RuminOmics Technologies and Methane

a mixture of metagenomics and metaproteomics

‘Meta-Omics’

- Characterisation of the function and composition of a microbial community
  - Environmental sample (Ruminal Digesta)

- CO_2 and Hydrogen producers
  - Bacteria
  - Ciliates

- Reduction of CO_2 to CH_4 (Hydrogenotrophic)
  - Methanogenic Archaea

- Metagenomics (Genes)
  - Illumina Hi Seq

- Metaproteomics (Peptides/Proteins)
  - 2D SDS PAGE
  - Shotgun metaproteomics

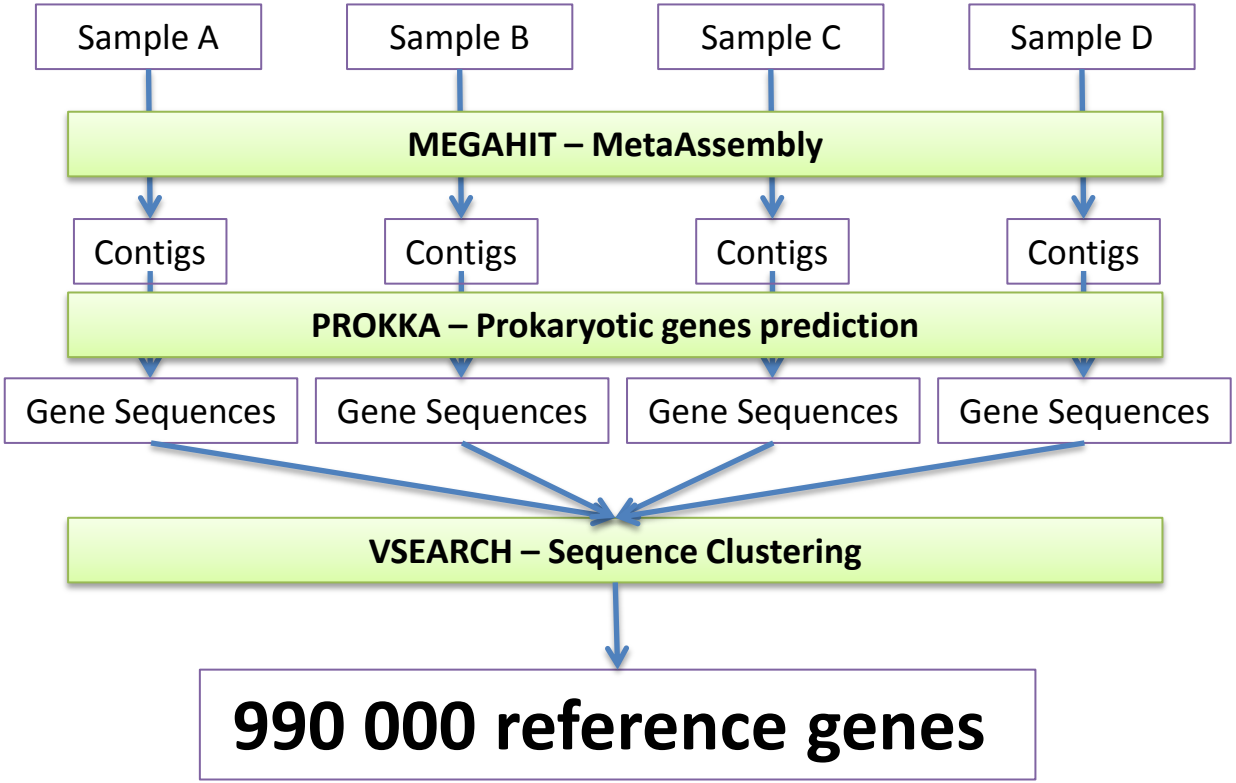
RuminOmics 1000

- Sampling
  - ‘On-farm’ via gastric tube
- Sample storage
  - Glycerol cryopreservation (McKain et al., 2013)
- DNA Extraction
  - Repeated bead beating +column filtration (Yu and Morrison, 2004)
- Selection of ‘High’ and ‘Low’ ranked CH_4 emitters
  - CH_4 = PMR(kg DMI) + BW

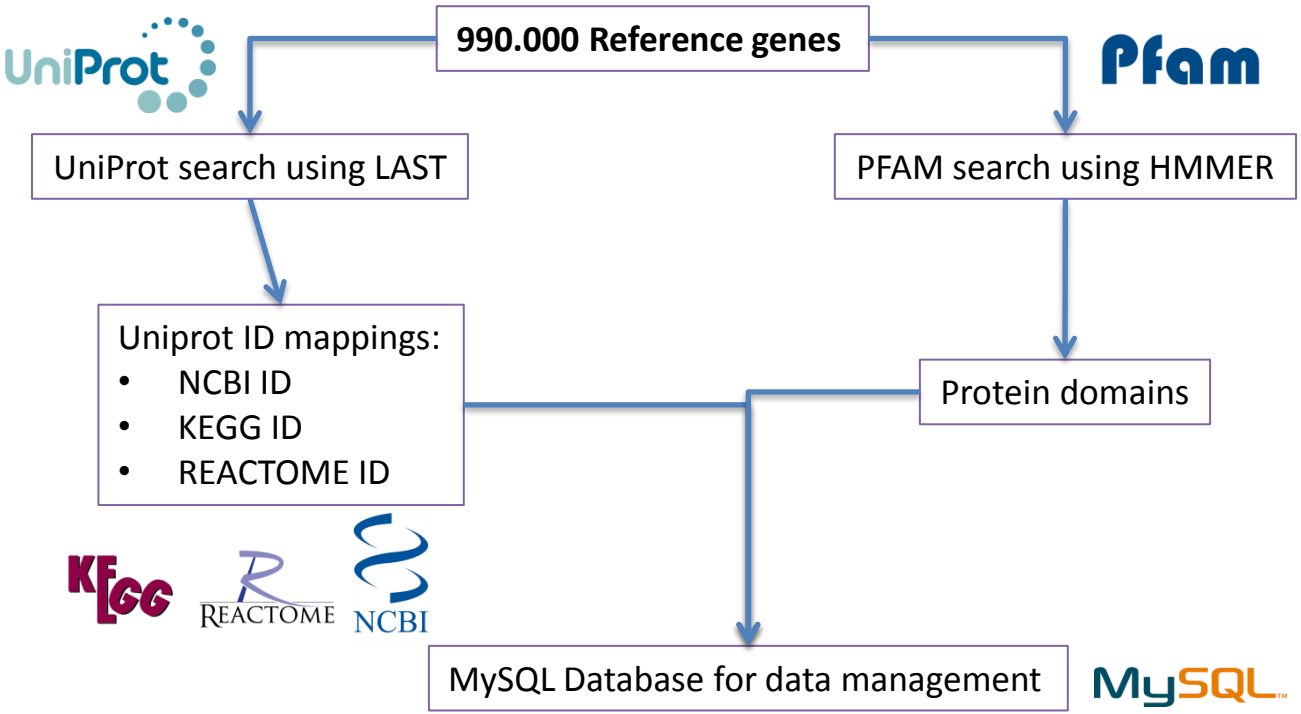
  - 15 samples from **Finland**
  - 14 samples from **Italy**
  - 15 samples from **Sweden**

- Illumina HiSeq2000
  - 100bp Paired End Reads
  - Sequence data ranging from 10Gb to 20Gb per sample

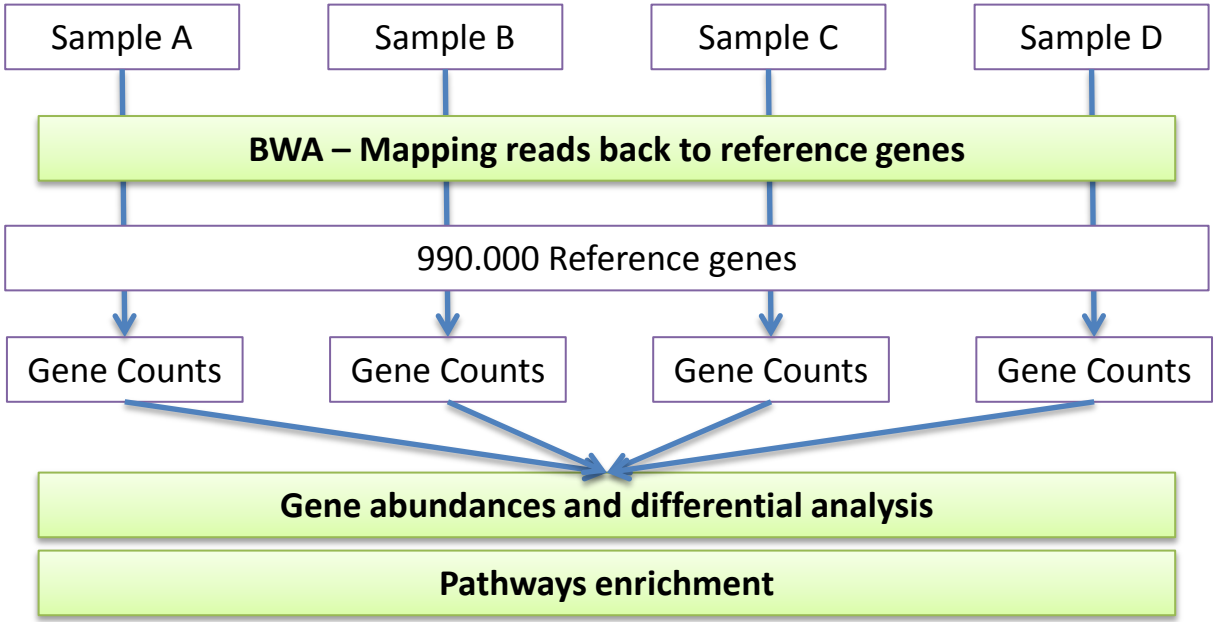
Finding genes



Annotation

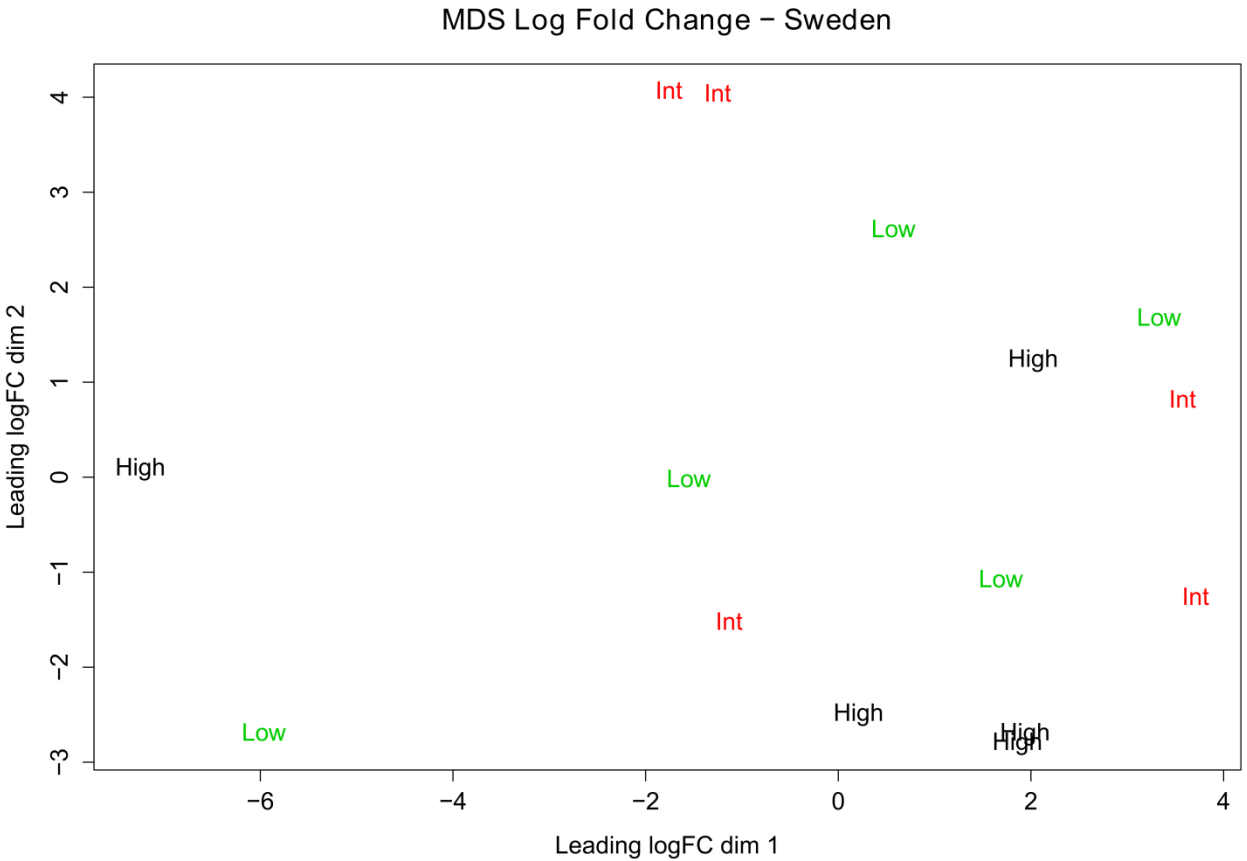

990.000 Reference genes
UniProt
Pfam
UniProt search using LAST
PFAM search using HMMER
Uniprot ID mappings:
• NCBI ID
• KEGG ID
• REACTOME ID
Protein domains
KEGG
REACTOME
NCBI
MySQL Database for data management
MySQL

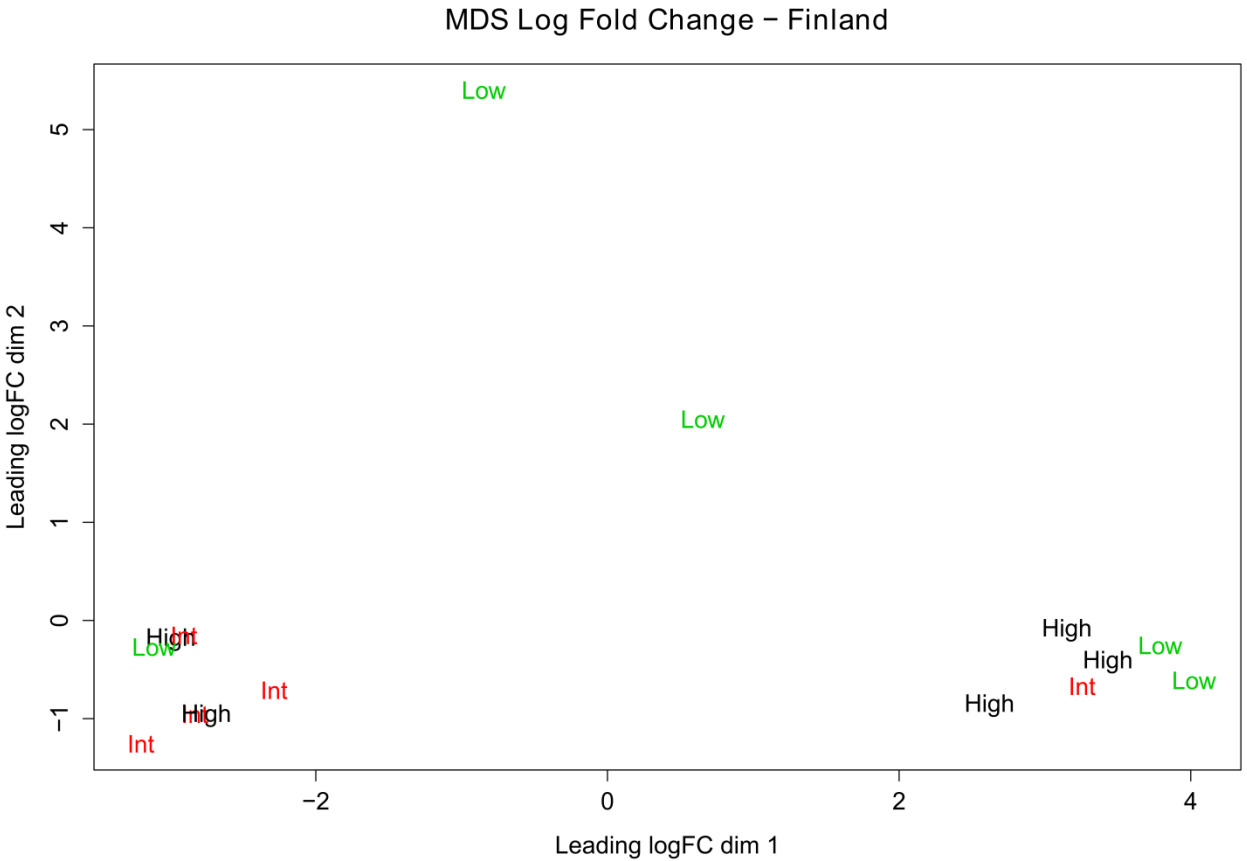
Analysis

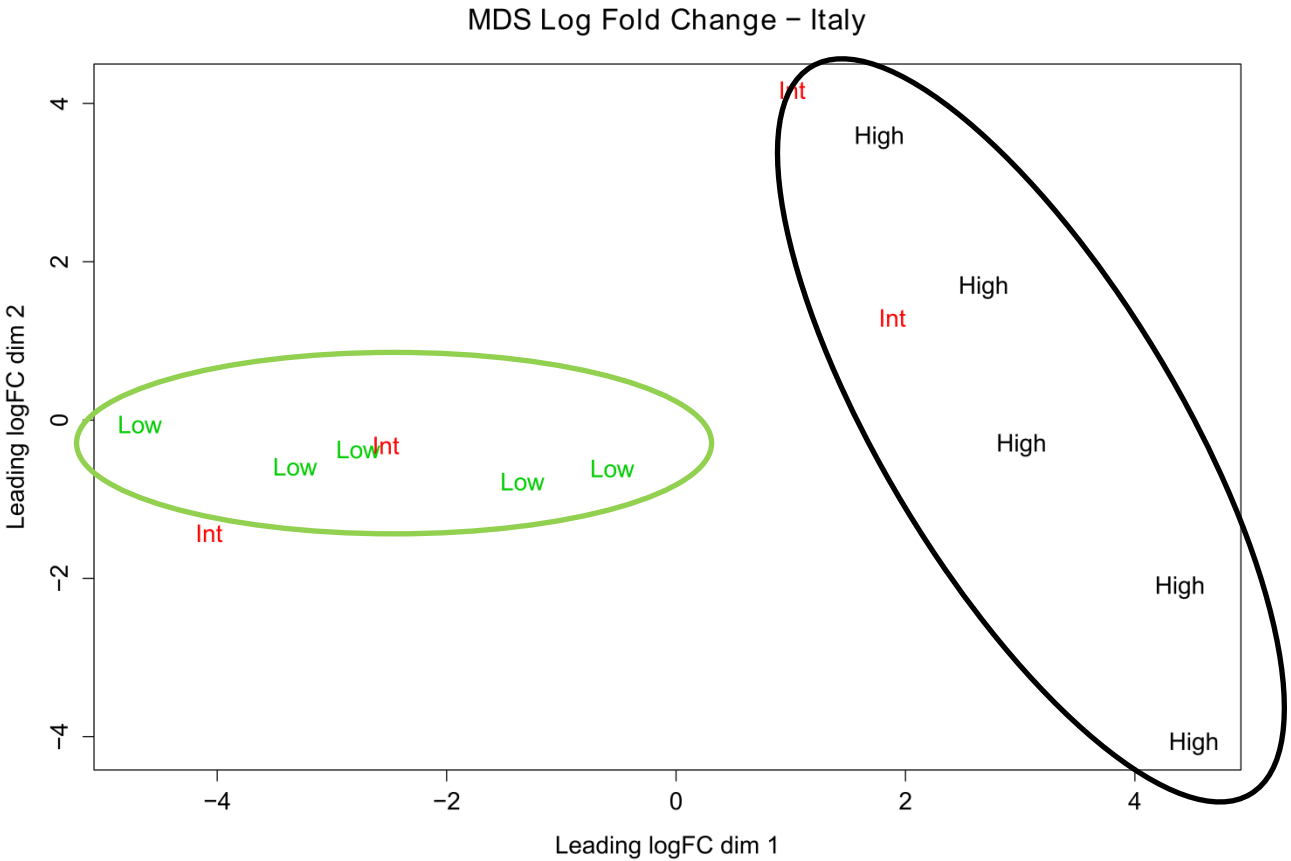

Sample A
Sample B
Sample C
Sample D
BWA – Mapping reads back to reference genes
990.000 Reference genes
Gene Counts
Gene Counts
Gene Counts
Gene Counts
Gene abundances and differential analysis
Pathways enrichment

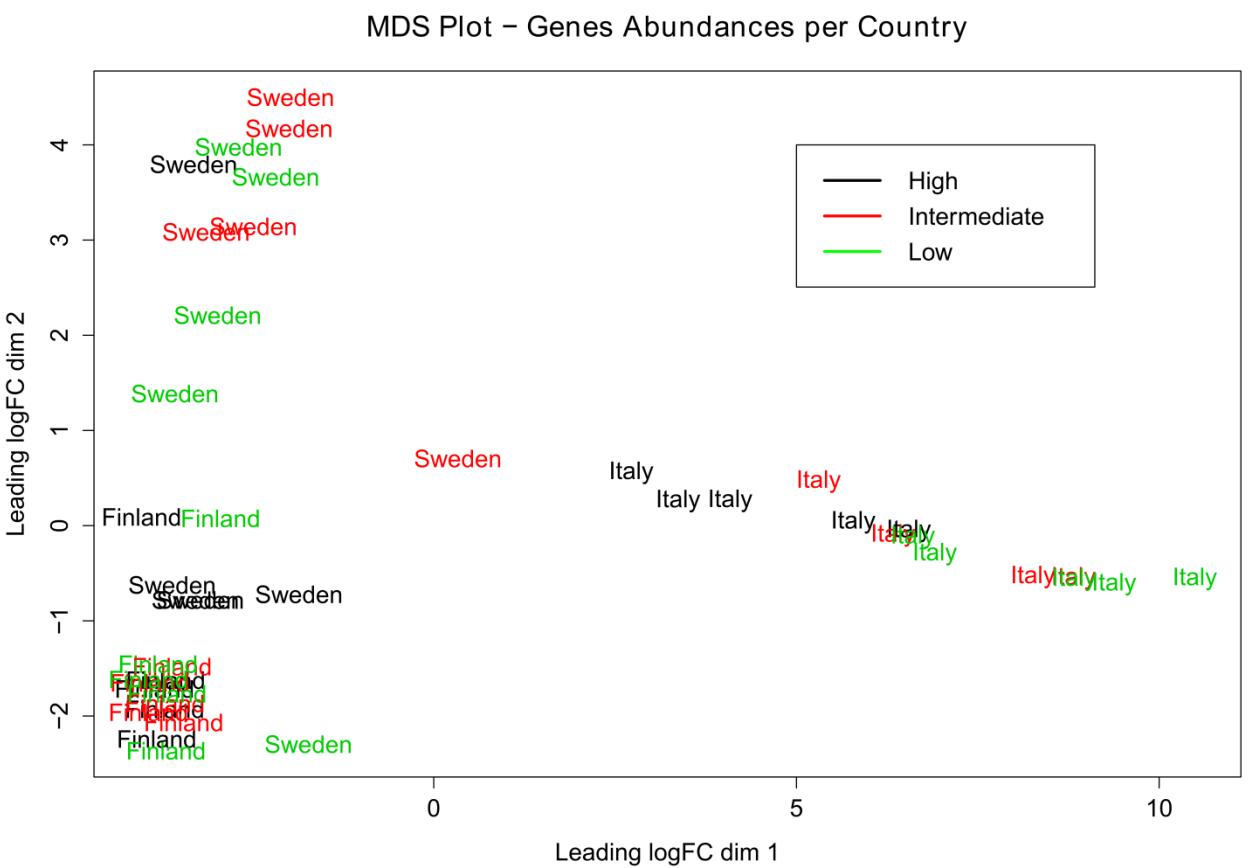
Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

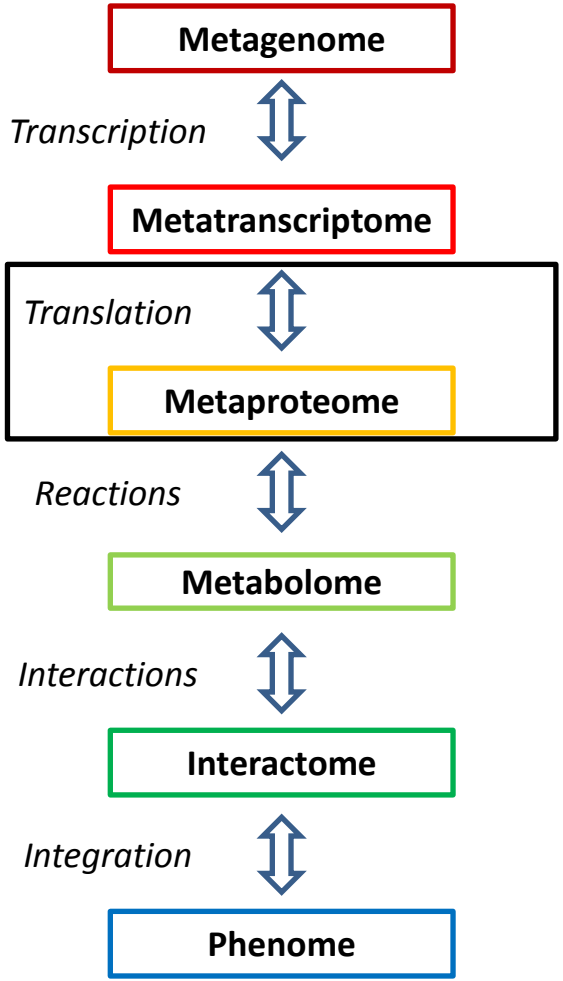
- **Issues**
  - Data processing
    - Large amount of data per sample
    - Large number of samples
  - Annotation
    - 40% of predicted proteins still remain unknown
    - Functional and pathway analysis strongly depend on reliable sequence annotations and information in reference databases
- **Solutions**
  - Continuing development of new assembly and sequence clustering software
  - Cataloguing the genomes of rumen microbiota (Hungate 1000)


MDS Log Fold Change − Sweden
Int
Int
Low
Low
High
Int
High
Low
Low
Int
Int
High
Low
High
High
Leading logFC dim 2
Leading logFC dim 1
4
3
2
1
0
−1
−2
−3
−6
−4
−2
0
2
4


MDS Log Fold Change − Finland
Leading logFC dim 2
Leading logFC dim 1
5
4
3
2
1
0
−1
−2
0
2
4
Low
Low
High
Low
Int
Int
High
Int
High
Low
High
Int
Low
High
Int


MDS Log Fold Change − Italy
Leading logFC dim 2
Leading logFC dim 1
4
2
0
−2
−4
−4
−2
0
2
4
Int
High
High
Int
High
High
Low
Low
Low
Int
Low
Low
Int
High
High


MDS Plot − Genes Abundances per Country
Leading logFC dim 2
Leading logFC dim 1
High
Intermediate
Low
Sweden
Finland
Italy
4
3
2
1
0
−1
−2
0
5
10

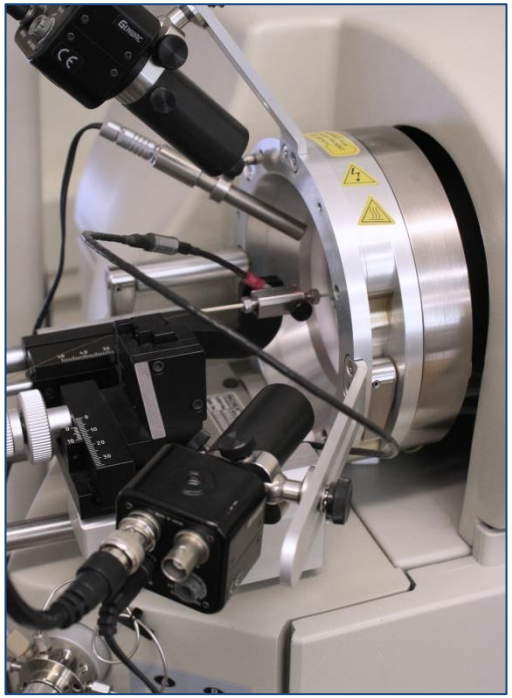


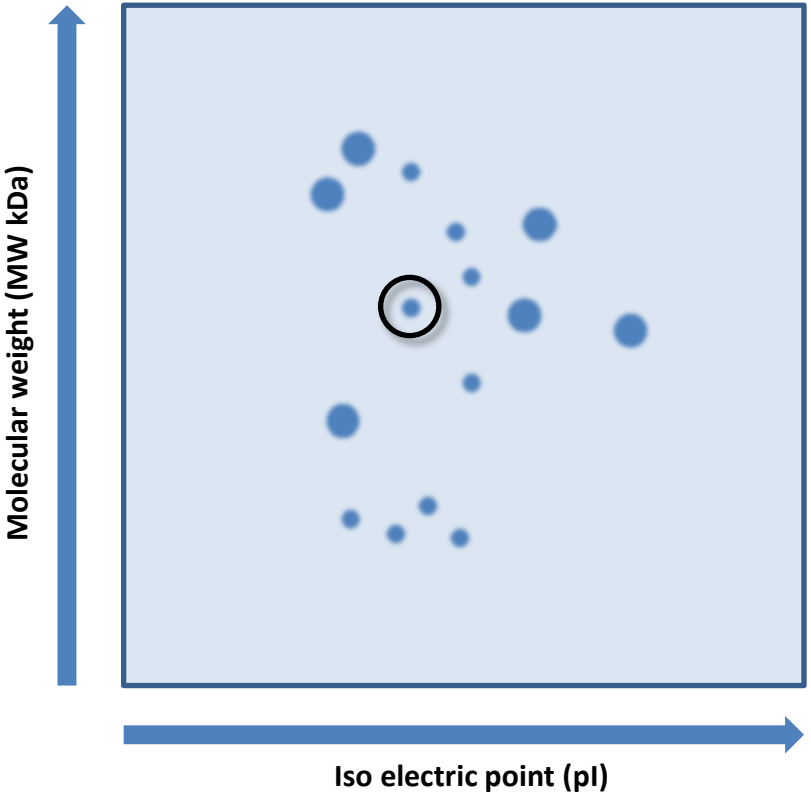
(Based on McSweeny and Mackie, 2012)

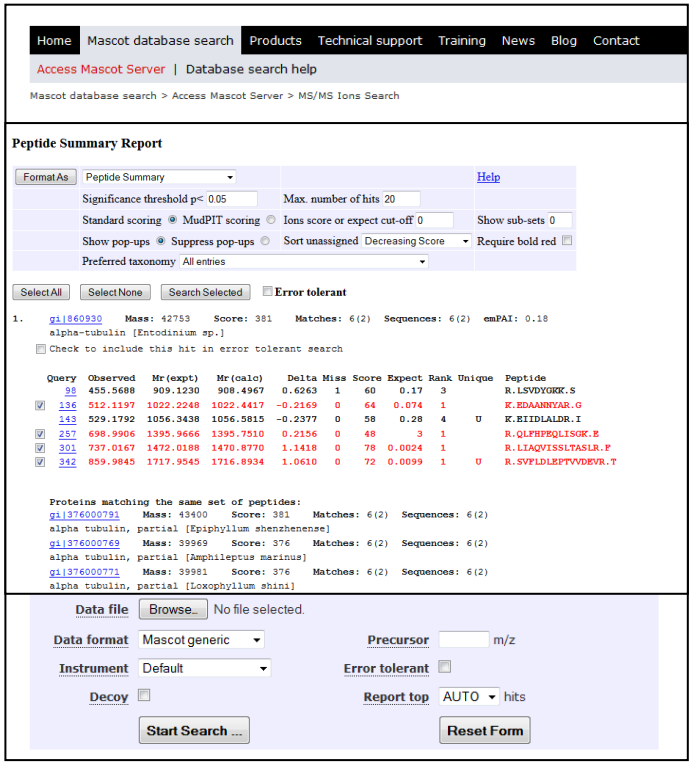
Metaproteomics

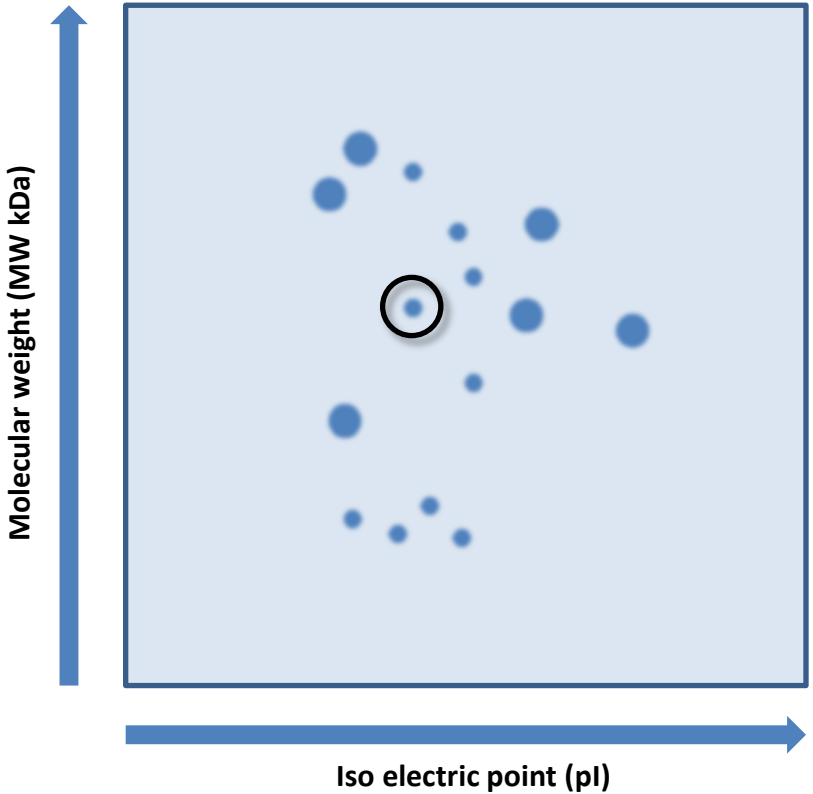
‘Large scale characterisation of the entire protein complement of an environmental microbiota at a given point in time’

(Wilmes and Bond, 2004)

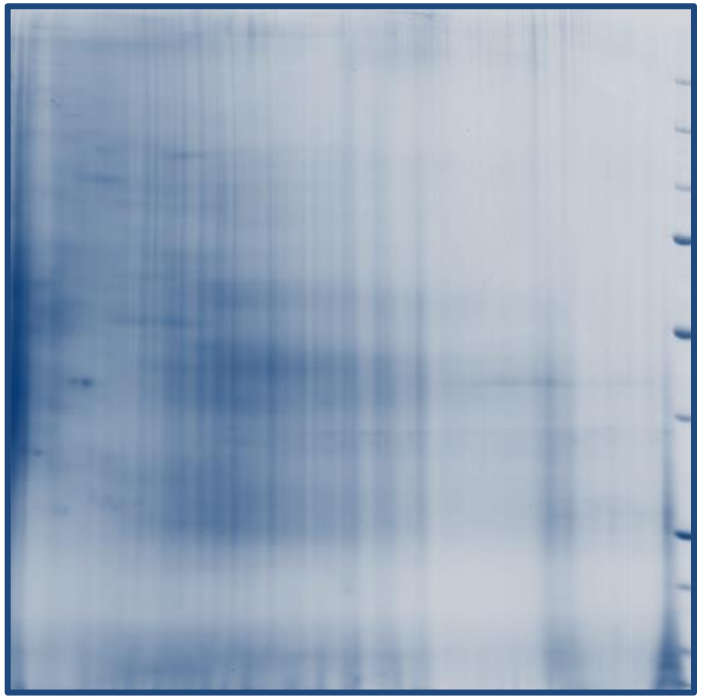



Molecular weight (MW kDa)
Iso electric point (pI)


Home
Mascot database search
Products
Technical support
Training
News
Blog
Contact
Access Mascot Server | Database search help
Mascot database search > Access Mascot Server > MS/MS Ions Search
Peptide Summary Report
Format As
Peptide Summary
Help
Significance threshold p<
0.05
Max. number of hits
20
Standard scoring
MudPIT scoring
Ions score or expect cut-off
0
Show sub-sets
0
Show pop-ups
Suppress pop-ups
Sort unassigned
Decreasing Score
Require bold red
Preferred taxonomy
All entries
Select All
Select None
Search Selected
Error tolerant
1. gi|860930 Mass: 42753 Score: 381 Matches: 6(2) Sequences: 6(2) emPAI: 0.18
alpha-tubulin [Entodinium sp.]
Check to include this hit in error tolerant search
Query Observed Mr(expt) Mr(calc) Delta Miss Score Expect Rank Unique Peptide
98 455.5688 909.1230 908.4967 0.6263 1 60 0.17 3 R.LSVDYGKK.S
136 512.1197 1022.2248 1022.4417 -0.2169 0 64 0.074 1 K.EDAANNYAR.G
143 529.1792 1056.3438 1056.5815 -0.2377 0 58 0.28 4 U K.EIIDLALDR.I
257 698.9906 1395.9666 1395.7510 0.2156 0 48 3 1 R.QLFHPEQLISGK.E
301 737.0167 1472.0188 1470.8770 1.1418 0 78 0.0024 1 R.LIAQVISSLTASLR.F
342 859.9845 1717.9545 1716.8934 1.0610 0 72 0.0099 1 U R.SVFLDLEPTVVDEVR.T
Proteins matching the same set of peptides:
gi|376000791 Mass: 43400 Score: 381 Matches: 6(2) Sequences: 6(2)
alpha tubulin, partial [Epiphyllum shenzhenense]
gi|376000769 Mass: 39969 Score: 376 Matches: 6(2) Sequences: 6(2)
alpha tubulin, partial [Amphileptus marinus]
gi|376000771 Mass: 39981 Score: 376 Matches: 6(2) Sequences: 6(2)
alpha tubulin, partial [Loxophyllum shini]
Data file
Browse...
No file selected.
Data format
Mascot generic
Precursor
m/z
Instrument
Default
Error tolerant
Decoy
Report top
AUTO
hits
Start Search ...
Reset Form

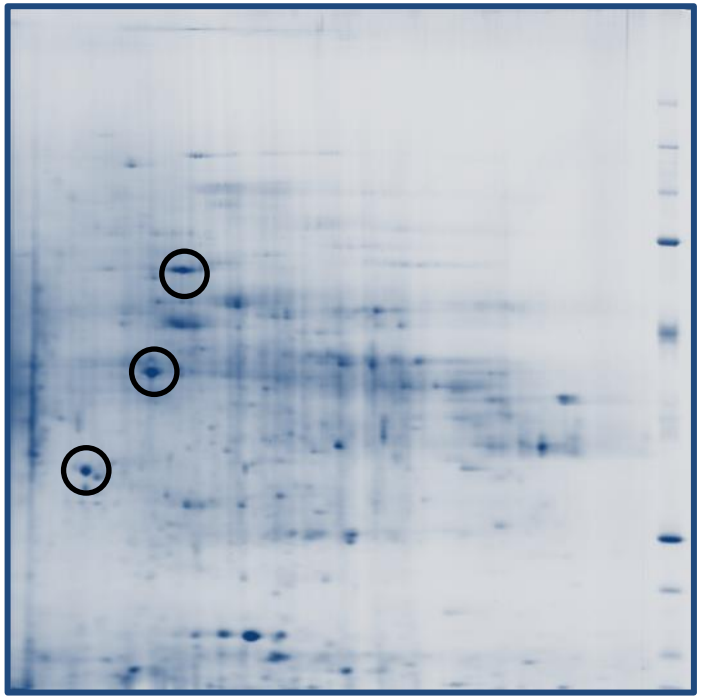

Molecular weight (MW kDa)
Iso electric point (pI)

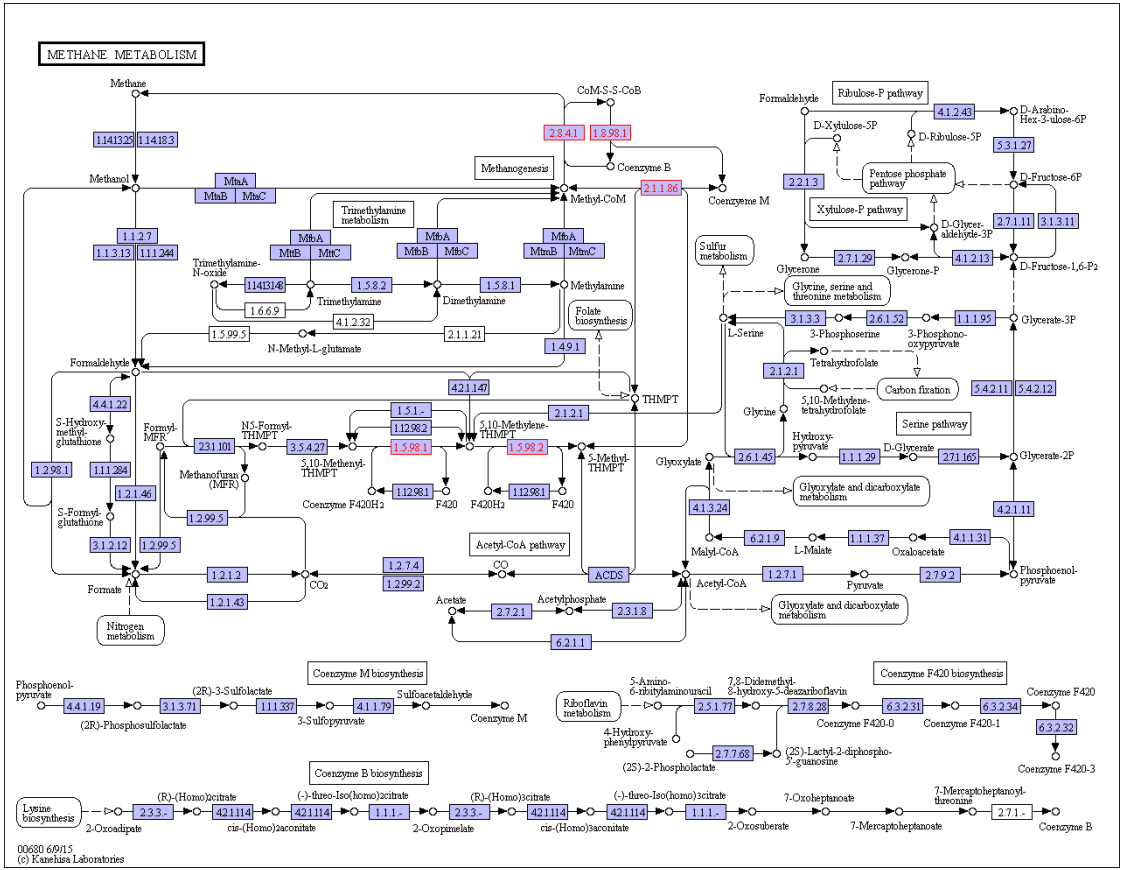
- Complexity of the metaproteome
- Low total protein yield
- Contaminants (eg. Humic compounds)
  - Cleanup unsuccessful
- Sample quality
  - Dietary fibre content
  - Ruminal cannula/Gastric intubation



- *Methanobrevibacter smithii*
  - Methyl coenzyme reductase subunit alpha (*mcr* A)

- *Methanobrevibacter ruminantium*
  - Methyl coenzyme M reductase subunit beta (*mcr* B)
  - 5,10-methylenetetrahydromethanopterin reductase *(mer)*




METHANE METABOLISM
Methane
Methanol
CoM-S-S-CoB
Coenzyme B
Methanogenesis
Methyl-CoM
Coenzyme M
MtaA
MtaB
MtaC
MtbA
MtbB
MtbC
MtmA
MtmB
MtmC
Trimethylamine metabolism
Trimethylamine-N-oxide
Trimethylamine
Dimethylamine
Methylamine
N-Methyl-L-glutamate
Formaldehyde
Formyl-MFR
S-Hydroxymethylglutathione
S-Formylglutathione
Methanofuran (MFR)
NS-Formyl-THMPT
5,10-Methenyl-THMPT
Coenzyme F420H2
F420
F420H2
5,10-Methylene-THMPT
5-Methyl-THMPT
THMPT
Folate biosynthesis
Formate
CO2
CO
Acetyl-CoA pathway
ACDS
Acetyl-CoA
Acetate
Acetylphosphate
Nitrogen metabolism
Sulfur metabolism
Ribulose-P pathway
D-Xylulose-5P
D-Arabino-Hex-3-ulose-6P
D-Ribulose-5P
Pentose phosphate pathway
D-Fructose-6P
Xylulose-P pathway
D-Glycer-aldehyde-3P
Glycerone
Glycerone-P
D-Fructose-1,6-P2
Glycine, serine and threonine metabolism
L-Serine
3-Phosphoserine
3-Phospho-oxypyruvate
Glycerate-3P
Tetrahydrofolate
5,10-Methylene-tetrahydrofolate
Carbon fixation
Glycine
Serine pathway
Hydroxy-pyruvate
D-Glycerate
Glycerate-2P
Glyoxylate
Glyoxylate and dicarboxylate metabolism
Malyl-CoA
L-Malate
Oxaloacetate
Phosphoenol-pyruvate
Pyruvate
Coenzyme M biosynthesis
Phosphoenol-pyruvate
(2R)-3-Sulfolactate
Sulfoacetaldehyde
(2R)-Phosphosulfolactate
3-Sulfopyruvate
Coenzyme M
Coenzyme F420 biosynthesis
Riboflavin metabolism
5-Amino-6-ribitylaminouracil
7,8-Didemethyl-8-hydroxy-5-deazariboflavin
Coenzyme F420-0
Coenzyme F420-1
Coenzyme F420
Coenzyme F420-3
4-Hydroxy-phenylpyruvate
(2S)-2-Phospholactate
(2S)-Lactyl-2-diphospho-5'-guanosine
Coenzyme B biosynthesis
Lysine biosynthesis
2-Oxoadipate
(R)-(Homo)citrate
cis-(Homo)aconitate
(-)-threo-Iso(homo)citrate
2-Oxopimelate
2-Oxosuberate
7-Oxoheptanoate
7-Mercaptoheptanoate
7-Mercaptoheptanoyl-threonine
Coenzyme B
00680 6/9/15
(c) Kanehisa Laboratories

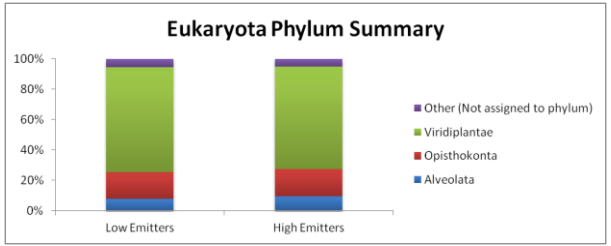
‘Shotgun’ Metaproteomics

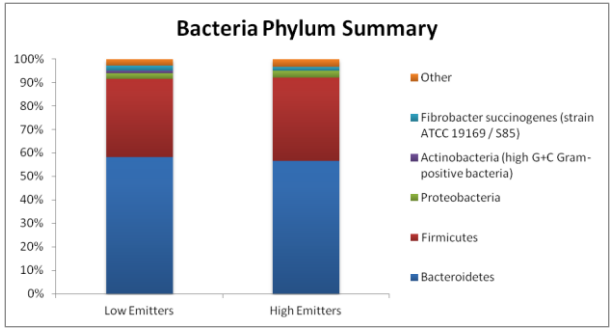
- Ultimate 3000 RS LC nano system
- LQT Orbitrap Velos/Q Exactive
- Proteome Discoverer software

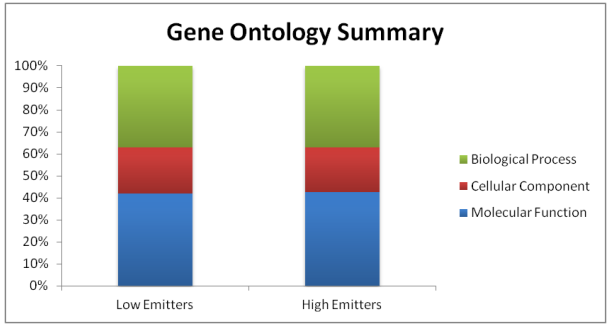
	A	B	C	D	E	
1	Accession	Description	Score	Coverage	# Proteins	# Uniq
2	860930	alpha-tubulin [Entodinium sp.]	578.30	25.39	1	
12	294674797	translation elongation factor Tu [Prevotella ruminicola 23]	408.34	29.70	32	
25	294674201	pyruvate, phosphate dikinase [Prevotella ruminicola 23]	392.54	17.33	1	
41	294674786	translation elongation factor G [Prevotella ruminicola 23]	361.00	29.52	1	
59	294674314	phosphoglycerate kinase [Prevotella ruminicola 23]	351.87	28.37	1	
69	294673934	pyruvate:ferredoxin (flavodoxin) oxidoreductase [Prevotella r	326.68	12.68	1	
82	294674921	phosphoenolpyruvate carboxykinase (ATP) [Prevotella ruminic	314.53	22.35	1	
91	461840	RecName: Full=Cruciferin CRU1; AltName: Full=11S globulin; A	295.78	20.04	1	
101	294673078	chaperonin GroEL [Prevotella ruminicola 23]	278.50	15.84	1	
108	496508379	elongation factor Tu [Bacteroidetes oral taxon 274]	267.57	15.70	68	
117	294674693	electron transfer flavoprotein subunit beta [Prevotella ruminic	266.06	33.10	1	
125	294674789	DNA-directed RNA polymerase subunit beta' [Prevotella rumini	265.08	7.76	1	
136	473750243	ATP synthase subunit alpha, chloroplastic [Triticum urartu]	262.36	14.37	14	
142	294674353	glyceraldehyde-3-phosphate dehydrogenase, type I [Prevote	259.36	26.39	1	
151	546392562	dNA-directed RNA polymerase subunit beta [Prevotella sp. CA	253.29	10.39	2	
161	390947271	translation elongation factor TU [Alistipes finegoldii DSM 17242	238.59	15.70	6	
169	546372837	elongation factor G [Prevotella sp. CAG:1058]	236.83	15.18	1	
179	294674529	triose-phosphate isomerase [Prevotella ruminicola 23]	236.21	36.36	7	
188	294673430	hypothetical protein PRU_0686 [Prevotella ruminicola 23]	233.42	49.36	1	
197	556509899	pyruvate, phosphate dikinase [Catonella morbi]	229.60	8.88	1	
205	302345656	glyceraldehyde-3-phosphate dehydrogenase, type I [Prevote	218.80	22.22	1	
213	294673915	glutamate dehydrogenase, NADP-specific [Prevotella ruminicol	212.97	18.43	1	
222	493326275	phosphoenolpyruvate carboxykinase [Prevotella bryantii]	211.28	12.13	1	
228	294674620	type III glutamine synthetase [Prevotella ruminicola 23]	210.42	11.94	1	
235	3341415	beta-tubulin [Colpoda sp.]	209.95	22.93	10	
244	395398505	transketolase [uncultured bacterium r_02]	207.49	10.15	1	
251	294673784	phosphoglucomutase/phosphomannomutase family protein [Pr	206.20	13.25	1	
259	294674338	family 57 glycosyl hydrolase [Prevotella ruminicola 23]	203.67	11.49	1	
264	294673431	hypothetical protein PRU_0687 [Prevotella ruminicola 23]	203.57	26.04	1	
273	294673443	glutamate synthase [Prevotella ruminicola 23]	198.81	9.80	1	
280	294673306	ketol-acid reductoisomerase [Prevotella ruminicola 23]	192.91	11.56	3	
286	294674623	glutamate synthase large subunit [Prevotella ruminicola 23]	192.88	5.54	1	
293	557097347	hypothetical protein EUTSA_v10029458mg [Eutrema salsugine	192.18	13.26	1	
301	294672852	chaperone protein DnaK [Prevotella ruminicola 23]	190.85	12.85	1	
309	546358564	putative uncharacterized protein [Blautia hydrogenotrophica C	190.53	17.83	2	
314	294674318	methylmalonyl-CoA mutase large subunit [Prevotella ruminicola	189.29	10.40	1	
321	294673366	translation elongation factor G [Prevotella ruminicola 23]	185.99	8.07	1	
327	294674941	tyrosine--tRNA ligase [Prevotella ruminicola 23]	185.23	8.82	1	
331	294673346	threonine--tRNA ligase [Prevotella ruminicola 23]	182.90	11.27	1	
338	17805	cruciferin cru4 subunit [Brassica napus]	182.46	15.01	1	
344	547907819	putative uncharacterized protein [Prevotella sp. CAG:474]	181.53	31.45	1	
353	383755284	putative flagellin [Selenomonas ruminantium subsp. lactilytica T	180.65	13.15	1	
359	294673930	carboxyl transferase domain-containing protein [Prevotella rur	179.63	7.87	1	
364	294675158	hypothetical protein PRU_2525 [Prevotella ruminicola 23]	178.03	13.86	1	
371	517750292	DNA-directed RNA polymerase subunit alpha [Prevotella oris]	176.90	17.35	10	

Metaproteome Summary

- UniProt
  - Taxonomy
  - Function
    - Gene Ontology (GO terms)





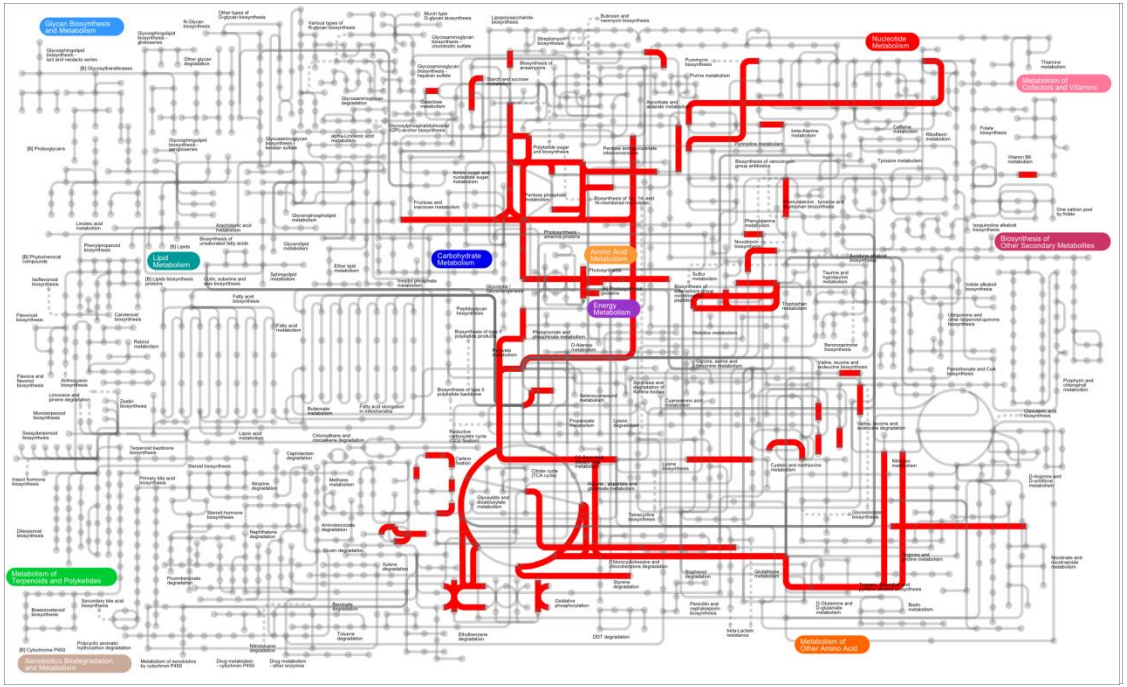


Metapeptidome Summary

- UniPept (Mesuere *et al*., 2012)

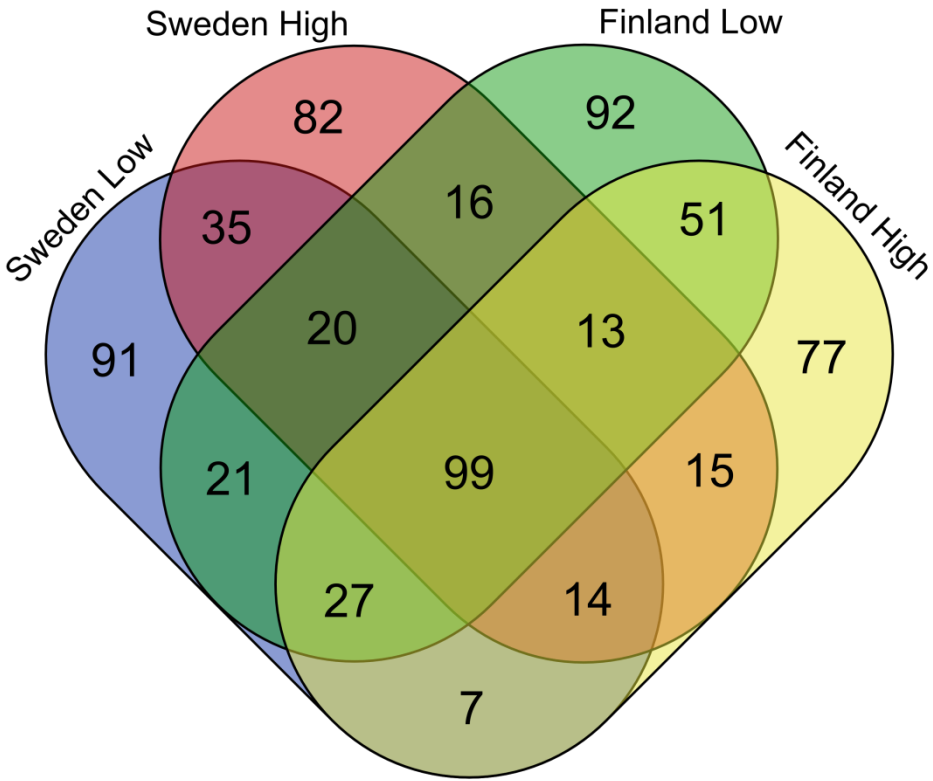
Functional Metapathways

- iPath (Letunic et al., 2008)
  - High Emitters
  - Low Emitters



Core metaproteome

- UniProt KB Accession Numbers



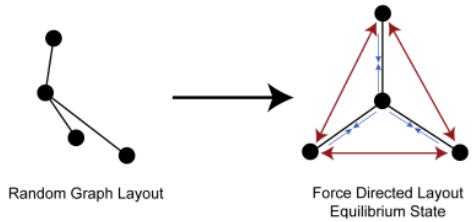
Finland – High Emitters

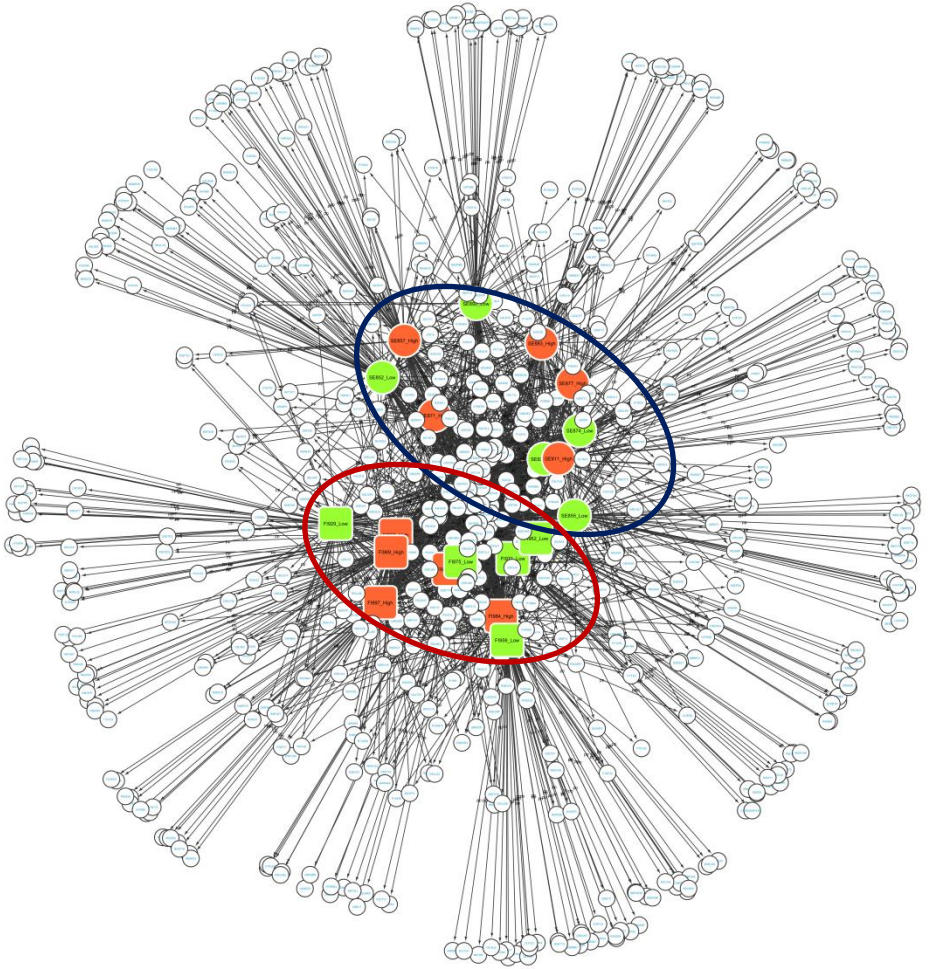
Finland – Low Emitters

Sweden – High Emitters

Sweden – Low Emitters

Proteins (UniProt KB)

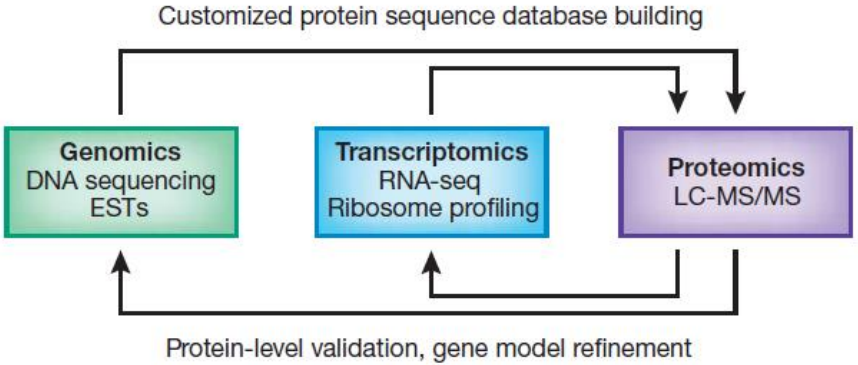




Mixing Metagenomics and Metaproteomics

- Metagenome data can be used to generate a sample specific reference protein database to identify novel peptides

- Metaproteome data can be used to provide evidence of gene expression and refine metagenome annotation



- **Rowett Institute, University of Aberdeen**
  - Nichol Fergus
  - Martin Reid
  - Louise Cantlay
- **Parco Tecnologico Padano (PTP), Italy**
- **Institute of Medical Sciences, University of Aberdeen**
  - David Stead
  - Ian Davidson
- **FingerPrints Proteomics, University of Dundee**
  - Douglas Lamont
  - Samantha Kosto
- **RuminOmics**
  - www.ruminomics.eu
  - @ruminomics