

A new metagenome mining strategy unlocks Glycoside Phosphorylases discovery

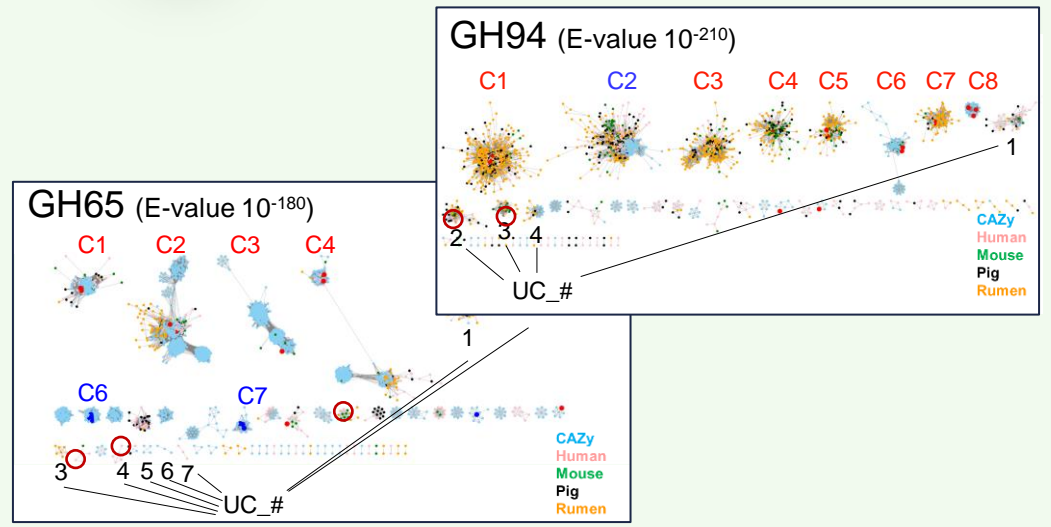
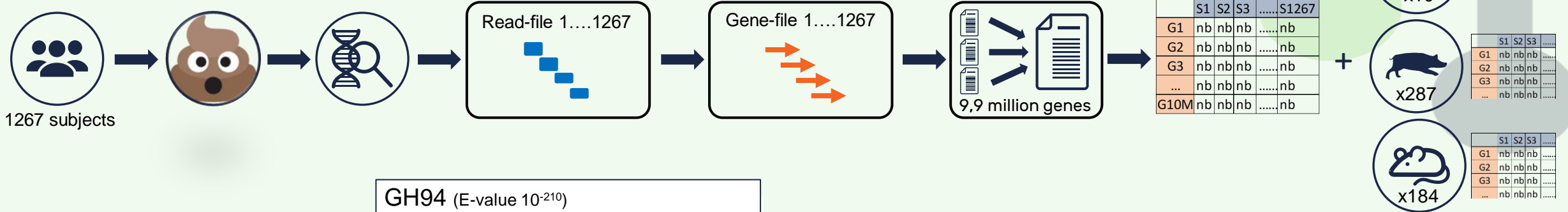
Dr. Simon LADEVEZE

Toulouse Biotechnology Institute, Bio & Chemical Engineering



- <https://radicalz.eu/>
- www.toulouse-biotechnology-institute.fr
- simon.ladeveze@insa-toulouse.fr
- veronese@insa-toulouse.fr

Mining Mammal Gut Metagenomes...



UC_# : meta-nodes with no characterized members
 Characterized members : ● GPs ● GHs

○
 Selection of Glycoside Phosphorylase candidates

26,800 non-redundant mammal gut metagenomic sequences

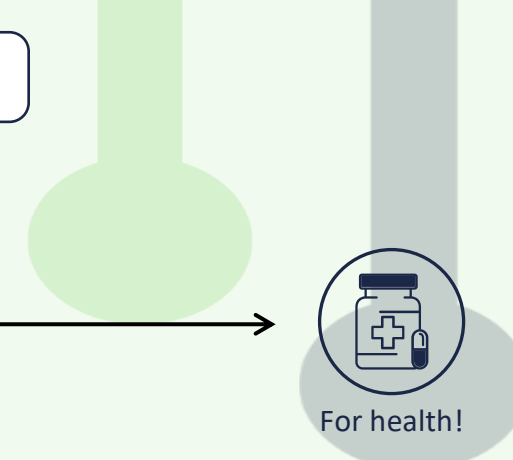
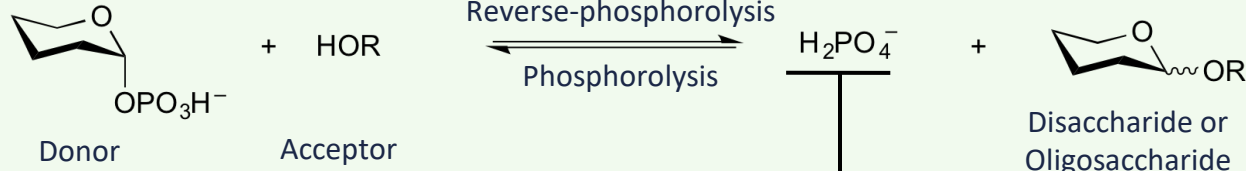
- GH65
- GH94
- GH112
- GH130
- GH149
- GH161



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000560

...for Glycoside Phosphorylases

Glycoside Phosphorylases ?



Putative GPs

- ✓ No signal peptide
- ✓ Conserved proton donor only (invertin)
- ✓ Conserved P_i binding residues
- ✓ Monomodular only

Putative GHs

- ✓ Signal peptide
- ✓ Conserved proton donor
- ✓ Conserved base
- ✓ Possibly multimodular

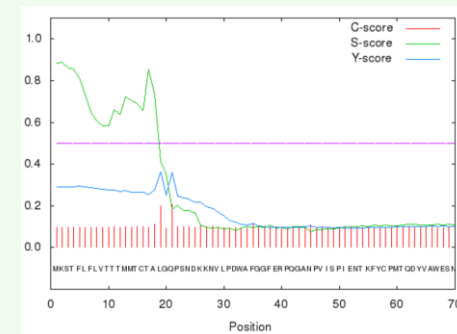
Multiple sequence alignments

```

DSPAINIGYTEDFIHFVKMNHFPF
-GPSIYIAETVDFKNEKMLPTGFL
-GPDIYIAKTTFDKRFDMIPTGFL
-GPDIYIAKTTFDKRFDMIPTGFL
-GPCHIAKSSDLKTFVLPSTGFL
G-AAIGVAKTQDFKTFVRIENPFL
GGPSLAIATHKDFKNEYERLANGFL
G-PTLGIGKTKDFKYFELIDNPFL
G-PTLGVGRTKDFKDFELIDNPFL
G-PTLGIGKTKDFKDFELLDNPYL
G-PTLGIGKTKDFKNEFELLDNPFL
G-PTLGVGKTKDFKDFELLDNPFL
G-PTLGIGKTKDFKDFELVDNPFL
G-PTLGIGKTKDFKYFELVDNPFL
G-PTLGIGKTKDFKYFELVDNPFL
G-PTISIAVTYDFKKEVVKLEAGFM
G-PTIAIAKTKDFKKEVKLDAPFM
G-PTISIAKTYDFKKEVKFDNPFL
G-PTISIAKTKDFKKEVKFDNPFL
    
```

- ➔ catalytic residues ?
- ➔ P_i binding residues ?
- ➔ +1 subsite residues ?

Signal peptide prediction



SignalP 4.1 Server



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000560



Come and see the results!
Teaser: 9 new GPs, new structure, new function

Acknowledgments

Cooper N.
Cioci G.,
Terrapon N.,
Lombard V.
Henrissat B.
Li A.
Laville E.
Esque J.
Potocki-Veronese G.

Poster #2

Sequence-similarity network approach
Identifies new Glycoside Phosphorylases

Logo: RADICALZ

Ladeveze S., Cooper N., Cioci G., Terrapon N., Lombard V., Henrissat B., Li A., Laville E., Esque J. & Potocki-Veronese G. Toulouse Biotechnology Institute (TBI), Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France

Background
The reversible reactions catalyzed by Glycoside Phosphorylases (GPs) have huge potential for glycochemistry. Although being hard to identify in sequence databases as they are so similar to GHs and GTs, GPs found in inverting GH families bear unique sequence specificities. Prepositing them requires the analysis of very large sequence alignments. Sequence Similarity Networks (SSN) allow the easy visualization of very large sequence alignments. Sequence Similarity Networks (SSN) specifically guiding the experimenter into unexplored sequence space, where new functions can be found. It is a valuable tool for the unbiased creation of CAZy subfamilies.

Strategy
From phylogenetic trees... to sequence similarity networks (SSN) ...For new Glycoside Phosphorylases

- Mammalian gut metagenomes (human, pig, mouse, cow rumen)
- Δ GP-containing inverting-GH families analyzed (GH65, 94, 112, 130, 149 and GH161)
- Monomodular enzymes only, with no signal peptide
- Only catalytic acid/base conserved (no nucleophile)
- Uncharacterized sequence clusters

Results
26,800 mammal gut metagenome sequences
19 selected candidates
11 soluble proteins
9 Active Glycoside Phosphorylases

Protein	GH65_UC3	GH64_UC2	GH64_UC3	GH112_SH2	GH149_UC1	GH149_UC2	GH161_SH2	GH161_SH3	GH161_SH5
Donor	zllc0237	zllc0237	zllc0237	zllc0237	zllc0237	zllc0237	zllc0237	zllc0237	zllc0237
Host	Human	Human	Human	Human	Human	Human	Human	Human	Human
Substrate	Galactose	Galactose	Galactose	Galactose	Galactose	Galactose	Galactose	Galactose	Galactose
Enzyme class	GH65	GH64	GH64	GH112	GH149	GH149	GH161	GH161	GH161
Enzyme activity	Galactose 4-epimerase	Galactose 4-epimerase	Galactose 4-epimerase	Galactose 4-epimerase	Galactose 4-epimerase	Galactose 4-epimerase	Galactose 4-epimerase	Galactose 4-epimerase	Galactose 4-epimerase
Enzyme activity	Galactose 4-epimerase	Galactose 4-epimerase	Galactose 4-epimerase	Galactose 4-epimerase	Galactose 4-epimerase	Galactose 4-epimerase	Galactose 4-epimerase	Galactose 4-epimerase	Galactose 4-epimerase

References:
1. Li A. et al. (2022). Discovery and Biotechnological Exploitation of Glycoside-Phosphorylases. *ACS* 23, 3043.
2. Anthonis W. J. et al. (2009). Using Sequence Similarity Networks for Visualization of Relationships Between Protein Subfamilies. *PLoS ONE* 4, e4245.
3. Li A. et al. (2020). Analysis of the diversity of the glycoside phosphorylase family 150 in mammalian gut metagenomes reveals a new mammalian glycoside phosphorylase subfamily. *Microbial Genomics* 6, e000100.
4. Hornung, B. V. H. & Terrapon, N. (2020). An objective criterion to evaluate sequence-similarity networks helps in finding the protein family sequence space. <https://doi.org/10.1101/2020.04.19.488343>.

Contact: simon.ladeveze@insa-toulouse.fr, veronese@insa-toulouse.fr
<https://radicalz.eu/>, www.toulouse-biotechnology-institute.fr



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000560