LA Le, TN Phung, NM van Straalen, D Roelofs, NH Truong (2018) Metagenomic insights into lignocellulose-degrading genes through Illumina-based de novo sequencing of the microbiome in Vietnamese native goats' rumen. *J Gen Appl Microbiol* 64: 108-116.

Nguyễn Minh Giang, Đỗ Thị Huyền, Phùng Thu Nguyệt, Trương Nam Hải (2018) Xây dựng probe để khai thác và chọn gen mã hóa endo 1-4 xylanase từ dữ liệu giải trình tự DNA metagenome. *Tạp chí Sinh học* 40(1): 39-50.

Kittelman S, PH Janssen (2011) Characterization of rumen ciliate community composition in domestic sheep, deer, and cattle, feeding on varying diets, by means of PCR-DGGE and clone libraries. *FEMS Microbiol Ecol* 75: 468-481.

Polizeli ML, AC Rizzatti, R Monti, HF Terenzi, JA Jorge, DS Amorim (2005) Xylanases from fungi: properties and industrial applications. *Appl Microbiol Biotechnol* 67: 577-591.

- Souza WRd (2013) Microbial degradation of lignocellulosic Biomass. In: Chandel AK and Silva SSd(eds.) Sustainable Degradation of Lignocellulosic Biomass - Techniques, Applications and Commercialization. InTech.
- Twomey LN, JR Pluske, JB Rowe, M Choct, W Brown, MF McConnell (2003) The effects of increasing levels of soluble non-starch polysaccharides and inclusion of feed enzymes in dog diets on faecal quality and digestibility. *Anim Feed Sci Technol* 108 (1-4): 71-82.
- Verma D, T Satyanarayana (2012) Molecular approaches for ameliorating microbial xylanases. *Bioresour Technol* 117: 360-367.
- Viikari L, A Kantelinen, J Sundquist, M Linko (1994) Xylanases in bleaching: From an idea to the industry. *FEMS Microbiol Rev* 13: 335-350.

## PROBE-MINING OF ENDO-1,4-BETA-XYLANASE FROM GOATS-RUMEN BACTERIAL METAGENOMIC DNA DATA

Dao Trong Khoa<sup>1,2</sup>, Do Thi Huyen<sup>1,2</sup>, Truong Nam Hai<sup>1,2</sup>

<sup>1</sup>*Institute of Biotechnology, Vietnam Academy of Science and Technology*

<sup>2</sup>*Graduate University of Science and Technology, Vietnam Academy of Science and Technology*

### SUMMARY

Endo-1,4-beta-xylanases (xylanases) are classified into 9 glycoside hydrolase families, GH5, 8, 10, 11, 30, 43, 51, 98, and 141 based on the CAZy database. The probe sequences representing the enzymes were constructed from published sequences of actual experimental studies with xylan decomposition activity. From online databases, we found one sequence belonging to the GH5 family, 6 sequences belonging to the GH8 family and 5 sequences belonging to the GH30 family exhibiting xylanase activity. Thus specific probes for xylanase GH8 and GH30 families were designed with the length of 351 and 425 amino acids respectively. The reference values for the probe of the GH8 family were defined as the sequences with maximum score greater than 168, the lowest coverage was 84%, the lowest similarity was 36%; for the probe GH30, the maximum score was greater than 316, the coverage was greater than 98%, the similarity was greater than 41%. Using the built probes, including the probe of the two GH10 and GH11 families, we found 41 xylanase-encoding sequences from the metagenomic DNA data of bacteria in Vietnamese goats' rumen. Of the 41 exploited sequences, 19 were identical to the BGI company's annotation result based on KEGG database, whereas there were 16 sequences that are not annotated by the BGI company. Total 28 of 41 exploited sequences were complete open reading frames, of which the predicted ternary structure was highly similar to the published structures of xylanase.

**Keywords:** endo-1,4-beta-xylanase, DNA metagenome, bacteria in goats-rumen, glycoside hydrolase (GH), probe