New Biorefinery technologies and new enzyme discovery tool: Optimized value from biomass conversion

Lene Lange, Professor, dr.scient.,
Research Director, Aalborg University,
Denmark

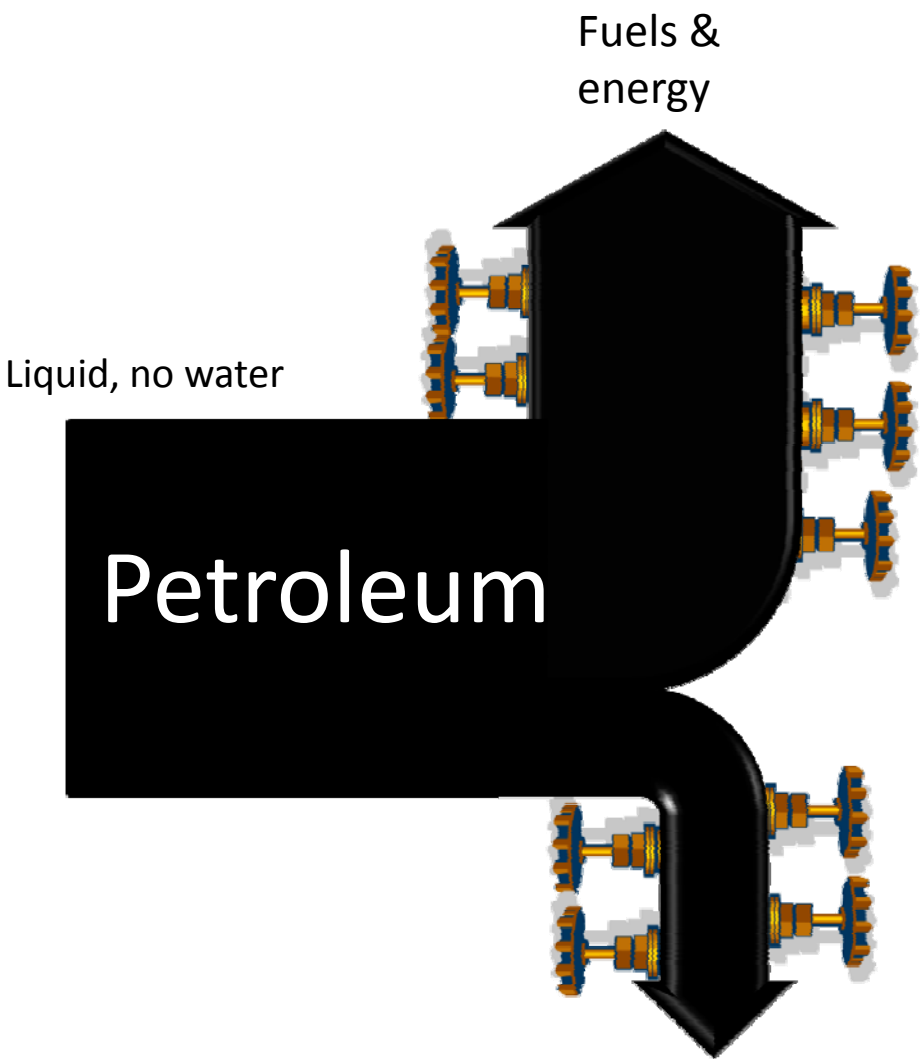
Aalborg University.
3 Campus locations in Denmark

Biorefinery focus too limited in scope

- Prime focus is on products of too low value
  - ex: biofuel; you need also higher value products to make the process ressource efficient and economically viable!
- Focus on substituting for fossils is too narrow
  - biomass has wider potentials than what we now get from crude oil

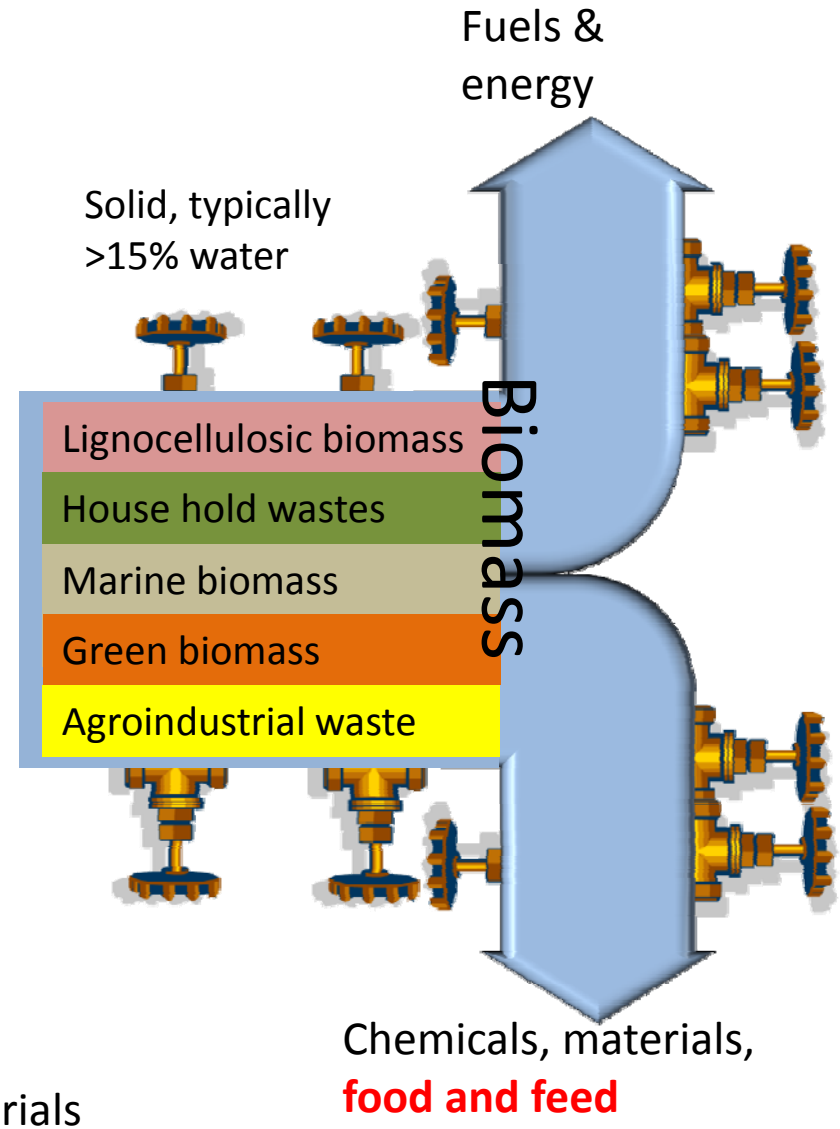
Oil-refinery

fossilization gives simplicity

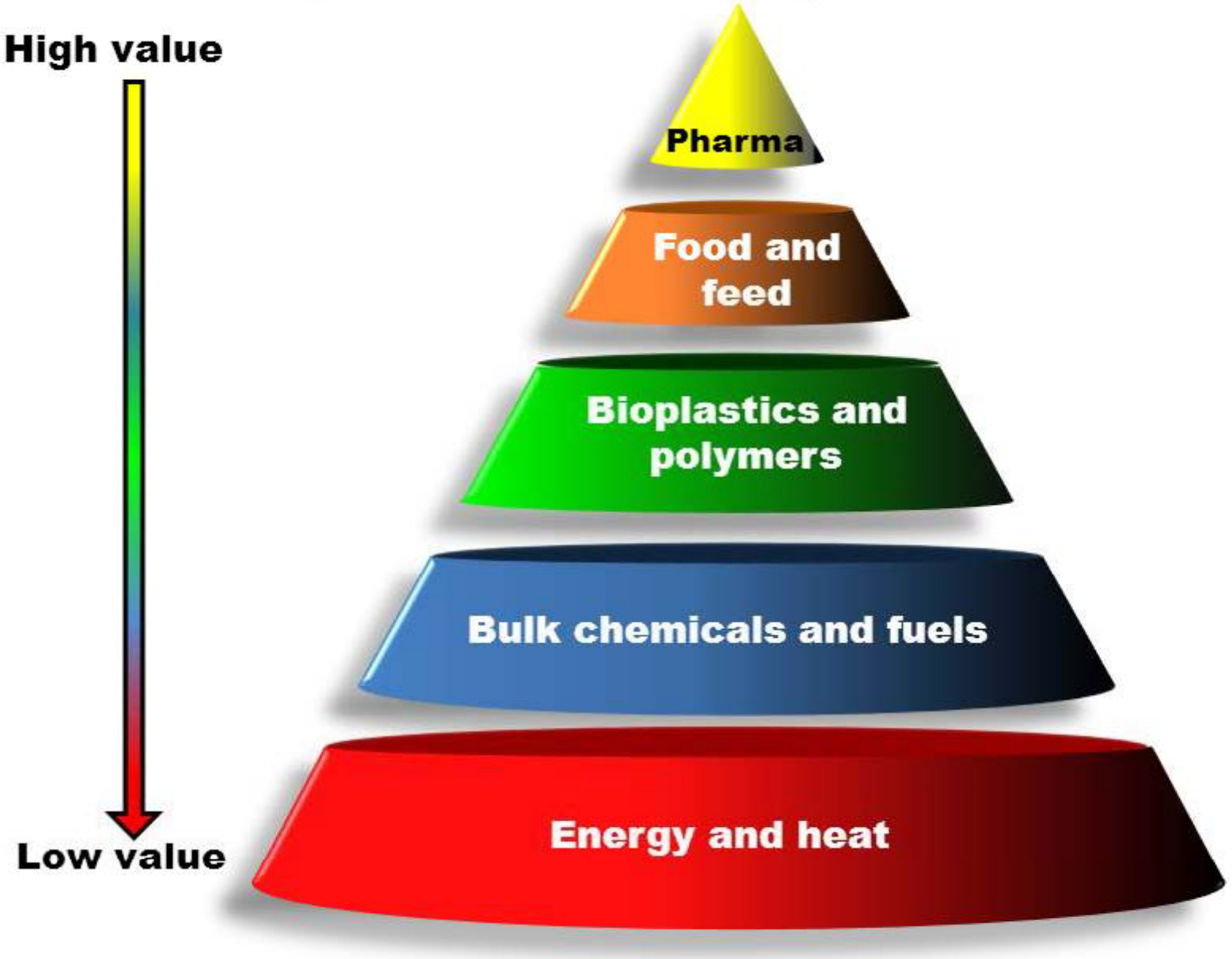


Bio-refinery

complexity intact!



Biomass cascade : High value compounds at the top, residual biomass for production of electricity and heat at the bottom



Many types of Biorefineries
-many new value chains (new enzymes needed)

***The Yellow biorefinery (straw and wood)**

***The Green Biorefinery (fresh leaves)**

The Brown biorefinery (from sludge)

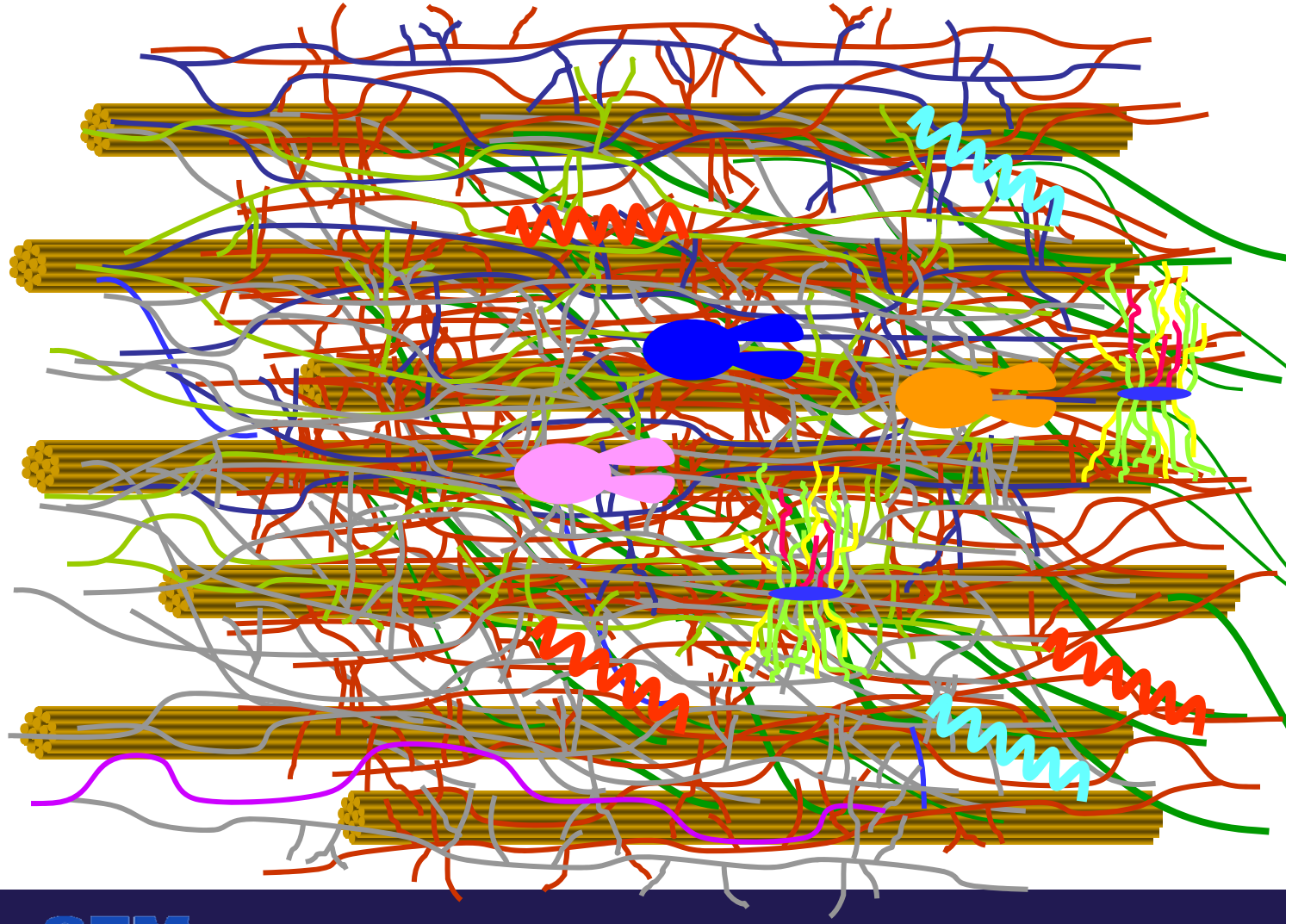
***The Blue Biorefinery (marin, seafood waste & algae)**

***The White Biorefinery (agroindustrial waste - !food grade!)**

***) bio-refineries with special potentials for food and feed!**

DO NOT BREAK IT ALL DOWN TO MONOMERS

MAKE *VALUE FROM NATURE´S COMPLEXITY*



Value chains from all three primary components of the plant cell wall biomass

- **Cellulose,**
- -C6 monomer sugar platform

  -For production of biofuel and biochemicals

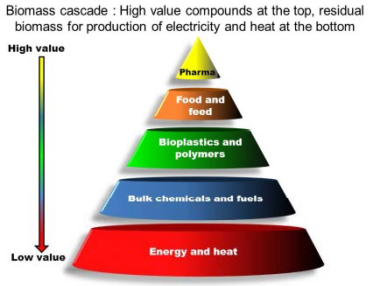
-

- **Hemi-cellulose,**
- **-**highly diverse polymer of C5 sugars, converted to active oligomers

  -for healthy feed and food ingredients

- **Lignin, polyphenolics**

  -carbon fibres, binders, coatings



Enzyme Discovery for development of new value chains from modified hemicellulose:

- The complexity of the C5 hemicellulose, incl
  - type of backbone,
  - branching pattern
  - substitutions

- harbours potentials for added value in the form of food and feed ingredients with prebiotic effect

New opportunity: The prebiotic effect of the hemicellulose-based food ingredients can be optimized enzymatically, guided by human gut biome studies

Value chain from C5 sugar polymers:

Discovery of many new hemicellulose modifying enzymes

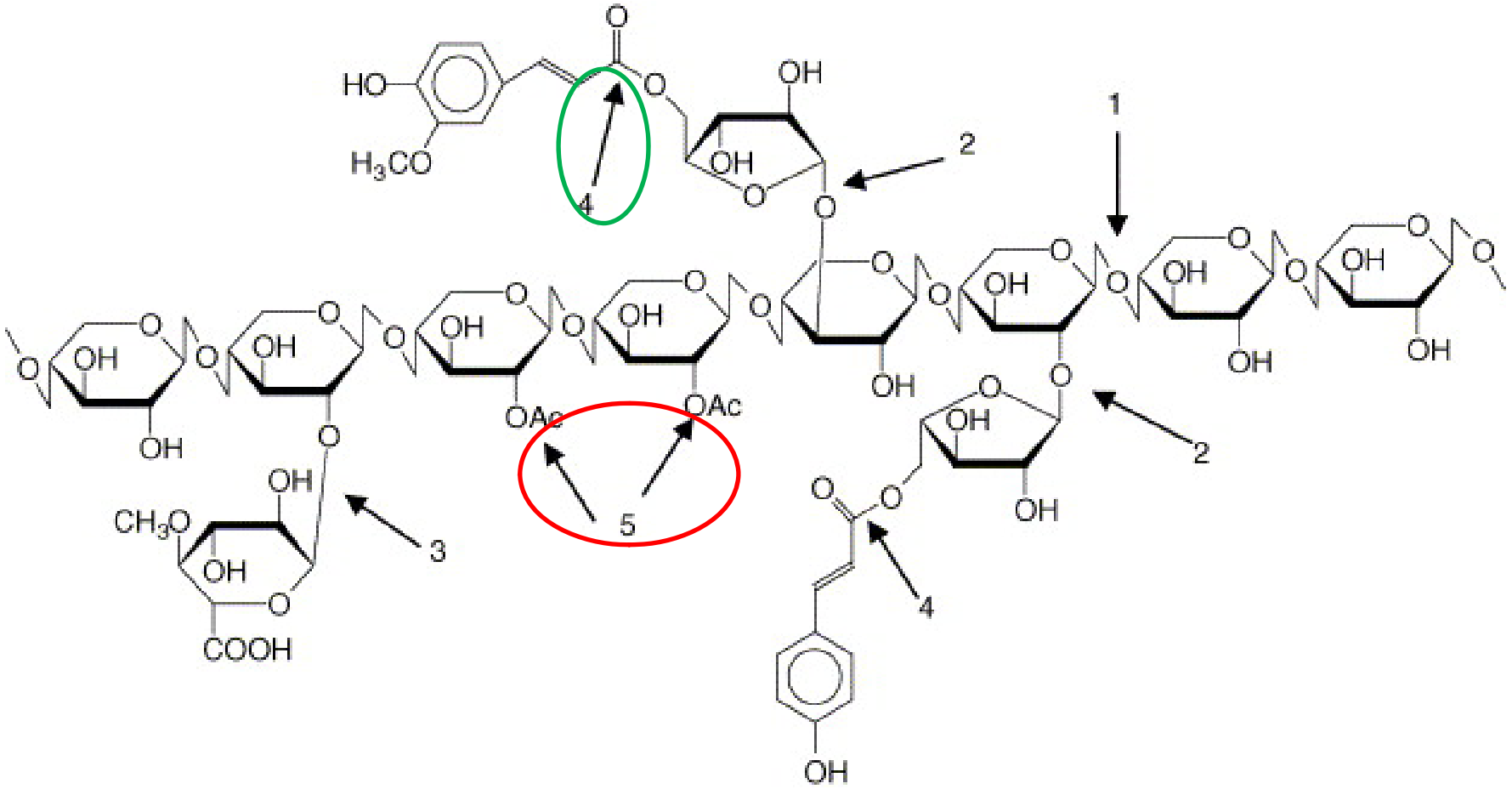


Figure. The structure of xylan and site of action of the enzymes of the xylanase complex. 1: endoxylanases; 2: α-L-arabinofuranosidases; 3: glucuronidases; 4: ferulic acid esterases; 5: acetyl xylan esterases. (Source: Chávez, R., Bull, P., Eyzaguirre, J., 2006.)

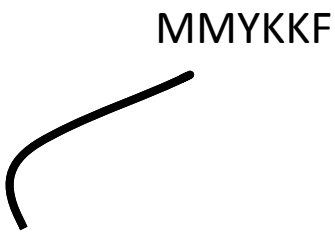
A new Tool now available: PPR, Peptide Pattern Recognition

- A non alignment based sequence analysis approach
  - providing a short cut to prediction of function from sequence

- *Busk & Lange, 2012 (patent published)*
- *Busk & Lange, 2013 (publication)*

Peptide Pattern Recognition

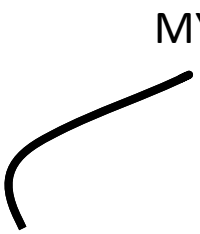
- A method for finding relationship between proteins
- Fast, easy, automatic!
- Predicts function of carbohydrate-modifying enzymes with 80 % - 97 % accuracy



MMYKKFAALAALVAGAAAQQACSLTTETHPRLTWKRCTSGGNCSTVN
GAVTIDANWRWTHTVSGSTNCYT

MRTAKFATLAALVASAAAQQACSLTTERHPSLSWKKCTAGGQCQTVQA
SITLDSNWRWTHQVSGSTNCYT

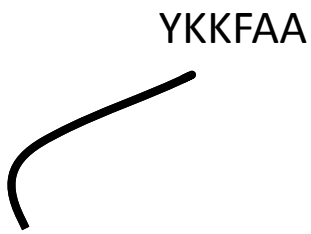
MVSAKFAALAALVASASAQQVCSLTPESHPPLTWQRCSAGGSCTNVAG
SVTLDSNWRWTHTLQGSTNCYS



MMYKKFAALAALVAGAAAQQACSLTTETHPRLTWKRCTSGGNCSTVN
GAVTIDANWRWTHTVSGSTNCYT

MRTAKFATLAALVASAAAQQACSLTTERHPSLSWKKCTAGGQCQTVQA
SITLDSNWRWTHQVSGSTNCYT

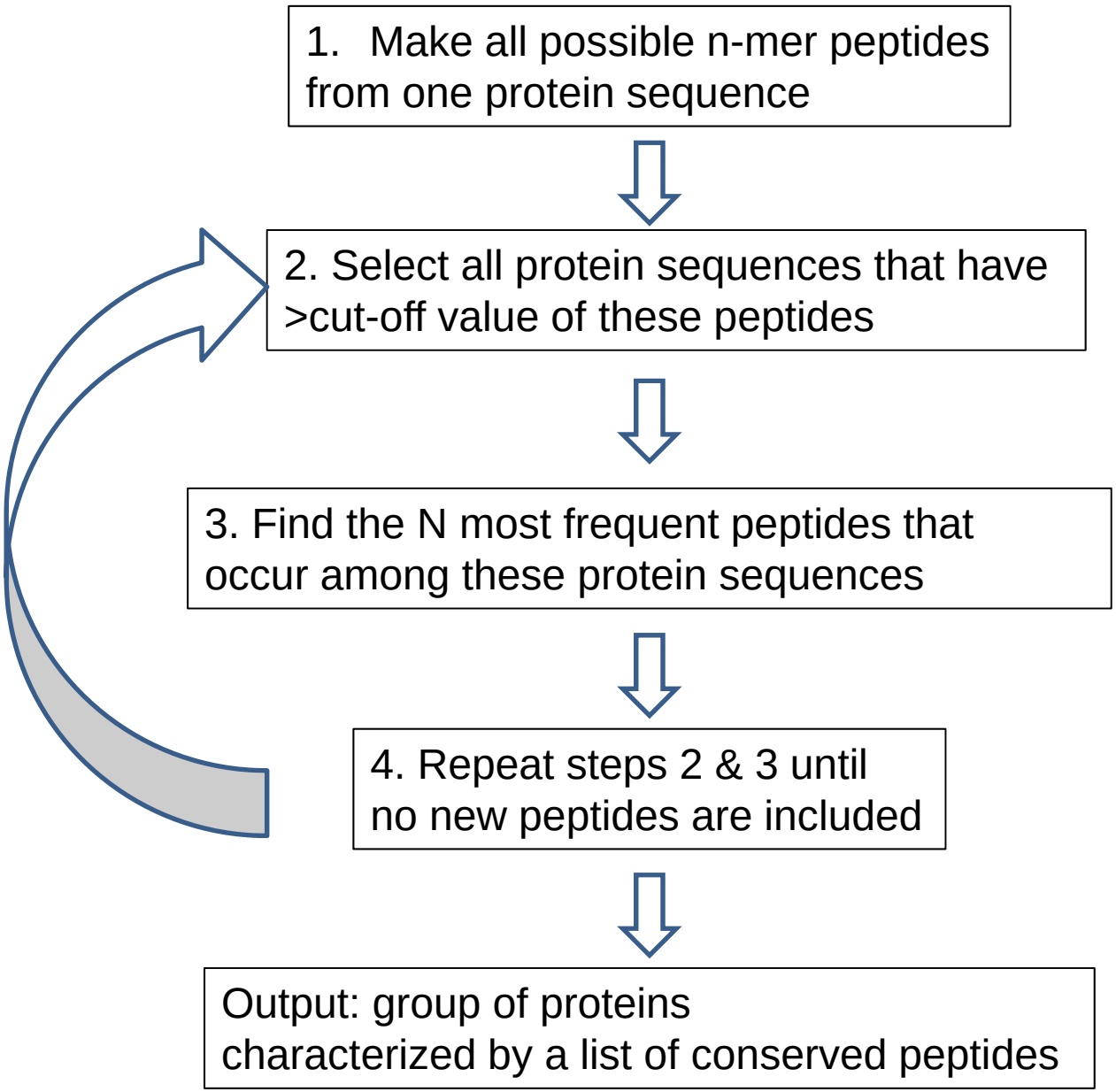
MVSAKFAALAALVASASAQQVCSLTPESHPPLTWQRCSAGGSCTNVAG
SVTLDSNWRWTHTLQGSTNCYS

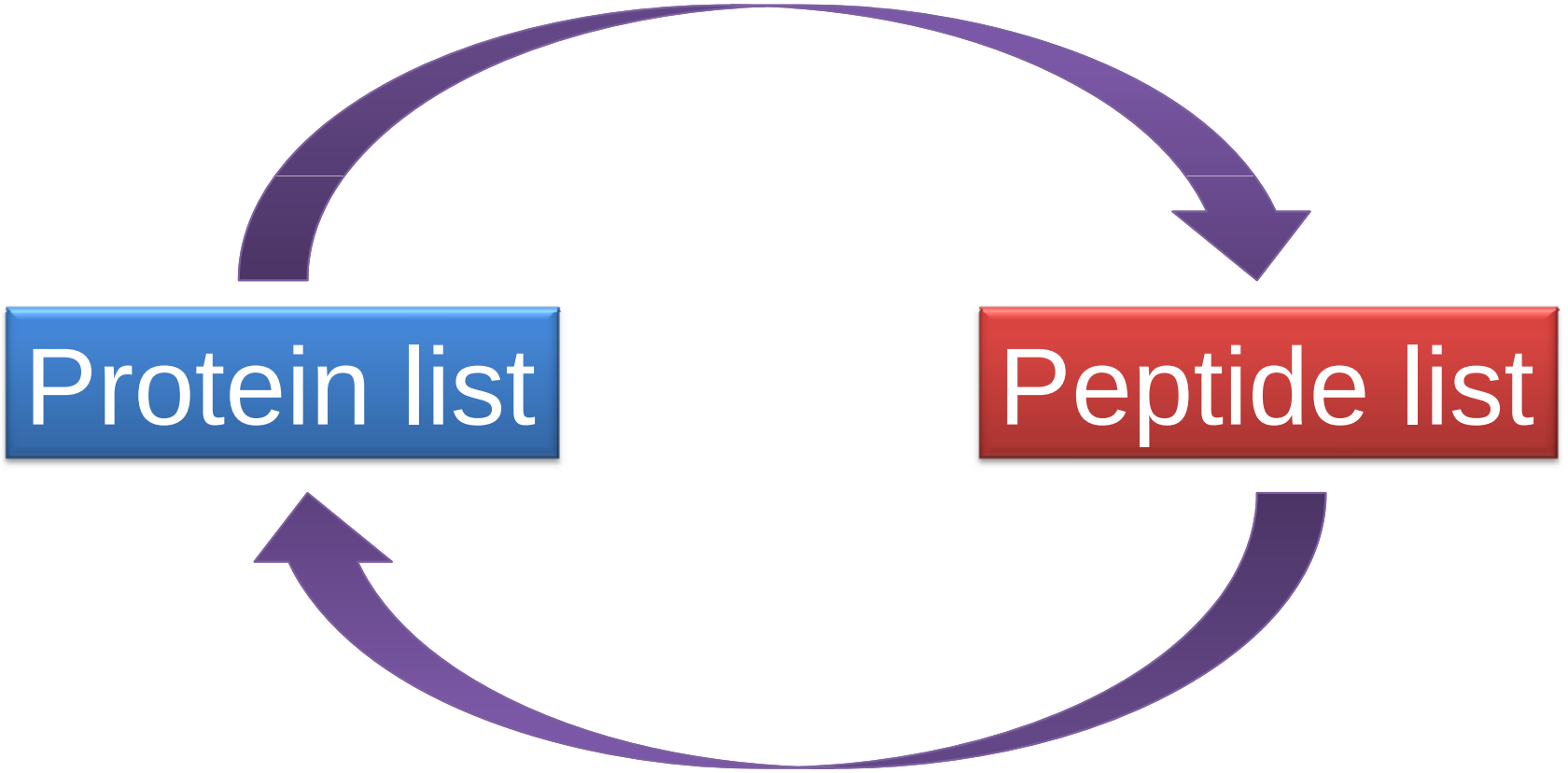


MMYKKFAALAALVAGAAAQQACSLTTETHPRLTWKRCTSGGNCSTVN
GAVTIDANWRWTHTVSGSTNCYT

MRTAKFATLAALVASAAAQQACSLTTERHPSLSWKKCTAGGQCQTVQA
SITLDSNWRWTHQVSGSTNCYT

MVSAKFAALAALVASASAQQVCSLTPESHPPLTWQRCSAGGSCTNVAG
SVTLDSNWRWTHTLQGSTNCYS


1. Make all possible n-mer peptides from one protein sequence
2. Select all protein sequences that have >cut-off value of these peptides
3. Find the N most frequent peptides that occur among these protein sequences
4. Repeat steps 2 & 3 until no new peptides are included
Output: group of proteins characterized by a list of conserved peptides



Output: group of proteins
and list of conserved peptides

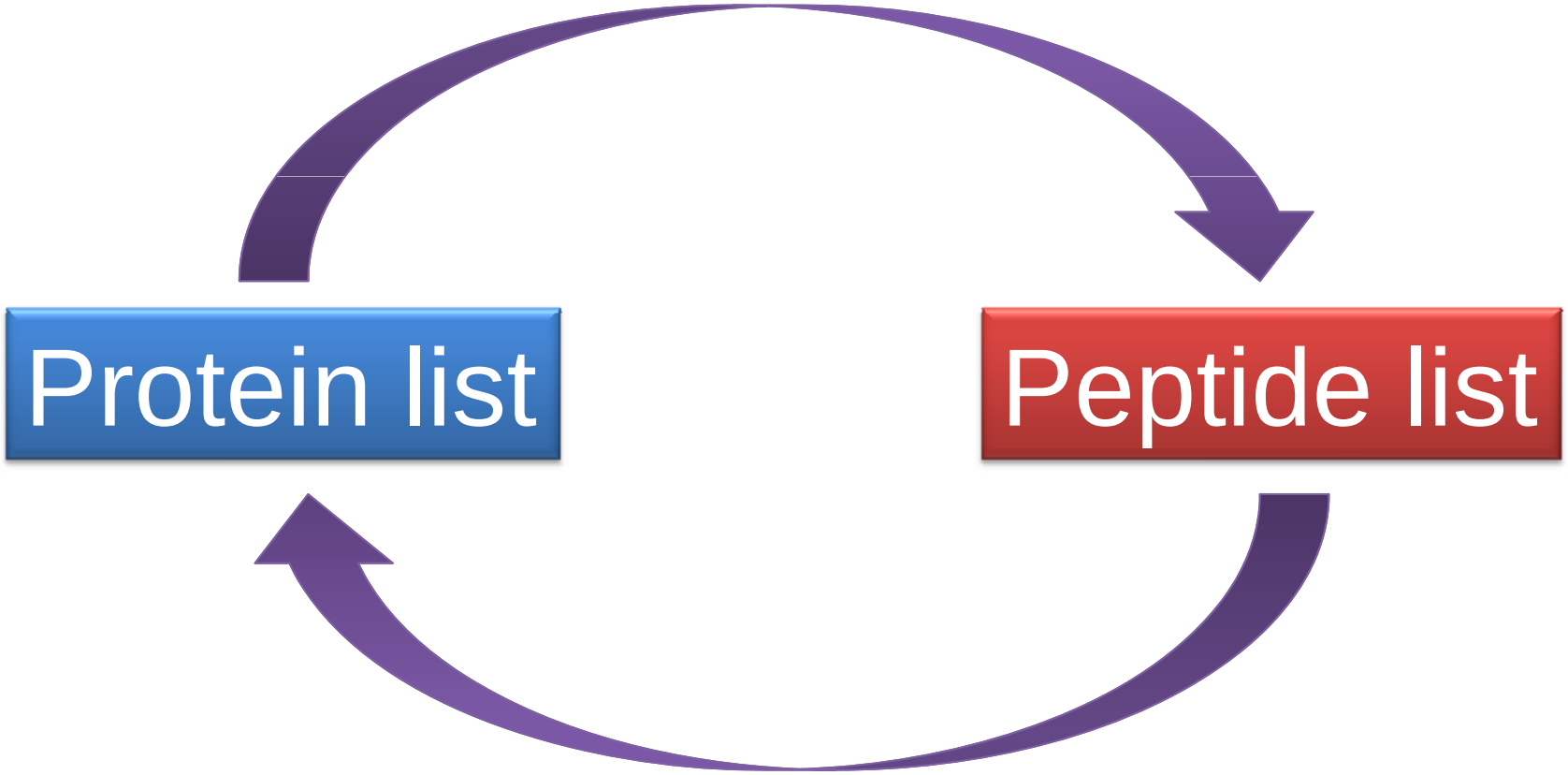
Validation of the new tool:
PPR analysis of Protein Family GH5

The GH5 family holds several functions

*cellulase (EC 3.2.1.4)
*glucan β-1,3-glucosidase (EC 3.2.1.58)
*glucan endo-1,6-β-glucosidase (EC 3.2.1.75)
*mannan endo-β-1,4-mannosidase (EC 3.2.1.78)

CAZy data base holds 115 characterized GH5
We mixed all GH5 sequences in silico and ran PPR

Can PPR sub-group sequences corectly to function?


Protein list
Peptide list

Prediction	Correct	Wrong
EC 3.2.1.4	59	1
EC 3.2.1.58	18	1
EC 3.2.1.75	7	0
EC 3.2.1.78	31	1
Total	**115**	**3**

PPR divided GH5 into subgroupings, correlating to function: 97 % correct!

=>

PPR predicted function from sequence with high precision

PPR analysis of
8138 GH13 proteins
(incl 204 characterized to function)

Can PPR sub-group in subfamilies correlated to function?

PPR\CAZY	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
17	57	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	252	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	70	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	190	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	102	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
38	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	1	0	0	104	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	724	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	280	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	1	712	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	48	9	26	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	0	2	0	0	0	0	0	0	0	0	0	0	18	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	46	186	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	389	3	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	3	0	0	0	0
41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	96	0	0	0	0	0	0	1	0	0	0	0	0	0	0	7	0	0	0	1	1
25	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26	0	0	0	0	0	24	0	0	0	0	0	0	0	13	0	0	0	1	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	82	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	56	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	189	86	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0
43	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	2	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	270	0	0	0	0	0	0	0	0	0	0
42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0	0	0	0	0	0	0	0	0
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	0	0	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	38	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	193	0	3	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	99	0	0	0	0	0	0	168	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	176	2	0	0	0	0	0	0	0	0	0	0	0	4	0	216	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	47	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	32	0	0	0
39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	31
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

185 of 204 experimentally characterized GH13 proteins

PPR analysis:
=> 91 % were classified to subfamilies with only one function

Use of the new PPR discovery tool

1. Finding suitable primer regions even when sequence identity is very low

2. Enabling protein classification at subfamily level, to guide functional discovery (=> predicting function from sequence)

3. Finding important amino acid motifs for functional analysis

4. Making targetted screening possible even of next generation sequenced metagenomic and metatranscriptomic libraries

Product development from Hemicellulose / C5 sugar polymer!
PPR Discovery of enzymes for hemicellulose modification

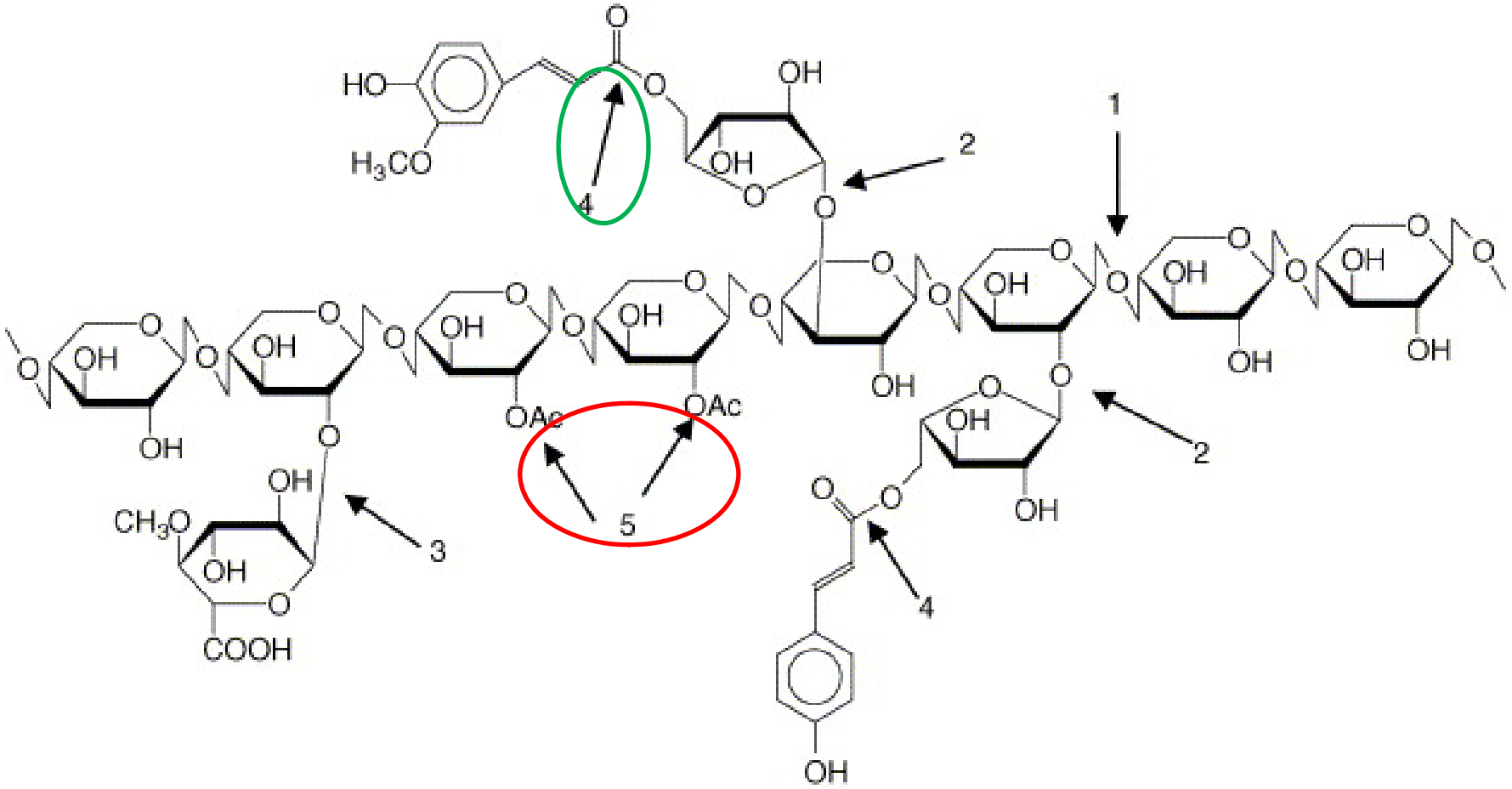


Figure. The structure of xylan and site of action of the enzymes of the xylanase complex. 1: endoxylanases; 2: α-L-arabinofuranosidases; 3: glucuronidases; 4: ferulic acid esterases; 5: acetyl xylan esterases. (Source: Chávez, R., Bull, P., Eyzaguirre, J., 2006.)

PPR discovery of C5 modifying enzymes: Acetyl xylan esterases

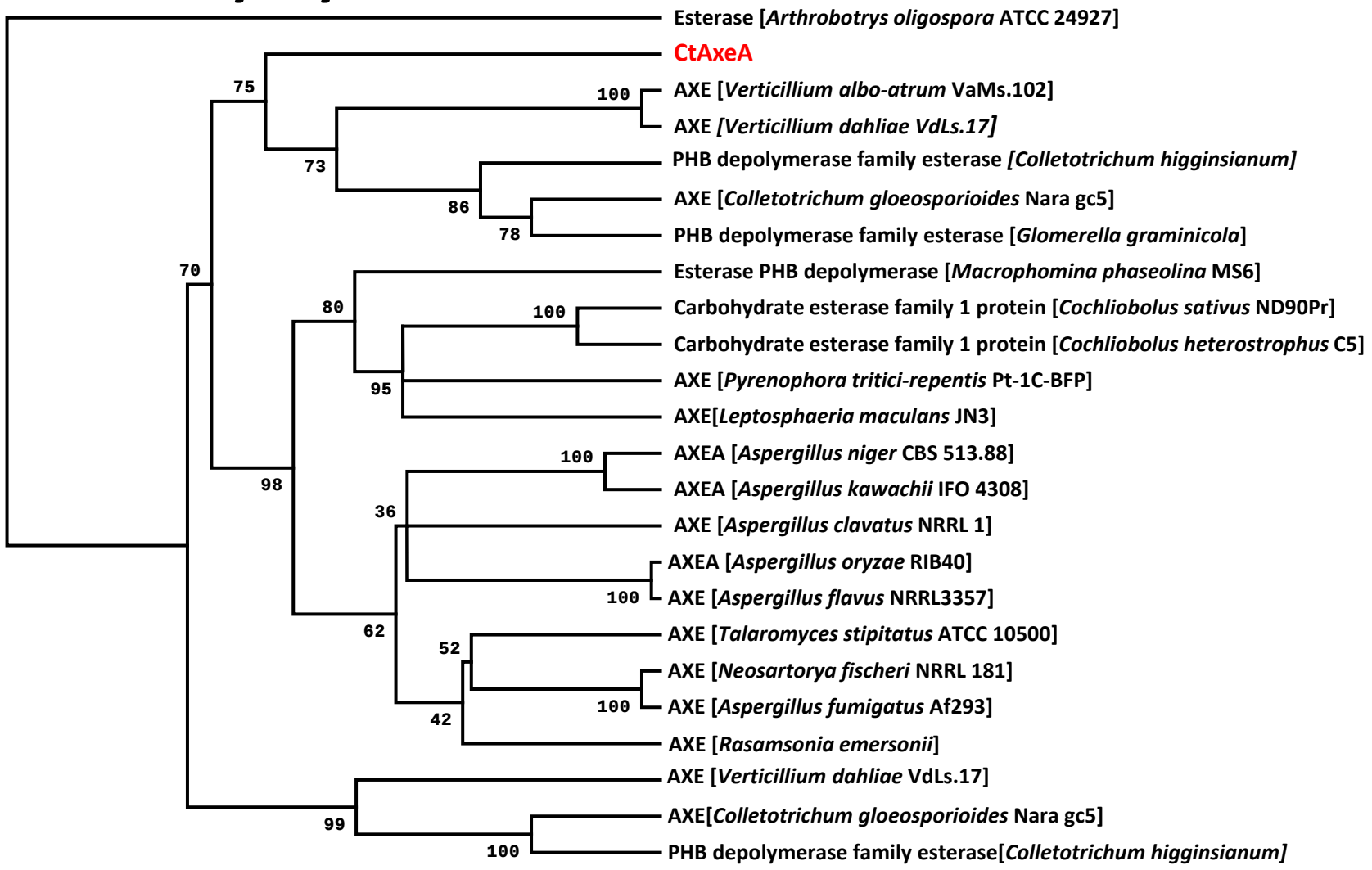


Fig. Neighbor-joining tree with 100 bootstrap replications.

Characterization of esterases: CtAxeA, CtaxeB and CtFaeA

Effect of temperature and pH on CtAxes and CtFaeA activity:

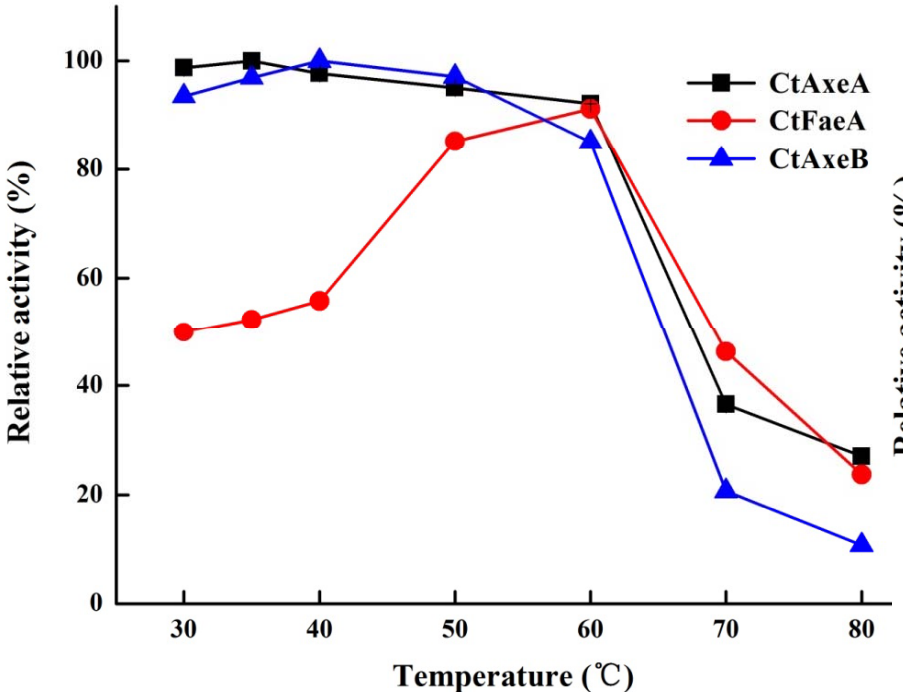


Fig. Effect of temperature on CtAxes and CtFaeA activity

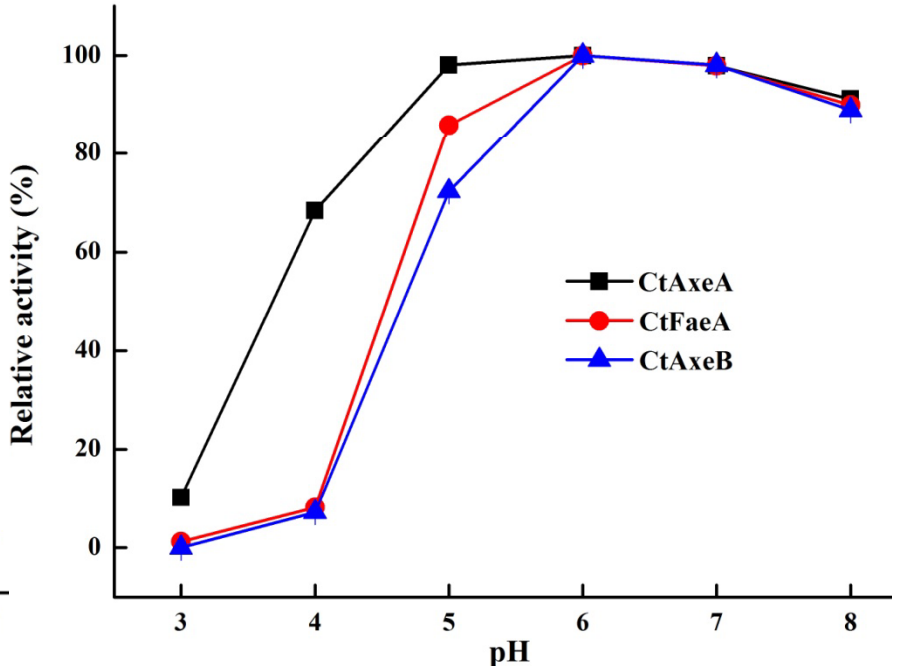
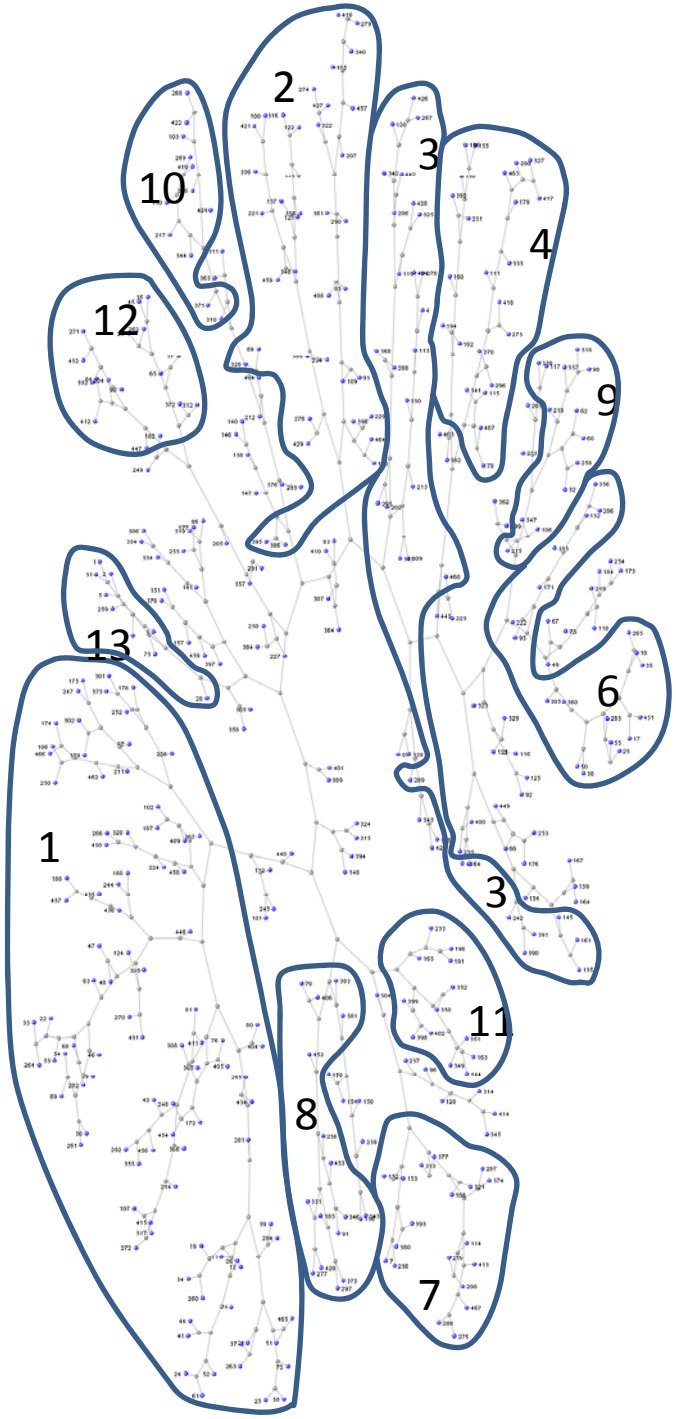
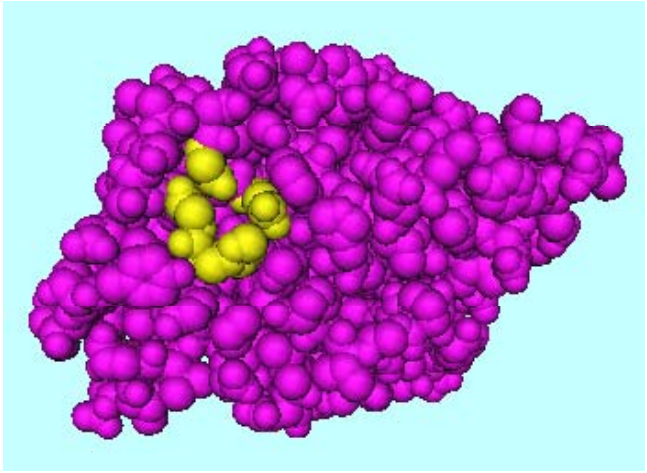


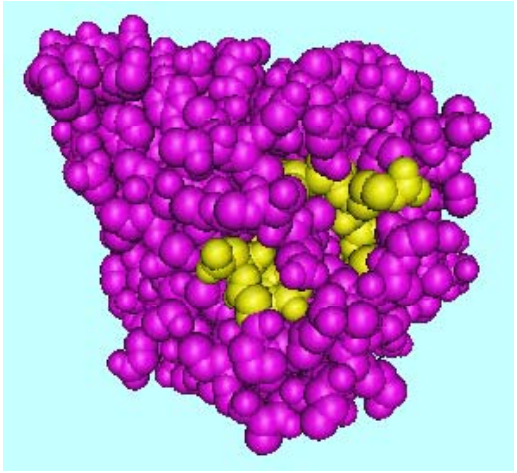
Fig. Effect of pH on CtAxes and CtFaeA activity

Phylogenetic tree of 467 GH61 proteins sequences with the subgroups indicated.
Each group was defined by 100 hexapeptides that were typically found in the proteins in the group.
Group 5 is too different from the other groups to be included in the tree.
The phylogentic tree was made with COBALT (Papadopoulos and Agarwala (2007) Bioinformatics 23, 1073-79.

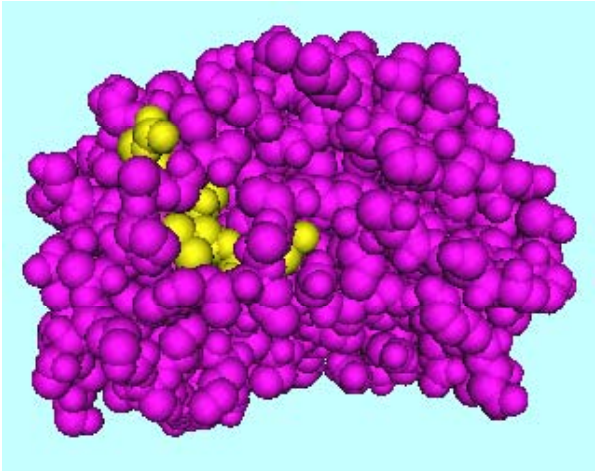




```
          114 118
group 1:  EFFKI
group 6:  DWFKI
group 7:  GWFKI
group 9:  DWFKV
```



```
          114  ..155
group 1:  E....GNY
group 2:  .....GQY
group 4:  .....GEY
group 5:  .....NRI
group 6:  D....GEY
group 7:  G....GNY
group 9:  D....GQY
group10:  .....GDY
```



```
          165         176
group 1:  HSA....GMQNY
group 2:  HSA....GAQFY
group 3:  HSA....GAQFY
group 7:  HQA....TPQFY
group 8:  H......GAQFY
group 9:  .......CAQIE
group11:  HLA....GAEFY
group12:  H......DPQFY
```

Top: Conserved amino acids mapping to the surface of group 1 proteins are indicated (yellow) on the crystal structure of a (Karkehabadi et al. (2008) J.Mol.Biol. 383, 144-54). **GH61/AAA9**

Discovery and characterization of a new family of lytic polysaccharide monooxygenases, AA11

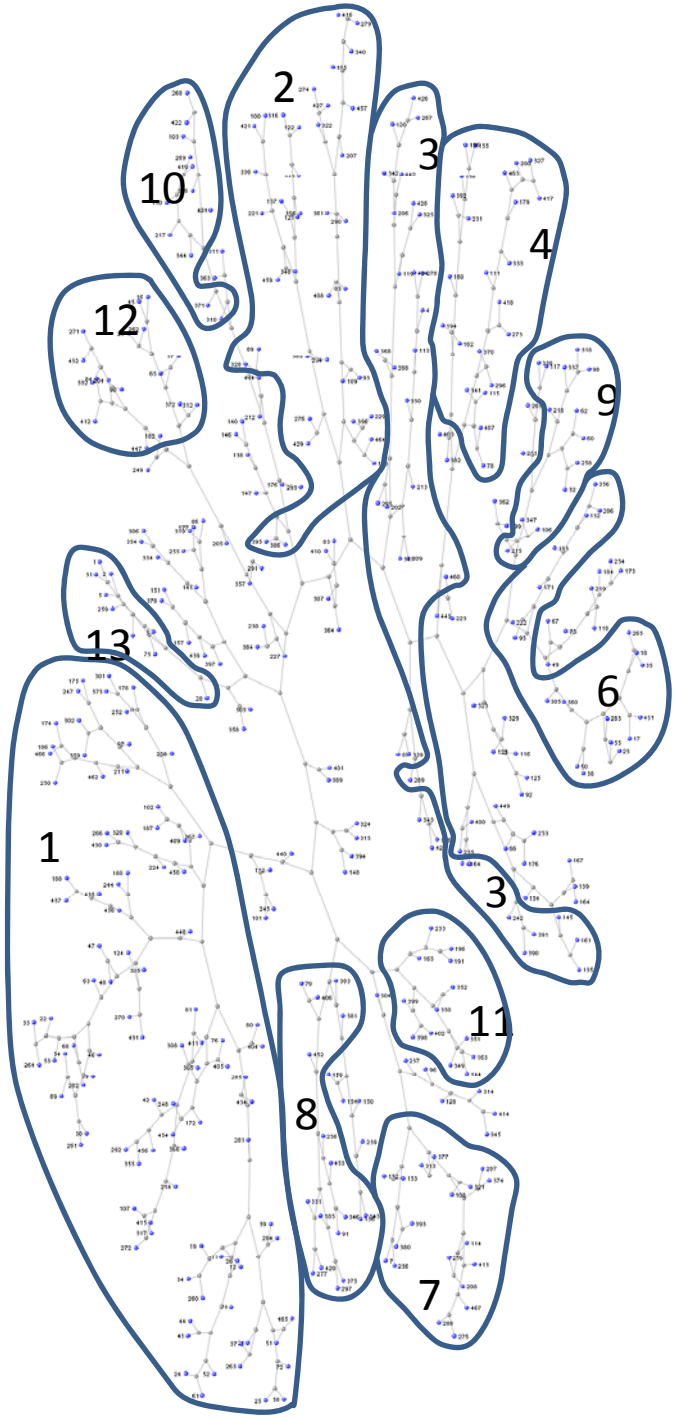
- By Glyn RHemsworth, Bernard Henrissat, Gideon J Davies & Paul H Walton, 2014:

- “Previous studies have identified two discrete sequence-based families of these enzymes termed AA9 (formerly GH61) and AA10 (formerly CBM33). Here, we report the discovery of a third family of LPMOs”

NOTE:

AA11 includes AA9/GH61 PPR subfamily 7 (Busk & Lange 2013), indicating the validity of the PPR tool –a short cut to identifying new subgroupings correlated to function

Phylogenetic tree of 467 GH61 proteins sequences with the subgroups indicated.
Each group was defined by 100 hexapeptides that were typically found in the proteins in the group.
Group 5 is too different from the other groups to be included in the tree.
The phylogentic tree was made with COBALT (Papadopoulos and Agarwala (2007) Bioinformatics 23, 1073-79.



Discovery from nature´s own biomass conversion hot spots!

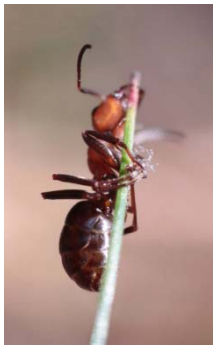
The Lange Lab the last 10 years,
Transcriptomics of the secretome!

Termite larvae (gut channel)

Cow (rumen of fistulated cow; +giraf!

Iberian snail (gut); duckweed (insect guts)

***Leaf cutter ant, fungal garden**



***Paxillus, Ectomycorrhizal fungus**
***Pandora, Enthomophthoralean**
insect pathogen

Learning from nature how to make a biorefinery!

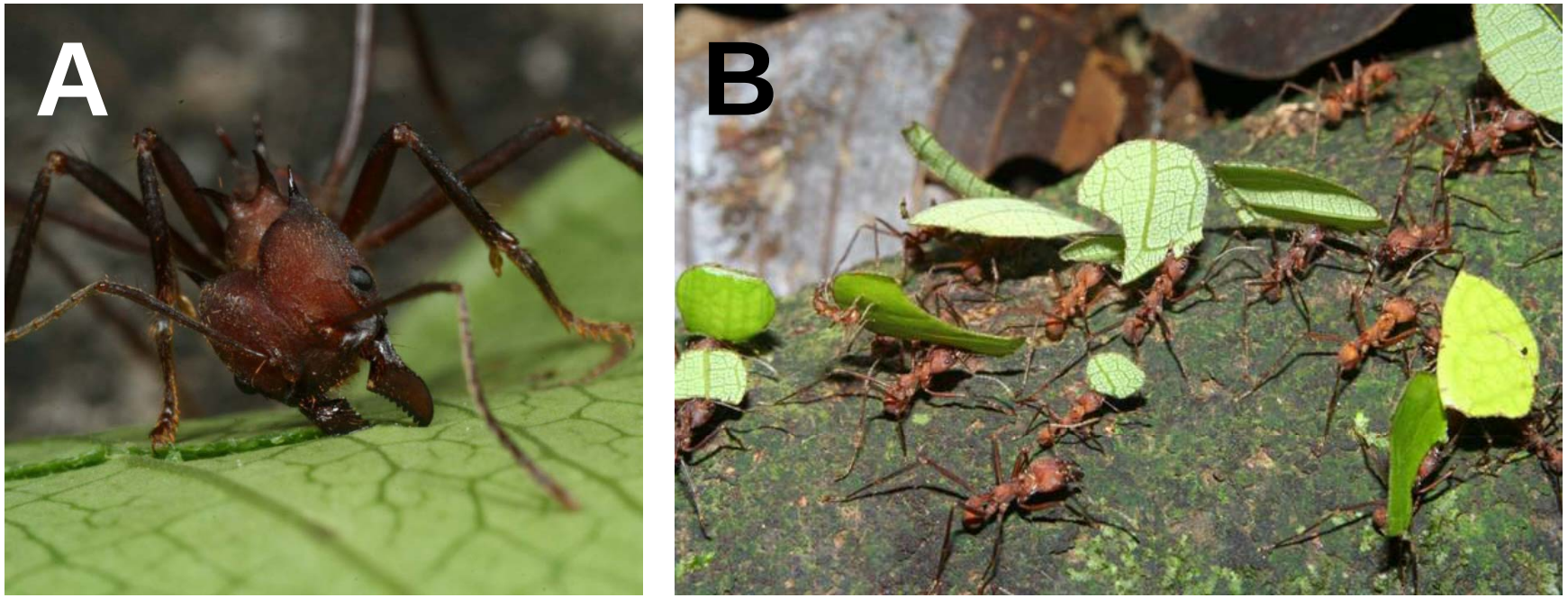
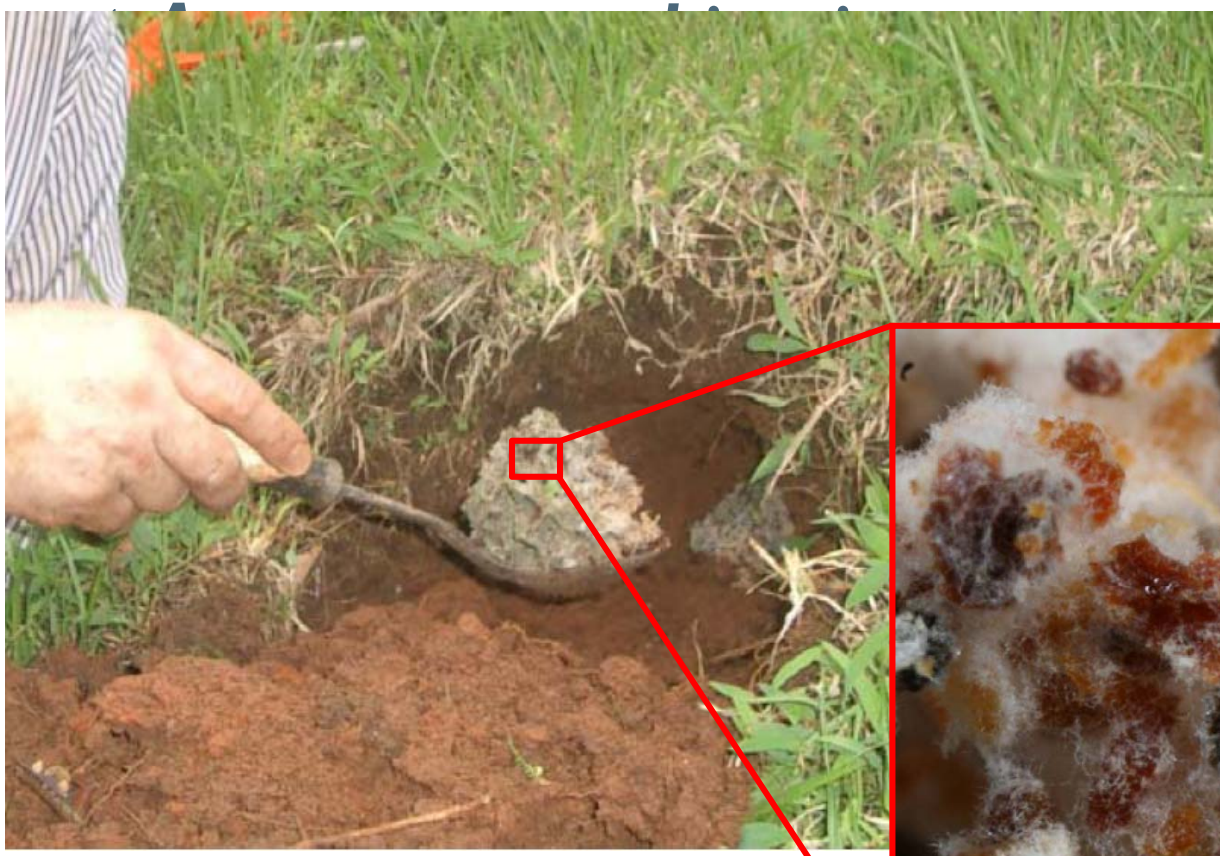


Fig. 1. Leaf-cutter ants. **A.** Ant worker cutting out leaf fragment using its heavy, serrated mandibles. **B.** Workers carrying leaf fragments back to the nest. Photos: Henrik H. de Fine Licht, University of Copenhagen, Denmark. (Grell & Lange)

Excavation of fungal garden of the leaf-cutter



Symbiont:
The basidiomycete fungus
Leucoagaricus gongylophorus

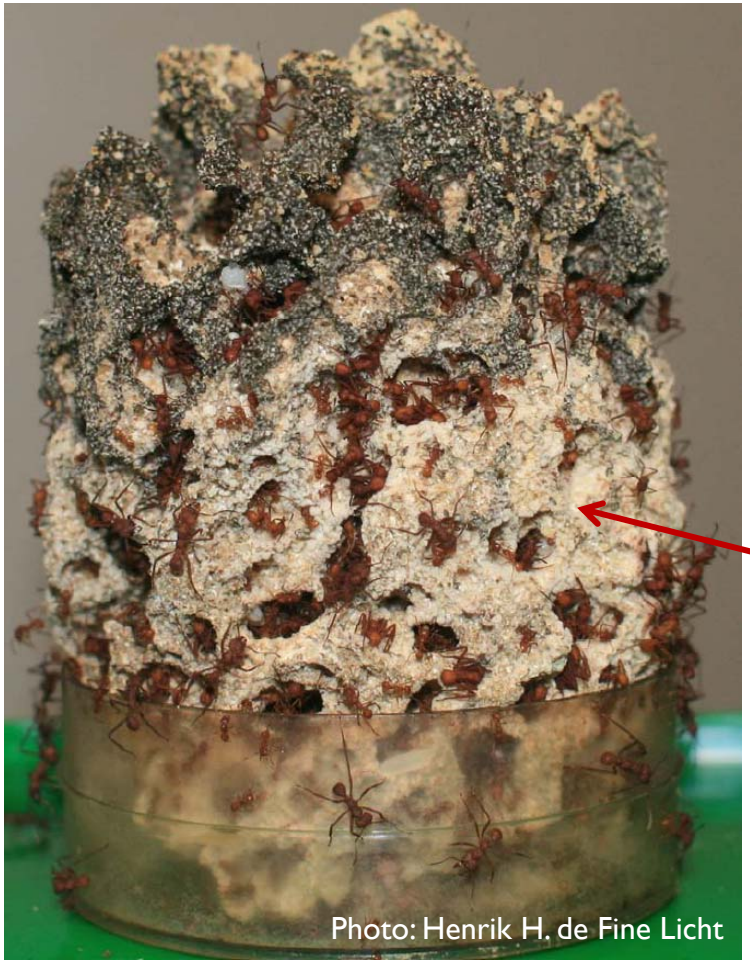
Photos: Henrik H. de Fine Licht,
Centre for Social Evolution

Fungal gardens in the lab

Photo: M.N. Grell

Fresh leaf fragments

Biomass conversion



Waste

Top – leaf fragments

Middle – gongylidia

Bottom – conversion completes

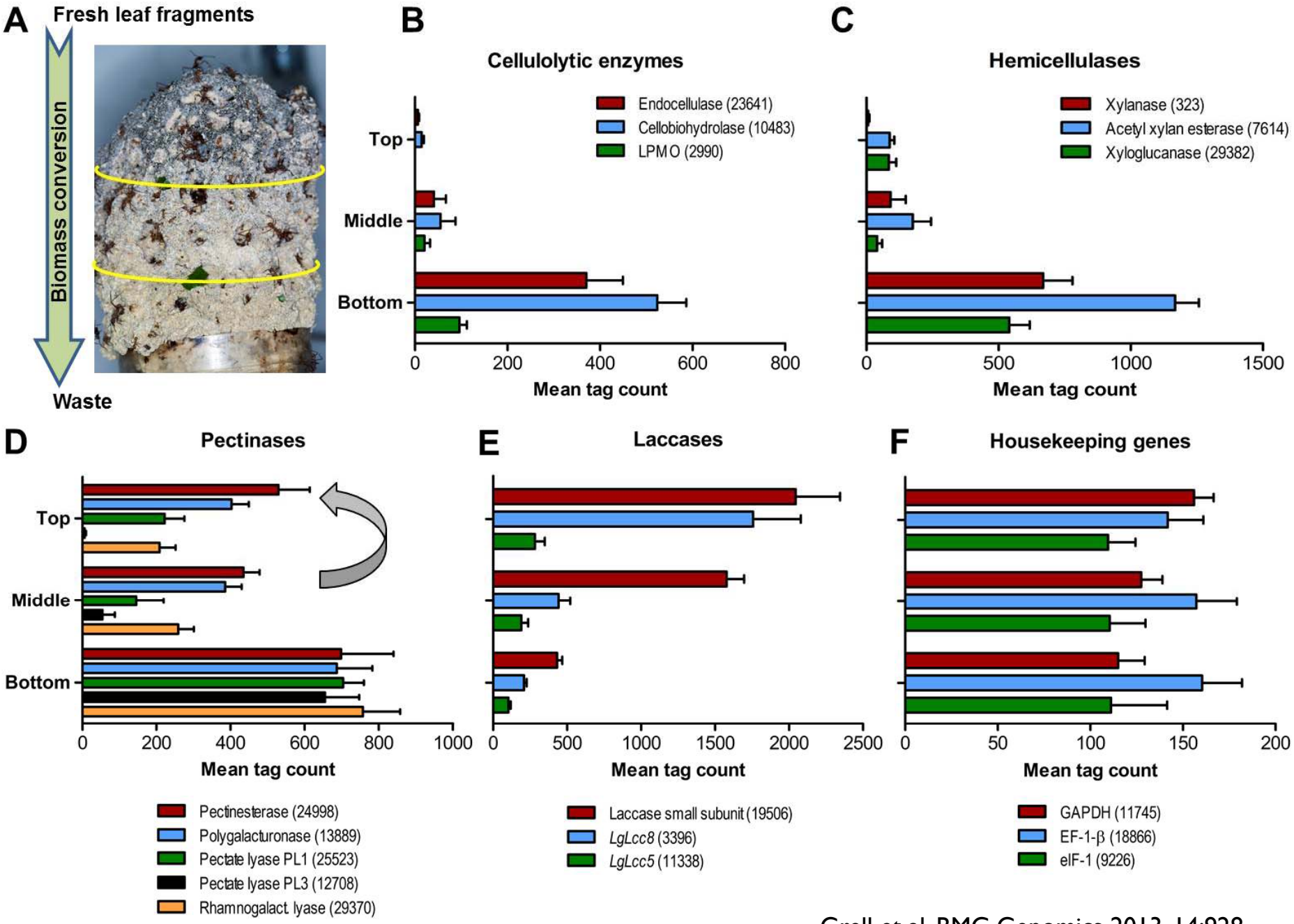
DeepSAGE analysis – summary

200,000 different mono-tags identified

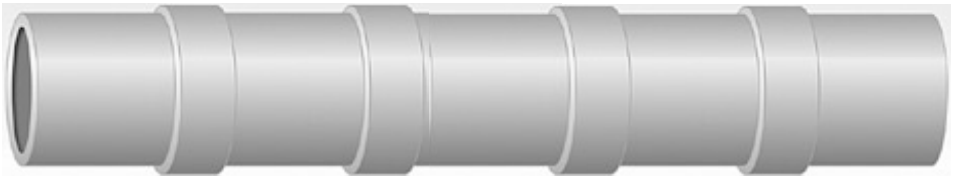
Mono-tags from each sample normalized to tags per million

Reduced data-set: 30,000 different tags (represented in at least 3 different samples, frequency ≥ 22)

9,5 mill tags totally from the 3 layers in the reduced data-set (14 samples)



Grell et al. BMC Genomics 2013, 14:928

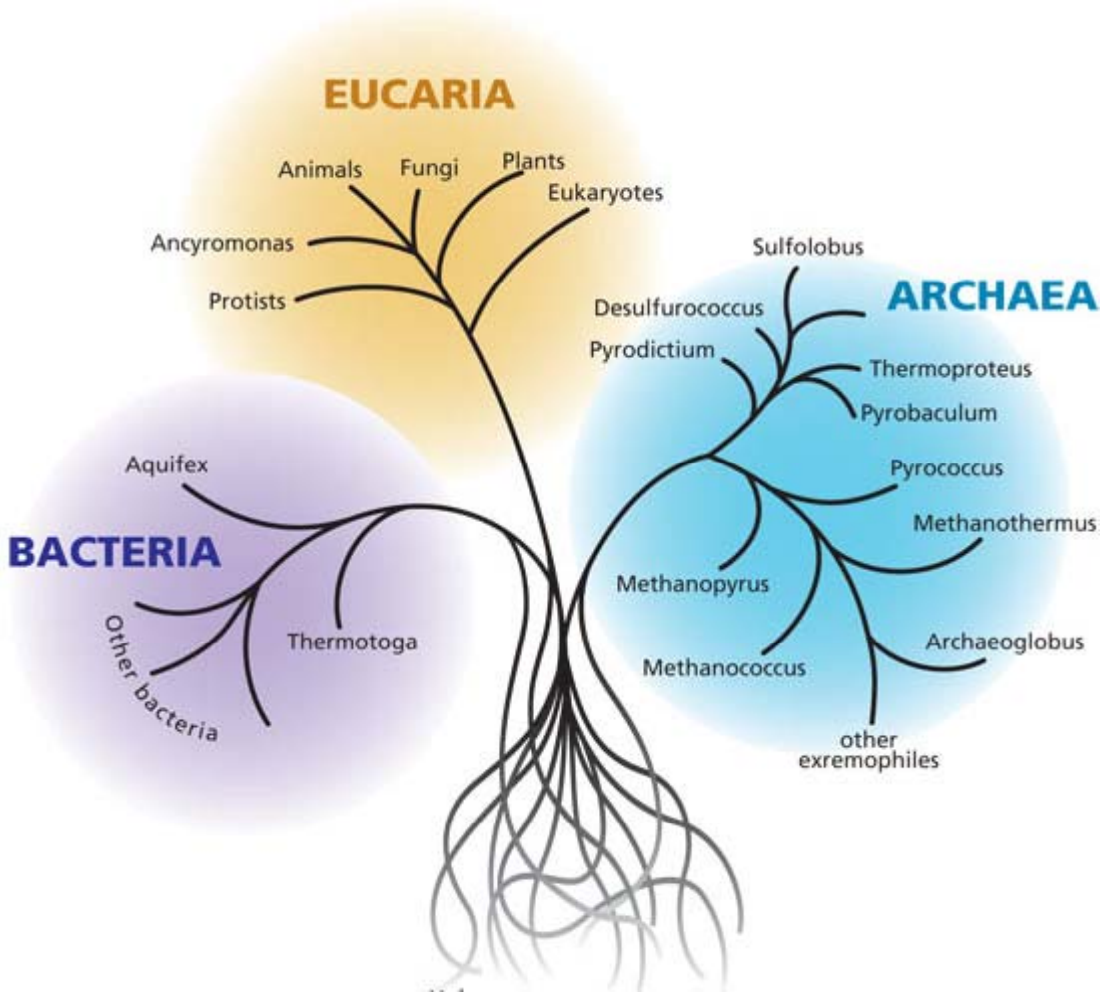


1. Find the interesting niche: Biology

2. Identify function and genes: Bioinformatics

3. Express enzymes/functions: Biochemistry

Find ecological niche/organisms




AALBORG UNIVERSITET
KØBENHAVN
aau.cph.dk

Hotpep

Finds all genes of interesting in a genome, transcriptome or metagenome

Speed is 30 megabases per hour per 100 protein families

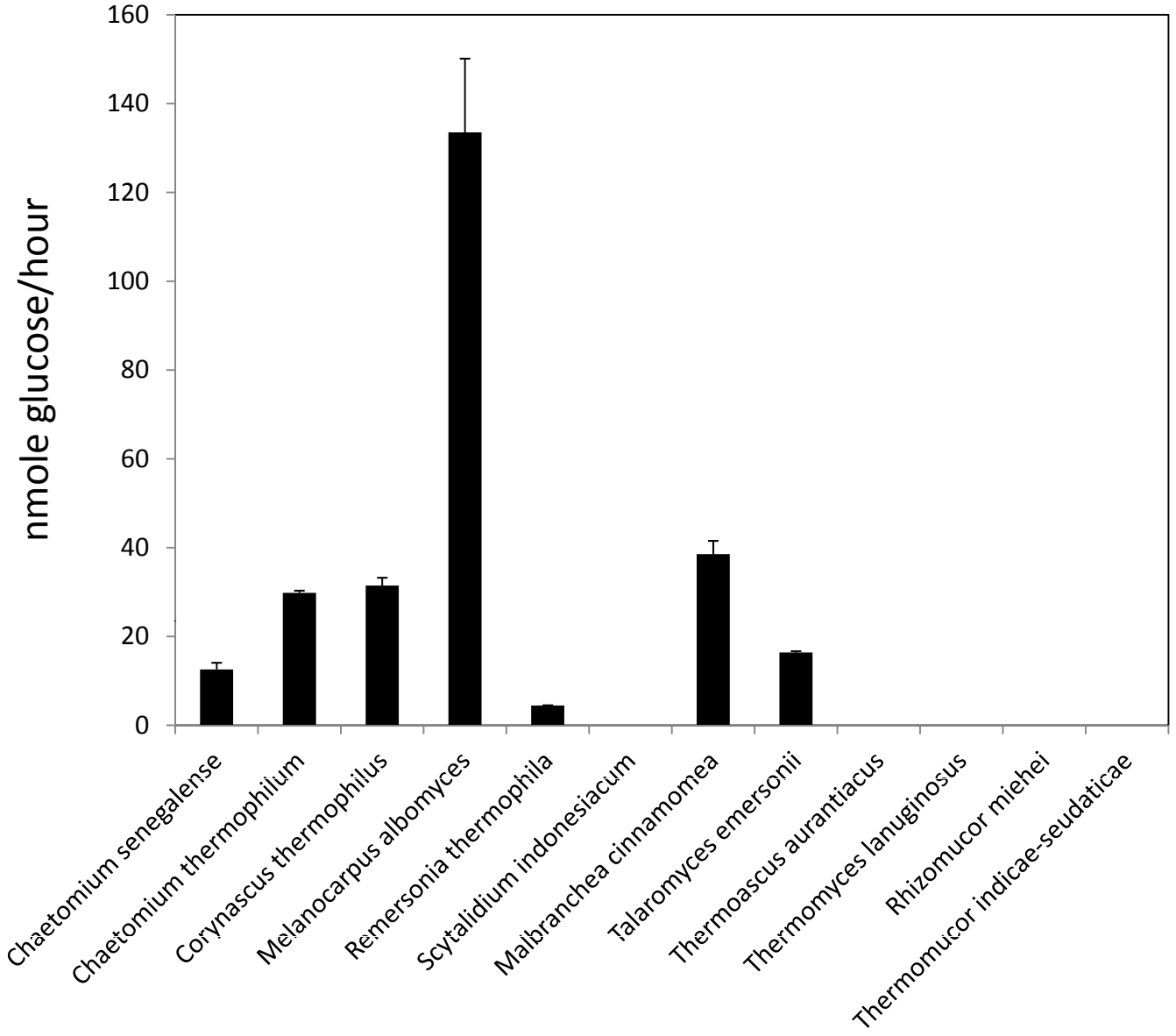
Latest development: HoTPeP, a version of PPR, designed for efficient data mining

HotPep -a short cut to mine genome databases:
-find list of GH families found
-predicts function and identify list of EC numbers
-in progress: discovery of a series of thermostable hemicellulose modifying enzymes

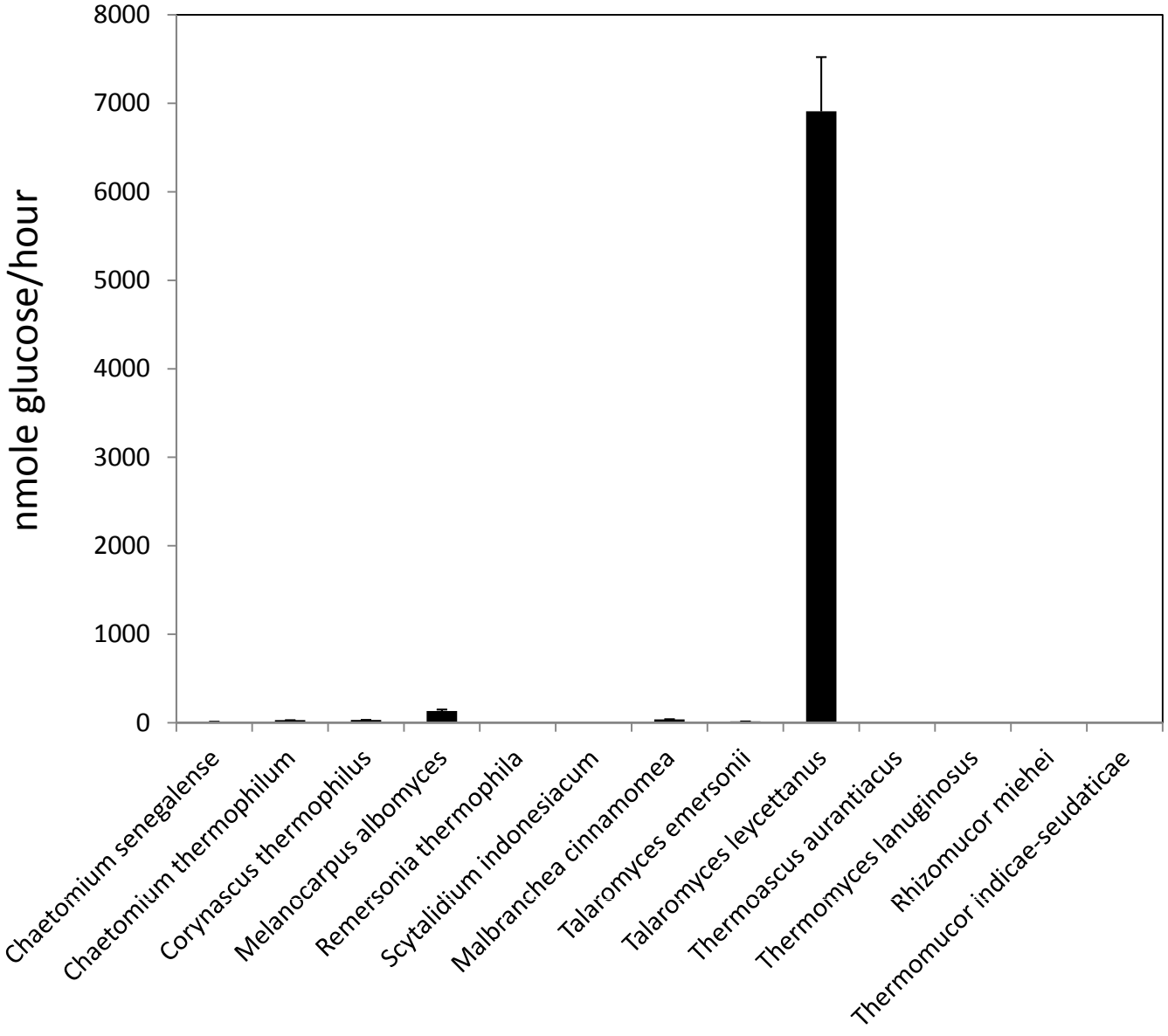
PPR is also a short cut to mining of metagenome databases!

EC number	Name	Hits			
3.2.1.4	cellulase	7	2 AA9	4 GH5	1 GH7
3.2.1.91	cellulose 1,4-beta-cellobiosidase (non-reducing end)	1	1 GH6		
3.2.1.176	cellulose 1,4-beta-cellobiosidase (reducing end)	2	2 GH7		
1.*.*.*	monooxygenase	4	4 AA9		
1.10.3.2	laccase	6	6 AA1		
3.2.1.1	alpha-amylase	4	4 GH13		
3.2.1.3	glucan 1,4-alpha-glucosidase	3	3 GH15		
3.2.1.6	endo-1,3(4)-beta-glucanase	1	1 GH16		
3.2.1.8	endo-1,4-beta-xylanase	5	3 GH10	2 GH11	
2.4.1.18	1,4-alpha-glucan branching enzyme	1	1 GH13		
2.4.1.25	4-alpha-glucanotransferase	1	1 GH13		
3.2.1.14	chitinase	12	12 GH18		
3.2.1.15	polygalacturonase	4	4 GH28		
3.2.1.20	alpha-glucosidase	6	2 GH13	4 GH31	
3.2.1.21	beta-glucosidase	11	1 GH0	2 GH1	8 GH3
3.2.1.22	alpha-galactosidase	5	4 GH27	1 GH36	
3.2.1.23	beta-galactosidase	4	4 GH35		
3.2.1.24	alpha-mannosidase	1	1 GH38		
3.2.1.25	beta-mannosidase	5	5 GH2		
3.2.1.28	alpha,alpha-trehalase	2	1 GH37	1 GH65	
3.2.1.31	beta-glucuronidase	1	1 GH79		
3.2.1.37	xylan 1,4-beta-xylosidase	4	2 GH3	2 GH43	
3.2.1.39	glucan endo-1,3-beta-D-glucosidase	4	3 GH16	1 GH81	
3.2.1.40	alpha-L-rhamnosidase	3	3 GH78		

Cellulolytic potential of fungi growing on cellulose



We have a winner!



Be inspired by processes in nature ex Lignin degradation

Giraf Rumen discovery: contribution to science

Be inspired by un-exploited potentials and un-met needs

***The Yellow biorefinery (straw and wood)**

***The Green Biorefinery (fresh leaves)**

The Brown biorefinery (from sludge)

***The Blue Biorefinery (marin, seafood waste & algae)**

***The White Biorefinery (agroindustrial waste - !food grade!)**

***) bio-refineries with special potentials for food and feed!**

Acknowledgements

Peter K Busk and Morten Grell
PhD students and post doc´s
Aalborg University, Denmark


Thank you for listening!
Lene
Aalborg University in
Copenhagen (new campus!)