

German-Vietnamese Science Day 2023

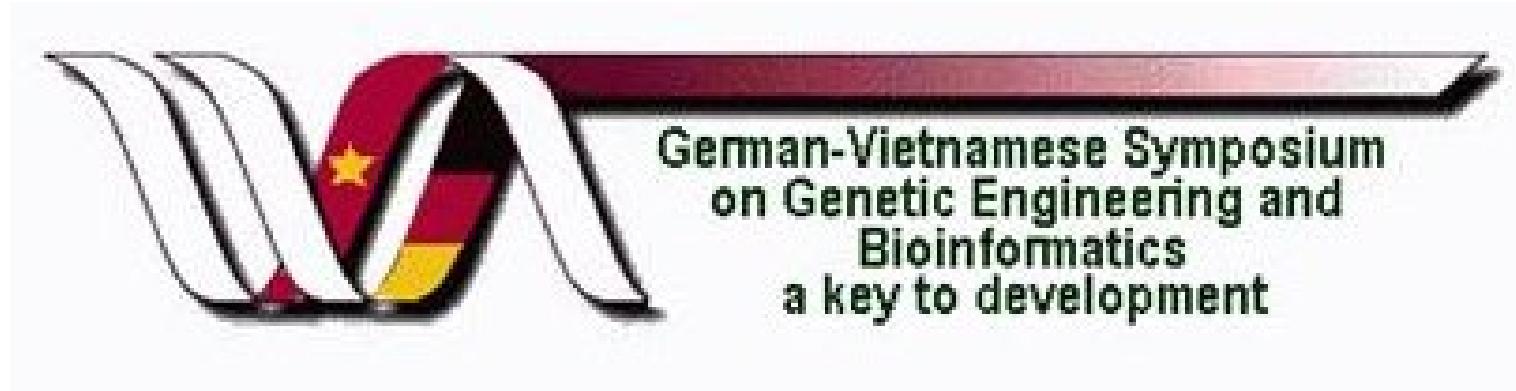
GERMAN-VIETNAMESE PROJECT:
**Mining metagenome DNA databases of microbiomes for
lignocellulose degradation enzymes by bioinformatics tools**
Metagenlig, NDT.50.GER/18, 2018-2022

IBT (VAST) – ITB (Stuttgart Univ) – IMB (Hamburg Univ)

Prof. Dr. Truong Nam Hai
Institute of Biotechnology
Vietnam Academy of Science and Technology

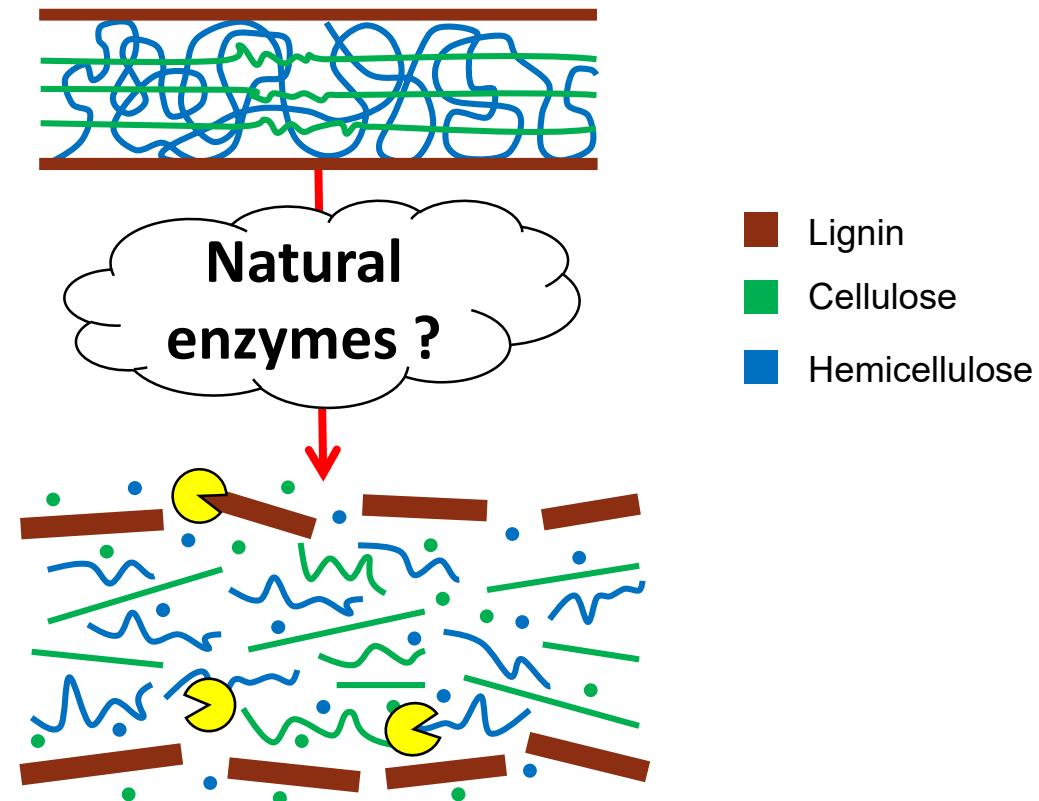
Da Nang, April 26, 2023

Story of collaboration



1. The First German-Vietnamese Workshop on Genetic Engineering and Bioinformatics (21.9 - 2.10.1998, funded by BMBF).
2. The first joint project: **Amylase Engineering Database** (1.3.2000 - 28.2.2003, funded by BMBF)
 - Creation of amylase mutants by directed evolution methods (Error-prone PCR and DNA shuffling method)
 - Modelling and simulation of proteins

Lignocellulose bioconversion



The added value of collaboration between German-Vietnamese research groups on metagenomics study of microbiomes

1. Efficient bioinformatics tools for mining candidate genes.
2. Fast and efficient expression systems for screening candidate genes.

Partners: Institute of Biotechnology (IBT, VAST), Institute of Technical Biochemistry (ITB, Stuttgart Univ) and Institute of Microbiology and Biotechnology (IMB, Hamburg Univ).



Target microbiomes



Native goat (Co)



Bach Thao



White rot fungi from National rain forest
Cuc Phuong, Ninh Binh

Analysis workflow of metagenome DNA databases

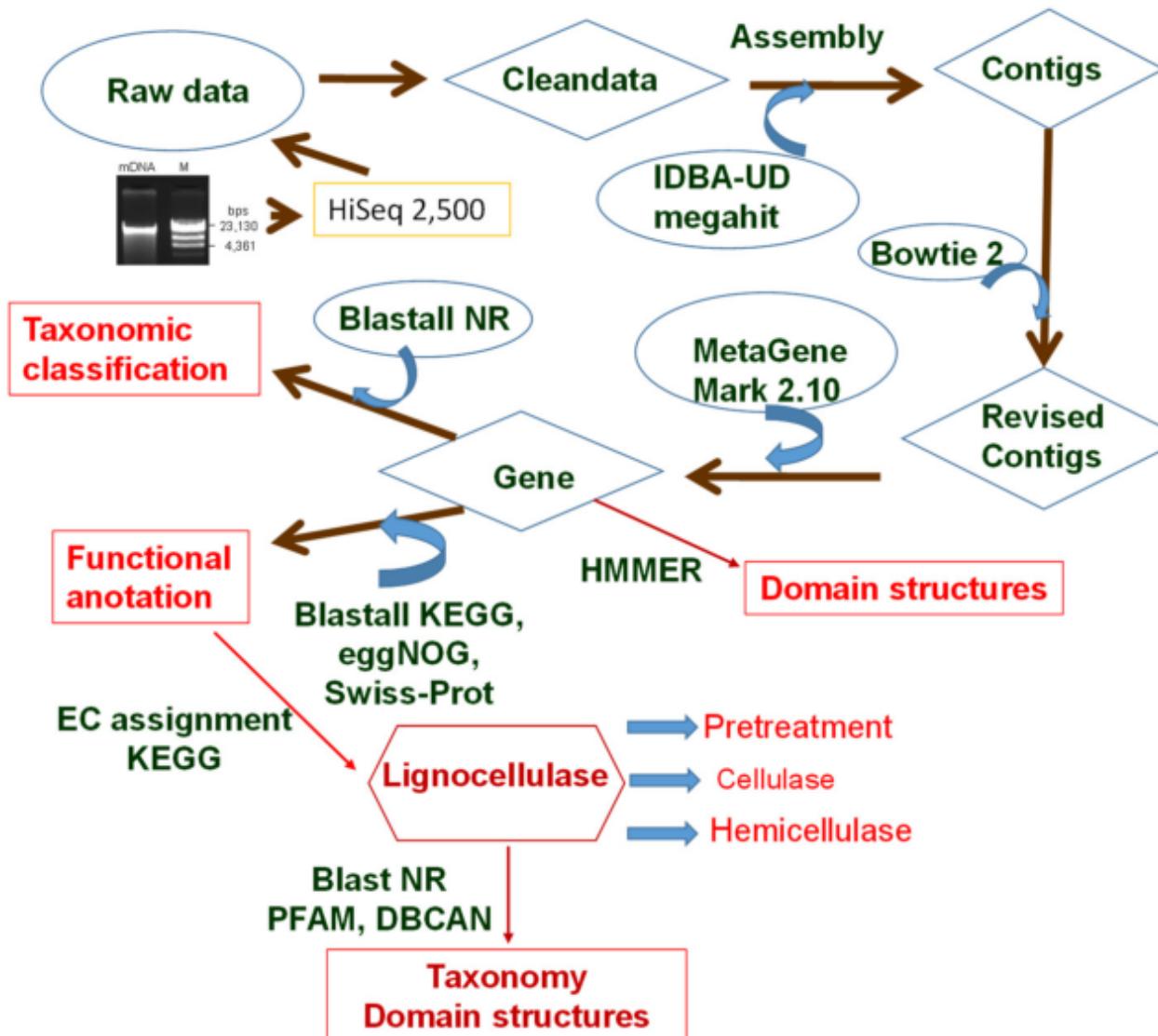


Figure 2. Workflow diagram for mining lignocellulose-degrading enzymes.

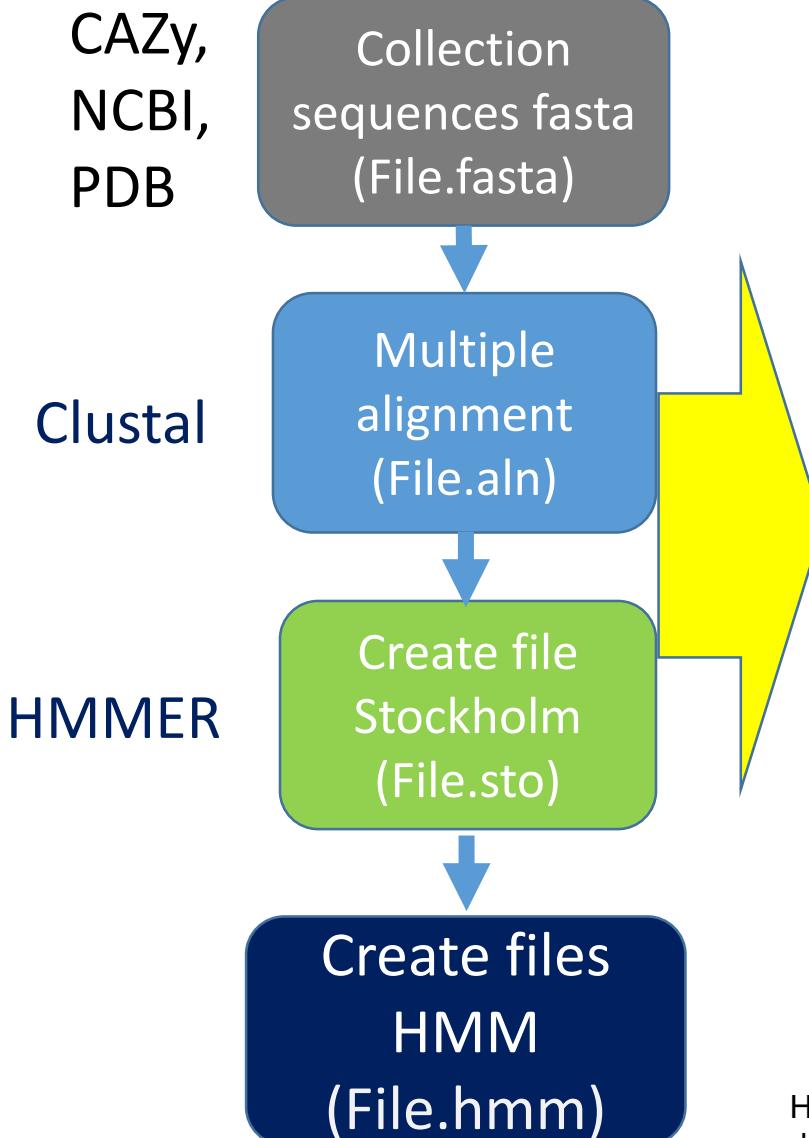
Results of gene prediction by MetaGeneMark 2.10

Samples	Clean data (Gbs)	(ORFs)
Humus	51,82	3,884,879
Goat rumen	48,66	4,222,317

1. Efficient bioinformatics tools for mining candidate genes

Prof. Dr. Juergen Pleiss
ITB, Stuttgart Univ

Creating files HMM for mining genes from metagenomic DNA databases



Clustal

HMMER

NN	Candidate genes	NN	Candidate genes
1	Lignin peroxidase (LiP)	16	Acetyl xylan esterase
2	Hydrogen peroxide oxidoreductase	17	Feruloyl esterase
3	Manganese peroxidase (MnP)	18	Xyloglucanase
4	Versatile peroxidase (VP)	19	Mannanase
5	Laccase	20	β -mannosidase
6	Expansin	21	Arabinase
7	Lytic polysaccharide monooxygenase	22	Galactanase
8	Cellobiohydrolase	23	Polygalacturonase
9	Endo-1,4-glucanase	24	β -glucuronidase
10	β -glucosidase	25	Glucuronyl esterase
11	Endo-xylanase	26	CBM(1-84)
12	β -xylosidase	27	Fibronectin 3-like domain (FN3)
13	α -L-arabinofuranosidase	28	Dockerin
14	α -glucuronidase	29	Immunoglobulin-like domain (Ig)
15	Licheninase		

HMMER is software package that provides tools for making probabilistic models of protein and DNA sequence domain families called profile hidden Markov models for using to annotate new sequences...

Comparison of mining results of goat rumen metagenome DNA database by SBS và HMM methods

	SBS	HMM
Cellulase	18028	5236
endoglucanase	7368	2009
cellobiohydrolase	216	1645
β-glucosidase	10444	1582
Hemicellulase	14723	19048
α-L-arabinofuranosidase	6229	2646
α glucurinindase (GH76N)	561	102
Arabinanase (GH43)	0	1894
Axetylxylanesterase (AXE1)	4	219
β-glucuronidase	888	4171
β-xylosidase	1018	3276
β-mannosidase (GH2)	659	1890
Feruloyl esterase	92	7

	SBS	HMM
galactanase	0	305
Glucuronyl esterase	0	165
HPOXRE catalase)	0	1
lichenase	0	350
mannanase	1237	263
polygalacturonase	635	211
Xylanase (GH44)	3400	3534
xyloglucanase	0	14
Pretreatment & others	53	15713
MnP, VerP, LiP	0	0
Laccase/MCO	9/0	9/341
LPMO	11	1
expansin	33	36
CBM	0	7285
dockerin	0	1734
Ig	0	2688
fn3	0	3655

Comparison of mining results of humus metagenome DNA database by SBS và HMM methods

	SBS	HMM
Cellulase	6488	1928
Endoglucanase	2216	557
Cellobiohydrolase	0	253
β -Glucosidase	4272	1118
Hemicellulase	3158	4842
α -L-Arabinofuranosidase	1016	431
α Glucurinindase (GH76N)	161	102
Arabinanase (GH43)	0	343
Axetylxylanesterase (AXE1)	1	79
β _Gluronidase	227	1044
β _Xylosidase	659	945
β -Mannosidase (GH2)	611	594
Feruloyl esterase	75	53
Galactanase	0	17
Glucuronyl esterase	0	22

	SBS	HMM
HPOXRE (catalase)	0	224
Lichenase	0	290
Mannanase	368	40
Polygalacturonase	38	45
Xylanase (GH44)	552	599
Xyloglucanase	0	14
Pretreatment & others	17	6809
MnP, VerP, LiP	0	0
Laccase/MCO	10/0	10/1105
LPMO	0	69
Expansin	7	0
CBM	0	3163
Dockerin	0	11
Ig	0	1178
Fn3	0	1273

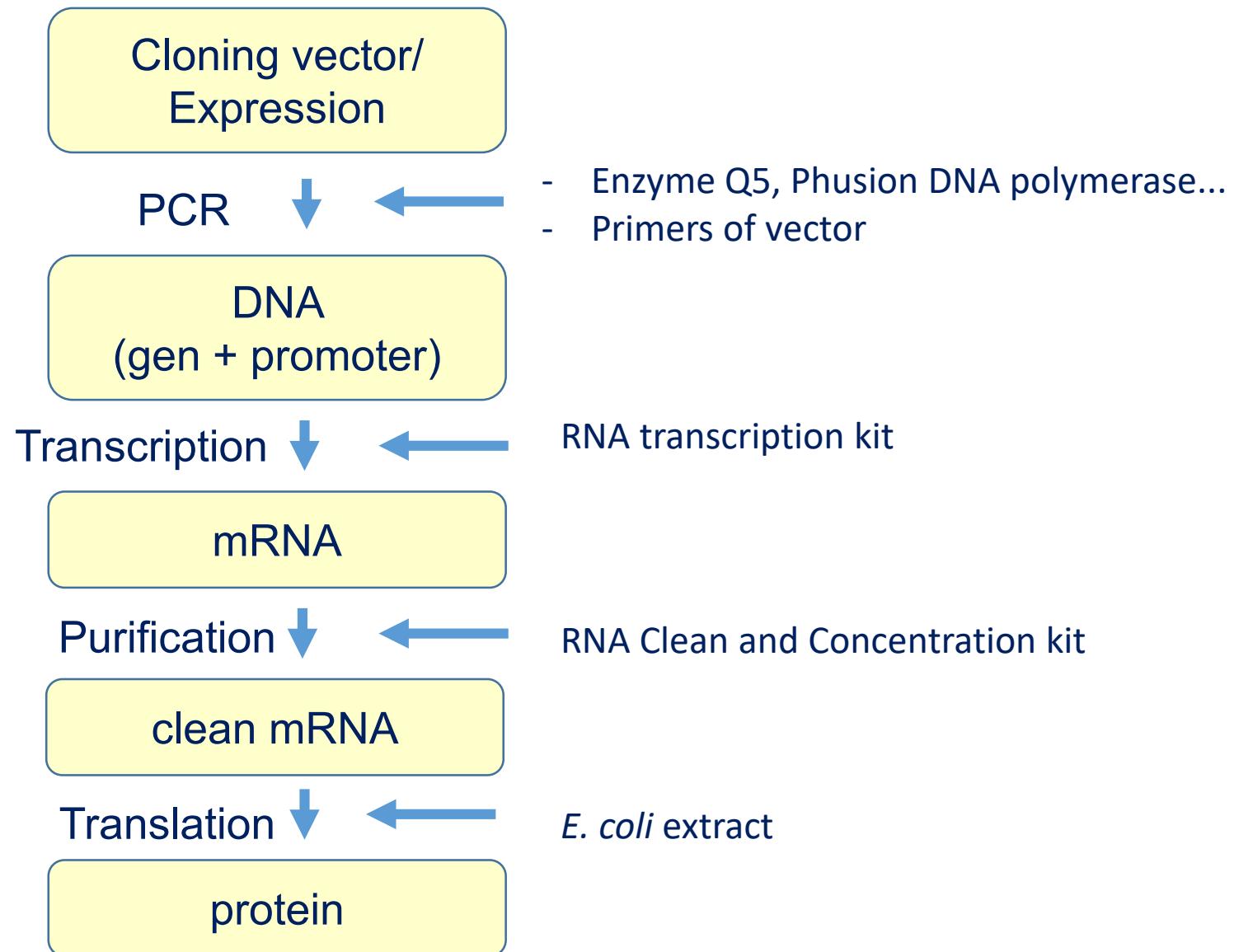
2. Fast and efficient expression systems for screening candidate genes

Prof. Dr. Wolfgang Streit
Institute of Microbiology and Biotechnology, Hamburg Univ

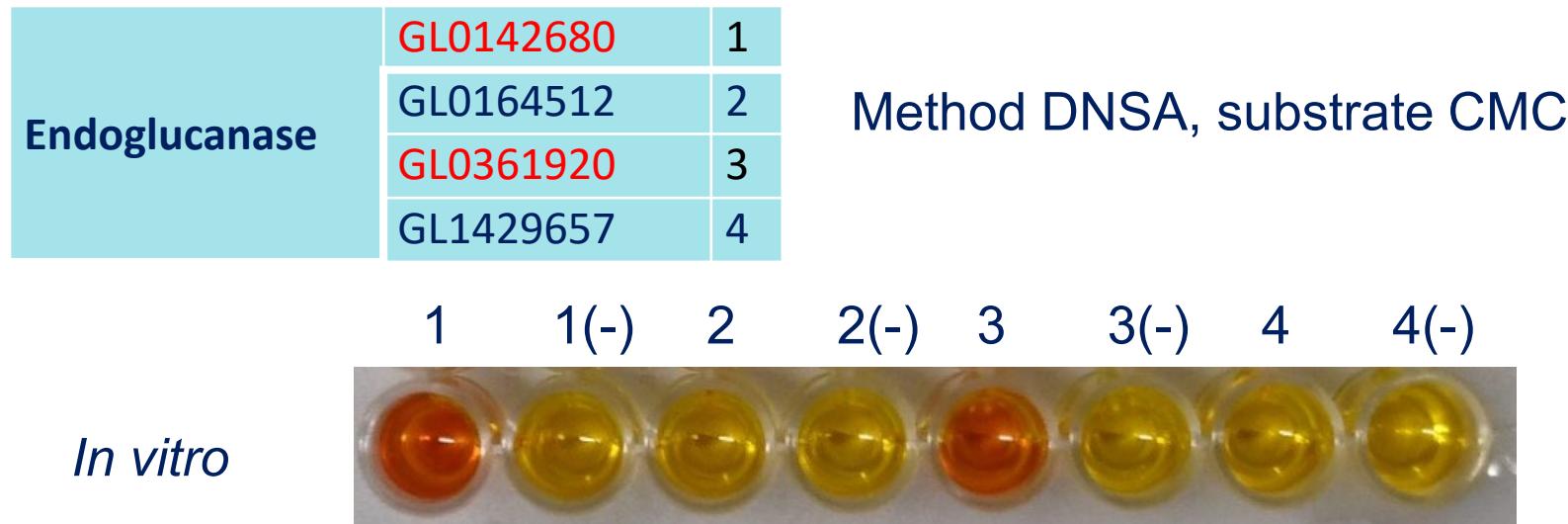
Cell-free expression system (Hamburg Univ)

Advantages:

- Easy to change conditions for optimization of protein folding.
- Unaffected by the toxicity of the products from cell expression;
- Suitable for detailed study with large number of samples;
- Allows automation of protein screening;
- Easy to scale-up industrial production of proteins in bioreactors.

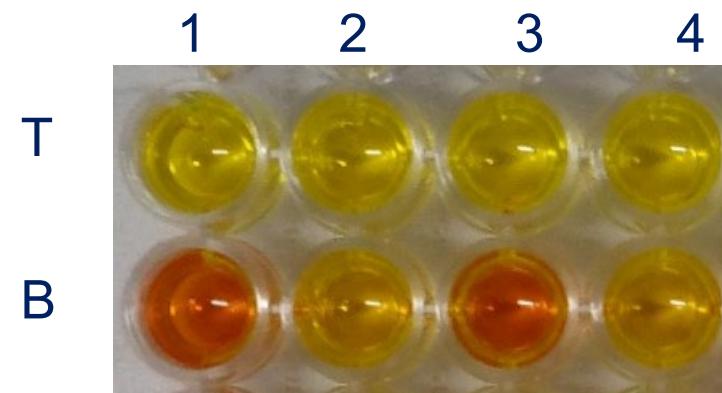


Cell-free expression of selected genes



In vivo

Expression in BL21 (B)
and T7 Shuffle (T),
medium LB, 0,1 mM
IPTG, 20°C



List of genes selected for expression by cell-free expression (Hamburg Univ)

STT	Protein (13 gene)	ORF ID	Vector	Length (bp)	Cloning sites
1	Laccase (5 genes)	GL0003597	Synthetic	2207 bp	
2		GL0020702	PCR	2351 bp	
3		GL0023947	Synthetic	2228 bp	
4		GL0197041	Synthetic	2213 bp	
5		GL0436272	PCR	2228 bp	
6	Endo-glucanase (4 genes)	GL0142680	Synthetic	1374 bp	<i>Nde</i> I + <i>Xba</i> I
7		GL0164512	Synthetic	pET21a(+)	
8		GL0361920	Synthetic	1545 bp	
9		GL1429657	Synthetic	1731 bp	
10	Lytic polysaccharide mono-oxygenase (LPMO) (4 genes)	GL0393374	Synthetic	1305 bp	
11		GL1004230	Synthetic	642 bp	
12		GL0883009	Synthetic	846 bp	
13		GL0293304	Synthetic	1119 bp	
				1380 bp	

List of genes selected for expression by cell-free expression (Vietnam)

NN	Enzymes	Source	ORF ID	domains	Lenth (aa)
14			GL0036730	BGCGH3-31	1259
15	Beta glucosidase		GL0073774	BGCGH16	465
16		Goat	GL0847998	BGCGH3	215
17	Xylan beta xylosidase		GL0916261	XBXGH43	287
18	Endo glucanase		GL0694641	EGCGH5CBM72	741
19	Beta glucosidase		GL0050362	BGCGH3S2	849
20			GL0278103	EGCGH64CBM13-6	884
21	Endo glucanase	Humus	GL0127466	EGCGH8G1	427
22			GL0183420	EGCGH8G2	465
23	Xylan beta xylosidase		GL0612568	XBXGH39	518
24	Expansin	Goat	GL1401173	EXP	284
25	LPMOs	Humus	GL0247266	LPMO1-GbpACBM	630
			GL0620585	LPMO2-IgCBM	583
			GL0183513	LPMO3-GbpA	478
			GL0391320	LPMO4.1-2FnCBM	599
			GL0555809	LPMO4.2-GbpA	474
			GL0772141	LPMO4.3-GbpA	482

Summary

- Metagenome databases:
 - Goat's rumen: 4.2 mil genes
 - Humus: 3.9 mil genes
- Protein database of lignocellulose degradation enzymes:
 - Goat's rumen:
 - Pretreatment proteins/enzymes: 5236 genes
 - Cellulases: 19048 genes
 - Hemicellulose: 15713 genes
 - Humus surrounding White rot fungi:
 - Pretreatment proteins/enzymes: 6809 genes
 - Cellulases: 1928 genes
 - Hemicellulose: 4802 genes
- Two PhD students.

Activities

Research stays

- 1.4. - 30.9.2019 Dao Trong Khoa (IBT,Hanoi) in Stuttgart
1.10.19 -31.3.2020 Dao Trong Khoa (IBT,Hanoi) in Hamburg
1.5.-5.7.2019 Marco Orlando (University of Milano-Bicocca) in Stuttgart

Project meetings, workshops, conferences

- 18.-20.6.2018 Kickoff meeting (Hanoi)
8.-10.10.2018 Project meeting (Stuttgart)
2.-6.7.2019 Project meeting (Stuttgart)
3.-5. 7.2019 International Workshop "Nature as a teacher: big data, complex systems, new chemistry" (Stuttgart)
30.10.2019 Poster at German Science Day in Hanoi
5.-7.10.2020 International Workshop "Systematic profiling of multicopper oxidases" (online)
9.11.2022 Final MetagenLig Workshop (online)

Publications

Scientific papers:

1. Le TTH, Nguyen TB, Nguyen HD, Nguyen HD, Le NG, Dao TK, Nguyen TQ, Do TH, **Truong NH** (2022) De novo metagenomic analysis of microbial community contributing in lignocellulose degradation in humus samples harvested from Cuc Phuong tropical forest in Vietnam. *Diversity* **14**: 220.
2. Zhang H, Pérez-García P, Dierkes RF, Danso D, **Pleiss J**, Almeida A, Höcker B, Schmitz-Streit RA, Chow J, **Streit WR** (2022) The Bacteroidetes *Aequorivita* sp. and *Kaistella jeonii* produce promiscuous esterases with PET-hydrolyzing activity. *Front Microbiol* **12**: 803896.
3. Buchholz PCF, Feuerriegel G, Zhang H, Pérez-García P, Nover LL, Chow J, **Streit WR**, **Pleiss J** (2022) Plastics degradation by hydrolytic enzymes: the Plastics-Active Enzymes Database - PAZy. *Proteins* **90**: 1443.
4. Chow J, **Pleiss J**, **Streit WR** (2022) PAZy – mikrobielle Enzyme für den Abbau künstlicher Polymere. *BIOspektrum* **4**: 451.
5. Antranikian G, **Streit WR** (2022) Microorganisms harbor keys to a circular bioeconomy making them useful tools in fighting plastic pollution and rising CO₂ levels. *Extremophiles* **26**:10.
6. Lohoff C, Buchholz PCF, Le Roes-Hill M, **Pleiss J** (2021) The Expansin Engineering Database: a navigation and classification tool for expansins and homologues. *Proteins* **89**: 149.
7. Orlando M, Buchholz PCF, Lotti M, **Pleiss J** (2021) The GH19 Engineering Database: sequence diversity, substrate scope, and evolution in glycoside hydrolase family 19. *PLoS One* **16**: e0256817.
8. Pérez-García P, Danso D, Zhang H, Chow J, **Streit WR** (2021) Exploring the global metagenome for plastic-degrading enzymes. *Methods in Enzymology* **648**:137
9. Dao TK, Do TH, Le NG, Nguyen HD, Nguyen TQ, Le TTH, **Truong NH** (2021) Understanding the role of *Prevotella* genus in the digestion of lignocellulose and other substrates in Vietnamese native goats' rumen by metagenomic deep sequencing. *Animals* **11**: 3257.
10. Bauer TL, Buchholz PCF, **Pleiss J** (2020) The modular structure of α/β-hydrolases. *FEBS J* **287**: 1035.
11. Danso D, Chow J, **Streit WR** (2019) Plastics: environmental and biotechnological perspectives on microbial degradation. *Appl Environ Microbiol* **85**:e01095-19

Challenges

- Screening for specific activity of enzymes;
- Screening for substrate affinity of enzymes;
- Screening for temperature, pH activities of enzymes.

Acknowledgments

- MOST Vietnam and BMBF Germany
- Prof. Jurgen Pleiss and Prof. Wolfgang Streit
- Institute of Biotechnology, VAST
- Researchers of GEL, IBT

Thanks for your attention!

Comparision of SBS and HMM annotations

NN	Gene	HMM	NR	eggNog	Kegg	Swissprot
1	GL0066553	LPMO10	NA-GbpA	Interact GlcNAc	NA-GbpA	GbpA
2	GL0089352	LPMO10	CBP	Interact GlcNAc	CBP	NA-GbpA
3	GL0125011	LPMO10	NA	NA	CBD-3P	Spheroidin-like P
4	GL0183513	LPMO10	NA	CBP	NA-GbpA	NA-GbpA
5	GL0200824	LPMO10	HP	NA	CBD-3P	NA-GbpA
6	GL0213284	LPMO10	NA	CBP	CBD-3P	CBP
7	GL0225724	LPMO10	CBP	CBD-3P	CBP	NA-GbpA
8	GL0247266	LPMO10	NA	Spindolin-RP	CBD-3P	Spindolin-RP 1
9	GL0251010	LPMO10	NA-GbpA	CBP	NA-GbpA	NA-GbpA
10	GL0293304	LPMO10	HP	NA	CBP, FN3, GbpA	Chitinase A1
11	GL0297948	LPMO10	NA	Interact GlcNAc	NA-GbpA	NA-GbpA
12	GL0338092	LPMO10	NA	NA	CBD-3P	NA-GbpA
13	GL0391320	LPMO10	NA	NA	CBD-3P	NA-GbpA
14	GL0393374	LPMO10	NA	Interact GlcNAc	CBP	NA-GbpA
15	GL0411524	LPMO10	CBP	Interact GlcNAc	CBP	NA-GbpA
16	GL0417116	LPMO10	NA-GbpA	ACT	NA-GbpA	NA-GbpA
17	GL0422153	LPMO10	NA-GbpA	CBP	NA-GbpA	NA-GbpA
18	GL0489328	LPMO10	NA	Interact GlcNAc	NA-GbpA	NA-GbpA
19	GL0507050	LPMO10	NA	Interact GlcNAc	NA-GbpA	NA-GbpA
20	GL0522565	LPMO10	NA	CBP	NA-GbpA	NA-GbpA
21	GL0555809	LPMO10	NA-GbpA	Interact GlcNAc	NA-GbpA	NA-GbpA
22	GL0589166	LPMO10	MDT	CBP	NA-GbpA	NA-GbpA
23	GL0600730	LPMO10	CBP	Interact GlcNAc	CBP	NA-GbpA
24	GL0620585	LPMO10	HP	Spindolin-RP	CBD-3P	NA-GbpA
25	GL0656018	LPMO10	CBP	CBD-3P	CBD-3P	NA-GbpA
26	GL0658692	LPMO10	NA	CBP	CBD-3p	CBP
27	GL0681738	LPMO10	CBP	Interact GlcNAc	CBD-3p	NA-GbpA
28	GL0772141	LPMO10	NA-GbpA	Interact GlcNAc	NA-GbpA	NA-GbpA
29	GL0875000	LPMO10	NA	Interact GlcNAc	NA-GbpA	NA-GbpA
30	GL0883009	LPMO10	HP	Interact GlcNAc	FN3, CBP	NA-GbpA
31	GL0999597	LPMO10	HP	CBD-3P	CBP	NA-GbpA
32	GL1004230	LPMO10	CBP	NA	CBP	NA-GbpA
33	GL1034380	LPMO10	Chitinase	CBD-3P	CBD-3P	Spindolin

- NA-GbpA: N-acetylglucosamine-binding protein A [*Shewanella xiamenensis*]
- CBP: Chitin-binding protein [*Pseudomonas* sp. CCOS 191]
- HP: Hypothetical protein [*Chryseobacterium* sp. CF365]
- MDT: Multidrug transporter [*Pseudomonas putida*]
- Interact GlcNAc: Probably interacts with GlcNAc residues.
- CBD-3P: Chitin-binding domain 3 protein
- Spindolin-RP: spindolin-related protein
- ACT: Aspartate carbamoyltransferase
- NA-GbpA: N-acetylglucosamine-binding protein GbpA

Cell-free expression system (Hamburg Univ)

In vitro expression for rapid screening of PETase candidates

