Acta Scientific MICROBIOLOGY (ISSN: 2581-3226)

Volume 4 Issue 7 July 2021

Research Article

Diversity of Glycoside Hydrolase 10 Family Xylanases Found in Rumen Metagenome and Selection of Sequences with Biotechnological Potential

Gabriella Cavazzini Pavarina[1], Natália Sarmanho Monteiro Lima[1,2], Claudio Damasceno Pavani[1], Eliana Gertrudes de Macedo Lemos[1,2], João Martins Pizauro Junior[1] and Adriano Marques Gonçalves[3,4]*

[1]Technology Department, Sao Paulo State University (Unesp), Faculty of Agricultural and Veterinary Sciences, Brazil

[2]Molecular Biology Laboratory, Bioenergy Research Institute (IPBEN), Jaboticabal, Sao Paulo, Unesp, Brazil

[3]Department of Biological and Health Sciences, University of Araraquara (UNIARA), Brazil

[4]Department of Biochemistry and Organic Chemistry, Chemistry Institute, Sao Paulo State University (UNESP), Brazil

***Corresponding Author:** Adriano Marques Gonçalves, Professor, Department of Biological and Health Sciences, University of Araraquara (UNIARA), Araraquara, SP, Brazil.

Received: June 01, 2021

Published: June 17, 2021

Abstract

Metagenomics is an important tool for mining and discovering new enzymes, making it possible to explore the diversity of environments. Therefore, this strategy can be used to prospect for xylanases, which degrades lignocellulosic biomass, an important source of renewable energy of great economic interest. The aim of the work was to identify and investigate the diversity of xylanases of the GH10 family present in the rumen metagenome of Nelore cattle and to prospect molecules with good potential for biotechnological application through *in silico* analyzes. Pfam was used for the initial selection of GH10 sequences, then the physical and chemical parameters were computed using the ProtParam tool, SignalP-5.0 server was used to predict signal peptides and cleavage location, transmembrane helices prediction was made in TMHMM server, version 2.0 and domain annotation was performed with dbCAN meta server. In addition, identity comparison was performed with NCBI BLAST webtool, sequences were aligned with ClustalW and Neighbor-Joining Tree and pairwise analysis were performed. The metagenomics analysis from Nelore cattle rumen returned 38 sequences with the GH10 domain, CE1, GH43 and CBM6 were also identified in these sequences. Analysis of Neighbor-Joining Tree and proteins identity enabled to differentiate 6 groups with, at least, two proteins with identity higher than 70%. Based on the analysis, 13 sequences were considered unappropriated for biotechnological application for either for being unstable or having transmembrane helices. In this sense, based on the *in silico* analyzes, 25 of the 38 sequences presented good characteristics for *in vitro* studies. Thus, in addition to the identification of the 38 sequences with GH10 domain, a workflow of *in silico* methodologies was suggested to assist the selection of sequences that will guide future *in vitro* studies.

Keywords: Metagenomics; Xylanase; *In silico* Studies; Lignocellulose; Protein Biotechnology; Lignocellulose Degrading Enzymes

Introduction

The use of metagenomic databases is valuable for discovering new enzymes making it possible to explore the diversity of environments [1,2]. In this way, this application presents itself as an alternative pathway that helps and enriches the search for new genes related to degradation of lignocellulosic material. Metagenomics provides industry an unprecedented chance to bring diverse biomolecules to industrial application [3,4].

Lignocellulosic biomass, an important source of renewable energy, has gained prominence of great economic interest [5]. In this sense, there are highly efficient microbial communities in the hydrolysis of lignocellulose, such as the digestive system of termites and ruminants [6,7]. The degradation of plant biomass by ruminants is a consequence of the symbiosis between these and the numerous microorganisms, present in the ruminal cavity, that specialize for this purpose [7]. The metagenomics of the rumen environment allows both understanding of the functioning of this microenvironment, as well as the identification of new enzymes for the hydrolysis of lignocellulosic biomass [8-10].

Among hemicellulose degrading enzymes, xylanases are responsible for cleaving xylans, the most abundant hemicellulose in nature. The biodegradation of xylans involves the action of several hydrolytic enzymes, those that act on the branches (acetyl xylan esterease, α-D-glucoronidase, α-L-arabinofuranosidase), and those acting in the main chain (endo-1,4-β-xylanases EC 3.2.1.8 and β-xilosidase, EC 3.2.1.37) [11], leading to the release of xylooligosaccharides, which will be hydrolyzed by the other hemicellulases. Hemicellulases, together with auxiliary enzymes, increase the yield of monosaccharide generation in the enzymatic hydrolysis process of the lignocellulosic complex [12]. Within the group of hemicellulases, xylanases are used in industry in xylooligosaccharides production processes to obtain xylitol [15], production of fermentable sugars for the production of second generation ethanol [13] and paper production [11].

The heterologous expression of proteins using *Escherichia coli* is one of the most common systems when high productivity of recombinant proteins is desired, *E. coli* is called a "microbial factory", due to its easy handling, low production cost and high efficiency [16]. Even though it is highly advantageous, it depends on the characteristics of each gene, often making the process difficult [17], and some strategies can be taken to increase the efficiency of the process and avoid further problems. For example, some proteins, such as membrane proteins and proteins of larger molecular size, are often not expressed in *E. coli*, or when they are, they form inclusion bodies [18]. The presence of signal peptides can also cause problems in the production of heterologous proteins, often leading to insoluble molecules [19]. Therefore, the use of tools that provide prior knowledge of the gene is extremely important for the overexpression of recombinant proteins. The use of bioinformatics through *in silico* prediction significantly increases the chances of protein expression success [20]. Most of these tools offer an alternative of optimization, restriction sites, removal of harmful codons or nucleotides, in addition, the great advantage is that they are often available in online servers [21].

In this sense, the present work, using *in silico* analysis, investigated the diversity of GH10 xylanases in the rumen of Nellore cattle and the characteristics of the identified sequences. These studies were carried out aiming to differentiate and indicate the sequences with the greatest potential for biotechnological application. Additionally, a workflow was proposed to systematize criteria that are essential for the selection of sequences with greater biotechnological potential.

Materials and Methods

Identification of xylanases

The metagenomic data used for prospecting the sequences were obtained from the sequencing llumina HiScanSQ sequencing data set from Nelore cattle rumen (public data access: SRX818104), which assembly and annotation of genes were stored in the metagenomic database of the Laboratory of Biochemistry and Plant Microorganisms (LBMP). For this purpose, a reference sequence of endo-β-1,4-xylanase from the GH10 family, obtained in The American database of the National Center for Biotechnology Information (NCBI) was used to infer the xylanases sequences, then the sequences were submitted to the database eggNOG orthologist [22]. The ORFs (open reading frames) were selected with an e-value of -30, aligned with ClustalW [23], and validated for enzymatic domains in the Pfam database [24].

Prediction of physical and chemical parameters, signal peptides, transmembrane helices and domains

The physical and chemical parameters were computed using the ProtParam tool [25], SignalP-5.0 server was used to predict signal peptides and cleavage location [26], transmembrane helices prediction was made in TMHMM server, version 2.0 [27] and domain annotation was performed with dbCAN meta server [28].

Sequence alignment and estimated divergence between sequences

The sequences were organized in a FASTA file and, in order to estimate identities between sequences, a protein-protein BLAST was used to align the obtained sequences [29]. In addition, sequences were aligned with NCBI patent database using the same BLAST tool.

Also, the sequences were aligned with multiple alignment tool ClustalW with Clustal X (version 2.0) software [30]. After the ClustalW alignment the number of amino acid differences per sequence from between sequences were calculated. Standard error estimates were obtained by a bootstrap procedure (1000 replicates). This analysis involved 38 amino acid sequences containing gh10 domain. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 1410 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [31].

Next, a Neighbor-Joining tree was inferred to evaluate the relationship between the protein sequences [32]. The bootstrap consensus tree inferred from 1000 replicates was taken to represent the relation of the sequences analyzed [33]. Branches corresponding to partitions reproduced in less than 75 bootstrap replicates were collapsed. Sequences distances were computed using the number of differences method [34] and are in the units of the number of amino acid differences per sequence. This analysis involved the 38 amino acid sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 1410 positions in the final dataset. The referred analysis was conducted in MEGA X [31]. The final tree was analyzed and generated with ITol (Interactive Tree of Life) tool [35].

Results and Discussion

The metagenomics analysis from *Nelore cattle* rumen returned 38 sequences with the GH10 domain, additionally, the sequences may contain other domains (Table 1), this type of analysis is important to infer GH10 diversity in this complex environment. Besides, other domain recognition may be used to choose between sequences to be investigated in further steps, to clone and express the molecules with best biotechnology potential. Xylanases have broad use in industrial processes [11,13-15], in this sense, this kind of study is important to identify new molecules, in order to improve enzyme utilization in industrial and biotechnological procedures.

ID	Sequence Length (aa)	Molecular Weight (kDa)	Domains	Instability Index*	TMHMM	pI	% of positive aa**	% of negative aa***	Aliphatic Index	SignalP****
Xyl_01	368	41.62	GH10(24-365)	32.36/Y	0	7,06	12.2	12.2	88.80	Y(1-21)
Xyl_02	719	81.28	GH10(34-364)+CE1(503-715)	32.41/Y	0	6,03	11.4	12.8	67.97	Y(1-20)
Xyl_03	387	44.52	GH10(33-383)	28.02/Y	0	5,22	11.4	15.8	83.10	N
Xyl_04	724	82.10	GH10(34-365)+CE1(505-720)	34.25/Y	0	6,44	12.2	12.8	72.32	Y(1-20)
Xyl_05	753	85.29	GH10(26-346)+CE1(537-748)	41.85/N	0	8,53	12.6	11.8	62.59	Y(1-24)
Xyl_06	731	82.81	GH10(33-367)+CE1(507-728)	34.69/Y	0	7,89	12.4	12.2	71.78	Y(1-20)
Xyl_07	711	80.42	GH10(36-369)+CE1(509-687)	32.92/Y	1	6,44	12.1	12.8	69.93	Y(1-23)
Xyl_08	723	81.65	GH10(35-368)+CE1(507-719)	30.34/Y	0	6,21	10.9	12.0	65.84	Y(1-21)
Xyl_09	727	82.39	GH10(34-368)+CE1(508-723)	35.81/Y	0	6,38	12.4	13.2	74.04	Y(1-21)
Xyl_10	368	41.91	GH10(25-365)	26.95/Y	0	6,72	12.2	12.5	90.35	Y(1-22)
Xyl_11	754	84.48	GH10(35-360)+CE1(499-750)	38.74/Y	0	5,69	10.5	12.7	70.04	Y(1-22)

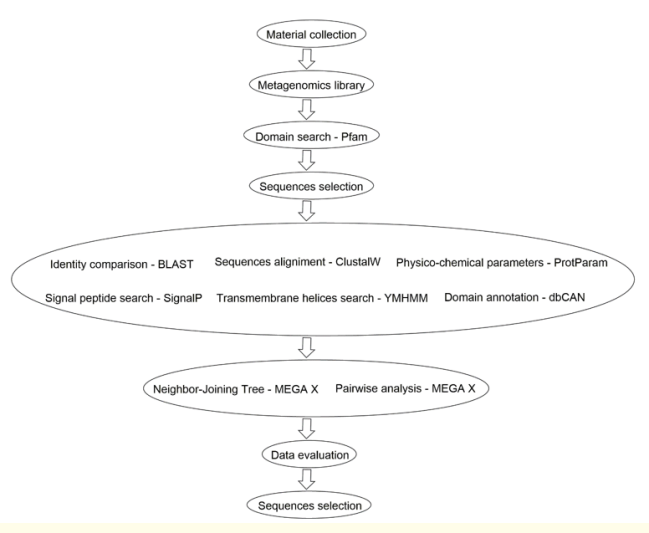
Xyl_12	368	42.19	GH10(27-365)	30.16/Y	0	6,21	11.7	10.6	80.65	Y(1-23)
Xyl_13	345	39.81	GH10(22-343)	36.00/Y	0	6,56	12.2	12.8	77.74	Y(1-20)
Xyl_14	652	72.26	GH10(60-386)	29.05/Y	1	6,36	12.0	12.4	77.81	Y(1-25)
Xyl_15	744	84.43	GH10(23-347)+CE1(528-738)	43.47/N	0	5,86	10.8	12.4	68.84	Y(1-21)
Xyl_16	379	42.99	GH10(35-376)	39.09/Y	1	7,15	12.1	12.1	83.88	Y(1-33)
Xyl_17	368	41.55	GH10(24-365)	36.65/Y	0	7,06	11.1	11.1	86.71	Y(1-21)
Xyl_18	728	82.92	GH10(34-370)+CE1(509-724)	37.57/Y	1	6,76	12.9	13.2	72.72	Y(1-22)
Xyl_19	714	81.61	GH10(24-355)+CE1(498-710)	37.72/Y	0	6,21	12.6	13.7	70.57	Y(1-20)
Xyl_20	727	82.22	GH10(34-368)+CE1(508-723)	35.98/Y	0	6,29	12.5	13.3	72.43	Y(1-22)
Xyl_21	746	84.00	GH10(37-371)+CE1(509-728)	40.04/N	0	6,89	11.7	11.9	72.96	Y(1-23)
Xyl_22	645	73.46	GH10(33-366)+CE1(506-639)	35.18/Y	0	7,98	12.6	12.2	70.59	Y(1-20)
Xyl_23	644	75.85	GH10(8-343)+CE1(483-634)	37.20/Y	0	6,33	13.2	14.1	68.99	N
Xyl_24	433	49.78	GH10(35-369)	38.62/Y	1	8,35	14.5	13.9	74.32	Y(1-24)
Xyl_25	382	43.20	GH10(37-373)	28.44/Y	0	9,37	15.4	11.8	77.15	Y(1-26)
Xyl_26	501	57.39	GH10(21-355)	39.42/Y	0	6	12.0	13.8	65.35	N
Xyl_27	725	82.40	GH10(34-370)+CE1(512-722)	38.70/Y	0	6,26	12.4	13.2	68.61	Y(1-21)
Xyl_28	723	82.04	GH10(30-360)+CE1(507-719)	38.91/Y	0	7,02	12.2	12.3	74.44	Y(1-19)
Xyl_29	385	44.51	GH10(49-383)	37.63/Y	1	8,03	13.0	12.5	67.12	Y(1-35)
Xyl_30	726	82.52	GH10(33-367)+CE1(507-722)	32.32/Y	0	7,24	12.9	12.9	71.43	Y(1-19)
Xyl_31	748	85.03	GH10(42-379)+CE1(518-743)	33.82/Y	1	6,42	12.4	13.1	70.40	Y(1-29)
Xyl_32	418	47.85	GH10(3-327)	35.54/Y	0	5,67	11.5	14.1	70.89	N
Xyl_33	389	45.23	GH10(36-371)	36.33/Y	1	8,47	14.1	13.1	65.42	Y(1-22)
Xyl_34	726	82.27	GH10(33-367)+CE1(507-722)	36.21/Y	0	7,9	13.1	12.8	70.25	Y(1-20)
Xyl_35	432	49.32	GH10(44-427)	44.81/N	0	5,51	10.9	13.2	71.83	N
Xyl_36	691	77.77	GH10(17-302)+CBM6(446-551)	29.65/Y	0	5,33	9.7	12.2	68.08	N
Xyl_37	849	95.33	GH43_7(26-295)+GH10(500-849)	42.50/N	0	4,99	9.7	12.8	75.18	Y(1-22)
Xyl_38	305	35.03	GH10(1-302)	32.88/Y	0	6,03	12.1	15.8	79.02	N

Table 1: Predicted physico-chemical parameters, domains, transmembrane helices and signal peptide of the 38 identified GH10 sequences.

*Y refers to stable proteins and N refers to unstable proteins. ** Percentage of positively charged residues (Arg and Lys). ***Percentage of negatively charged residues (Asp and Glu). **** Y indicates presence of signal peptide and N indicates absence of signal peptide.

In order to characterize and analyze the sequences returned by prospecting, *in silico* analyzes were carried out to search for the most promising molecules to be tested in future *in vitro* heterologous expression assays. Despite not substituting laboratory tests, in silico analysis contributes to the choice of molecules with the best potential, leading to savings in human and financial resources and time. Thus, a workflow (Figure 1) was used to choose, among the 38 sequences, those with better conditions for heterologous expression.

Figure 1: Workflow for protein sequence selection.

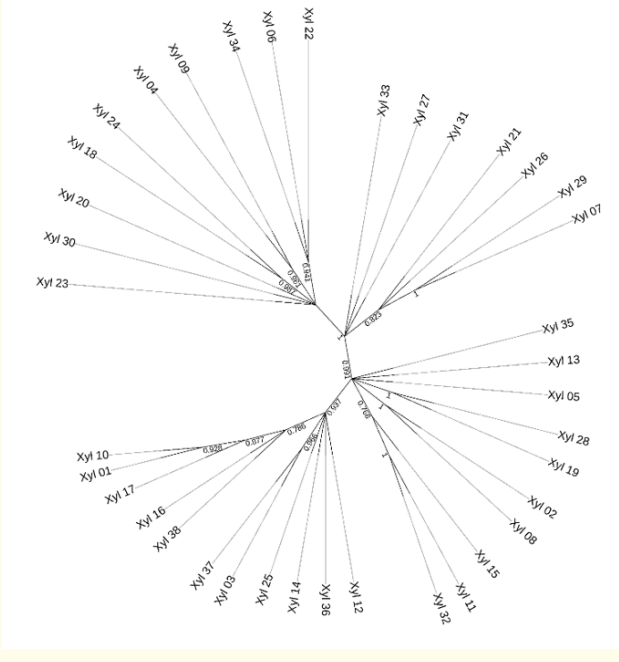


The alignment of the 38 sequences, followed by an analysis by Neighbor-Joining Tree, made it possible to establish relationships between the sequences. When Neighbor-Joining Tree (Figure 2) and proteins identity (Table 2) are compared it is possible to differentiate 6 groups with, at least, two proteins with identity higher than 70%. These groups are divided in branches with Xyl_23 to Xyl_22 and Xyl_33 to Xyl_07 (17 proteins with high identity); branch with Xyl_38 to Xyl_10 (5 proteins with high identity); Xyl_37 and Xyl_03; Xyl_11 and Xyl_32; Xyl_02 and Xyl_08; Xyl_28 and 19. Proteins Xyl_35, Xyl_13, Xyl_05, Xyl_15, Xyl_12, Xyl_36, Xyl_14 and Xyl_25 do not have high identity with any of the identified sequences (Table 2).

These groups, together with the degree of identity between the sequences, provide important relationships for knowing the diversity of proteins that may have different properties. Another analysis that contributed in this regard was the BLAST of the 38 sequences against the NCBI patent database. The results in table 4 demonstrated the presence of sequences with up to 68.79% identity (Xyl_07 and AKY00359.1). The observed values of identity are not high and do not exclude future studies with the herein identified sequences. In this sense, sequences obtained by metagenomic analysis of rumen can contribute in an important way to the discovery of new molecules with biotechnological potential.

Figure 2: Neighbor-Joining Tree inferred after ClustalW alignment of the 38 identified proteins.

The percentage of replicate trees in which the associated proteins clustered together in the bootstrap test (1000 replicates) are shown next to the branches.



ID	Xyl_01	Xyl_02	Xyl_03	Xyl_04	Xyl_05	Xyl_06	Xyl_07	Xyl_08	Xyl_09	Xyl_10	Xyl_11	Xyl_12	Xyl_13	Xyl_14	Xyl_15	Xyl_16	Xyl_17
Xyl_38	84	37	43	39	39	39	38	37	38	85	37	51	38	39	36	86	85
Xyl_37	41	34	92	36	34	35	35	33	36	42	33	35	34	33	34	41	41
Xyl_36	28	33	23	32	30	26	27	33	26	27	31	28	28	29	33	28	30
Xyl_35	35	50	32	54	49	51	51	49	50	33	46	33	50	27	47	35	36
Xyl_34	40	65	35	88	57	90	78	64	87	38	56	37	58	31	55	40	41
Xyl_33	39	66	34	70	64	71	79	66	73	39	57	37	61	31	57	38	40
Xyl_32	37	62	34	60	61	58	64	60	60	36	92	35	56	31	61	36	38
Xyl_31	38	65	34	79	58	83	78	64	80	38	55	37	59	29	55	37	40
Xyl_30	40	65	34	87	66	90	78	64	88	39	56	37	59	30	56	39	41
Xyl_29	38	67	33	70	62	69	95	67	71	37	60	35	62	29	60	36	39
Xyl_28	38	61	32	61	63	61	63	62	62	37	57	36	57	32	56	35	38
Xyl_27	38	66	34	75	70	77	73	66	76	40	59	37	59	30	57	38	38
Xyl_26	39	70	34	79	62	77	86	68	80	39	62	36	61	30	58	38	40
Xyl_25	44	33	38	36	38	36	37	35	36	42	36	36	38	41	37	43	42
Xyl_24	40	65	34	83	51	84	72	62	85	40	54	37	57	31	49	40	41
Xyl_23	40	65	35	86	62	90	79	66	89	40	57	37	59	31	56	39	41
Xyl_22	39	64	35	88	55	92	77	64	84	40	55	35	58	32	54	39	40
Xyl_21	37	66	33	75	64	75	78	66	78	37	56	38	59	30	57	36	37
Xyl_20	39	65	34	89	67	89	78	64	90	40	57	36	57	31	55	37	38
Xyl_19	34	60	32	59	62	60	62	60	60	35	58	38	56	31	58	34	35
Xyl_18	39	65	32	86	58	85	78	64	84	39	55	38	58	31	54	39	38
Xyl_17	90	37	40	40	39	40	39	37	38	85	36	49	36	40	37	80	
Xyl_16	85	34	40	37	38	37	35	35	37	79	35	48	36	38	35		80
Xyl_15	36	60	32	56	68	55	58	58	56	35	60	36	60	32		35	37
Xyl_14	39	31	32	32	33	31	28	30	30	38	33	34	32		32	38	40
Xyl_13	37	55	34	58	58	55	62	55	60	36	53	36		32	60	35	36
Xyl_12	49	35	35	35	37	35	35	36	36	48	36		36	34	36	48	49
Xyl_11	36	61	34	56	57	55	57	60	56	36		36	55	33	60	35	36
Xyl_10	92	35	38	38	38	38	37	35	37		36	48	36	38	35	79	85
Xyl_09	39	65	35	90	58	84	78	64		37	56	36	60	38	32	37	38
Xyl_08	36	93	33	64	71	63	65		64	36	60	37	57	31	58	34	38
Xyl_07	38	65	33	79	62	77		65	78	37	57	35	62	29	58	36	40
Xyl_06	38	64	33	86	67		77	64	85	38	54	36	58	31	55	37	40
Xyl_05	40	56	33	56		56	57	56	58	38	57	37	60	33	69	38	39
Xyl_04	39	65	34		67	86	79	64	90	38	56	35	58	32	56	37	40
Xyl_03	40	33		34	33	33	33	32	35	38	34	35	34	32	32	40	40
Xyl_02	36		34	65	70	64	65	93	65	35	61	35	58	31	61	34	37
Xyl_01		36	39	38	39	38	37	36	38	88	36	48	37	40	36	82	87

38	36	39	37	39		41	47	38	39	36	38	40	38	37	38	40	35														
34	33	37	35	35	36	37	41	35	36	32	34	36	35	33	35	36	32														
27	26	32	28	25	25	24	27	27	26	28	27	25	28	31	28	26	27														
49	47	49	48	49	51	49	31	54	48	48	49	52	49	49	56	50															
89	60	91	77	92	91	88	36	81	77	62	69	92	81	59	73																
71	61	72	69	72	78	71	37	77	79	63	79	76	77	60		73	50														
58	60	59	62	60	60	56	37	64	61	60	62	60	60		60	59	49														
80	60	82	73	80	85	77	37	77	84	61	72	83		60	77	81	49														
90	61	92	76	87	92	88	37	81	78	62	71		83	60	74	91	50														
67	62	70	76	69	72	69	37	87	73	65		72	72	63	80	69	49														
61	80	62	62	60	63	60	36	63	61		65	63	61	60	63	62	48														
76	60	77	70	77	80	76	38	74		61	73	78	84	61	79	77	48														
81	61	81	81	79	81	77	38		74	63	87	81	77	64	77	81	54														
35	38	36	39	35	37	35		38	36	35	37	37	37	37	36	35	30														
91	57	88	70	84	91		36	77	77	60	68	90	78	56	71	88	49														
89	62	91	77	86		91	37	81	80	63	73	92	85	60	78	91	51														
85	58	86	75		86	84	35	79	77	60	68	88	80	60	72	92	49														
75	61	76		75	77	69	40	81	70	62	76	77	73	62	69	77	48														
89	61		76	86	91	88	36	81	77	62	70	93	82	59	72	91	49														
61		61	62	58	62	57	38	61	61	80	62	62	60	60	64	60	47														
	61	89	75	85	89	90	36	81	76	61	67	90	80	58	74	89	49														
38	35	38	37	40	41	40	44	40	38	38	38	41	40	38	39	41	36														
39	34	37	36	39	39	40	45	38	38	35	35	40	37	36	38	40	35														
54	58	55	57	54	56	48	38	58	57	56	60	56	55	61	57	55	47														
31	31	31	30	32	31	31	42	30	30	32	29	30	30	31	31	31	27														
55	55	55	56	55	59	55	38	61	56	56	61	59	59	56	61	56	48														
38	38	36	38	35	37	37	38	36	37	36	35	37	37	35	37	37	33														
55	58	57	56	55	57	53	37	62	59	58	60	57	55	92	57	56	46														
39	35	40	37	40	40	40	43	39	40	37	36	41	38	36	39	38	33														
84	60	90	78	84	89	85	36	80	76	62	71	88	80	60	73	87	50														
64	60	64	66	64	65	62	35	68	66	62	67	65	65	60	66	64	49														
78	62	78	78	77	80	72	37	86	74	63	94	79	78	64	80	78	51														
86	60	89	75	92	91	84	36	77	77	61	69	91	83	58	71	91	51														
57	56	56	56	55	59	50	39	55	61	63	62	66	58	56	64	57	49														
86	59	89	75	88	86	83	36	79	75	61	70	88	79	60	70	88	50														
32	32	34	33	35	35	34	40	34	34	32	33	34	34	34	34	35	32														
65	61	65	66	65	66	65	34	70	67	61	67	66	65	62	66	65	50														
37	34	38	36	37	40	39	44	39	37	37	37	40	37	37	39	38	35														
Xyl_18	Xyl_19	Xyl_20	Xyl_21	Xyl_22	Xyl_23	Xyl_24	Xyl_25	Xyl_26	Xyl_27	Xyl_28	Xyl_29	Xyl_30	Xyl_31	Xyl_32	Xyl_33	Xyl_34	Xyl_35														

Xyl_36	Xyl_37	Xyl_38	>70%
28	41	84	4
34	35	37	2
23	92	43	1
32	36	39	15
30	34	39	0
26	35	39	16
27	35	39	15
33	34	38	2
26	36	38	16
27	42	85	4
31	33	37	1
28	35	51	0
28	34	38	0
29	33	39	0
33	34	36	0
28	41	86	4
30	41	85	4
27	34	38	15
26	33	36	1
32	37	39	15
28	35	37	14
25	35	39	15
25	36	39	16
24	37	41	14
28	40	47	0
27	34	38	16
26	36	39	17
28	32	36	1
28	34	38	11
25	36	40	16
28	35	38	16
31	33	37	1
29	35	38	15
26	36	40	16
27	32	35	0
	23	31	0
23		43	1
31	43		4

Table 2: Percentage of identity between the 38 sequences identified in Metagenomic analysis and counting of sequences with identity higher than 70%.

In addition to GH10, the analysis of the domains present in the 38 sequences on the dbCAN platform made it possible to identify three other domains (Table 1). The CE1 domain, Carbohydrate esterase family 1, corresponds to an extensive class of enzymes capable of degrading xylan [36], and is present in most of the identified sequences: Xyl_02; Xyl_04; Xyl_05; Xyl_06; Xyl_07; Xyl_08; Xyl_09; Xyl_11; Xyl_15; Xyl_18; Xyl_19; Xyl_20; Xyl_21; Xyl_22; Xyl_23; Xyl_27; Xyl_28; Xyl_30; Xyl_31 e Xyl_34. The CBM6 domain, carbohydrate-binding modules, is common in xylanases, having an auxiliary role in the hydrolysis of insoluble substrates [37] and was identified in Xyl_36. Xyl_37 contains the GH43 domain, glycoside hydrolase, related on the degradation of hemicellulose and pectin [38]. In this sense, the presence of complementary domains can contribute to the choice of multifunctional molecules, something desired in biotechnological applications [39].

Additionally, ProtParam (Table 1) and Pairwise analysis (Table 3) show that the identified sequences are distinct based on amino acids residues number. According to table 3 the differences between amino acids residues number did not interfere in sequences identity (Table 2) and Neighbor-Joining Tree clustering (Figure 2). Besides, ProtParam returned the prediction of physical-chemical parameters of the identified sequences (Table 1). It was possible to observe the presence of 33 stable proteins (values lower than 40), and the unstable ones were considered inadequate for future studies. The aliphatic index is related to the thermostability of globular proteins, higher values indicate possibly more thermostable proteins [25]. This is an important characteristic when concerning biotechnological applications, therefore, sequences Xyl_01, Xyl_03, Xyl_10, Xyl_12, Xyl_16 and Xyl_17 presented values above 80, in the other hand, the lowest observed value was 62.59 for Xyl_05 (Table 1).

In addition, the analysis conducted on the webtool SignalP identified the presence of signal peptide in 31 of the 38 sequences (Table 1). The presence of signal peptide may hinder studies of cloning and expression of proteins in vectors, despite this, signal peptides can be removed when starting in vitro studies, so this is not a determining factor in discarding a particular sequence. Another point considered was the absence of transmembrane helices, since these structures make it difficult to express and purify proteins, in this sense, only 8 of the 38 sequences have this undesired characteristic (Table 1).

According to the excluding criteria (instability index and transmembrane helices) the sequences Xyl_05, Xyl_07, Xyl_14, Xyl_15, Xyl_16, Xyl_18, Xyl_21, Xyl_24, Xyl_29, Xyl_31, Xyl_33, Xyl_35 and Xyl_37 were considered not viable for cloning and expression. Instability index is an excluding criterion since stability is directly related to the production of heterologous enzymes, in systems like *E. coli*, and unstable proteins can not be applicated in biotechnological sectors. The presence of transmembrane helices in the prediction of the secondary structure may also be a determining factor that makes heterologous production unfeasible, since these

Xyl_16	Xyl_15	Xyl_14	Xyl_13	Xyl_12	Xyl_11	Xyl_10	Xyl_09	Xyl_08	Xyl_07	Xyl_06	Xyl_05	Xyl_04	Xyl_03	Xyl_02	Xyl_01	
68	217	223	211	192	223	44	217	218	212	213	201	214	221	223		**Xyl_01**
227	256	480	151	228	256	223	247	50	238	250	262	248	233		13.8	**Xyl_02**
225	228	241	224	240	230	224	229	230	226	230	222	231		14.0	13.6	**Xyl_03**
214	279	474	151	229	291	211	73	255	147	98	262		14.2	13.8	13.9	**Xyl_04**
211	218	458	145	209	242	206	257	261	241	257		14.0	13.8	14.8	13.3	**Xyl_05**
214	283	475	149	230	293	212	109	257	161		14.2	9.8	14.2	14.1	14.0	**Xyl_06**
222	253	484	135	229	265	214	150	240		12.1	14.0	11.2	14.1	14.0	13.6	**Xyl_07**
226	262	479	151	223	264	214	256		13.9	14.4	14.8	13.9	14.0	6.8	13.6	**Xyl_08**
219	274	477	144	228	289	216		14.0	11.7	10.3	14.0	8.3	14.0	13.8	14.0	**Xyl_09**
79	218	221	216	192	224		13.8	13.3	13.5	13.8	13.3	13.8	13.7	13.6	6.7	**Xyl_10**
228	230	475	165	225		13.8	14.8	14.5	14.5	15.0	14.4	14.7	14.0	14.4	13.9	**Xyl_11**
192	221	242	222		14.0	13.2	14.1	14.1	14.2	14.1	13.8	14.2	14.1	14.1	13.4	**Xyl_12**
219	141	231		14.2	11.7	13.7	11.2	11.6	11.1	11.5	11.5	11.3	13.9	11.5	13.8	**Xyl_13**
229	467		13.8	14.2	18.3	14.1	18.0	18.6	18.4	18.2	17.9	18.1	14.4	18.6	14.2	**Xyl_14**
222		18.3	11.3	13.8	13.7	13.8	14.8	14.6	13.8	15.1	13.7	14.9	13.8	14.7	13.7	**Xyl_15**
	13.8	14.0	13.7	13.4	13.9	8.8	13.7	13.6	13.8	13.7	13.6	13.6	13.7	13.7	8.2	**Xyl_16**
8.4	13.6	14.0	13.8	13.2	13.7	7.4	13.7	13.4	13.5	13.8	13.2	13.7	13.5	13.5	6.9	**Xyl_17**
13.6	15.1	18.3	11.5	14.2	14.9	13.7	10.6	14.5	11.4	9.8	14.1	9.8	14.2	14.3	13.9	**Xyl_18**
14.1	14.9	18.4	11.9	14.3	15.1	14.2	14.9	15.2	14.6	15.1	15.1	15.2	14.3	15.0	14.2	**Xyl_19**
13.6	15.0	18.0	11.4	14.2	15.1	13.6	8.0	14.2	11.6	8.4	14.2	8.6	14.2	13.9	13.8	**Xyl_20**
13.9	14.6	18.2	11.5	14.0	14.6	13.6	11.3	13.8	11.8	12.4	14.1	11.8	14.2	13.8	13.8	**Xyl_21**
13.7	14.5	17.8	11.4	14.2	14.0	13.8	9.4	13.4	11.3	7.0	13.9	8.6	14.1	13.2	13.9	**Xyl_22**
13.3	14.2	18.2	11.0	13.7	14.2	13.4	8.4	13.7	11.0	7.4	13.6	9.2	13.8	13.5	13.7	**Xyl_23**
13.7	12.7	15.7	11.6	14.2	12.2	13.6	7.8	11.8	10.6	8.3	11.7	8.3	14.1	11.5	13.8	**Xyl_24**
13.6	13.8	13.4	13.0	14.0	13.9	14.1	13.9	13.9	13.7	13.8	13.2	13.9	13.8	14.0	13.8	**Xyl_25**
13.5	12.3	16.9	10.8	13.8	12.6	13.3	10.0	12.0	8.1	10.5	12.0	9.6	13.8	11.9	13.5	**Xyl_26**
14.0	15.0	18.5	11.2	14.2	14.8	13.9	12.3	14.2	12.3	12.2	14.2	12.7	14.0	13.8	14.0	**Xyl_27**
14.2	14.8	18.4	11.8	14.3	15.0	14.0	15.0	14.8	14.4	15.0	14.7	15.0	14.3	14.9	14.1	**Xyl_28**
13.8	11.0	14.6	11.1	14.1	11.0	13.5	10.1	10.2	4.1	10.1	10.8	10.0	13.9	10.2	13.6	**Xyl_29**
13.5	15.0	18.2	11.2	14.1	15.0	13.6	9.2	14.3	11.5	8.1	14.4	9.5	14.1	14.1	13.9	**Xyl_30**
14.1	15.0	18.3	11.4	14.2	14.9	13.9	11.2	14.2	12.1	10.6	13.7	11.5	14.0	14.0	14.0	**Xyl_31**
13.7	10.9	15.8	11.4	13.8	5.8	13.6	11.7	12.0	11.3	11.9	11.7	11.8	13.7	11.7	13.9	**Xyl_32**
13.9	11.6	14.8	11.5	14.2	11.6	13.6	10.2	10.9	8.6	10.0	10.8	10.2	14.0	10.6	13.7	**Xyl_33**
13.6	14.9	18.1	11.4	14.1	14.8	13.6	9.8	14.4	11.5	7.7	14.1	9.0	14.1	14.1	13.9	**Xyl_34**
14.0	11.6	14.4	11.8	14.0	11.8	13.6	11.6	11.6	11.7	11.6	11.3	11.6	13.6	11.6	13.9	**Xyl_35**
15.1	18.8	18.8	14.7	15.0	18.7	15.0	18.7	18.8	18.8	18.7	18.5	18.8	15.3	18.8	15.2	**Xyl_36**
13.8	13.8	14.5	13.6	14.1	14.1	13.6	14.1	14.0	14.4	14.2	13.6	14.2	6.3	14.1	13.7	**Xyl_37**
6.5	12.7	12.8	12.6	11.6	12.7	6.9	12.6	12.2	12.2	12.7	12.5	12.5	12.3	12.5	7.2	**Xyl_38**

Citation: Adriano Marques Gonçalves, *et al.* "Diversity of Glycoside Hydrolase 10 Family Xylanases Found in Rumen Metagenome and Selection of Sequences with Biotechnological Potential". *Acta Scientific Microbiology* 4.7 (2021): 76-89.

Xyl_30	Xyl_29	Xyl_28	Xyl_27	Xyl_26	Xyl_25	Xyl_24	Xyl_23	Xyl_22	Xyl_21	Xyl_20	Xyl_19	Xyl_18	Xyl_17
210	212	215	209	208	209	212	200	211	219	214	223	214	49
245	120	280	239	156	230	153	228	222	237	248	269	254	215
229	225	241	225	219	221	232	219	228	226	230	235	236	217
93	114	280	175	100	223	71	92	77	175	75	280	101	205
261	130	256	241	164	204	167	229	237	259	259	258	263	198
69	117	277	165	114	221	74	57	52	186	75	271	104	209
151	17	258	184	69	217	127	136	148	154	157	255	153	207
255	124	281	245	165	227	167	231	227	242	257	274	266	209
90	110	277	172	101	224	64	76	98	154	70	276	116	210
210	215	220	212	208	208	213	196	210	218	214	224	215	58
285	143	278	273	184	217	191	258	260	286	284	268	291	220
230	227	217	226	220	230	231	211	231	222	230	215	229	186
144	136	148	141	128	203	154	136	150	146	149	154	151	216
477	250	480	474	376	215	304	456	460	476	474	479	481	221
280	136	278	273	179	212	187	254	253	265	281	266	289	214
210	226	224	215	212	208	215	199	214	225	216	222	215	76
205	206	215	207	201	210	208	190	206	219	210	222	211	
77	125	281	176	96	224	41	72	91	180	83	279		13.8
265	128	138	271	183	212	177	247	252	265	268		15.1	14.2
57	116	273	166	95	223	52	62	82	170		15.0	8.9	13.6
168	92	280	212	97	214	132	156	159		11.6	15.1	12.0	13.7
76	117	251	144	108	222	67	80		11.3	8.8	14.5	9.1	13.8
52	96	246	136	92	205	37		8.5	11.3	7.8	14.3	8.3	13.4
50	120	176	104	91	211		6.1	7.9	10.8	7.1	12.1	6.3	13.7
222	204	215	216	208		13.5	13.4	13.8	13.8	13.9	13.9	13.8	13.8
93	47	189	130		13.4	8.9	9.5	10.0	9.5	9.2	12.7	9.3	13.3
159	107	281		10.8	13.6	9.8	11.3	11.5	13.3	12.1	15.0	12.3	13.8
271	131		14.9	12.8	13.7	12.2	14.0	14.2	15.2	14.8	11.3	14.9	13.9
108		10.8	9.8	6.7	13.4	10.1	9.1	10.0	8.9	10.0	10.8	10.3	13.4
	9.6	14.8	12.0	9.5	13.8	7.0	7.3	8.5	11.5	7.5	14.9	8.7	13.7
10.2	10.1	14.9	9.9	10.3	13.7	9.3	9.5	10.7	13.0	10.5	14.6	10.8	13.8
11.6	10.3	12.1	11.7	11.4	13.5	11.7	11.5	11.7	11.6	11.9	11.9	11.9	13.5
9.9	8.3	11.2	9.2	9.0	13.3	10.2	8.7	10.0	10.5	10.3	11.2	10.2	13.5
8.0	10.1	14.9	12.5	9.5	13.8	7.1	7.8	6.8	11.4	8.3	15.0	8.5	13.8
11.4	12.0	11.6	11.9	11.0	13.8	11.7	10.7	11.4	12.0	11.5	11.6	11.8	13.7
18.8	14.7	18.9	18.7	17.2	14.7	16.0	18.7	18.4	18.8	18.7	18.7	18.7	15.1
14.1	14.4	14.1	14.1	13.9	13.9	14.1	13.5	14.2	14.4	14.1	14.2	14.3	13.3
12.5	12.2	13.0	12.7	12.4	12.4	12.5	12.6	12.6	12.6	12.5	12.9	12.6	6.8

Citation: Adriano Marques Gonçalves., *et al.* "Diversity of Glycoside Hydrolase 10 Family Xylanases Found in Rumen Metagenome and Selection of Sequences with Biotechnological Potential". *Acta Scientific Microbiology* 4.7 (2021): 76-89.

	Xyl_31	Xyl_32	Xyl_33	Xyl_34	Xyl_35	Xyl_36	Xyl_37	Xyl_38
Xyl_01	213	205	209	210	218	267	225	48
Xyl_02	243	155	131	248	155	533	238	175
Xyl_03	228	215	229	229	228	270	38	182
Xyl_04	141	166	117	85	163	544	232	169
Xyl_05	248	151	127	258	154	540	221	168
Xyl_06	117	169	113	65	161	550	235	172
Xyl_07	162	146	77	153	164	550	235	167
Xyl_08	251	164	135	257	160	536	236	168
Xyl_09	135	165	113	96	163	543	232	173
Xyl_10	212	208	212	210	217	265	223	47
Xyl_11	292	35	160	287	176	547	242	175
Xyl_12	225	210	224	229	232	264	241	147
Xyl_13	144	142	141	148	156	252	223	173
Xyl_14	479	298	257	475	249	504	263	173
Xyl_15	279	138	154	285	166	552	226	177
Xyl_16	219	206	216	211	229	271	235	42
Xyl_17	209	200	202	205	213	265	217	45
Xyl_18	137	175	118	78	170	548	241	172
Xyl_19	266	163	143	272	163	554	233	177
Xyl_20	125	168	117	67	161	550	233	170
Xyl_21	193	158	122	167	167	547	240	173
Xyl_22	119	166	110	49	160	490	236	170
Xyl_23	97	162	86	64	140	531	210	169
Xyl_24	97	167	117	51	166	315	236	170
Xyl_25	217	196	207	220	220	258	229	158
Xyl_26	116	149	93	96	144	383	223	174
Xyl_27	104	163	89	167	170	547	235	171
Xyl_28	273	164	145	278	167	549	240	179
Xyl_29	114	119	73	118	171	255	248	167
Xyl_30	117	163	108	64	159	550	234	168
Xyl_31		164	94	130	172	555	244	173
Xyl_32	11.8		136	167	140	322	212	177
Xyl_33	9.4	11.0		112	165	270	238	166
Xyl_34	11.1	11.8	10.2		161	548	236	170
Xyl_35	12.1	10.7	11.7	11.5		257	246	174
Xyl_36	18.8	16.4	15.3	18.8	14.7		270	223
Xyl_37	14.2	13.5	14.1	14.2	14.1	15.3		178
Xyl_38	12.7	12.7	12.5	12.6	12.6	14.1	12.3	

Table 3: Estimates of divergence between sequences, the number of amino acid differences per sequence from between sequences are shown. Standard error estimates are shown above the diagonal and all ambiguous positions were removed for each sequence pair (pairwise deletion).

ID	Identity	Acession
Xyl_01	57,10%	AKY00358.1
Xyl_02	67,49%	AKY00359.1
Xyl_03	51,72%	ABZ41703.1
Xyl_04	66,33%	AKY00359.1
Xyl_05	64,42%	AKY00359.1
Xyl_06	66,24%	AKY00359.1
Xyl_07	68,79%	AKY00359.1
Xyl_08	67,10%	AKY00359.1
Xyl_09	66,33%	AKY00359.1
Xyl_10	54,91%	AKY00358.1
Xyl_11	64,46%	AKY00359.1
Xyl_12	49,43%	AKY00358.1
Xyl_13	57,01%	AKY00359.1
Xyl_14	42,53%	AER76185.1
Xyl_15	66,89%	AKY00359.1
Xyl_16	55,65%	AKY00374.1
Xyl_17	56,03%	AKY00358.1
Xyl_18	64,99%	AKY00390.1
Xyl_19	63,61%	AKY00359.1
Xyl_20	65,47%	AKY00359.1
Xyl_21	67,38%	AKY00359.1
Xyl_22	65,58%	AKY00359.1
Xyl_23	67,18%	AKY00359.1
Xyl_24	62,25%	AKY00359.1
Xyl_25	53,45%	ACP87342.1
Xyl_26	68,41%	AKY00390.1
Xyl_27	66,24%	AKY00359.1
Xyl_28	64,48%	AKY00359.1
Xyl_29	66,27%	AKY00359.1
Xyl_30	66,38%	AKY00359.1
Xyl_31	64,74%	AKY00359.1
Xyl_32	69,34%	AKY00390.1
Xyl_33	62,82%	AKY00359.1
Xyl_34	66,47%	AKY00359.1
Xyl_35	48,10%	AKY00359.1
Xyl_36	48,48%	AER76187.1
Xyl_37	59,87%	AKY00367.1
Xyl_38	57,70%	AKY00358.1

Table 4: Percentage of identity between the 38 sequences identified in Metagenomic analysis and NCBI Patent database.

proteins normally participate in specific processes and functions in biological systems [40], consequently making it difficult to obtain their soluble form.

In this sense, based on the set of in silico analyzes performed, 25 of the 38 sequences showed good characteristics for study, since the presence of the signal peptide is not an excluding criterion. Therefore, the choice for future in vitro analyzes was made based on the desirable characteristics present in the aforementioned sequences: absence of transmembrane helices, structural stability, thermal stability, identity with biotechnological application molecules, presence of domains of interest and difference in their amino acid sequences.

It is important to emphasize that one of the selected sequences has already presented promising laboratory results regarding its applicability and bifunctionality [41]. The other selected molecules will be investigated in future studies. Additionally, our work also contributes with a workflow that can be used to select enzymes of interest, starting from a database (Figure 1). It is important to note that this workflow can be used for other proteins and types of sequences once it is possible to analyze general characteristics of amino acid sequences.

Conclusion

The present work analyzed the diversity of sequences containing the GH10 domain present in the rumen of Nelore cattle. In addition to the identification of the 38 sequences, a workflow of *in silico* methodologies was suggested to select sequences that will guide future *in vitro* studies.

Acknowledgements

This work was supported by grants from grant 2018/12885-0, São Paulo Research Foundation (FAPESP), CNPq, and CAPES.

Conflict of Interest

The authors declare that there is not any conflict of interest within this paper.

Bibliography

1. Madhavan Aravind., *et al*. "Metagenome Analysis: A Powerful Tool for Enzyme Bioprospecting". *Applied Biochemistry and Biotechnology* 183.2 (2017): 636-651.

2. Handelsman Jo. "Metagenomics: Application of Genomics to Uncultured Microorganisms". *Microbiology and Molecular Biology Reviews: MMBR* 68.4 (2004): 669-685.

3. von Mering C., *et al*. "Quantitative Phylogenetic Assessment of Microbial Communities in Diverse Environments". *Science (New York, N.Y.)* 315.5815 (2007): 1126-1130.

4. Chistoserdova Ludmila. "Recent Progress and New Challenges in Metagenomics for Biotechnology". *Biotechnology Letters* 32.10 (2010): 1351-1359.

5. Sharma Hem Kanta., *et al*. "Biological Pretreatment of Lignocellulosic Biomass for Biofuels and Bioproducts: An Overview". *Waste and Biomass Valorization* 10.2 (2019): 235-251.

6. Lazuka Adèle., *et al*. "Anaerobic Lignocellulolytic Microbial Consortium Derived from Termite Gut: Enrichment, Lignocellulose Degradation and Community Dynamics". *Biotechnology for Biofuels* 11.1 (2018): 284.

7. Moraïs Sarah and Itzhak Mizrahi. "The Road Not Taken: The Rumen Microbiome, Functional Groups, and Community States". *Trends in Microbiology* 27.6 (2019): 538-549.

8. Deusch Simon., *et al*. "A Structural and Functional Elucidation of the Rumen Microbiome Influenced by Various Diets and Microenvironments". *Frontiers in Microbiology* 8 (2017).

9. Gruninger Robert J., *et al*. "Contributions of a Unique β-Clamp to Substrate Recognition Illuminates the Molecular Basis of Exolysis in Ferulic Acid Esterases". *The Biochemical Journal* 473.7 (2016): 839-849.

10. Ogunade Ibukun M., *et al*. "Effects of Live Yeast on Differential Genetic and Functional Attributes of Rumen Microbiota in Beef Cattle". *Journal of Animal Science and Biotechnology* 10.1 (2019): 68.

11. Collins T., *et al*. "Xylanases, Xylanase Families and Extremophilic Xylanases". Undefined (2005).

12. Gao Dahai., *et al*. "Hemicellulases and Auxiliary Enzymes for Improved Conversion of Lignocellulosic Biomass to Monosaccharides". *Biotechnology for Biofuels* 4 (2011): 5.

13. Beg Q K., *et al*. "Microbial Xylanases and Their Industrial Applications: A Review". *Applied Microbiology and Biotechnology* 56.3-4 (2001): 326-338.

14. de Freitas Caroline., *et al*. "Xylooligosaccharides Production Process from Lignocellulosic Biomass and Bioactive Effects". *Bioactive Carbohydrates and Dietary Fibre* 18 (2019): 100184.

15. Venkateswar Rao Linga., *et al*. "Bioconversion of Lignocellulosic Biomass to Xylitol: An Overview". *Bioresource Technology* 213 (2016): 299-310.

16. Jhamb Kamna and Debendra K Sahoo. "Production of Soluble Recombinant Proteins in Escherichia Coli: Effects of Process Conditions and Chaperone Co-Expression on Cell Growth and Production of Xylanase". *Bioresource Technology* 123 (2012): 135-143.

17. Hannig G and S C Makrides. "Strategies for Optimizing Heterologous Protein Expression in Escherichia Coli". *Trends in Biotechnology* 16.2 (1998): 54-60.

18. Peti Wolfgang and Rebecca Page. "Strategies to Maximize Heterologous Protein Expression in Escherichia Coli with Minimal Cost". *Protein Expression and Purification* 51.1 (2007): 1-10.

19. Leow Thean Chor., *et al*. "A Thermoalkaliphilic Lipase of Geobacillus Sp. T1". *Extremophiles* 11.3 (2007): 527-535.

20. Chang Catherine Ching Han., *et al*. "Bioinformatics Approaches for Improved Recombinant Protein Production in Escherichia Coli: Protein Solubility Prediction". *Briefings in Bioinformatics* 15.6 (2014): 953-962.

21. Gaspar Paulo., *et al*. "EuGene: Maximizing Synthetic Gene Design for Heterologous Expression". *Bioinformatics* 28.20 (2012): 2683-2684.

22. Powell Sean., *et al*. "EggNOG v4.0: Nested Orthology Inference across 3686 Organisms". *Nucleic Acids Research* 42 (2014): D231-239.

23. Thompson J D., *et al*. "CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice". *Nucleic Acids Research* 22.22 (1994): 4673-4680.

24. Finn Robert D., *et al*. "Pfam: The Protein Families Database". *Nucleic Acids Research* 42 (2014): D222-230.

25. Gasteiger Elisabeth., *et al*. "Protein Identification and Analysis Tools on the ExPASy Server". *The Proteomics Protocols Handbook*, edited by John M. Walker, Humana Press (2005): 571-607.

26. Almagro Armenteros José Juan., *et al*. "SignalP 5.0 Improves Signal Peptide Predictions Using Deep Neural Networks". *Nature Biotechnology* 37.4 (2019): 420-423.

27. Krogh A., *et al*. "Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes". *Journal of Molecular Biology* 305.3 (2001): 567-580.

28. Huang Le., *et al*. "DbCAN-Seq: A Database of Carbohydrate-Active Enzyme (CAZyme) Sequence and Annotation". *Nucleic Acids Research* 46.D1 (2018): D516-521.

29. Altschul S F., *et al*. "Basic Local Alignment Search Tool". *Journal of Molecular Biology* 215.3 (1990): 403-410.

30. Larkin M A., *et al*. "Clustal W and Clustal X Version 2.0". *Bioinformatics (Oxford, England)* 23.21 (2007): 2947-2948.

31. Kumar Sudhir., *et al*. "MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms". *Molecular Biology and Evolution* 35.6 (2018): 1547-1549.

32. Saitou N and M Nei. "The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees". *Molecular Biology and Evolution* 4.4 (1987): 406-425.

33. Felsenstein Joseph. "CONFIDENCE LIMITS ON PHYLOGENIES: AN APPROACH USING THE BOOTSTRAP". *International Journal of Organic Evolution* 39.4 (1985): 783-791.

34. Masatoshi Nei and Sudhir Kumar. Molecular Evolution and Phylogenetics. Oxford University Press (2000).

35. Letunic Ivica and Peer Bork. "Interactive Tree Of Life (ITOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation". *Nucleic Acids Research* (2021).

36. Kmezik Cathleen., *et al*. "A Polysaccharide Utilization Locus from the Gut Bacterium Dysgonomonas Mossii Encodes Functionally Distinct Carbohydrate Esterases". *Journal of Biological Chemistry* 296 (2021): 100500.

37. Qiao Weibo., *et al*. "Biochemical Characterization of a Novel Thermostable GH11 Xylanase with CBM6 Domain from Caldicellulosiruptor Kronotskyensis". *Journal of Molecular Catalysis B: Enzymatic* 107 (2014): 8-16.

38. Mewis Keith., *et al*. "Dividing the Large Glycoside Hydrolase Family 43 into Subfamilies: A Motivation for Detailed Enzyme Characterization". *Applied and Environmental Microbiology* 82.6 (2016): 1686-1692.

39. Glasgow Evan., *et al*. "Multifunctional Cellulases Are Potent, Versatile Tools for a Renewable Bioeconomy". *Current Opinion in Biotechnology* 67 (2021): 141-148.

40. Möller Steffen., *et al*. "Evaluation of Methods for the Prediction of Membrane Spanning Regions". *Bioinformatics* 17.7 (2001): 646-653.

41. Pavarina Gabriella Cavazzini., *et al*. "Characterization of a New Bifunctional Endo-1,4-β-Xylanase/Esterase Found in the Rumen Metagenome". *Scientific Reports* 11.1 (2021): 10440.

Volume 4 Issue 7 July 2021