Mining the wheat grain proteome

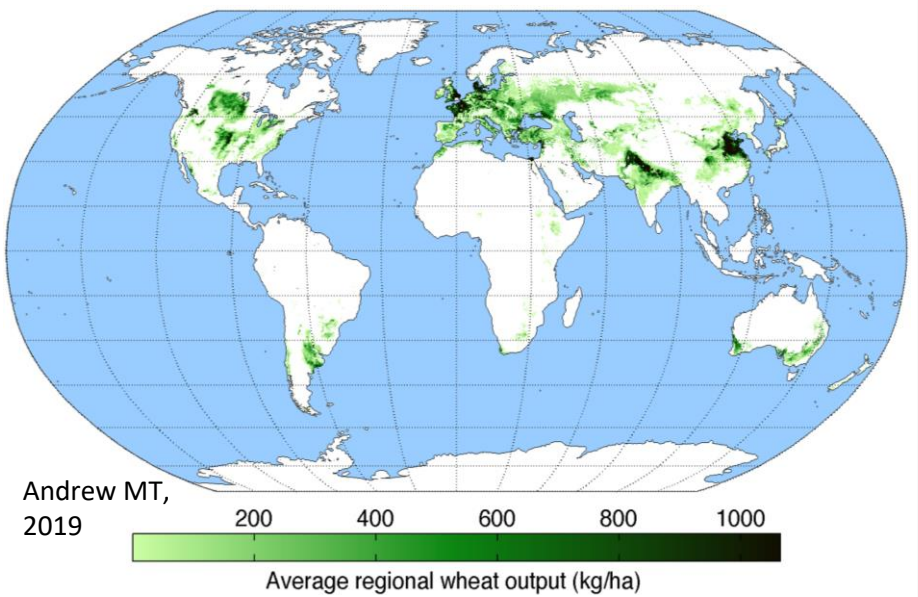
Dr Delphine Vincent

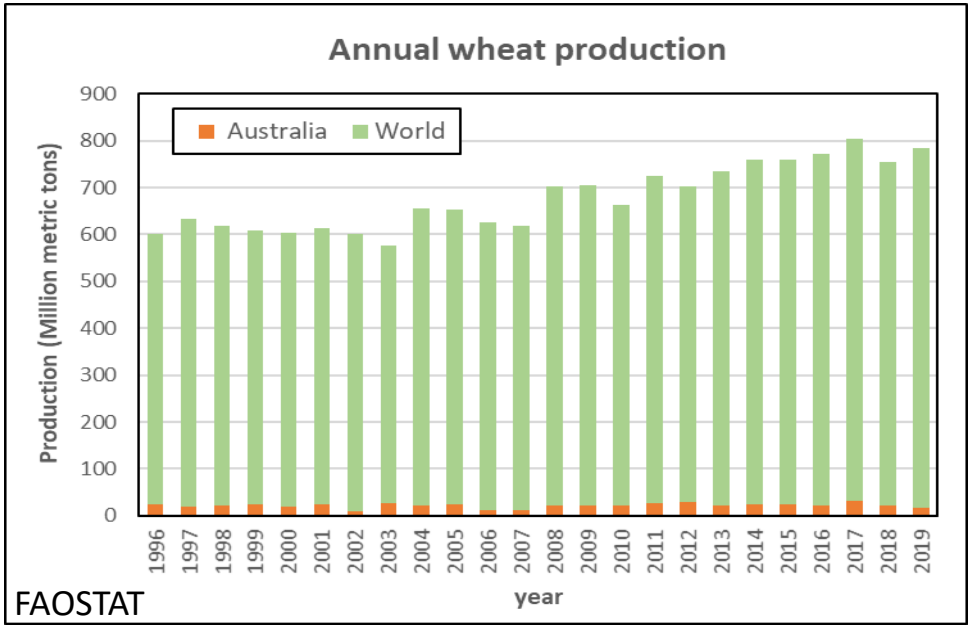
02/04/2022


Introduction

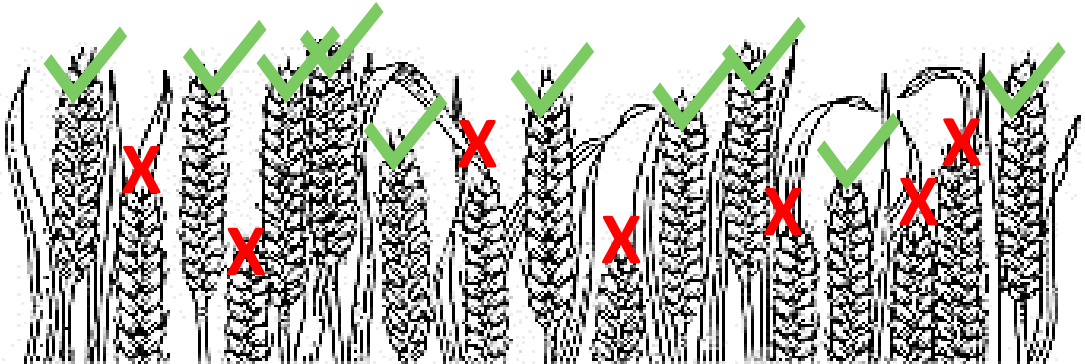
Context

Bread wheat (*Triticum aestivum*) is the most widely cultivated crop worldwide, used for human and animal food products. In 2019, Australia produced 18 million metric tons (2.3% of worldwide production).

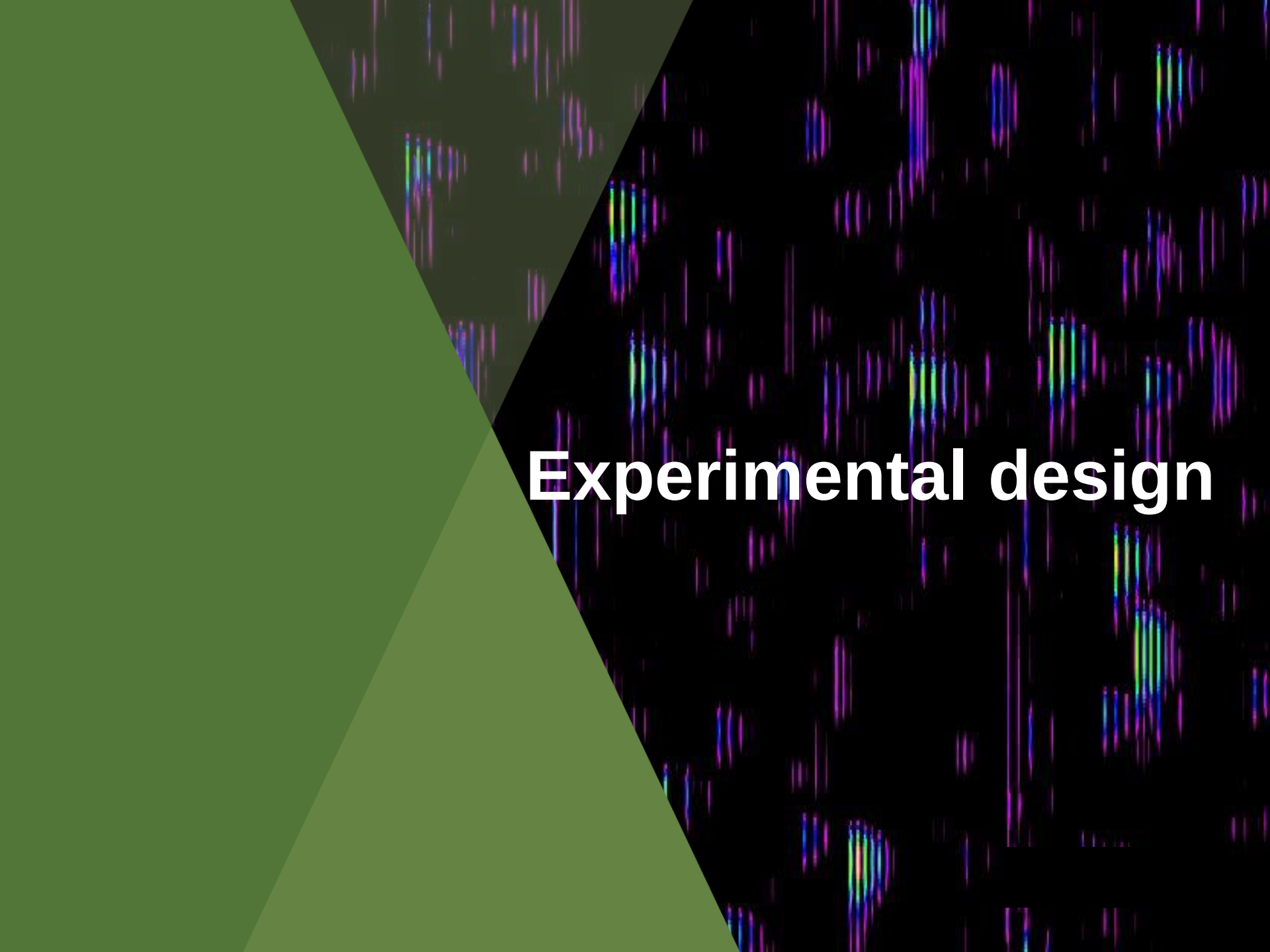




Selection tools applied early in the breeding cycle are needed to accelerate genetic gain for increased wheat production while maintaining or improving grain quality.
Proteomics screening assays of wheat flour can assist breeders to select the best performing breeding lines and discard the worst lines.



Aim: develop a robust shotgun proteomics method to screen thousands of wheat genotypes.


Experimental design

Experimental design

Grinding

Material	*Triticum aestivum* grains	½ tspn wheat grains, Genogrinder (1500 rpm, 2 x 60 sec)

Weighing*

Flour weight (mg)	10	20	30

Extraction*

Urea buffer	Gnd-HCl buffer
6M urea 10mM DTT 5.37 mM sodium citrate tribasic 2H2O 0.1 M Bis-Tris	6M Gdn-HCl 10mM DTT 5.37 mM sodium citrate tribasic 2H2O 0.1 M Bis-Tris

Digestion*

Protease name	Protease code	AA targeted	Terminus targeted	Selectivity	Cost
Glu-C	G	E, D	C-term	intermediate	2.5x
Chymotrypsin	C	F, W, Y	C-term	low	1.5x
Trypsin/Lys-C	TL	R, K	C-term	intermediate	1x

Desalting

Solid Phase Extraction	C18 gravity 96-well plate

LC*

Flow rate	0.1 ml/min	0.2 mL/min		
Solvent gradient	3-40B	3-11-40B	3-15-40B	6-36B
Run duration	60 min	45 min	38 min	43 min
Online desalting time	6 min	2.5 min		
LC column	BioZen XB-C18	Aeris XB-C18		

ESI-MS

ESI tune	MS1	MS2
positive polarity 3.9 kV 275 °C MS1 AGC target 10^5 MSn AGC target 5000	FTMS + p norm 15000 resolution m/z 300.0-2000.0	ITMS + c norm MS/MS of 10 most intense ion CID fragmentation mode signal intensity threshold 3000 isolation width 2 normalised collision energy 35 z +1 excluded

IT

Software	Xcalibur QualBrowser	Genedata Expressionist	Proteome Discoverer 1.4	Excel
Tools	LC-MS maps	PCAs	Venn diagrams	Charts
Data mining	UniprotKB Retrieve/ID mapping	AgBase-GOanna, AgriGO, SEA, REVIGO	BlastGUI, Pathway tools	KEGG

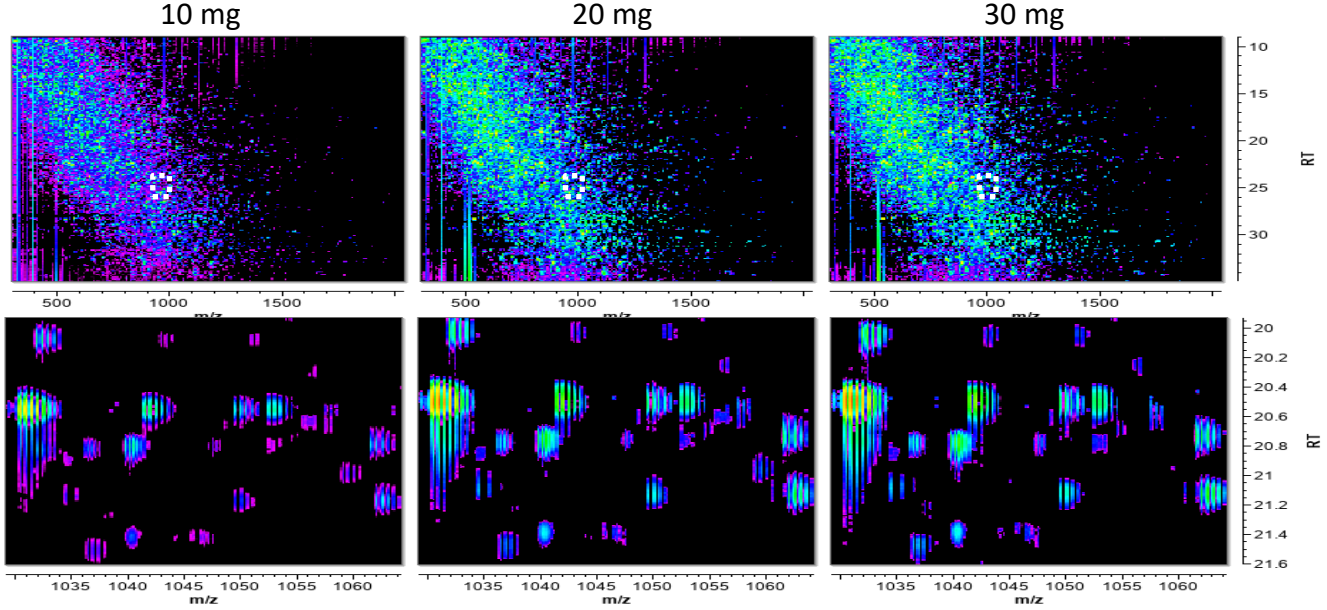
*steps where optimisation occurred

Wheat grain bags sent by Joe Panozzo

Testing set: 6 varieties
Validation set: 96 varieties
Total wheat lines: 4061


Results
Method development

Testing flour weight



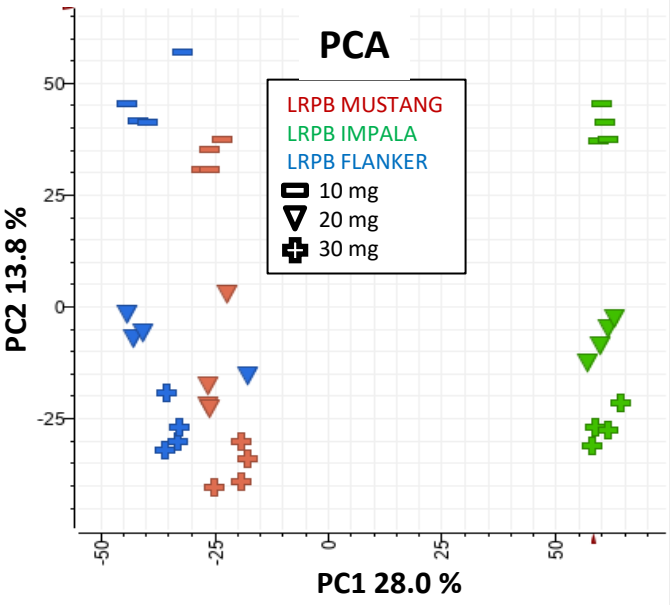
3 amounts

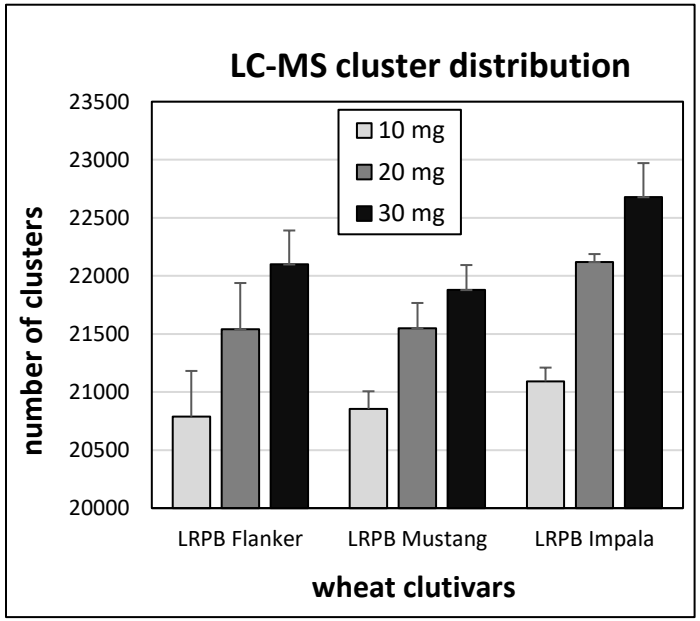
- 10 mg
- 20 mg
- 30 mg

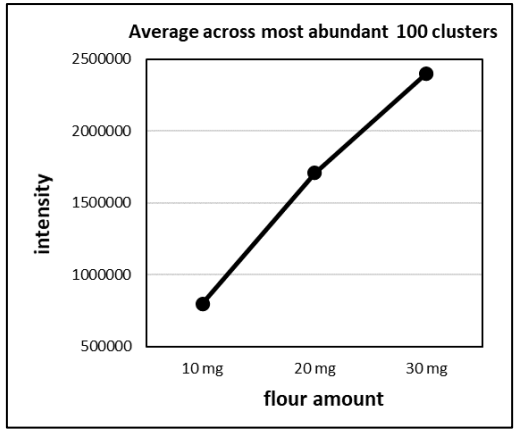
Conclusions

Good reproducibility.

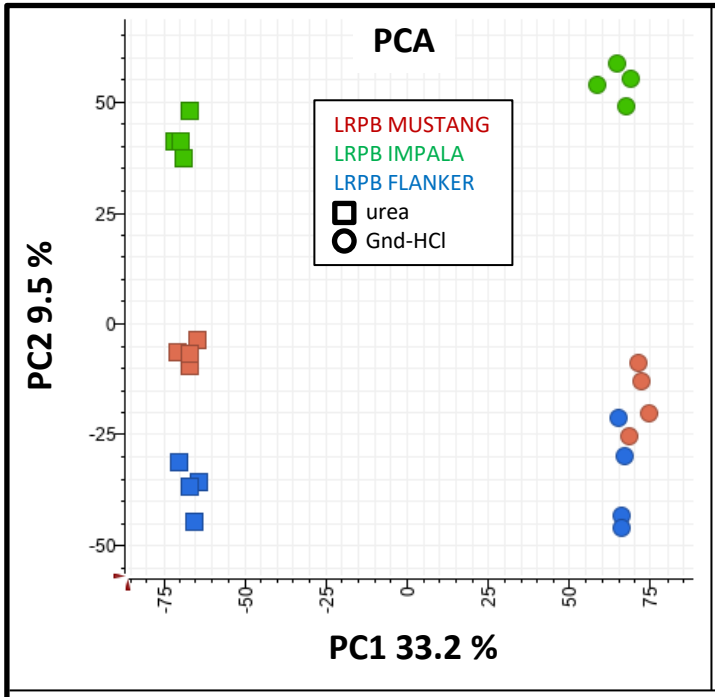
The number of LC-MS clusters was comparable between wheat cultivars and increased with the amount of flour but tapered off when 30 mg were used, indicating incomplete resuspension of the flour

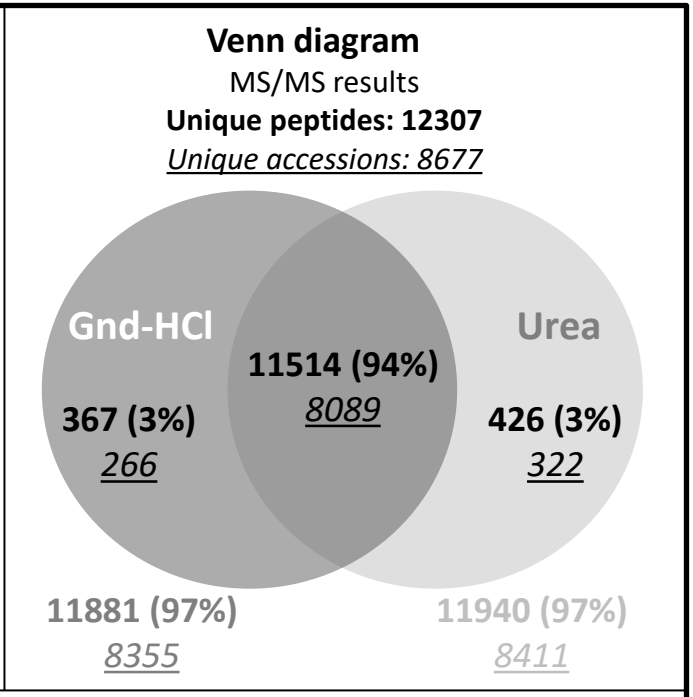


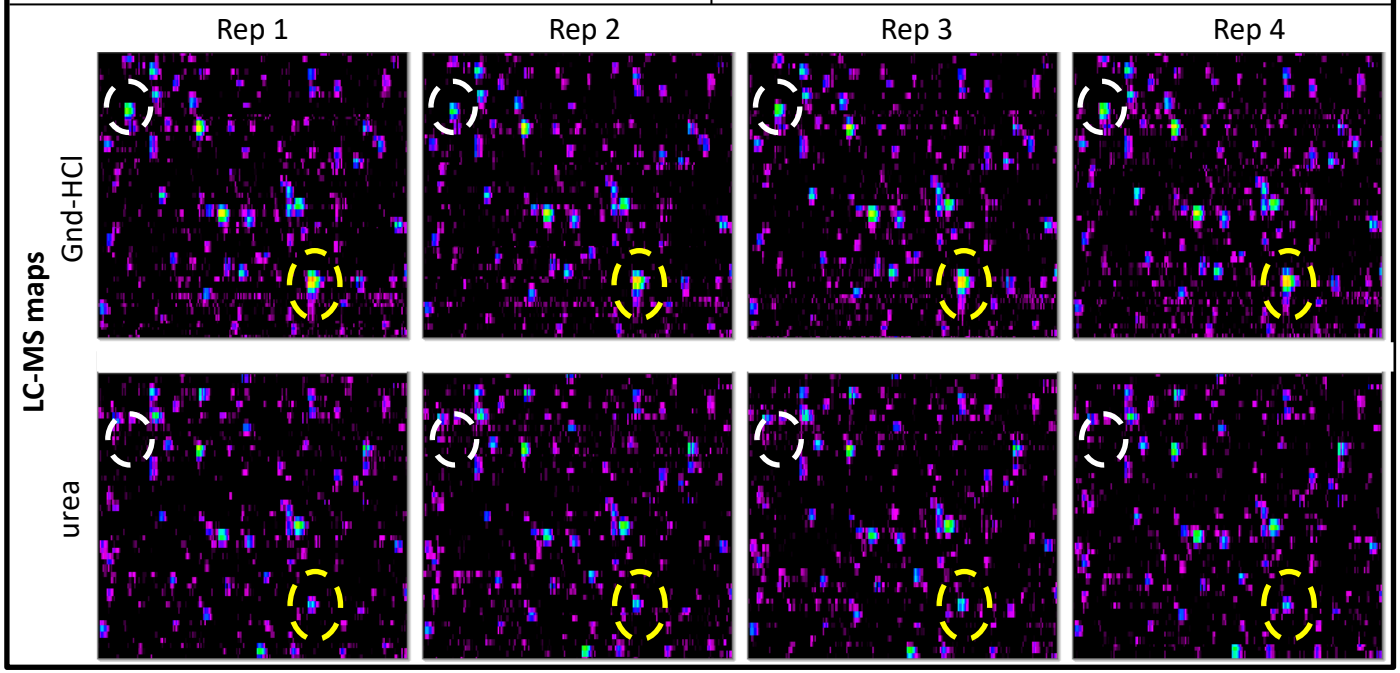




Testing extraction buffers







2 buffers

- urea
- Gnd-HCl

0.5 mL volume

Conclusions

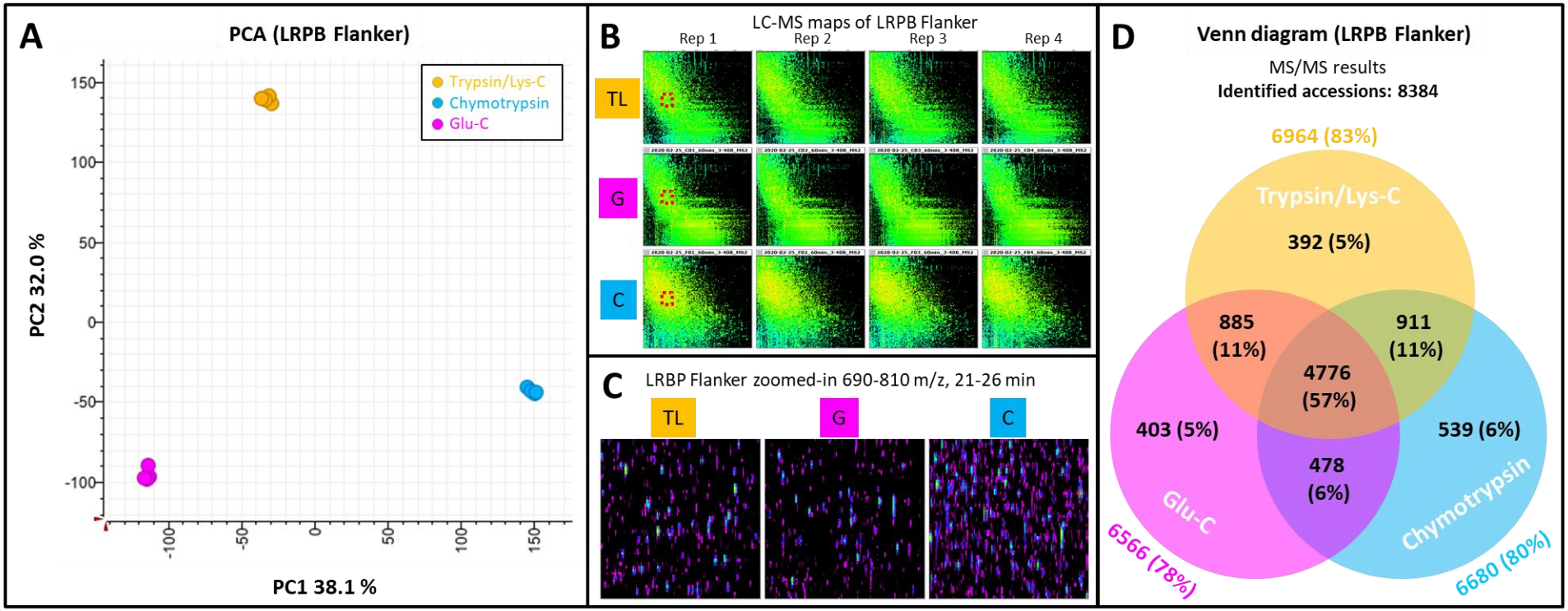
Both urea and Gnd-HCl buffers produced comparable and reproducible results.

We selected Gnd-HCl because it is cheaper which is an important factor for large scale experiments

Testing proteases

3 enzymatic digestions

- Trypsin/Lys-C (positively charged AAs: R,K)
- Chymotrypsin (hydrophobic AAs: Y, F, W)
- Glu-C (negatively charged AAs: E,D)



Conclusions

Good reproducibility.

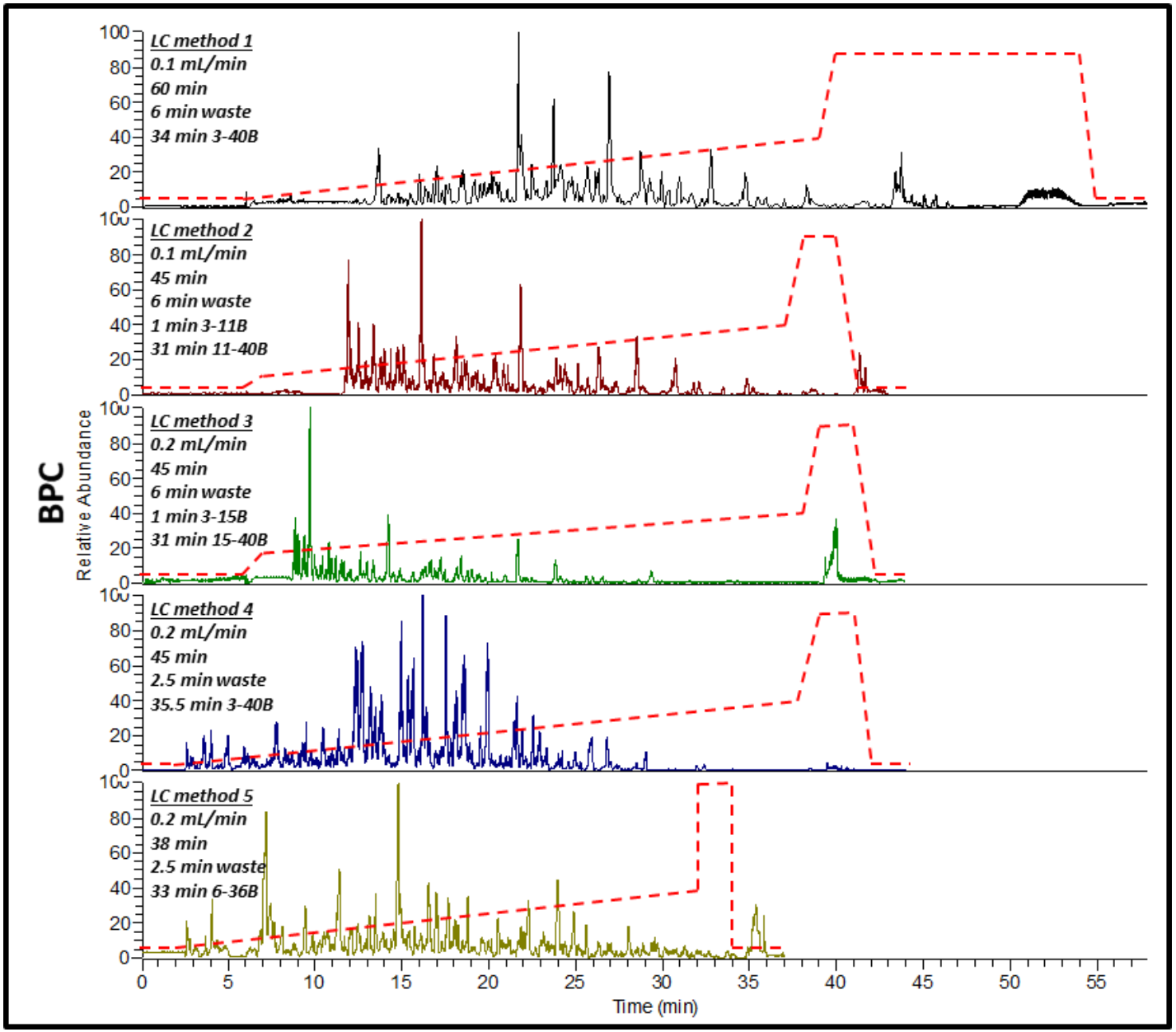
8,384 T. aestivum protein accessions identified.

targeting distinct AA residues via orthogonal proteases increases proteome coverage.

Too expensive on large scale experiment.

Chose trypsin/Lys-C is the cheaper protease.

Testing LC parameters



5 parameters

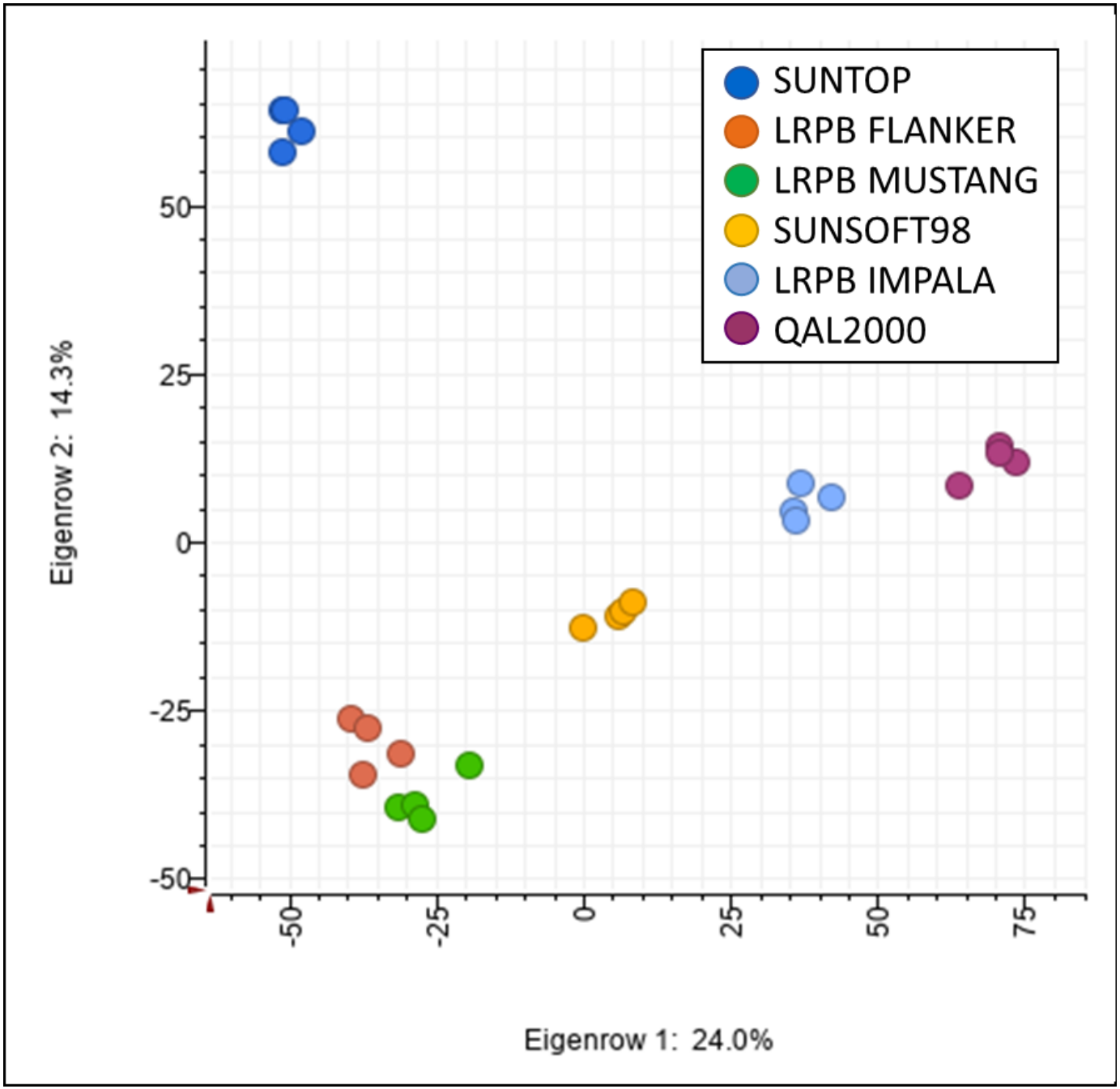
- total duration
- solvent gradient
- online desalting time
- flow rate
- separation columns

Conclusions

Chosen LC method applied 6-36% ACN gradient for 33 min and was 43 min long, including washing and equilibration steps.

→ 33 samples/day

Chosen method on 6 cultivars



Our method

- 20 mg
- Gnd-HCl
- Trypsin/Lys-C
- 3-36%B, 43 min

Conclusions

4 replicates clustered

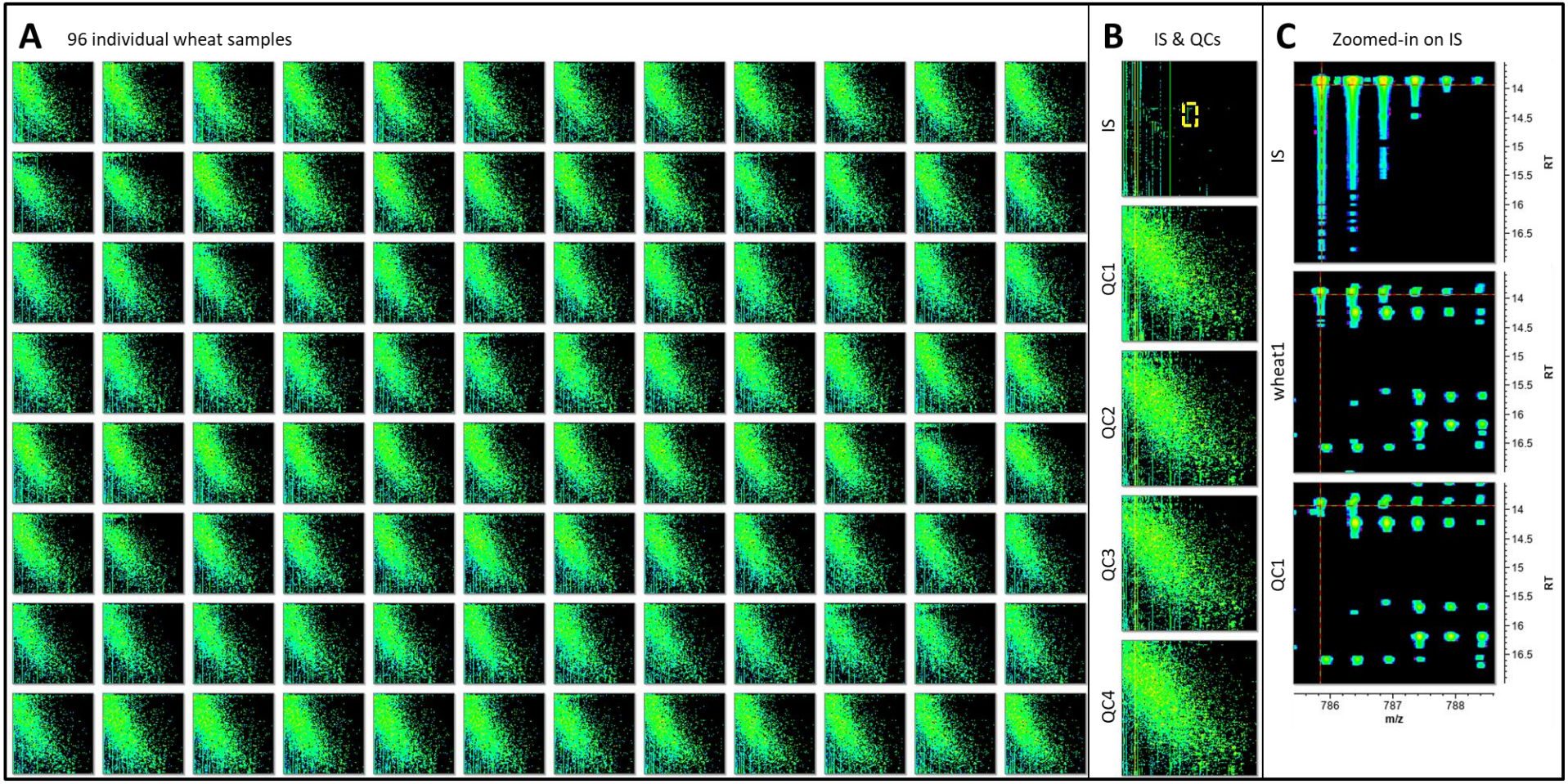
Separation of the 6 cultivars

Results
Method validation

Upscaling to 96 samples (1 plate)

Validation set

- 96 wheat samples
- Quality Control (QC)
- Internal Standard (IS)

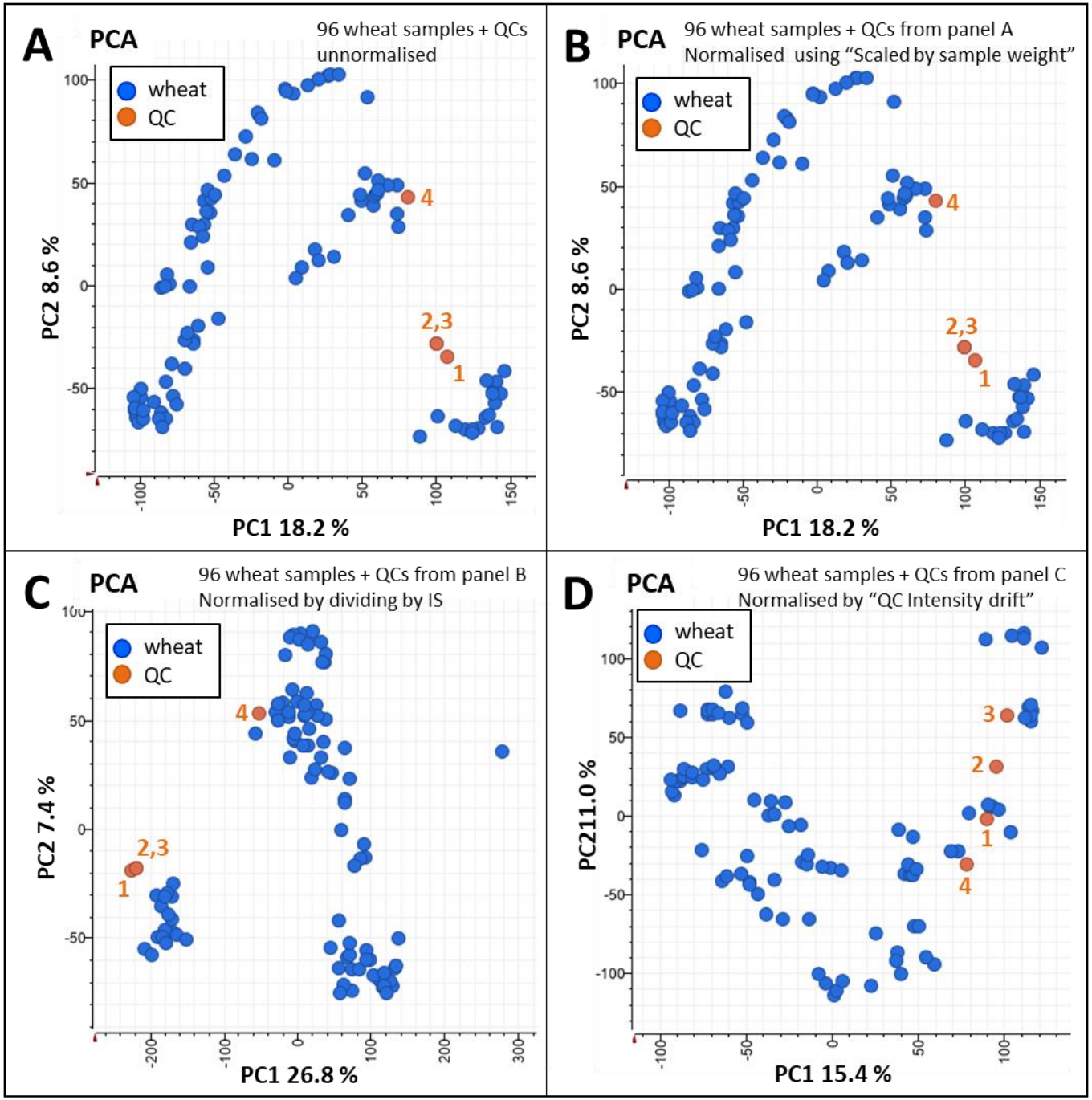


Conclusions

Good reproducibility.

QC and IS are used for normalisation purposes.

Normalisation



3 steps

- Sample weight
- IS
- QC

Conclusions

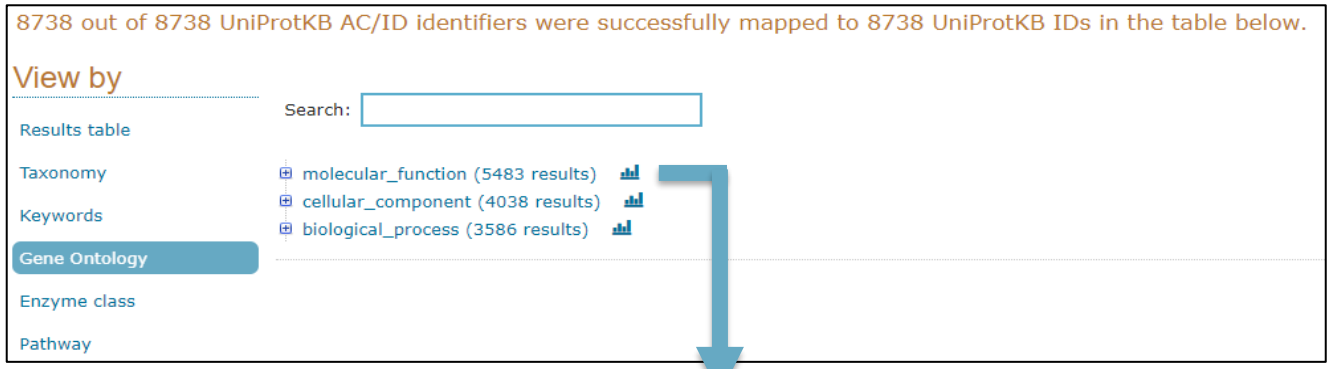
Weighing accuracy (1%) ensures reproducibility (no impact of normalising using weight).

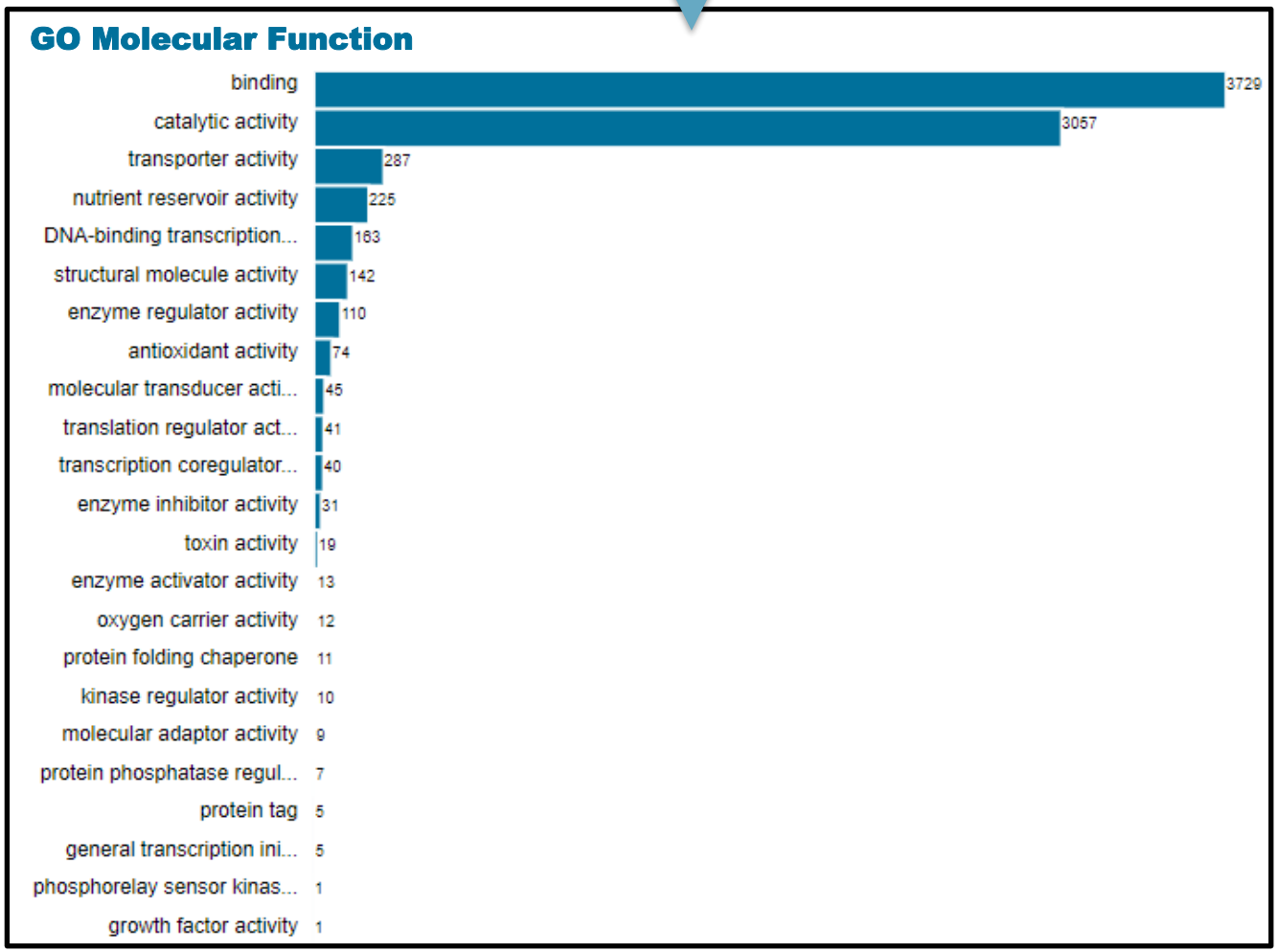
IS normalisation creates tighter groups.

QC normalisation eliminates the 2 groups thus minimising uncontrollable technical variations


Results
Data mining

UniprotKB





URL

https://www.uniprot.org/uploadlists/

Conclusions

- ✓ UniprotKB allows you to retrieve relevant information in a tabulated format that can be exported.
- ✓ It links to other protein DBs.
- ✓ It's quick and easy to use.
- ✗ The viewing tools are very rudimentary.

KEGG mapper

KEGG Mapper Reconstruction Result

Pathway (381) | Brite (45) | Brite Table (5) | Module (14)

Show matched objects

Metabolism

Global and overview maps
- 01100 Metabolic pathways (340)
- 01110 Biosynthesis of secondary metabolites (188)
- 01120 Microbial metabolism in diverse environments (66)
- 01200 Carbon metabolism (43)
- 01210 2-Oxocarboxylic acid metabolism (10)
- 01212 Fatty acid metabolism (11)
- 01230 Biosynthesis of amino acids (37)
- 01240 Biosynthesis of cofactors (42)
- 01220 Degradation of aromatic compounds (3)

Carbohydrate metabolism
- 00010 Glycolysis / Gluconeogenesis (26)
- 00020 Citrate cycle (TCA cycle) (9)
- 00030 Pentose phosphate pathway (10)
- 00040 Pentose and glucuronate interconversions (9)
- 00051 Fructose and mannose metabolism (10)
- 00052 Galactose metabolism (11)
- 00053 Ascorbate and aldarate metabolism (11)
- 00500 Starch and sucrose metabolism (23)
- 00520 Amino sugar and nucleotide sugar metabolism (19)
- 00620 Pyruvate metabolism (20)
- 00630 Glyoxylate and dicarboxylate metabolism (12)
- 00640 Propanoate metabolism (7)
- 00650 Butanoate metabolism (4)
- 00660 C5-Branched dibasic acid metabolism (1)
- 00562 Inositol phosphate metabolism (16)

Energy metabolism
- 00190 Oxidative phosphorylation (24)
- 00195 Photosynthesis (8)
- 00196 Photosynthesis - antenna proteins (3)
- 00710 Carbon fixation in photosynthetic organisms (13)
- 00720 Carbon fixation pathways in prokaryotes (7)
- 00680 Methane metabolism (10)
- 00910 Nitrogen metabolism (9)
- 00920 Sulfur metabolism (4)

Lipid metabolism
- 00061 Fatty acid biosynthesis (5)
- 00062 Fatty acid elongation (2)
- 00071 Fatty acid degradation (11)
- 00073 Cutin, suberine and wax biosynthesis (4)
- 00100 Steroid biosynthesis (7)
- 00140 Steroid hormone biosynthesis (1)
- 00561 Glycerolipid metabolism (15)
- 00564 Glycerophospholipid metabolism (15)
- 00565 Ether lipid metabolism (4)
- 00600 Sphingolipid metabolism (5)
- 00590 Arachidonic acid metabolism (3)
- 00591 Linoleic acid metabolism (3)
- 00592 alpha-Linolenic acid metabolism (9)
- 01040 Biosynthesis of unsaturated fatty acids (4)

Nucleotide metabolism
- 00230 Purine metabolism (19)
- 00240 Pyrimidine metabolism (7)

Amino acid metabolism
- 00250 Alanine, aspartate and glutamate metabolism (13)
- 00260 Glycine, serine and threonine metabolism (9)
- 00270 Cysteine and methionine metabolism (18)
- 00280 Valine, leucine and isoleucine degradation (6)
- 00290 Valine, leucine and isoleucine biosynthesis (2)
- 00300 Lysine biosynthesis (1)
- 00310 Lysine degradation (11)
- 00220 Arginine biosynthesis (10)
- 00330 Arginine and proline metabolism (8)
- 00340 Histidine metabolism (5)
- 00350 Tyrosine metabolism (7)
- 00360 Phenylalanine metabolism (7)
- 00380 Tryptophan metabolism (13)
- 00400 Phenylalanine, tyrosine and tryptophan biosynthesis (8)

Metabolism of other amino acids
- 00410 beta-Alanine metabolism (7)
- 00430 Taurine and hypotaurine metabolism (2)
- 00440 Phosphonate and phosphinate metabolism (1)
- 00450 Selenocompound metabolism (4)
- 00460 Cyanoamino acid metabolism (8)
- 00471 D-Glutamine and D-glutamate metabolism (1)
- 00480 Glutathione metabolism (14)

Glycan biosynthesis and metabolism
- 00510 N-Glycan biosynthesis (8)
- 00513 Various types of N-glycan biosynthesis (7)
- 00514 Other types of O-glycan biosynthesis (1)
- 00531 Glycosaminoglycan degradation (3)
- 00563 Glycosylphosphatidylinositol (GPI)-anchor biosynthesis (5)
- 00603 Glycosphingolipid biosynthesis - globo and isoglobo series (2)
- 00604 Glycosphingolipid biosynthesis - ganglio series (1)
- 00540 Lipopolysaccharide biosynthesis (1)
- 00541 O-Antigen nucleotide sugar biosynthesis (4)
- 00511 Other glycan degradation (6)

Metabolism of cofactors and vitamins
- 00730 Thiamine metabolism (3)
- 00740 Riboflavin metabolism (4)
- 00750 Vitamin B6 metabolism (4)
- 00760 Nicotinate and nicotinamide metabolism (4)
- 00770 Pantothenate and CoA biosynthesis (5)
- 00780 Biotin metabolism (2)
- 00785 Lipoic acid metabolism (1)
- 00790 Folate biosynthesis (4)
- 00670 One carbon pool by folate (2)
- 00830 Retinol metabolism (2)
- 00860 Porphyrin and chlorophyll metabolism (6)
- 00130 Ubiquinone and other terpenoid-quinone biosynthesis (7)

Metabolism of terpenoids and polyketides
- 00900 Terpenoid backbone biosynthesis (7)
- 00902 Monoterpenoid biosynthesis (2)
- 00909 Sesquiterpenoid and triterpenoid biosynthesis (3)
- 00904 Diterpenoid biosynthesis (8)
- 00906 Carotenoid biosynthesis (5)
- 00905 Brassinosteroid biosynthesis (3)
- 00981 Insect hormone biosynthesis (1)
- 00908 Zeatin biosynthesis (5)
- 00903 Limonene and pinene degradation (2)
- 00281 Geraniol degradation (1)
- 01051 Biosynthesis of ansamycins (1)

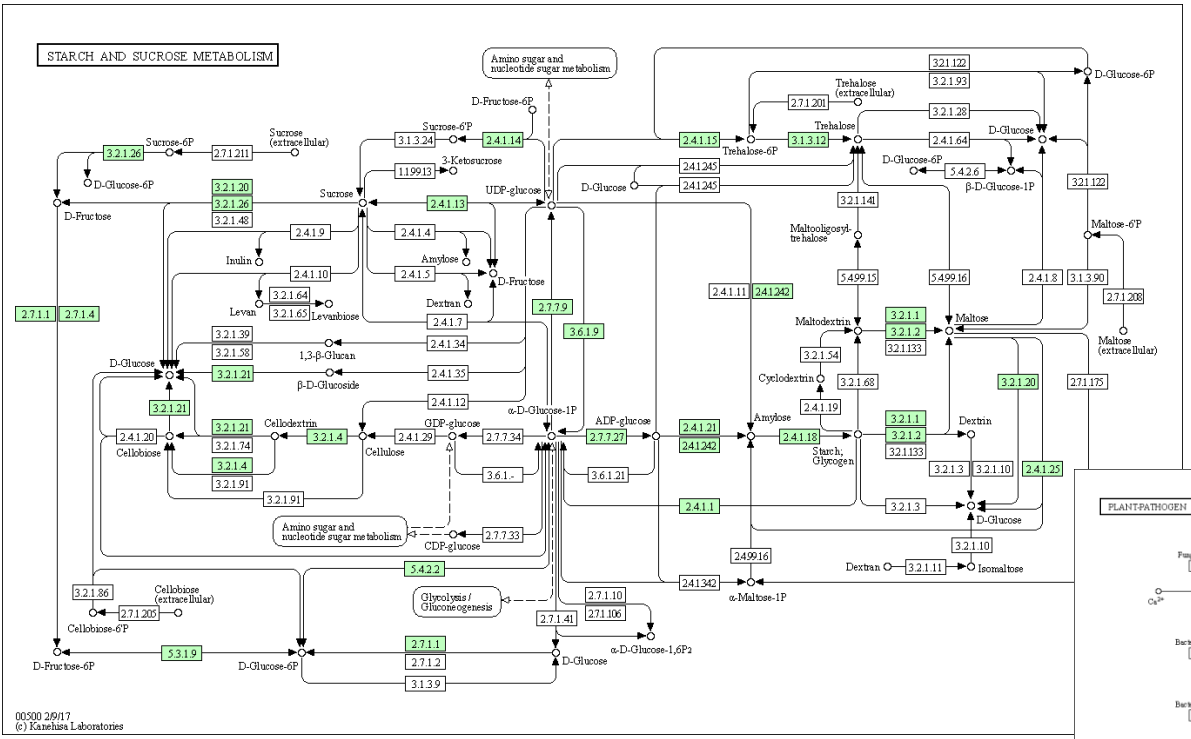
URL

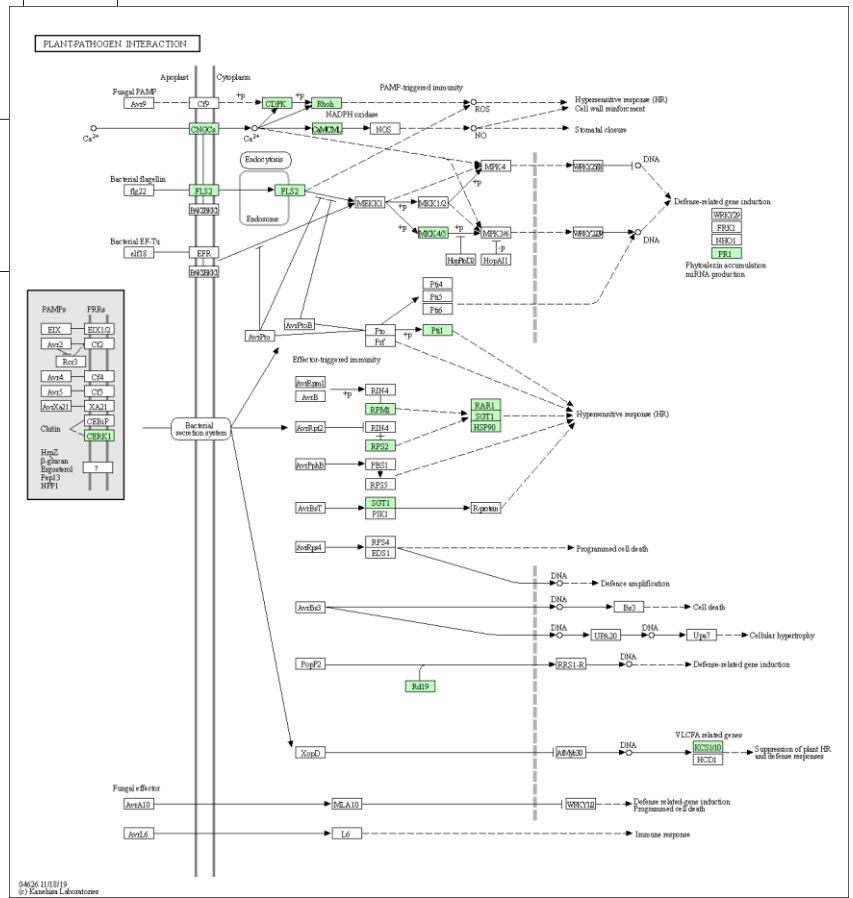
https://www.genome.jp/kegg/mapper/reconstruct.html

Conclusions

- ✓ Great mapping tools for metabolisms and pathways.
- ✓ Perfect for enzymes.
- ✗ Not so good for proteins that are not enzymes.
- ✗ Need to convert accessions into KEGG orthologs (KO).

KEGG mapper





In green, enzymes identified in our study, displayed with their E.C. numbers.

AgriGO v2

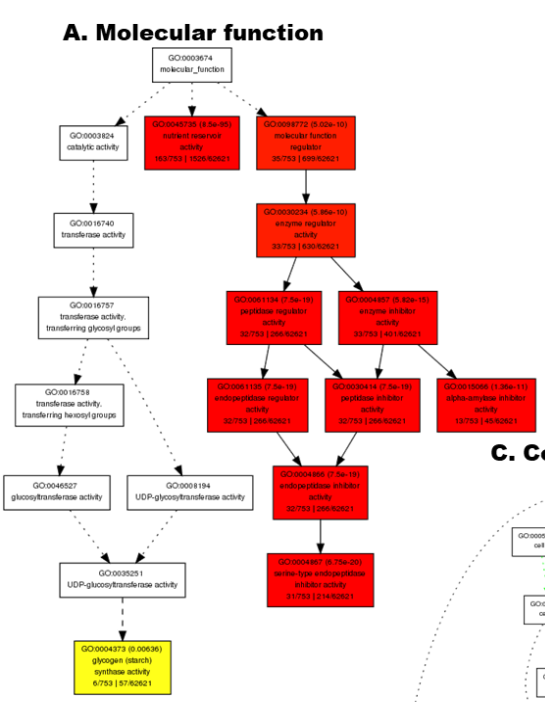
URL

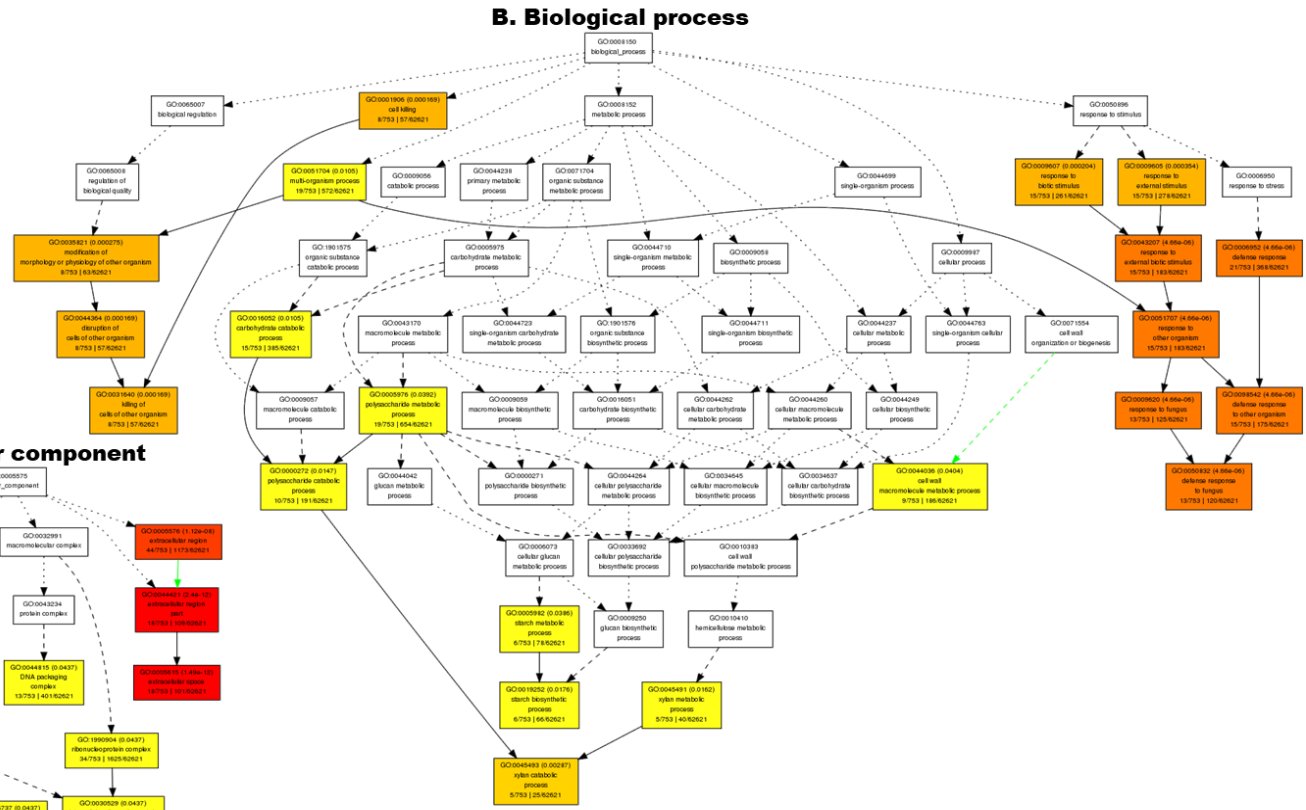
http://systemsbiology.cau.edu.cn/agriGOv2/

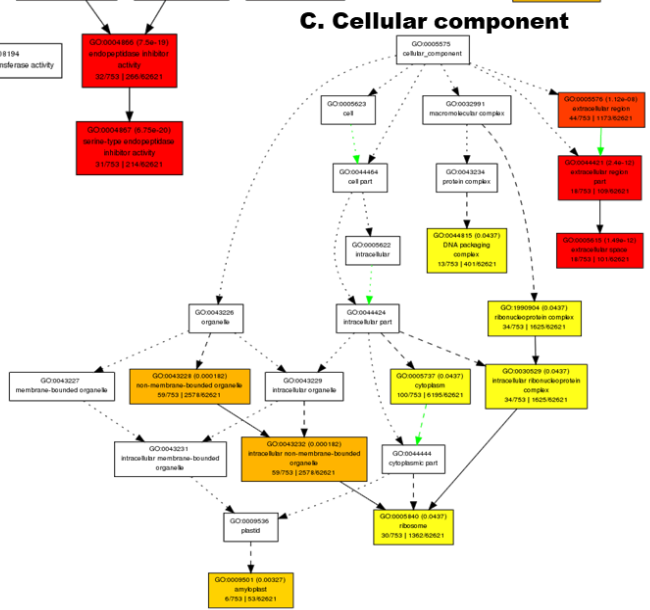
Singular Enrichment Analysis (SEA)

Conclusions

- ✓ Accepts Uniprot accessions and retrieves gene ontology (GO) terms.
- ✓ Good display tools (hieratical graph).
- ✓ SEA highlights relevant pathways (e.g. defense response), exportable.
- ✗ Gets very busy when lots of GO terms are uploaded.







Yellow-red gradient indicates p-value significance.

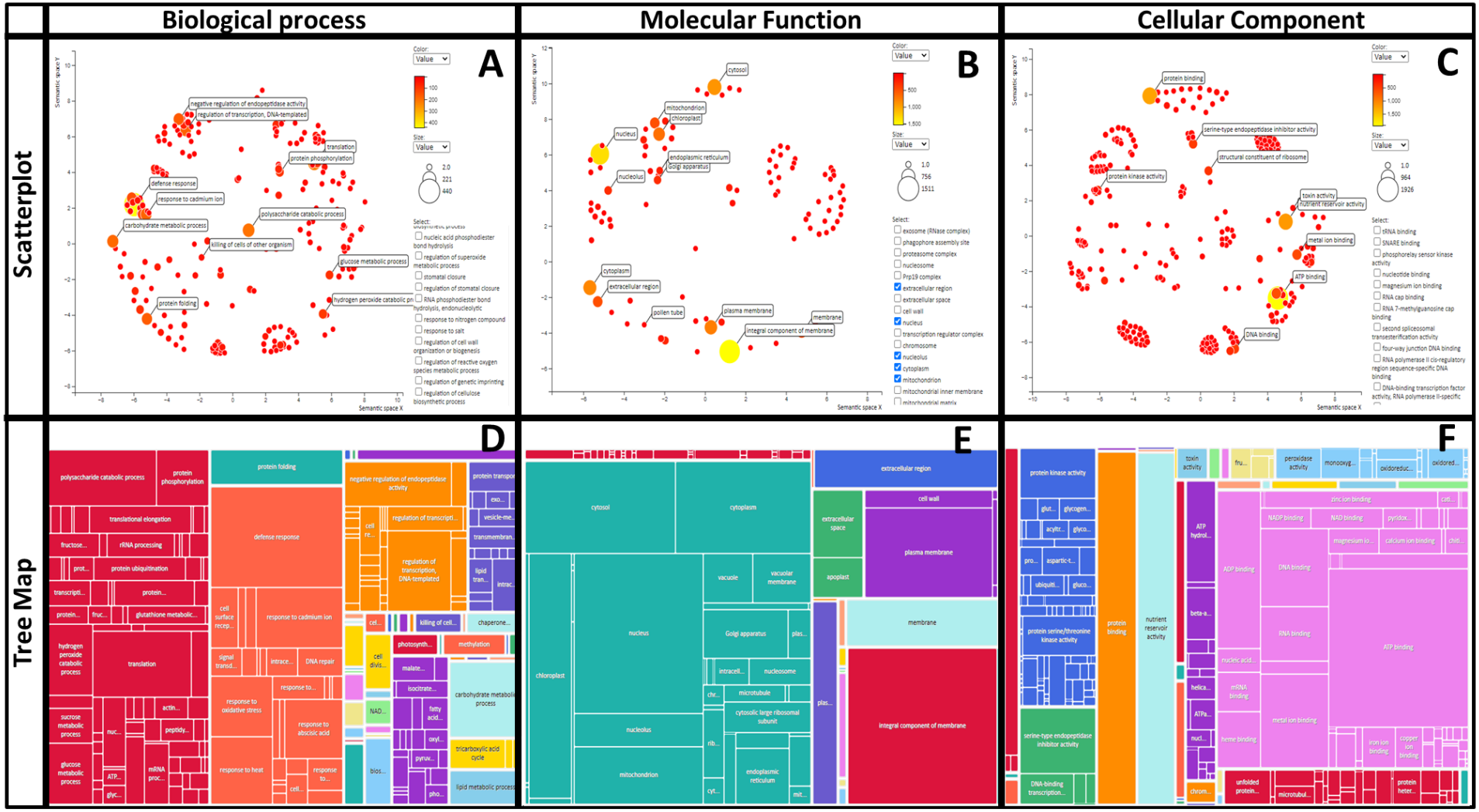
REVIGO

URL

http://revigo.irb.hr/
Reduction tool

Conclusions

- ✓ Uploads GO terms from AgriGO (or other sources), applies enrichment analysis.
- ✓ Some flexibility in the display tools; results are exportable (R scripts provided).
- ✓ Different viewing tools (scatterplots and tree maps).
- ✗ Can also get very busy and barely legible.



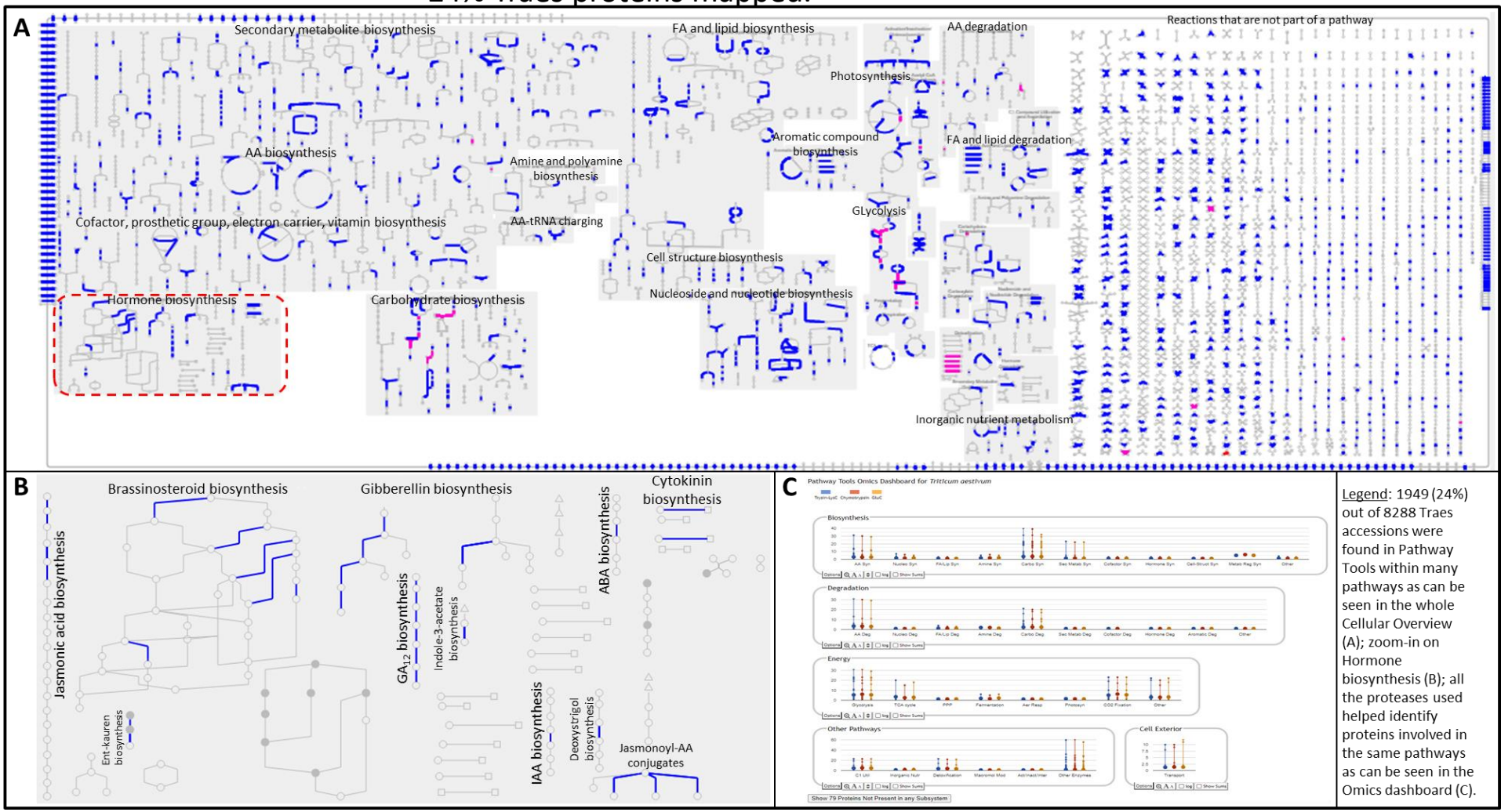
Pathway Tools

URL

https://pmn.plantcyc.org/organism-summary?object=BREADWHEAT

Conclusions

- ✓ PT is the most comprehensible biological DB (Karp *et al.*, 2016).
- ✓ Best visualisation tools, also displaying quantitative data.
- ✓ Can reorganise whole pathways and their nodes for figure export purpose.
- ✗ Doesn't recognise Uniprot accessions, must use "Traes" accessions; only 24% Traes proteins mapped.



Legend: 1949 (24%) out of 8288 Traes accessions were found in Pathway Tools within many pathways as can be seen in the whole Cellular Overview (A); zoom-in on Hormone biosynthesis (B); all the proteases used helped identify proteins involved in the same pathways as can be seen in the Omics dashboard (C).

Conclusions

Take-home messages

Conclusions

- ✓ We have devised a high-throughput proteomics method suitable for screening lots of wheat flour samples.
- ✓ 20 mg could be fully resuspended in a 0.5 mL volume of buffer.
- ✓ Urea and Gnd-HCl buffers yielded similar results.
- ✓ Using 3 sets of orthogonal proteases helped dig down deeper into the wheat proteome.
- ✓ The LC method we selected applied a 6-36% ACN gradient for 33 min, total duration of 43 min.
- ✓ 8,738 *T. aestivum* proteins were identified.
- ✓ Essential aspects of the workflow were the accurate weighing of the flour, and the inclusion of IS and QCs to ensure reproducibility and robustness of the method over time.
- ✓ Many data mining tools are available online; the ones we presented (KEGG, UniprotKB, AgriGO, REVIGO, and Pathway Tools) allowed for a rapid and powerful exploration of the data under different angles, thus not only confirming the presence of the expected storage proteins and associated enzymes (starch and sucrose) but also highlighting novel results (pathogen response).

Article published (Vincent, D.; et al. Mining the Wheat Grain Proteome. Int. J. Mol. Sci. 2022, 23, 713. https://doi.org/10.3390/ijms23020713)

Next step

- ✓ The 4061 lines have been screened.
- ✓ Manuscript in preparation.


Thank you