IN SILICO MINING FOR ALKALINE ENZYMES FROM METAGENOMIC DNA DATA OF GUT MICROBES OF THE LOWER TERMITE *Coptotermes gestroi* IN VIETNAM

Nguyen Minh Giang[1]*, Do Thi Huyen[2], Truong Nam Hai[2]

[1]Ho Chi Minh University of Pedagogy
[2]Institute of Biotechnology, Vietnam Academy of Science and Techology

ASBTRACT: The high alkaline proteases, lipases, cellulases and hemicellulases are important enzymes in research and industries. In this study, using the Alcapred software, the metagenomic DNA sequences of the gut flora of *Coptotermes gestroi* were analyzed to identify the enzymes that were specifically adapted to alkaline condition. The results show that 737 of 943 ORFs (accounting for 72%) encoded proteases, 154 of 214 ORFs (holding 78%) encoded lipases and 338 of 575 ORFs (accounting for 59%) encoded cellulase and hemicelluase. All those enzymes were predicted to be alkaline enzymes. This study provide an overview picture of the alkaline enzyme groups of the gut flora of *C. gestroi*, and provide a good database for mining, isolation of the genes to produce recombinant enzymes.

Keywords: Coptotermes gestroi, alkaline enzyme, cellulase, gut, hemicellulase, lipase, metagenome, protease.

INTRODUCTION

Termites contribute substantially to the turnover of carbon and nitrogen in tropical ecosystems. Their diet consists exclusively of lignocellulose of various stages of decomposition, ranging from sound wood to humus. The digestion of this recalcitrant diet relies on the metabolic activities of a dense and diverse intestinal microbiota. In the gut of many lower termite *Zootermopsis nevadensis*, *Reticulitermes lucifugus* and *R. flvipes,* the pH was neutral to slightly acidic throughout, ranging from 5.5 to 7.5. In many higher termites, the hindgut is compartmentalized to form several consecutive microbial bioreactors [1], and the pH of the special anterior hindgut is highly alkaline. In soil-feeding Termitinae such as *Nasutitermes nigriceps* and *N. corniger*, the pH increases sharply at the mixed segment and reaches its maximum pH of 12 in hindgut [3].

Since the hindgut of some termites are extremely alkaline pH, we assumed that termite gut is a highly convenient mini ecosystem to exploit the alkaline enzymes of the intestinal microbiota. Therefore, we conducted a survey for alkaline enzymes such as proteases, lipases, cellulases and hemicellulases in the DNA sequences metagenome of microbiota in the gut of *C. gestroi*. If we could find many novel enzymes, they might be interested in both academic and industrial aspects.

MATERIALS AND METHODS

Genomic DNA was extracted from gut flora of *C. gestroi* extracted from the free-living microorganism in gut of *C. gestroi* collected from wood-nesting colonies in Hanoi and Hung Yen province in Vietnam according to the method described by Sambrook et al. [21] and sequenced using HiSeq2000 sequencing system (Illumina, San Diego, USA). Metagenomic DNA sequence data was analyzed using standard bioinformatics approach. A range of bioinformatic softwares such as BLAST, MEGAN, SOAP and the data in NCBI, KEGG, eggNOG were used to identify and account for ORFs [6].

Alkaline and acidic enzymes analysis: Based on the predicted ORFs annotated using KEGG and eggNOG, we used a sequence-based tool to discriminate acidic and alkaline enzymes. A feature selection technique was used to pick out a number of informative features. Based on these features, the support vector machine (SVM) analysis was performed to establish a prediction model. Prediction results demonstrate that the proposed method is reliable. Then, a free online database called AcalPred was built to provide a useful tool for basic academic study and industrial application of acidic and alkaline enzymes [12]. An overall accuracy of 96% was achieved, demonstrating that the proposed model is a powerful tool for the study on the adaptation of enzymes to acidic or alkaline environment.

RESULTS AND DISCUSSION

We present the metagenomic analysis of a large data set (5.4 Gb) generated by Illumina-based de novo sequencing of genomic DNA of the gut flora of lower termite, *C. gestroi*. To our best our knowledge, this study is the first successful application of high-throughput sequencing for the investigation of the gut flora

of the lower termite, *C. gestroi*.

Metagenomic sequence analysis of the genomic DNA (8.5 mg) extracted from the gut flora of *C. gestroi* Illumina platform yielded 5.6 Gb of sequence reads. Meta Gene Annotator [19] identified 125,431 putative ORFs. The functional profile of the metagenome of the flora was determined by the classification of predicted genes based on the eggNOG [20] and KEGG [16] databases. We found that the metagenomic sequences were distributed among typically prokaryotic eggNOG functional categories [20]. Among 125,431 ORFs, 36,477 ORFs were classified into enzyme families and 65,536 ORFs were predicted to be functional. The functional properties were determined using deeper levels of the eggNOG and KEGG classes [16, 20].

Putative alkaline enzymes in the gut flora of *C. gestroi* termite

We predicted alkaline and acidic enzymes such as proteases, lipases, cellulases and hemicellulases using Alcapred software. Among 125,341 ORFs, 943 were annotated to encode proteases, 214 encode lipases and 579 encode cellulases and hemicellulases (table 1).

Table 1. The total predicted ORF, protease, lipase, cellulase and hemicellulase of metagenome of the termite gut microbiota

Enzymes	Total ORF	Proteases	Lipases	Cellulases and hemicellulases
Number of sequences	125341	943	214	578

The number of proteinases sequences (934) was the highest. This is reasonable because proteases are present in all microorganisms in the gut, even flagellate protozoan symbionts [22]. Proteases are critically important in a diverse biological processes including the regulation of the metabolism of cells and are essential constituents of all forms of life on earth [23]. So it is reasonable that protease is the largest enzyme group in cells compared with other enzymes.

In the gut of low and high termites, the most outstanding microbe in the community is bacteria. They dominate not only on the

intestinal wall of termite gut but also symbiotic flagellate or protozoa [26]. Metagenomic DNA sequence analysis of the gut flora of *C. gestroi* revealed that 80% of the total ORFs are belong to bacteria.

Lignocellulose is main constituent of termite food with high percentage of cellulose and hemicellulose. Trend is emerging suggesting collaboration among termite-derived genes expressed in the salivary gland/foregut, midgut and symbiont genes expressed in the hindgut. However, many studies demonstrated that cellulolytic and hemicellulolytic enzymes from microorganism in termite gut is quiet rich

[11] such as: 73% cellulases and xylanases were known in *Microcerotermes* sp. [27]; the series of cellulase in the foregut/salivary gland; and a rich diversity of derived GHFs in the hindgut. Of the various exoglucanases are undeniably the most diverse [23]. In our results, we identified 578 cellulases and hemicellulases sequences, which is just behind the number of proteases and much higher than the number of lipase sequences (214)

Putative alkaline protease and lipase

For the sequnce data encoding different enzyme groups, we used the Alcapred software to predict proteases and lipases that are tolerant to alkaline condition (table 2).

As shown in table 2, 737 (78%) out of 943 proteinase ORFs are alkaline proteases and 154 (72%) of 214 lipase ORFs are alkaline lipases.

Table 2. Summary of predicted alkaline and acidic protease and lipase in the gut of *C. gestroi*

Enzymes	Total sequences	Alkaline sequences	Percentage (%)	Acidic sequences	Percentage
Protease	943	737	78%	206	22%
Lipase	214	154	72%	60	28%

The hindgut of termites is the anaerobic fermentation tank containing a variety of different microorganisms having various enzymes that tolerate alkaline environment. Origin and distribution of proteases have been studied in detail in higher termites [REF]. However, little information is available about proteases in lower termites [24]. Only few publications about the proportion of alkaline proteases and lipases in the gut flora of the termites. Alkaline proteases and alkaline lipases were found originated from higher termite, *Nasutitermes corniger* and its enteric flora [10] . Our present results give the overall clarification of the high alkaline proteases and lipases of the gut flora of lower termite, *C. gestroi*. Among the ORFs encoding alkaline enzymes, 175 of 737 alkaline proteases ORFs and 33 of 154 alkaline lipases ORFs have alkaline index of higher than 0.99 (table 3, 4).

Table 3. Summary of the gene sequences encoding proteases having alkaline index of 0.99

STT	Code gene	Alkaline index	Acidic index	Protease
1	GL0054846	1.000000	0.000000	regulator of sigma E protease
2	GL0054872	1.000000	0.000000	carboxyl-terminal processing protease
3	GL0054905	1.000000	0.000000	ATP-dependent Lon protease
4	GL0054967	1.000000	0.000000	putative protease
5	GL0055331	1.000000	0.000000	Lon-like ATP-dependent protease
6	GL0056476	1.000000	0.000000	regulator of sigma E protease
7	GL0056477	1.000000	0.000000	ATP-dependent Lon protease
8	GL0057752	1.000000	0.000000	ATP-dependent Lon protease
9	GL0057900	1.000000	0.000000	zinc protease
10	GL0058152	1.000000	0.000000	cell wall-associated protease
11	GL0058203	1.000000	0.000000	major intracellular serine protease
12	GL0058234	1.000000	0.000000	hydrogenase 1 maturation protease
13	GL0058638	1.000000	0.000000	ATP-dependent Lon protease
14	GL0058989	1.000000	0.000000	carboxyl-terminal processing protease
15	GL0059141	1.000000	0.000000	cell wall-associated protease
16	GL0059540	1.000000	0.000000	ATP-dependent Lon protease
17	GL0059860	1.000000	0.000000	ATP-dependent Lon protease
18	GL0060204	1.000000	0.000000	Lon-like ATP-dependent protease
19	GL0060412	1.000000	0.000000	carboxyl-terminal processing protease
20	GL0060449	1.000000	0.000000	regulator of sigma E protease

21	GL0060495	1.000000	0.000000	zinc protease
22	GL0060843	1.000000	0.000000	zinc protease
23	GL0060923	1.000000	0.000000	ATP-dependent Lon protease
24	GL0061012	1.000000	0.000000	ATP-dependent Lon protease
25	GL0061226	1.000000	0.000000	carboxyl-terminal processing protease
26	GL0061851	1.000000	0.000000	carboxyl-terminal processing protease
27	GL0061932	1.000000	0.000000	ATP-dependent Lon protease
28	GL0005083	1.000000	0.000000	ATP-dependent Lon protease
29	GL0062064	1.000000	0.000000	ATP-dependent Lon protease
30	GL0062383	1.000000	0.000000	carboxyl-terminal processing protease
31	GL0062512	1.000000	0.000000	zinc protease
32	GL0063609	1.000000	0.000000	putative metalloprotease
33	GL0063636	1.000000	0.000000	zinc protease
34	GL0064020	1.000000	0.000000	carboxyl-terminal processing protease
35	GL0064465	1.000000	0.000000	zinc protease
36	GL0064477	1.000000	0.000000	ATP-dependent Lon protease
37	GL0064544	1.000000	0.000000	zinc protease
38	GL0064904	1.000000	0.000000	regulator of sigma E protease
39	GL0065106	1.000000	0.000000	putative metalloprotease
40	GL0065431	1.000000	0.000000	tricorn protease
41	GL0066206	1.000000	0.000000	ATP-dependent Lon protease
42	GL0067180	1.000000	0.000000	carboxyl-terminal processing protease
43	GL0067487	1.000000	0.000000	tricorn protease
44	GL0068015	1.000000	0.000000	carboxyl-terminal processing protease
45	GL0068234	1.000000	0.000000	putative protease
46	GL0068607	1.000000	0.000000	carboxyl-terminal processing protease
47	GL0069912	1.000000	0.000000	regulator of sigma E protease
48	GL0070217	1.000000	0.000000	regulator of sigma E protease
49	GL0070881	1.000000	0.000000	carboxyl-terminal processing protease
50	GL0070891	1.000000	0.000000	ATP-dependent Lon protease
51	GL0071299	1.000000	0.000000	carboxyl-terminal processing protease
52	GL0071667	1.000000	0.000000	ATP-dependent Lon protease
53	GL0071998	1.000000	0.000000	ATP-dependent Lon protease
54	GL0072105	1.000000	0.000000	putative protease
55	GL0072492	1.000000	0.000000	putative protease
56	GL0072552	1.000000	0.000000	regulator of sigma E protease
57	GL0072785	1.000000	0.000000	zinc protease
58	GL0072957	1.000000	0.000000	carboxyl-terminal processing protease
59	GL0073275	1.000000	0.000000	putative protease
60	GL0073638	1.000000	0.000000	carboxyl-terminal processing protease
61	GL0073667	1.000000	0.000000	Lon-like ATP-dependent protease
62	GL0074179	1.000000	0.000000	tricorn protease
63	GL0074314	1.000000	0.000000	carboxyl-terminal processing protease
64	GL0074377	1.000000	0.000000	ATP-dependent Lon protease
65	GL0074576	1.000000	0.000000	carboxyl-terminal processing protease
66	GL0074589	1.000000	0.000000	subtilase-type serine protease
67	GL0075722	1.000000	0.000000	ATP-dependent Lon protease
68	GL0076003	0.999999	0.000001	ATP-dependent Lon protease
69	GL0076249	0.999999	0.000001	carboxyl-terminal processing protease
70	GL0077278	0.999999	0.000001	cell wall-associated protease
71	GL0077302	0.999999	0.000001	ATP-dependent Lon protease
72	GL0077595	0.999999	0.000001	ATP-dependent Lon protease
73	GL0078352	0.999999	0.000001	carboxyl-terminal processing protease
74	GL0078771	0.999999	0.000001	ATP-dependent Lon protease
75	GL0079155	0.999999	0.000001	carboxyl-terminal processing protease

76	GL0079245	0.999999	0.000001	ATP-dependent Lon protease
77	GL0080059	0.999999	0.000001	zinc protease
78	GL0080468	0.999999	0.000001	Lon-like ATP-dependent protease
79	GL0080662	0.999999	0.000001	carboxyl-terminal processing protease
80	GL0080682	0.999999	0.000001	carboxyl-terminal processing protease
81	GL0081284	0.999999	0.000001	regulator of sigma E protease
82	GL0081653	0.999999	0.000001	zinc protease
83	GL0081869	0.999999	0.000001	putative protease
84	GL0082183	0.999999	0.000001	subtilase-type serine protease
85	GL0082905	0.999998	0.000002	zinc protease
86	GL0082905	0.999998	0.000002	zinc protease
87	GL0006359	0.999998	0.000002	carboxyl-terminal processing protease
88	GL0006895	0.999998	0.000002	putative protease
89	GL0007102	0.999998	0.000002	putative protease
90	GL0085754	0.999998	0.000002	carboxyl-terminal processing protease
91	GL0085819	0.999998	0.000002	regulator of sigma E protease
92	GL0086188	0.999998	0.000002	putative protease
93	GL0086783	0.999998	0.000002	tricorn protease
94	GL0087433	0.999998	0.000002	zinc protease
95	GL0087476	0.999998	0.000002	putative protease
96	GL0088709	0.999998	0.000002	Lon-like ATP-dependent protease
97	GL0089124	0.999998	0.000002	hydrogenase 3 maturation protease
98	GL0089539	0.999998	0.000002	ATP-dependent Lon protease
99	GL0089692	0.999997	0.000003	regulator of sigma E protease
100	GL0090100	0.999997	0.000003	subtilase-type serine protease
101	GL0090101	0.999997	0.000003	subtilase-type serine protease
102	GL0090767	0.999997	0.000003	tricorn protease
103	GL0091105	0.999997	0.000003	putative protease
104	GL0091196	0.999997	0.000003	regulator of sigma E protease
105	GL0091332	0.999997	0.000003	ATP-dependent Lon protease
106	GL0092578	0.999997	0.000003	putative protease
107	GL0093355	0.999997	0.000003	ATP-dependent Lon protease
108	GL0093891	0.999997	0.000003	carboxyl-terminal processing protease
109	GL0094050	0.999997	0.000003	Lon-like ATP-dependent protease
110	GL0094306	0.999996	0.000004	putative protease
111	GL0094337	0.999996	0.000004	ATP-dependent Lon protease
112	GL0094379	0.999996	0.000004	spore protease
113	GL0094531	0.999996	0.000004	putative protease
114	GL0094533	0.999996	0.000004	putative protease
115	GL0095247	0.999996	0.000004	hydrogenase 3 maturation protease
116	GL0095457	0.999995	0.000005	Lon-like ATP-dependent protease
117	GL0096224	0.999995	0.000005	regulator of sigma E protease
118	GL0096533	0.999995	0.000005	regulator of sigma E protease
119	GL0097138	0.999995	0.000005	putative protease
120	GL0097401	0.999995	0.000005	putative protease
121	GL0097824	0.999995	0.000005	putative protease
122	GL0098212	0.999994	0.000006	carboxyl-terminal processing protease
123	GL0098352	0.999994	0.000006	putative protease
124	GL0098435	0.999994	0.000006	zinc protease
125	GL0099896	0.999994	0.000006	hydrogenase 3 maturation protease
126	GL0100379	0.999994	0.000006	ATP-dependent Lon protease
127	GL0100766	0.999994	0.000006	ATP-dependent Lon protease
128	GL0100872	0.999994	0.000006	putative protease
129	GL0100931	0.999993	0.000007	ATP-dependent Lon protease
130	GL0101204	0.999993	0.000007	ATP-dependent Lon protease

131	GL0101716	0.999993	0.000007	ATP-dependent Lon protease
132	GL0101933	0.999993	0.000007	putative metalloprotease
133	GL0102081	0.999993	0.000007	regulator of sigma E protease
134	GL0102493	0.999992	0.000008	ATP-dependent Lon protease
135	GL0102586	0.999992	0.000008	carboxyl-terminal processing protease
136	GL0102946	0.999991	0.000010	carboxyl-terminal processing protease
137	GL0103635	0.999991	0.000010	zinc protease
138	GL0103640	0.999990	0.000010	putative protease
139	GL0103741	0.999990	0.000010	ATP-dependent Lon protease
140	GL0104021	0.999989	0.000011	ATP-dependent Lon protease
141	GL0008440	0.999989	0.000011	Lon-like ATP-dependent protease
142	GL0104067	0.999989	0.000011	muramoyltetrapeptide carboxypeptidase
143	GL0104382	0.999988	0.000012	D-alanyl-D-alanine carboxypeptidase
144	GL0105276	0.999988	0.000012	aminoacylhistidine dipeptidase
145	GL0105590	0.999987	0.000013	tripeptide aminopeptidase
146	GL0105694	0.999987	0.000013	aminoacylhistidine dipeptidase
147	GL0105744	0.999987	0.000013	putative endopeptidase
148	GL0106295	0.999986	0.000014	Aminopeptidase
149	GL0107933	0.999986	0.000014	acylaminoacyl-peptidase
150	GL0108253	0.999986	0.000014	putative endopeptidase
151	GL0108449	0.999985	0.000015	Aminopeptidase
152	GL0108581	0.999985	0.000015	Aminopeptidase
153	GL0109350	0.999985	0.000015	putative endopeptidase
154	GL0109963	0.999984	0.000016	X-Pro aminopeptidase
155	GL0110057	0.999983	0.000017	X-Pro dipeptidase
156	GL0110079	0.999983	0.000017	IgA-specific serine endopeptidase
157	GL0110198	0.999982	0.000018	X-Pro dipeptidase
158	GL0110319	0.997478	0.002522	gamma-glutamyltranspeptidase
159	GL0110421	0.997472	0.002528	D-aminopeptidase
160	GL0110658	0.997459	0.002541	D-alanyl-D-alanine carboxypeptidase
161	GL0110664	0.997457	0.002543	leucyl aminopeptidase
162	GL0110690	0.997444	0.002556	prolyl oligopeptidase
163	GL0110743	0.997443	0.002557	Aminopeptidase
164	GL0110837	0.997418	0.002582	methionyl aminopeptidase
165	GL0110963	0.997417	0.002583	X-Pro aminopeptidase
166	GL0111249	0.997389	0.002611	O-sialoglycoprotein endopeptidase
167	GL0111426	0.997349	0.002651	tripeptide aminopeptidase
168	GL0111488	0.997329	0.002671	D-alanyl-D-alanine carboxypeptidase
169	GL0111629	0.997233	0.002767	Aminopeptidase
170	GL0111698	0.997222	0.002778	putative endopeptidase
171	GL0111759	0.997181	0.002819	glutamyl endopeptidase
172	GL0112043	0.997118	0.002882	Aminopeptidase
173	GL0112548	0.997098	0.002902	O-sialoglycoprotein endopeptidase
174	GL0112732	0.997001	0.002999	X-Pro aminopeptidase
175	GL0113031	0.996930	0.003070	X-Pro aminopeptidase

Table 4. Summary of the gene sequences encoding lipases having alkaline index of 0.99

STT	Code gene	Alkaline index	Acidic index	Enzyme lipase
1	GL0094408	1.000000	0.000000	esterase / lipase
2	GL0095714	1.000000	0.000000	triacylglycerol lipase
3	GL0098504	1.000000	0.000000	triacylglycerol lipase
4	GL0100660	1.000000	0.000000	Lysophospholipase
5	GL0102502	1.000000	0.000000	triacylglycerol lipase

6	GL0103848	1.000000	0.000000	esterase / lipase
7	GL0115777	1.000000	0.000000	phospholipase A1
8	GL0028122	1.000000	0.000000	phospholipase A1
9	GL0052713	1.000000	0.000000	phospholipase A1
10	GL0057522	1.000000	0.000000	phospholipase A1
11	GL0091897	1.000000	0.000000	phospholipase A1
12	GL0097086	1.000000	0.000000	phospholipase A1
13	GL0102371	1.000000	0.000000	phospholipase A1
14	GL0102961	1.000000	0.000000	phospholipase A1
15	GL0130369	1.000000	0.000000	phospholipase C
16	GL0019568	1.000000	0.000000	phospholipase C
17	GL0092972	1.000000	0.000000	phospholipase C
18	GL0113097	1.000000	0.000000	phospholipase C
19	GL0017374	0.999999	0.000001	phospholipase D
20	GL0018116	0.999999	0.000001	phospholipase D
21	GL0033413	0.999999	0.000001	phospholipase D
22	GL0056310	0.999999	0.000001	phospholipase D
23	GL0071465	0.999999	0.000001	phospholipase D
24	GL0071465	0.999999	0.000001	phospholipase D
25	GL0076794	0.999998	0.000002	phospholipase D
26	GL0082514	0.999996	0.000004	phospholipase D
27	GL0087982	0.999996	0.000004	phospholipase D
28	GL0008869	0.999994	0.000006	phospholipase D
29	GL0104048	0.999993	0.000007	phospholipase A1
30	GL0108498	0.999991	0.000009	phospholipase A1
31	GL0108499	0.999991	0.000009	phospholipase A1
32	GL0108547	0.999991	0.000009	phospholipase A1
33	GL0113097	0.999985	0.000015	phospholipase D

Table 5. Summary of predicted alkaline and acidic cellulases and hemicellulases in the gut flora of *C. gestroi*

Enzymes	Total sequences	Alkaline sequences	Percentage (%)	Acidic sequences	Percentage (%)
Cellulases and hemicellulases	578	338	59%	240	41%

Putative alkaline and acidic cellulases and hemicellulases

Among a total of 575 cellulolytic and hemicellulolytic enzymes, 338 (59%) were predicted to be alkaline enzymes. The number of alkaline cellulases and hemicellulases are lower than those of alkaline proteases and lipases,.

It is already well-known that cellulases and hemicellulases are abundant in the symbiotic organisms in the gut of termites to degrade cellulose and hemicellulose [14, 17]. Diversity and of lignocellulose-degrading alkaline enzymes and their function in the termite gut microbial community have been reported [17]. Cellulases and hemicellulases from the gut flora of termites such as *R. flavipes, R. speratus* and *Macrotermes subhyalinus* have optimal pH around 5-7, and those from *Microcerotermes* sp. have wider optimal pH range of 5.0-10.0. In case of cellulolytic and hemmicellulolytic enzymes of *Sarocladium kiliense* and *Trichoderma virens* isolated from the gut of the lower termite, *R. santonensis*, optimal pH range was pH 9-10 [27]. In the higher termite,

Nasutitermes corniger, pH of the gut reaches as high as 11 [10]. In this study, there is not much difference between the proportion of the predicted alkaline (59%) and acidic (41%) cellulases and hemicellulases of the gut flora of *C. gestroi* In contrast, in cases of proteinases and lipases, the proportion of alkaline enzymes was much higher than acidic ones. Presence of huge number of alkaline enzymes in the gut flora suggest that those microbiota are suitable to suvive alkaline environment of the gut of *C. gestroi*. Extracellular enzymes produced/released from such microbiota are likely to have their optimum pH of alkaline range [7].

In this study we are interested in the ability of the cellulase and hemicellulase enzymes that can resist alkaline environment. Using Alcapred software, we found that the majority of alkaline cellulases and hemicellulases have very high alkaline index; 41 alkaline cellulases sequences and 40 alkaline hemicellulases have alkaline index of >0.99. All of them belong to beta-glucosidase and alpha-N-arabinofuranosidase.

CONCLUSION

Using Alcapred software, high percentages of proteases, lipases, cellulases and hemicellulases of the gut flora of *C. gestroi* were predicted as alkaline enzymes. These results might be useful for the effective utilization of novel alkaline enzymes in the industries. This is the first prediction of the alkaline enzyme groups of the gut flora of *C. gestroi* termites. The results of this study provide a comprehensive picture of alkaline tolerance of various enzyme groups which has not been reported previously.

REFERENCES

1. Bignell D. E., Eggleton P., 1995. On the elevated intestinal pH of higher termites (Isoptera: Termitidae). Insect Soc., 42(1): 57-69.

2. Brune A., Kühl M., 1996. pH profiles of the extremely alkaline hindguts of soil-feeding termites (Isoptera: Termitidae) determined with microelectrodes. J. Insect Physiol., 42(11): 1121-1122.

3. Brune A., Emerson D., Breznak J. A., 1995. The termite gut microflora as an oxygen sink: microelectrode determination of oxygen and pH gradients in guts of lower and higher termites. Appl. Environ. Microbiol., 61(7): 2681-2687.

4. Chandrasekharaiah M., Thulasi A., Bagath M., Kumar D. P., Santosh S. S., Palanivel C., Jose V. L., Sampath K. T., 2011. Molecular cloning, expression and characterization of a novel feruloyl esterase enzyme from the symbionts of termite (*Coptotermes formosanus*) gut. BMB Rep., 44(1): 52-57.

5. Cherif S., Mnif S., Hadrich F., Abdelkafi S., Sayadi S., 2011. A newly high alkaline lipase: an ideal choice for application in detergent formulations. Lipids Health Dis. 10: 221.

6. Do T. H., Nguyen T. T., Nguyen T. N., Le Q. G., Nguyen C., Kimura K., Truong N. H., 2014. Mining biomass-degrading genes through Illumina-based denovo sequencing and metagenomic analysis of free-living bacteria in the gut of the lower termite *Coptotermes gestroi* harvested in Vietnam. J Biosci Bioeng., 118(6): 665-671.

7. Gessesse A., Gashe B. A., 1997. Production of alkaline xylanase by alkaliphic *Bacillus* sp isolated alkaline soda lake. J. Appl. Microbiol., 83(4): 402-406.

8. Hongoh Y., Ohkuma M., Kudo T., 2003. Molecular analysis of bacterial microbiota in the gut of the termite *Reticulitermes speratus* (Isoptera: Rhinotermitidae). FEMS Microbiol. Ecol., 44(2): 231-242.

9. Huson D. H., Auch A. F., Qi J., Schuster S. C., 2007. MEGAN analysis of metagenomic data, Genome Res., 17(3): 377-386.

10. Köhler T., Dietrich C., Scheffrahn RH., Brune A., 2012. High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). Appl. Environ. Microbiol., 78(13): 4691-4701.

11. Kudo T., 2009. Termite-microbe symbiotic system and its efficient degradation of lignocellulose. Biosci. Biotechnol. Biochem., 73(12): 2561-2567.

12. Lin H., Chen W., Ding H., 2013. AcalPred: A sequence-based tool for discriminating between acidic and alkaline enzymes. PLoS ONE 8(10): e75726.

13. Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., et al., 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler, GigaScience, 1(1): 18.

14. Mattéotti C., Thonart P., Francis F., Haubruge E., Destain J., Brasseur C., Bauwens J., De Pauw E., Portetelle D., Vandenbol M., 2011. New glucosidase activities identified by functional screening of a genomic DNA library from the gut microbiota of the termite *Reticulitermes santonensis*. Microbiol Res., 166(8): 629-42.

15. Michael J., Liszka E., Schneider E., Clark D. S., 2012. Nature versus nurture: developing enzymes that function under extreme conditions. Annu. Rev. Chem. Biomol. Eng., 3: 77-102.

16. Mitra S., Rupek P., Richter D. C., Urich T., Gilbert J. A., Meyer F., Wilke A., Huson D. H., 2011. Functional analysis of metagenomes and metatranscriptomes using SEED and KEGG. BMC Bioinformatics. 12(Suppl 1): S21.

17. Ni J., Tokuda G., Takehara M., Watanabe H. 2007. Heterologous expression and enzymatic characterization of β-glucosidase from the drywood-eating termite, *Neotermes koshunensis*. Appl. Entomol. Zool., 42(3): 457-463.

18. Nimchua T., Thongaram T., Uengwetwanit T., Pongpattanakitshote S., Eurwilaichitr L., 2012. Metagenomic analysis of novel lignocellulose-degrading enzymes from higher termite guts inhabiting microbes. J. Microbiol. Biotechnol., 22(4): 462-9.

19. Noguchi H., Taniguchi T., Itoh T., 2008. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. DNA Res., 15: 387-396.

20. Powell S., Szklarczyk D., Trachana K., Roth A., Kuhn M., Muller J., Arnold R., Rattei T., Letunic I., Doerks T., 2011. eggNOGv3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. Nucleic Acids Res., 40: D284-289.

21. Sambrook J., Russell D. W., 2008. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press.

22. Scharf M. E., Karl Z. J., Sethi A., Boucias D. G., 2011. Multiple levels of synergistic collaboration in termite lignocellulose digestion, PLoS One. 6(7): e21709.

23. Scharf M. E., Tartar A., 2008. Termite digestomes as sources for novel lignocellulases. Biofuels, Bioprod. Bioref. DOI: 10.1002/bbb.107.

24. Smith A., Scharf M. E., Roberto M., Philip G., Koehler G., 2009. pH Optimization of gut cellulase and xylanase activities from the Eastern subterranean termite, *Reticulitermes flvipes* (Isoptera: Rhinotermidae). Sociobiol., 54: 199-210.

25. Sethi A., Xue Q. G., La Peyre J. F., Delatte J., Husseneder C., 2011. Dual origin of gut proteases in Formosan subterranean termites (*Coptotermes formosanus* Shiraki) (Isoptera: Rhinotermidae). Comp Biochem Physiol A Mol Integr Physiol., 159(3): 261-267.

26. Szalanski A. L., Austin J. W., Scheffrahn R. H., Messenger M. T., 2004. Molecular diagnostics of the formosan subterranean

termite (Isoptera: Rhinotermitidae). Fla. Entomol., 87: 145-151.

27. Tarayre C., Bauwens J., Brasseur C., Mattéotti C., Millet C., Guiot PA., Destain J., Vandenbol M., Portetelle D., De Pauw E., Haubruge E., Francis F., Thonart P., 2015. Isolation and cultivation of xylanolytic and cellulolytic *Sarocladium kiliense* and *Trichoderma virens* from the gut of the termite *Reticulitermes santonensis*. Environ Sci Pollut Res Int., 22(6): 4369-4382.

28. Tartar A., Wheeler M. M., Zhou X., Coy M. R., Boucias D. G., Scharf, M. E., 2009. Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite *Reticulitermes flavipes*. Biotechnol. Biofuels., 2: 25.

29. Watanabe H., Nakamura M., Tokuda G., Yamaoka I., Scrivener AM., Noda H., 1997.

30. Site of secretion and properties of endogenous endo-beta-1,4-glucanase components from Reticulitermes speratus (Kolbe), a Japanese subterranean termite. Insect Biochem Mol Biol., 27(4): 305-313.

31. Warnecke F., Luginbühl P., Ivanova N., Ghassemian M., Richardson T. H., Stege J. T., Cayouette M., McHardy A. C., Djordjevic G, Aboushadi N, Sorek R, Tringe S. G., Podar M., Martin H. G., Kunin V., Dalevi D., Madejska J., Kirton E., Platt D., Szeto E., Salamov A., Barry K., Mikhailova N., Kyrpides N. C., Matson E. G., Ottesen E. A., Zhang X., Hernández M., Murillo C., Acosta L. G., Rigoutsos I., Tamayo G., Green B. D., Chang C., Rubin E. M., Mathur E. J., Robertson D. E., Hugenholtz P., Leadbetter J. R., 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature, 450: 560-565.

SỬ DỤNG VI TÍNH KHAI THÁC CÁC ENZYME CHỊU KIỀM TỪ DỮ LIỆU DNA METAGENOME VI SINH VẬT SỐNG TRONG RUỘT MỐI BẬC THẤP *coptotermes gestroi* Ở VIỆT NAM

Nguyễn Minh Giang[1], Đỗ Thị Huyền[2], Trương Nam Hải[2]

[1]Trường Đại học Sư phạm tp. Hồ Chí Minh
[2]Viện Công nghệ sinh học, Viện Hàn lâm KH & CN Việt Nam

TÓM TẮT

Phân tích trình tự DNA metagenome của vi sinh vật sống trong ruột mối *Coptotermes gestroi* để xác định và tìm kiếm enzyme chịu được môi trường kiềm, nguồn vật liệu quan trọng để khai thác và ứng dụng trong nghiên cứu và sản xuất. Kết quả sử dụng phần mềm Alcapred để dự đoán khả năng chịu kiềm và axit của các nhóm enzyme protease, lipase, cellulase và hemicellulase từ dữ liệu metagenome của vi sinh vật trong ruột mối bao gồm: có 737 trình tự mã hóa protease chịu kiềm trong 943 trình tự và 154 trình tự mã hóa lipase chịu kiềm trong 214 trình tự từ DNA metgenome cho thấy tỷ lệ phần trăm của protease kiềm và lipase rất cao, chiếm 72% và 78%. Có 338 trong tổng số 575 trình tự đã được dự đoán thuộc về nhóm enzyme chịu kiềm phân giải cellulose và hemicellulose, chiếm 59%. Đây là những kết quả công bố chi tiết đầu tiên về các chuỗi gen mã hóa các enzyme chịu kiềm có nguồn gốc từ vi sinh vật sống tự do trong ruột mối của *C. gestroi* và là nguồn dữ liệu để khai thác, phân lập gen để sản xuất enzyme tái tổ hợp.

Từ khóa: Coptotermes gestroi, cellulase, enzyme chịu kiềm, hemicellulase, lipase, metagenome, protease, ruột.