

Analysis and interpretation of
Next Generation Sequencing data
from the nitrogen fixing
bacterium *Klebsiella oxytoca*
under different environmental
conditions



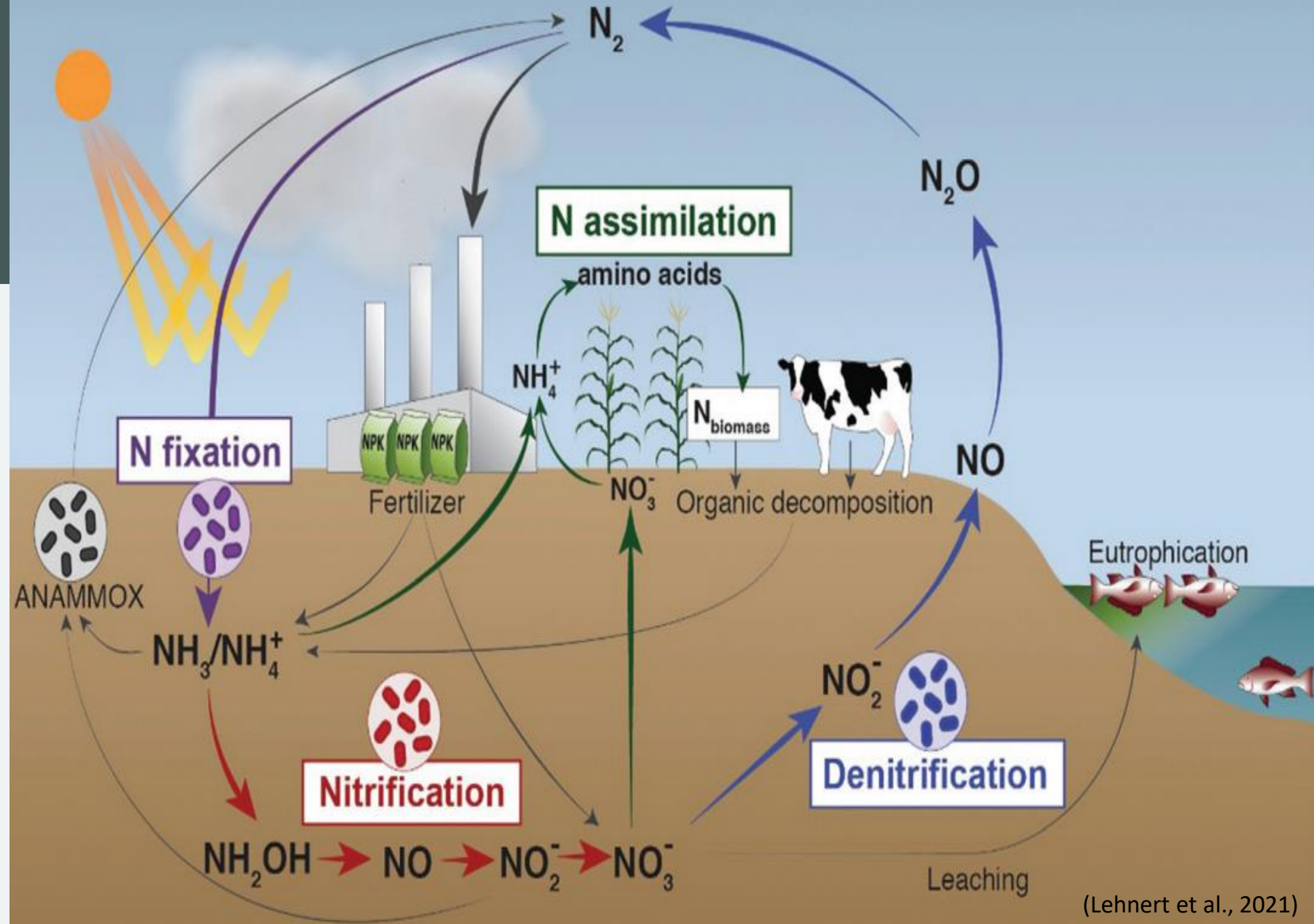
Michalis Gkouvelis

University of Patras

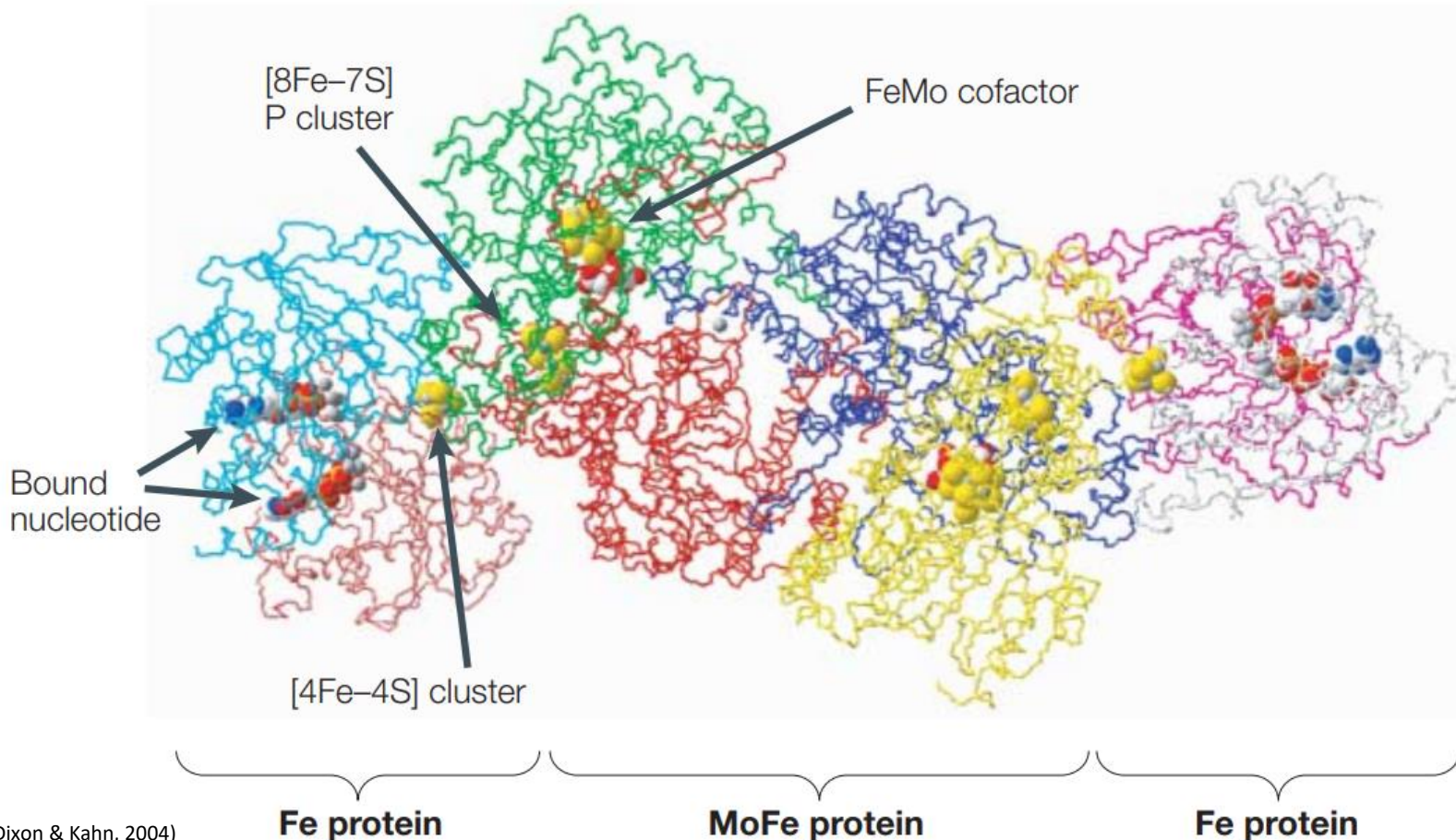
Informatic for Life Sciences

The nitrogen cycle

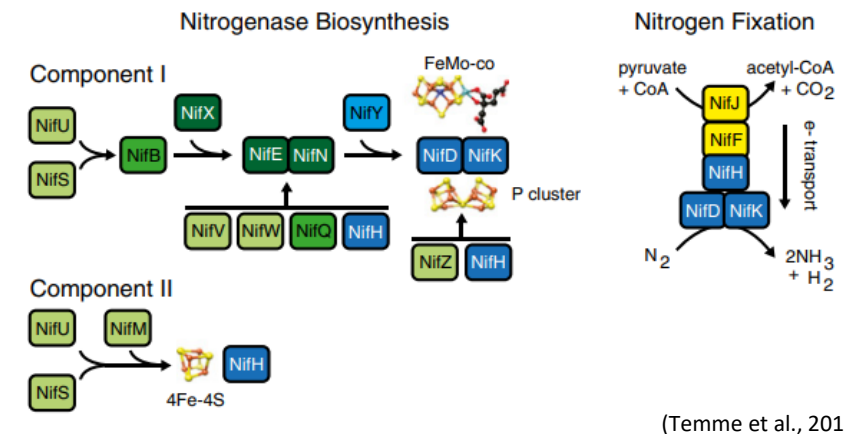
- Molecular nitrogen
- Organisms
- Atmosphere - Biosphere
- Transformation flows

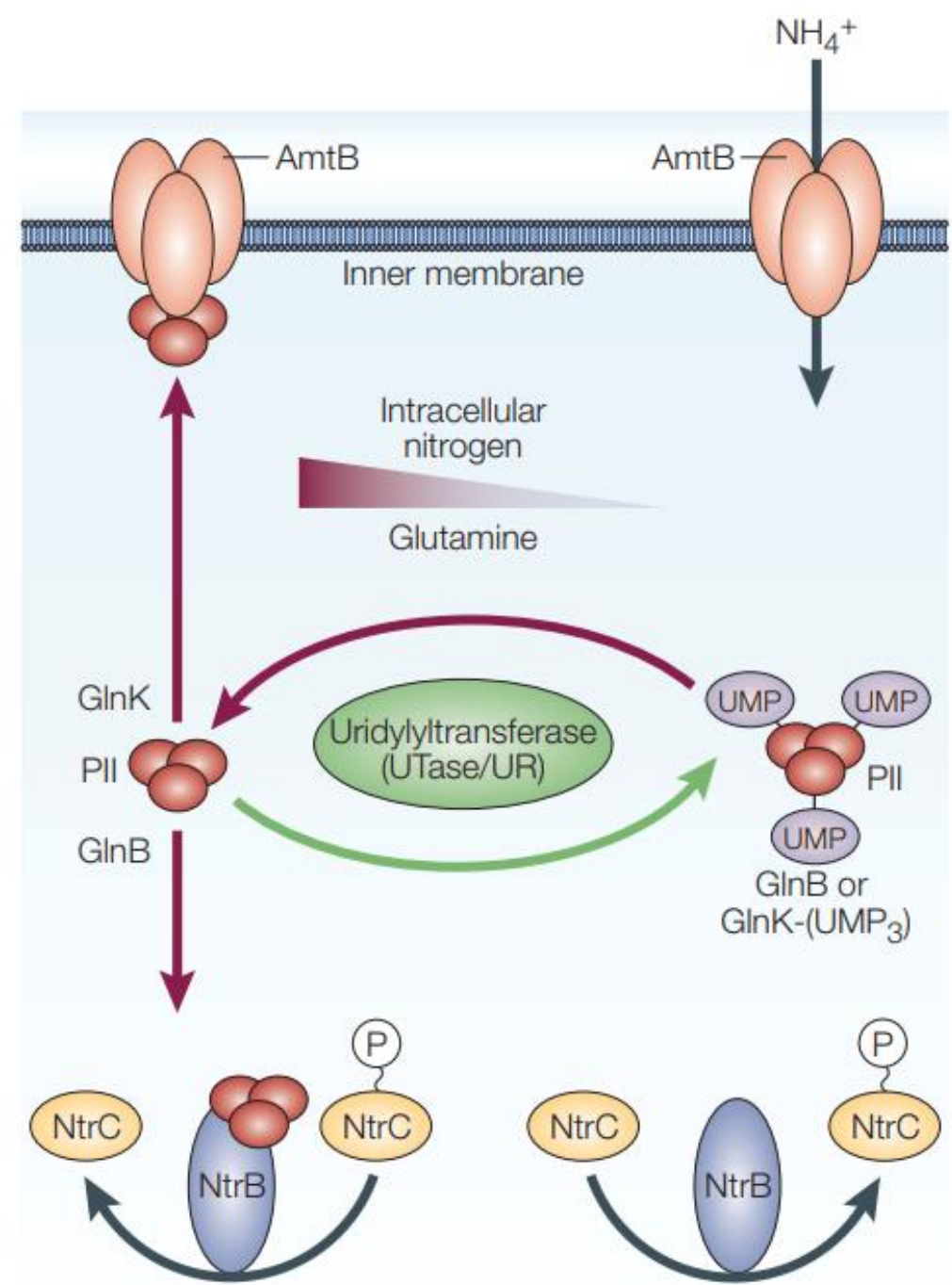


Nitrogen metabolism and regulation in *K. oxytoca*

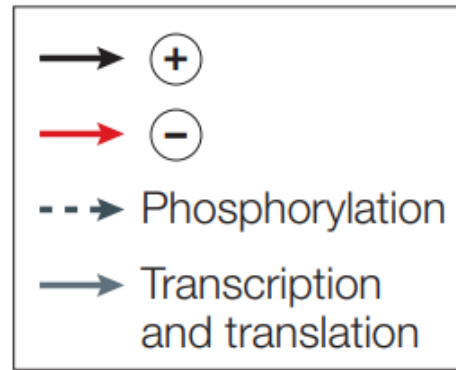


- Model organism
- Nitrogenase
- *nif* regulon

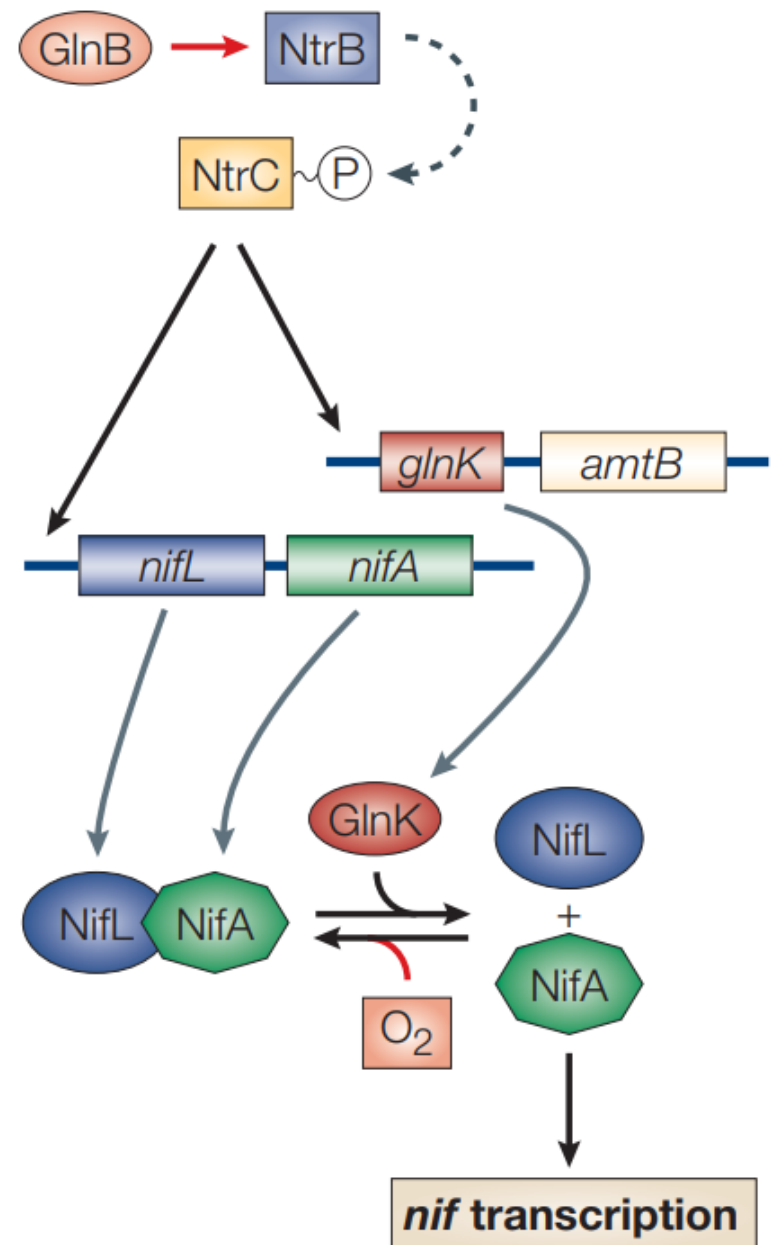




(Dixon & Kahn, 2004)



Klebsiella pneumoniae



(Rubio & Ludden, 2005)

Hypotheses and research questions

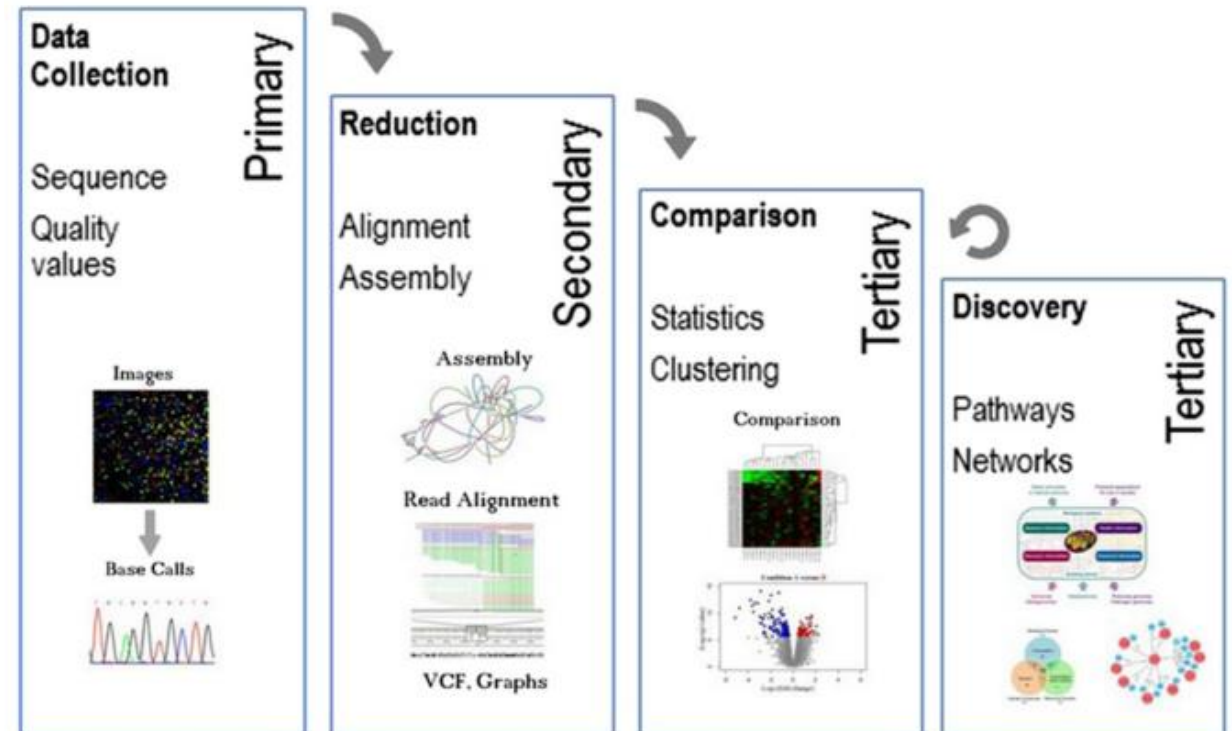
Diazotrophic conditions:

- 40% of proteome consists of nitrogenase
- 50% of genes differentially expressed
- Nitrogen related genes/processes ↑
- Genes/processes related to energy consumption ↓
- Regulatory mechanisms allow the cells to withstand the metabolic and translational stress (Waite et al., 2021)

Materials and methods

Klebsiella oxytoca next-generation sequencing transcriptomic data (FASTQ files)

No.	Sample name	Sample description
1	WT-highN-rep1	wild type strain in medium with high nitrogen concentration
2	WT-highN-rep2	wild type strain in medium with high nitrogen concentration
3	WT-lowN-rep1	wild type strain in medium with low nitrogen concentration
4	WT-lowN-rep2	wild type strain in medium with low nitrogen concentration
5	nifH-lowN-rep1	<i>nifH</i> knock out strain in medium with low nitrogen concentration
6	nifH-lowN-rep2	<i>nifH</i> knock out strain in medium with low nitrogen concentration



Steps - Tools

- Quality control- Preprocessing of raw data (FastQC, Trimmomatic)
- Alignment to a reference genome (Bowtie2)
- Generation of count tables (Gffread, Htseq-count)
- Differential gene expression analysis (DESeq2)
- Data visualization (ggplot2, pheatmap, plotPCA)
- Retrieval of GO terms (Blast2GO)
- Hypergeometric tests (phyper)
- Visualization of the enriched GO terms (REVIGO)

Results

Quality control – Preprocessing

Before preprocessing

Basic Statistics

Measure	Value
Filename	WT-highN-rep1_fastq_gz.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	10808384
Sequences flagged as poor quality	0
Sequence length	75
%GC	50

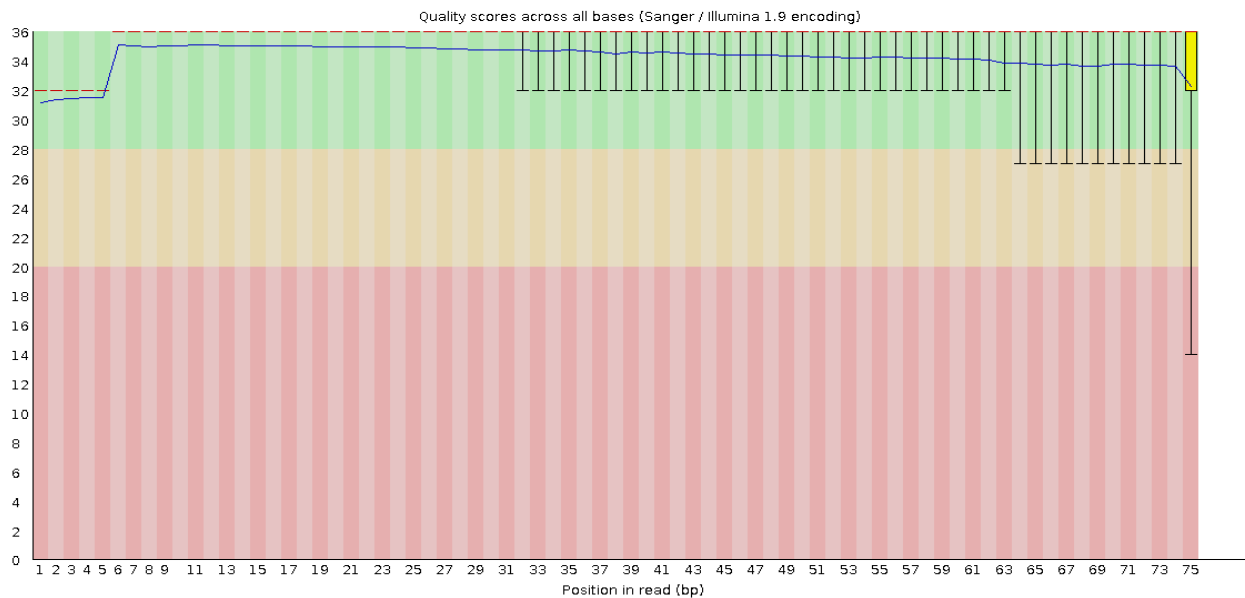
After preprocessing

Basic Statistics

Measure	Value
Filename	WT-highN-rep1_fastq_gz.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	10263748
Sequences flagged as poor quality	0
Sequence length	50-75
%GC	50

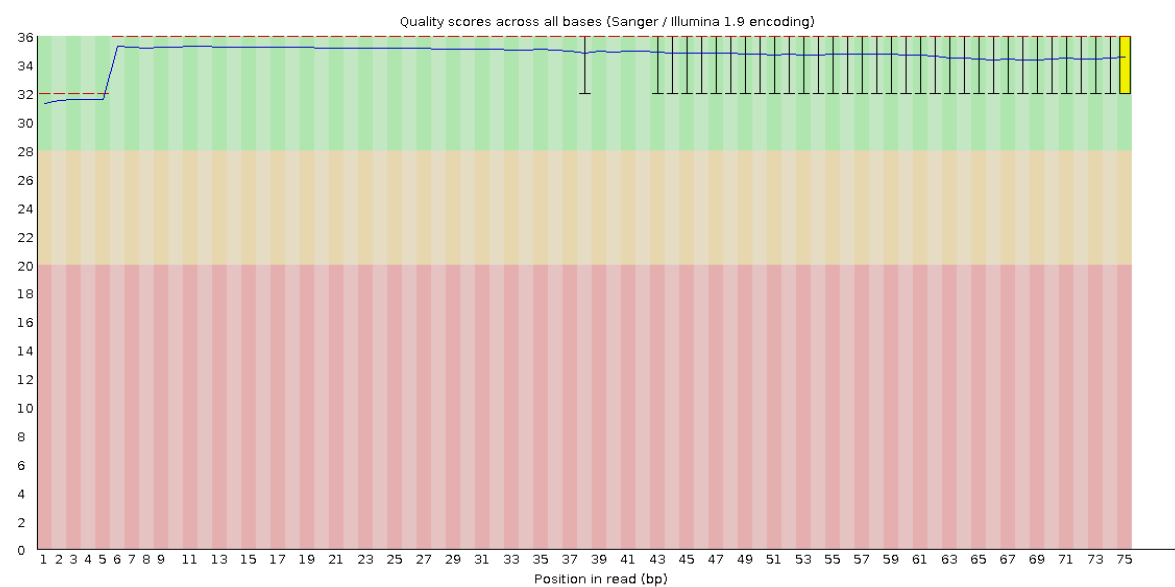
Before preprocessing

Per base sequence quality

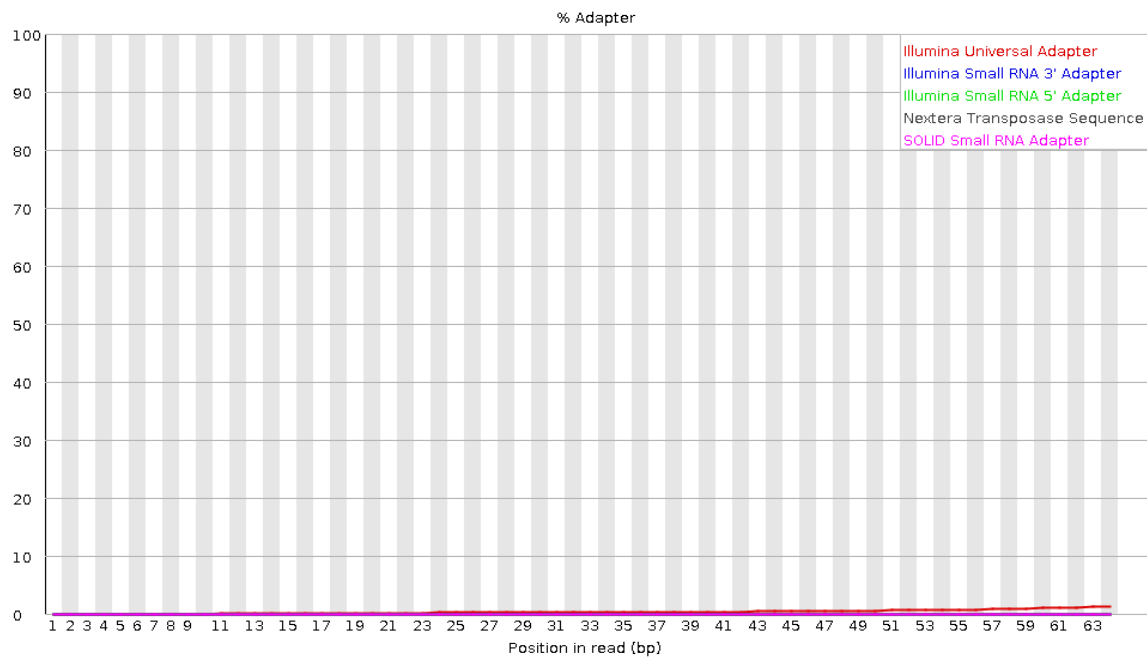


After preprocessing

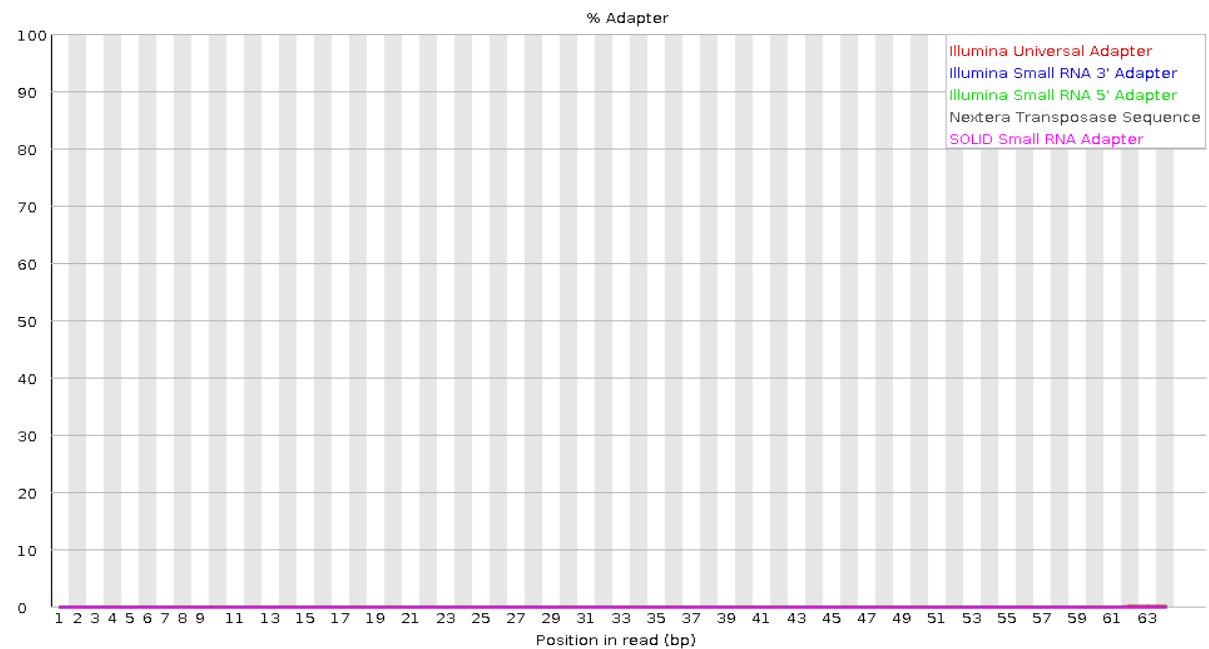
Per base sequence quality



Adapter Content

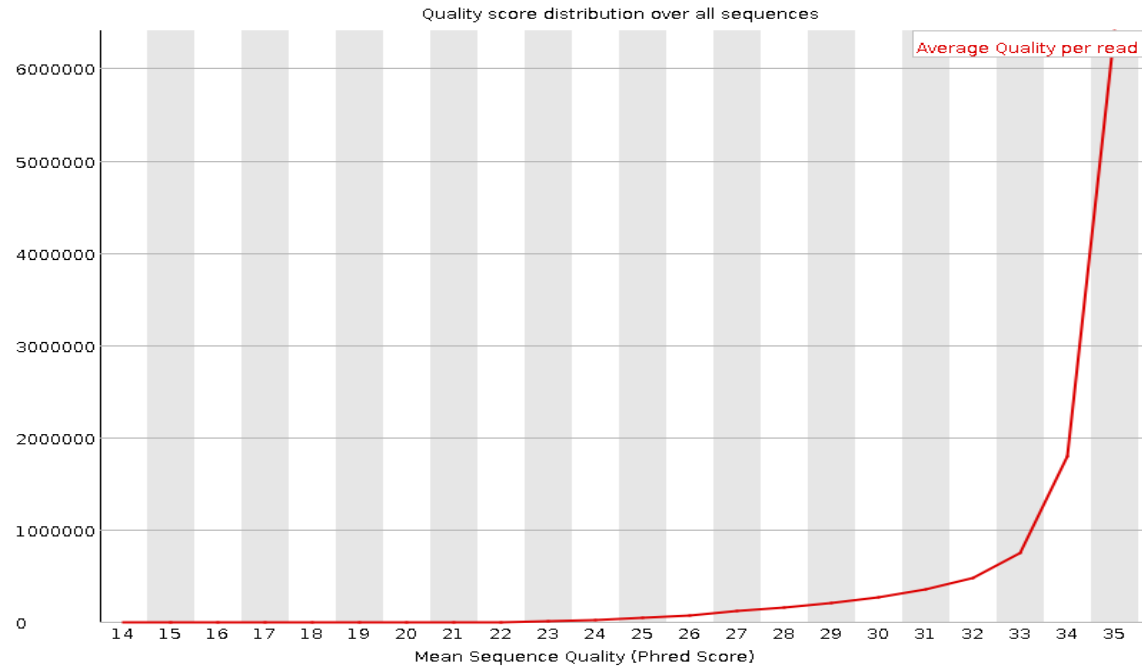


Adapter Content



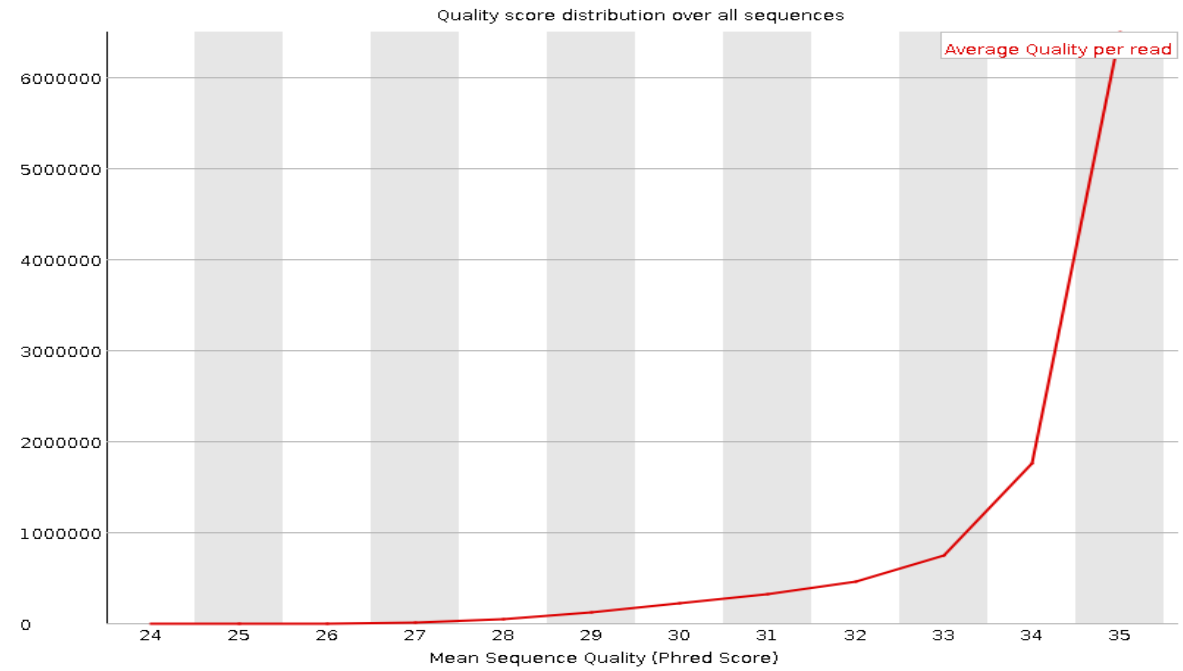
Before preprocessing

✔ Per sequence quality scores

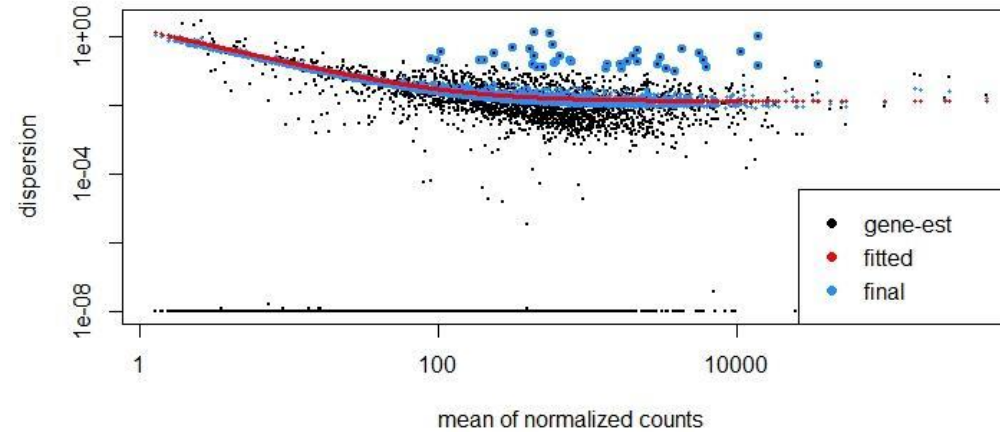


After preprocessing

✔ Per sequence quality scores



DESeq2 dispersion plot



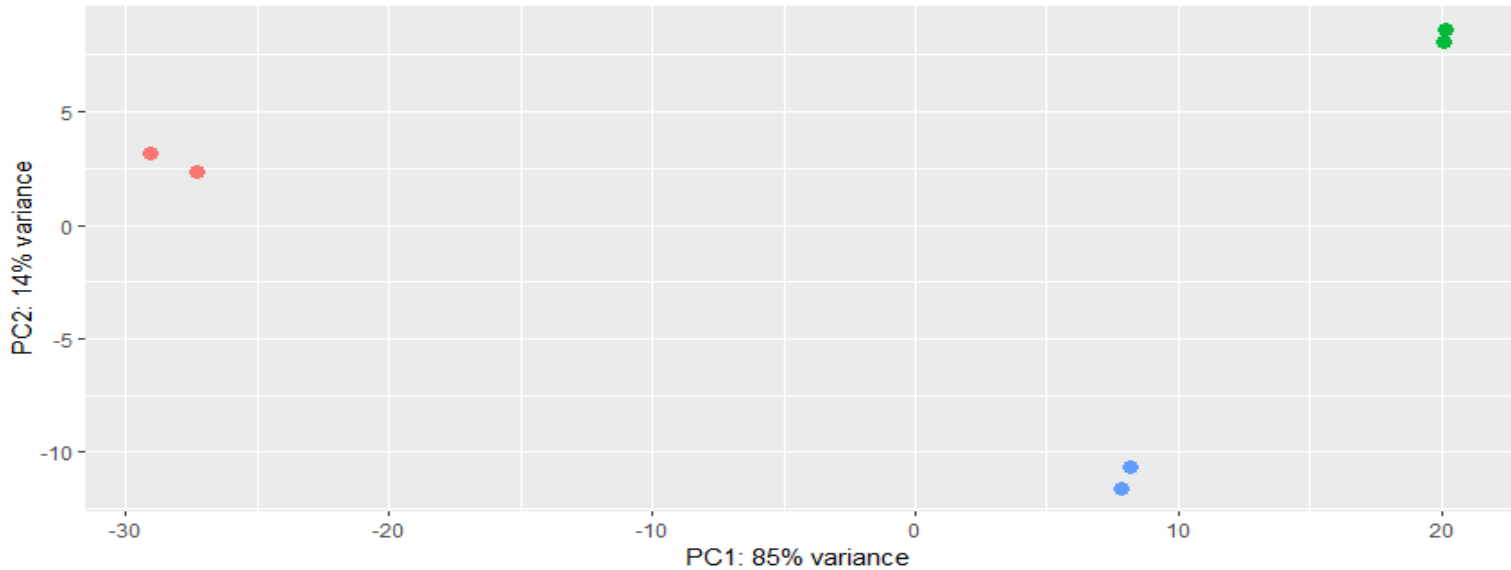
Alignment statistics

	WT-LowN-Rep1	WT-LowN-Rep2	WT-HighN-Rep1	WT-HighN-Rep2	NifH-LowN-Rep1	NifH-LowN-Rep2
Total Reads	13057294	15822528	10263748	12668858	11278147	13975491
Unpaired reads	13057294	15822528	10263748	12668858	11278147	13975491
	(100.00%)	(100.00%)	(100.00%)	(100.00%)	(100.00%)	(100.00%)
Reads Aligned 0 times	577419	636598	524606	719016	423903	537812
	(4.42%)	(4.02%)	(5.11%)	(5.68%)	(3.76%)	(3.85%)
Reads aligned 1 time	12137917	14695504	9208889	11090158	10611942	13204964
	(92.96%)	(92.88%)	(89.72%)	(87.54%)	(94.09%)	(94.49%)
Reads aligned >1 times	341958	490426	530253	859684	242302	232715
	(2.62%)	(3.10%)	(5.17%)	(6.79%)	(2.15%)	(1.67%)
Overall alignment rate	95.58%	95.98%	94.89%	94.32%	96.24%	96.15%

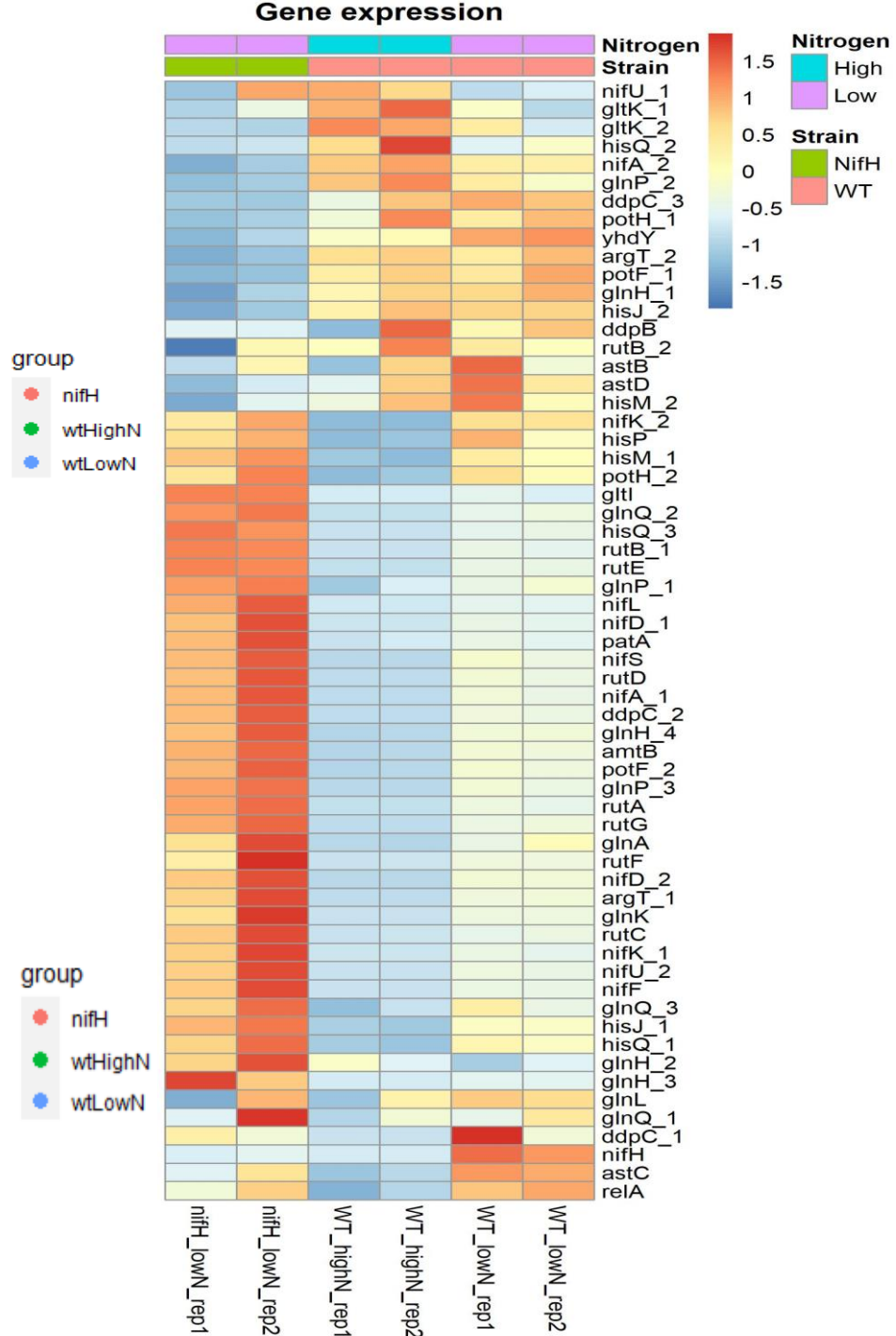
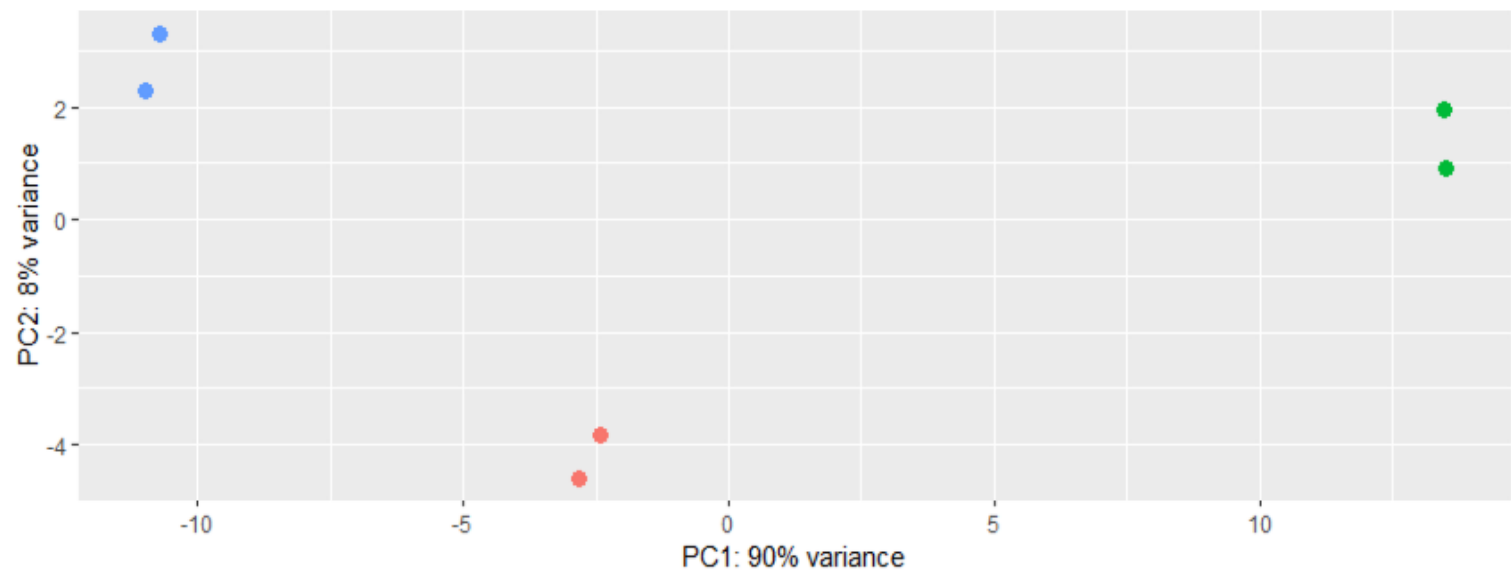
Count statistics

	WT-LowN-Rep1	WT-LowN-Rep2	WT-HighN-Rep1	WT-HighN-Rep2	NifH-LowN-Rep1	NifH-LowN-Rep2
No_feature	6227378	8549958	3530023	4765077	7088551	8599313
Ambiguous	112872	120119	118864	132058	52875	74898
Too_low_aQual	506403	740455	660001	966732	365232	389121
Not_aligned	577419	636598	524606	719016	423903	537812
Alignment_not_unique	0	0	0	0	0	0
Reads	13057294	15822528	10263748	12668858	11278147	13975491
Not_aligned	7424072	10047130	4833494	6582883	7930561	9601144
Aligned	5633222 (43,14%)	5775398 (36,5%)	5430254 (52,9%)	6085975 (48,03%)	3347586 (29,68%)	4374347 (31,3%)

PCA with all genes



PCA with nitrogen related genes



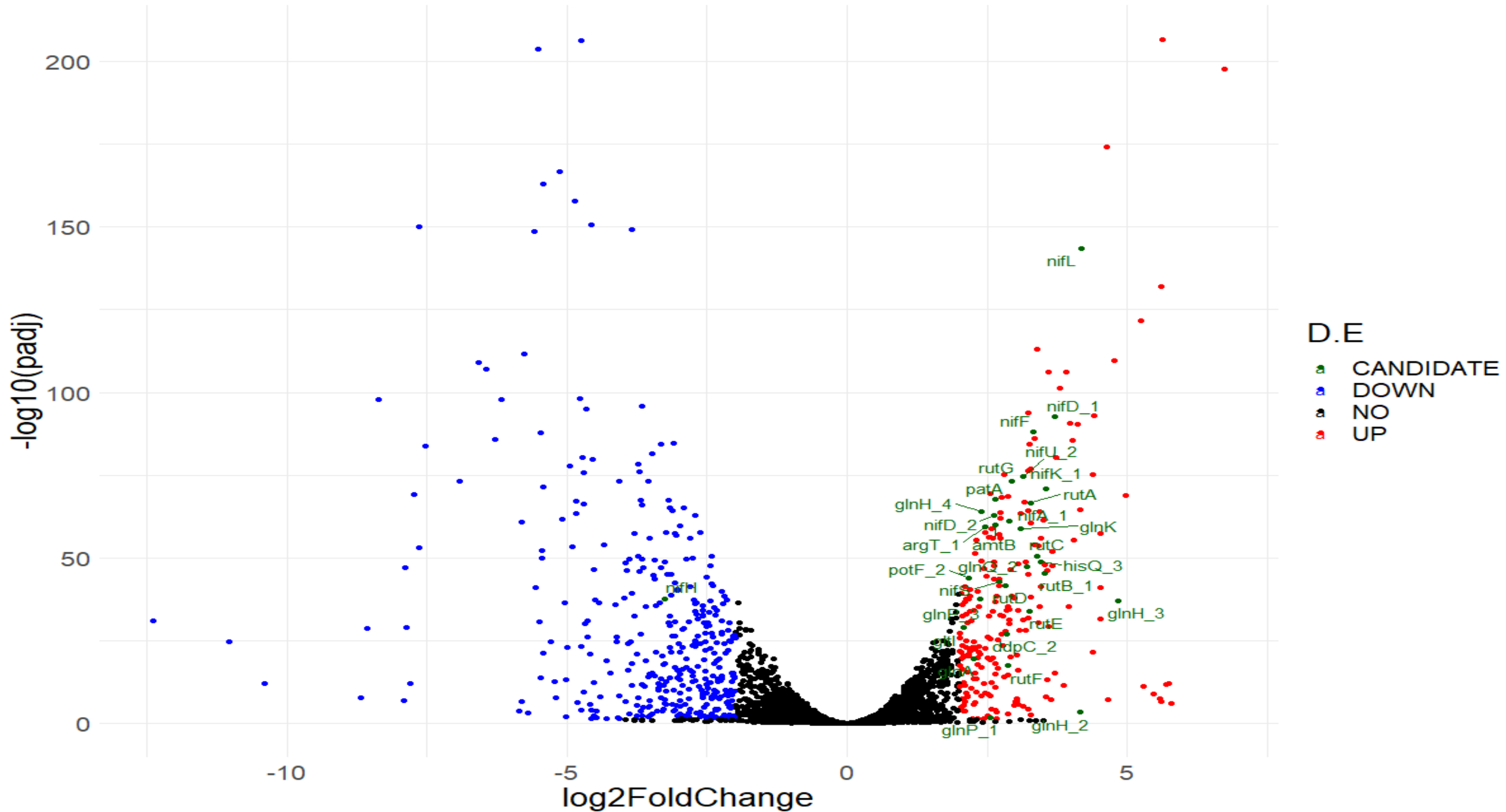
Differential gene expression analysis

	WT-LowN vs WT-HighN	NifH-LowN vs WT-LowN	NifH-LowN vs WT-HighN
Total count of DEG	235	611	741
LFC > 2 (up)	195 (83%)	240 (39%)	351 (47%)
LFC < -2 (down)	40 (17%)	371 (61%)	390 (53%)
Outliers	0 (0%)	0 (0%)	0 (0%)
Low counts	0 (0%)	0 (0%)	0 (0%)

WT_LowN_vs_WT_HighN



nifH_LowN_vs_WT_LowN

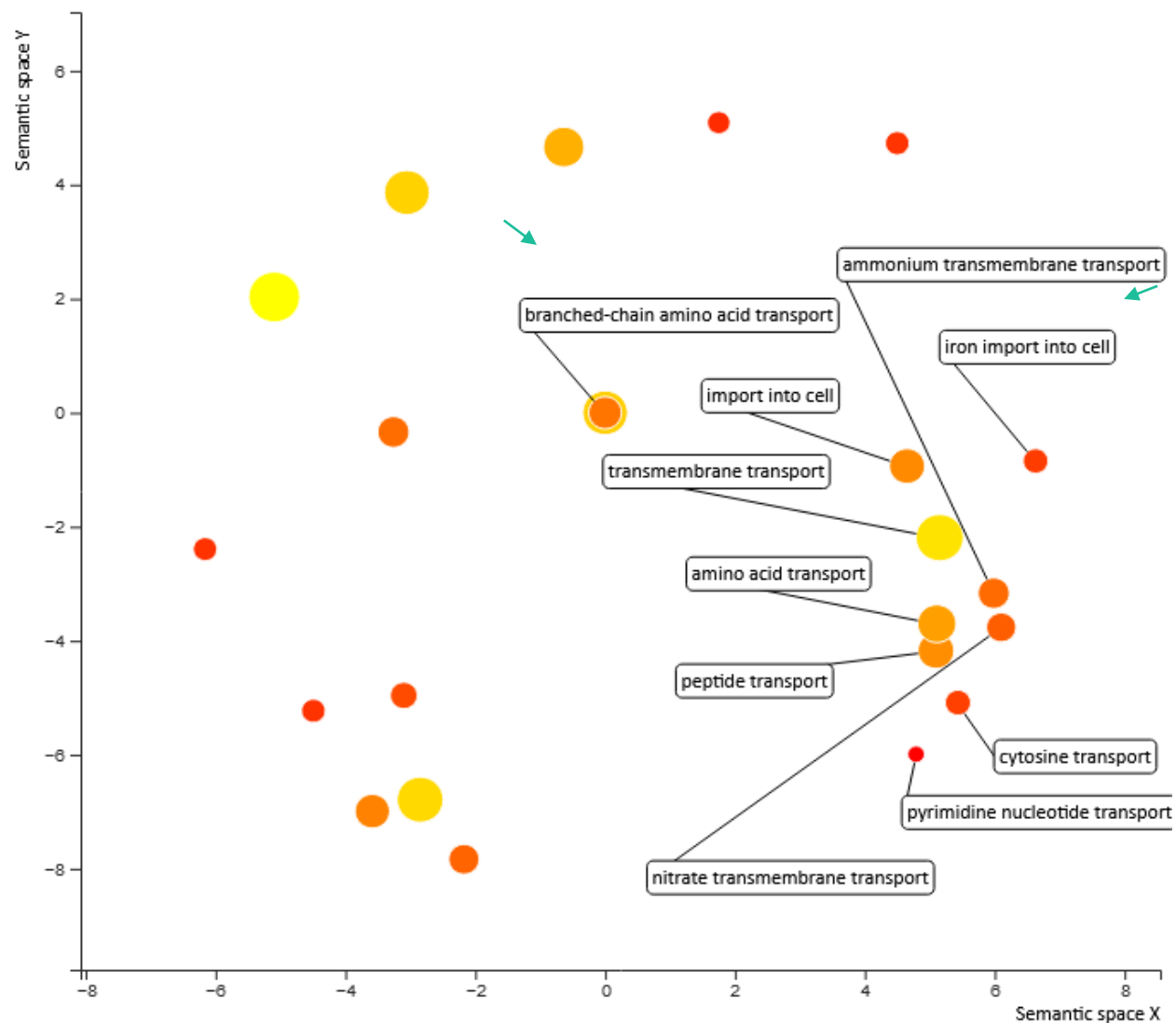


Gene ontology enrichment analysis

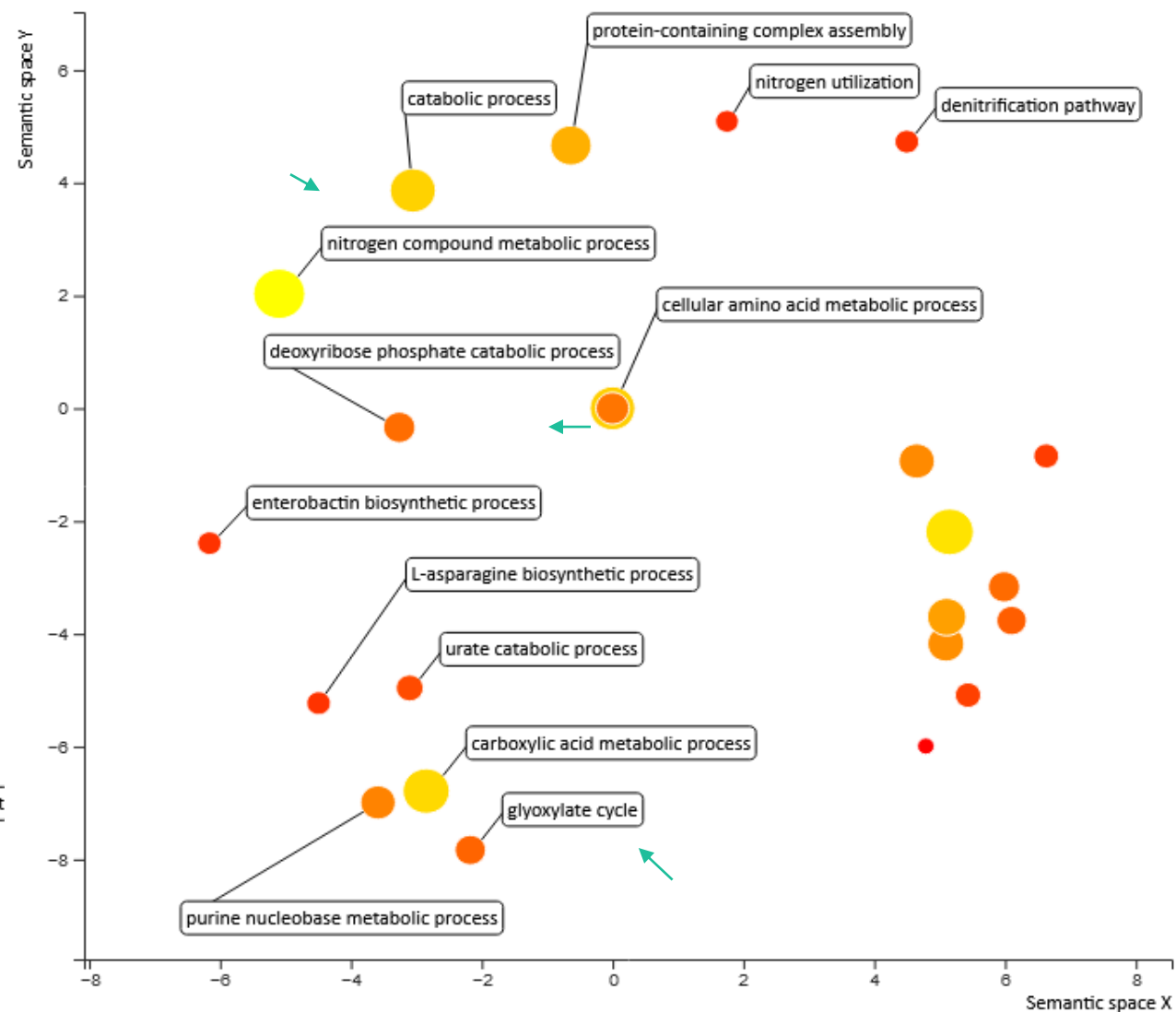
Pairwise comparison	Enriched GO terms for up and down regulated genes	Enriched GO terms for up regulated genes	Enriched GO terms for down regulated genes
WT-LowN vs WT-HighN	81	68	24
NifH-LowN vs WT-LowN	59	50	53
NifH-LowN vs WT-HighN	63	50	52

WT-LowN vs WT-HighN (all genes)

BP



BP

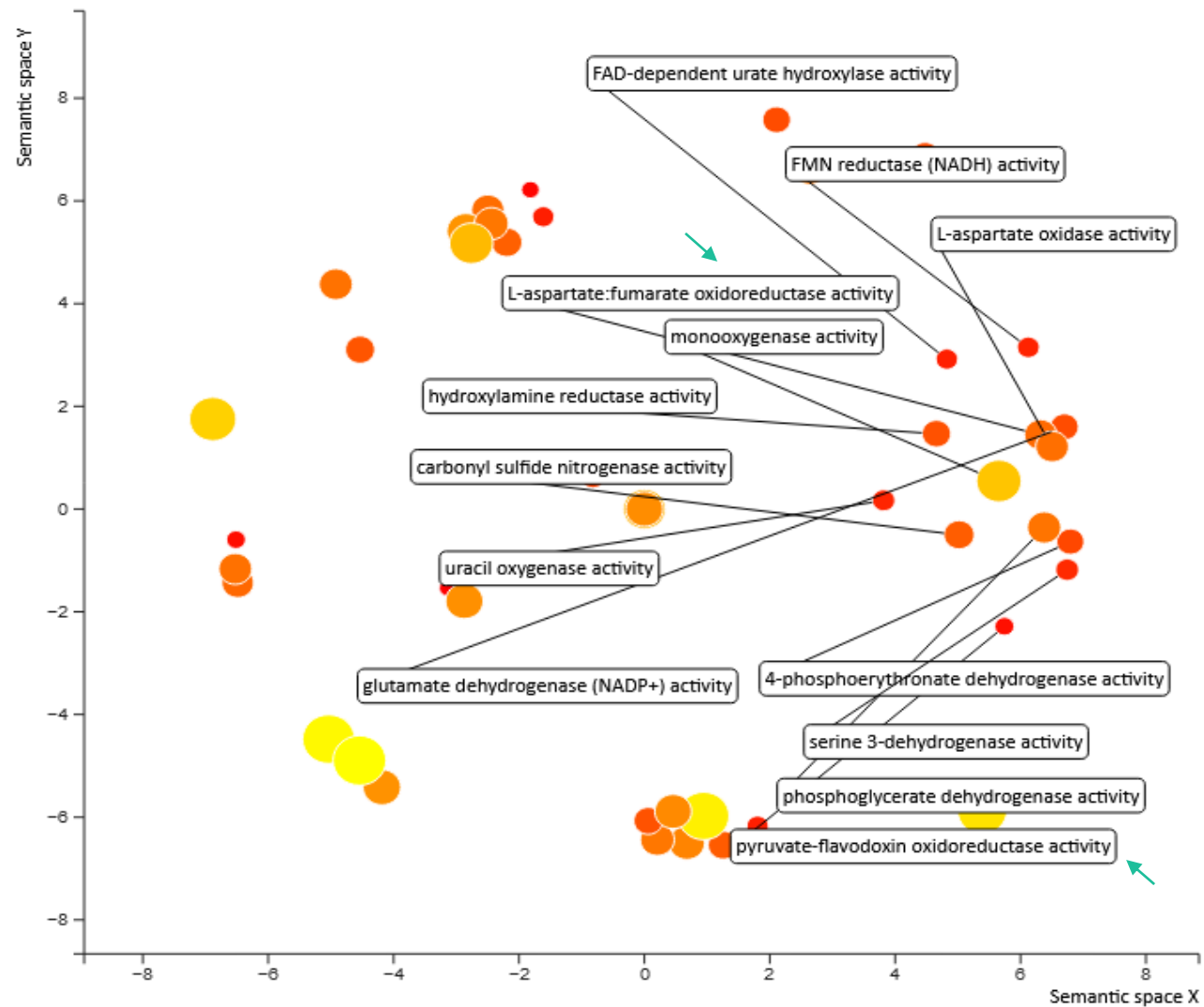
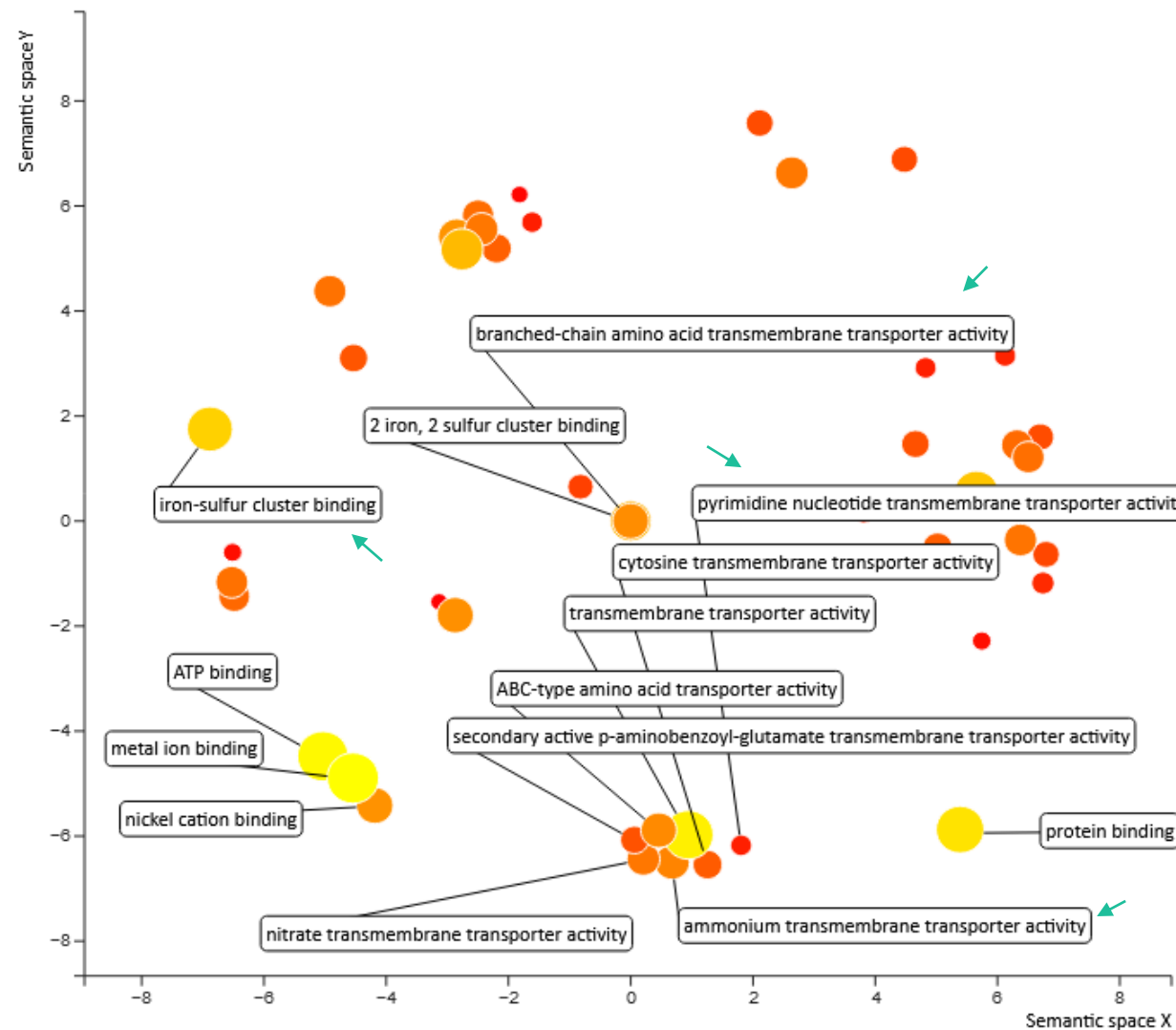


WT-LowN vs WT-HighN (all genes)

CC: molybdenum-iron nitrogenase complex.

MF

MF

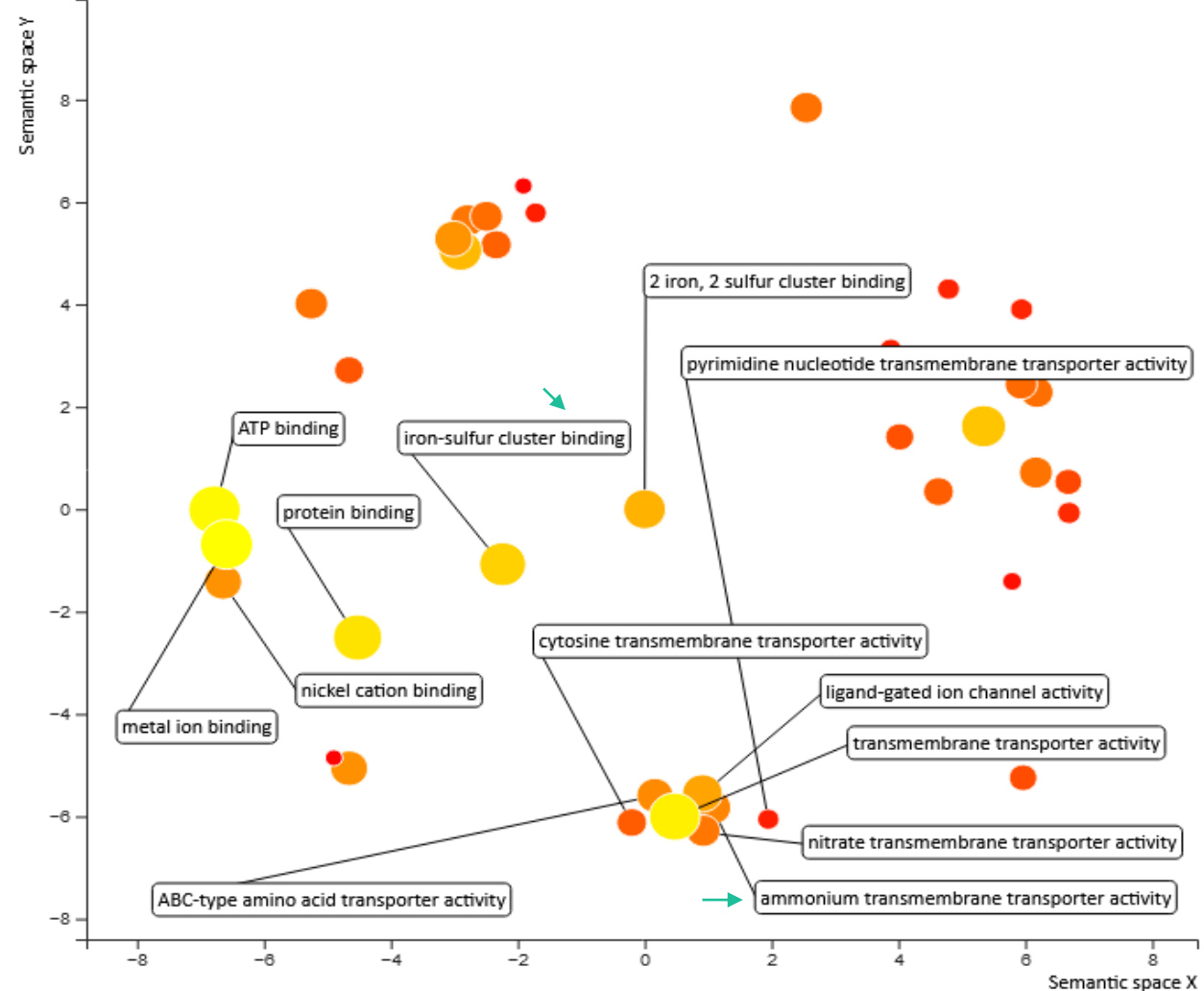
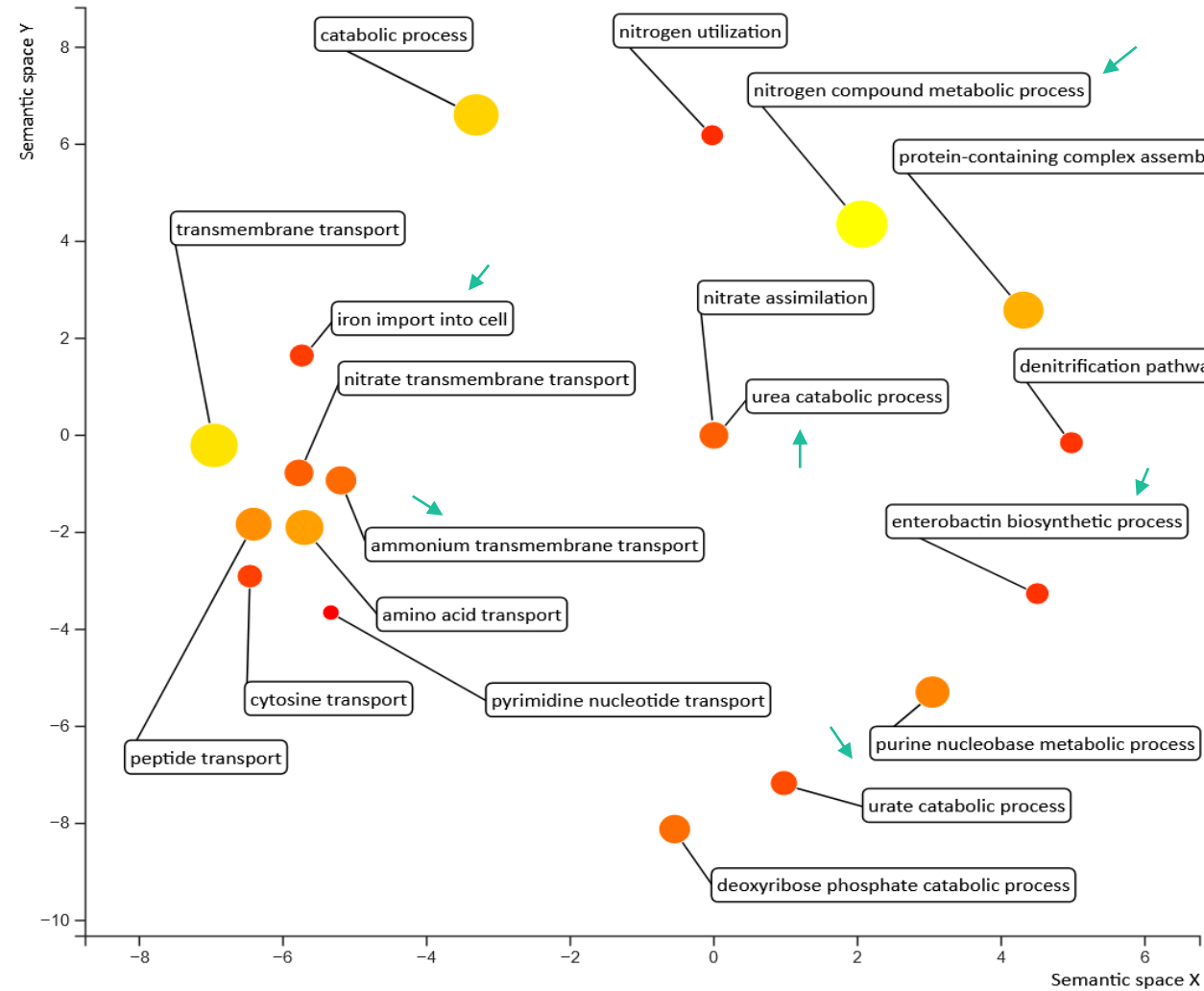


WT-LowN vs WT-HighN (up regulated genes)

CC: molybdenum-iron nitrogenase complex.

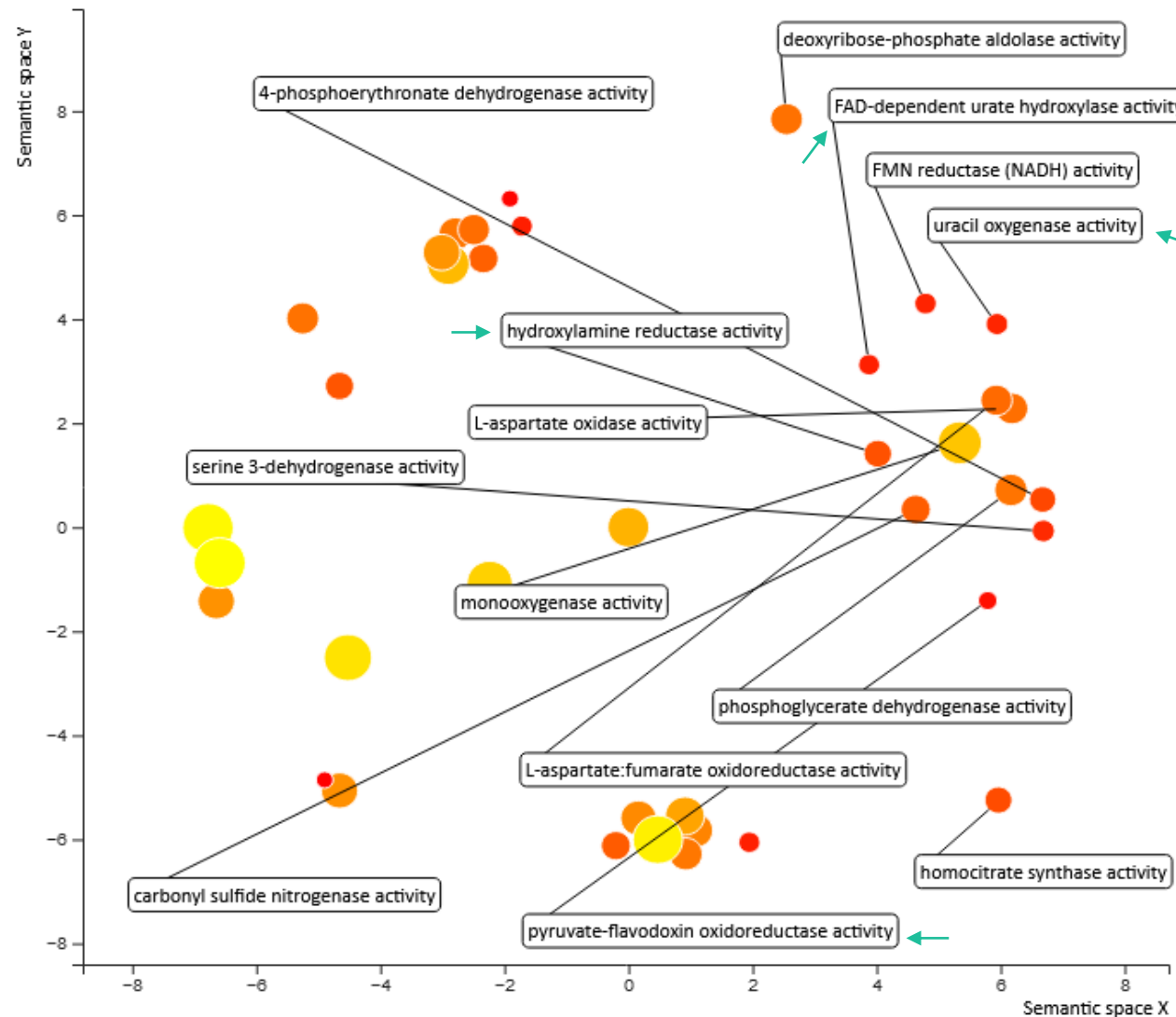
BP

MF

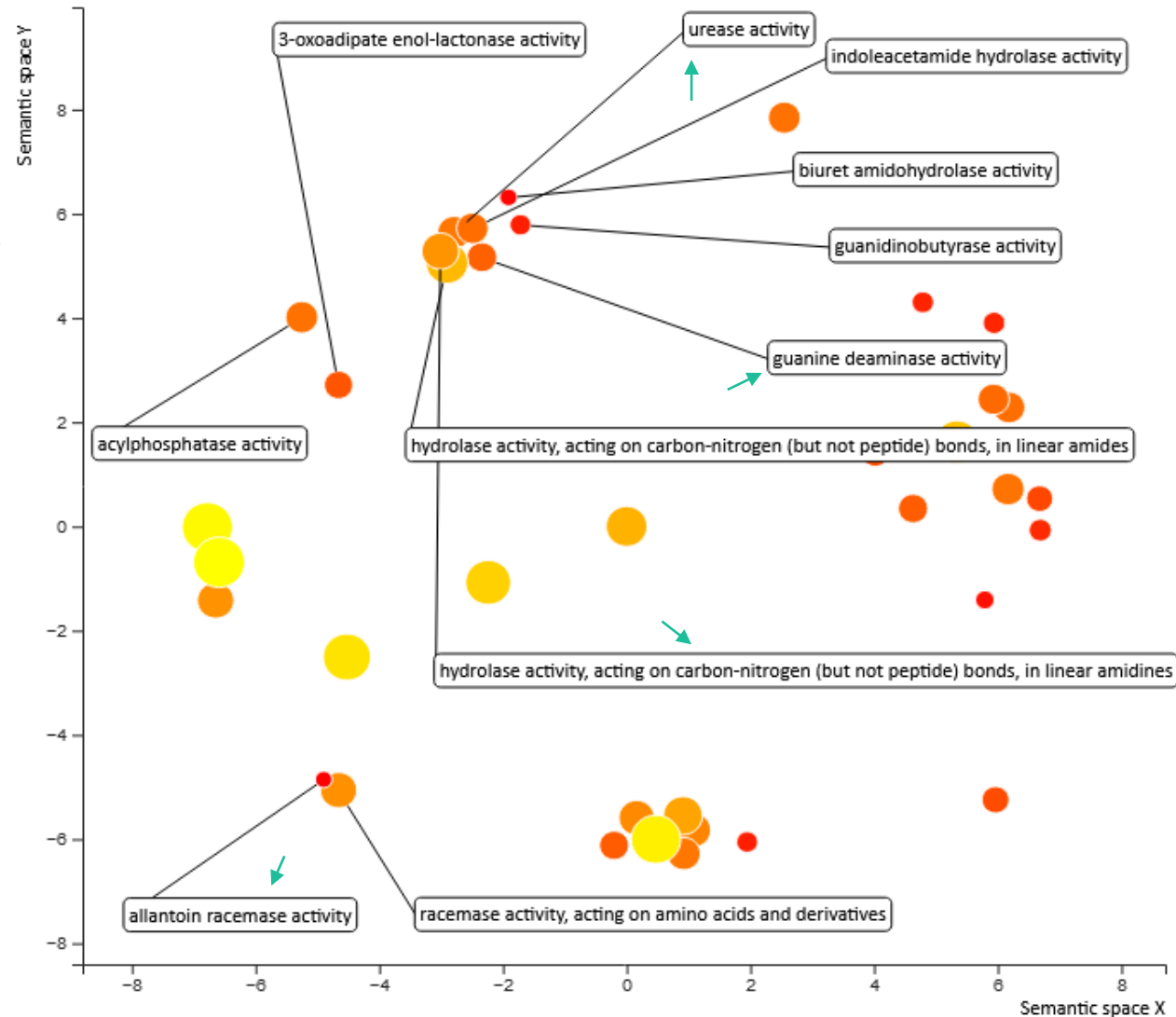


WT-LowN vs WT-HighN (up regulated genes)

MF



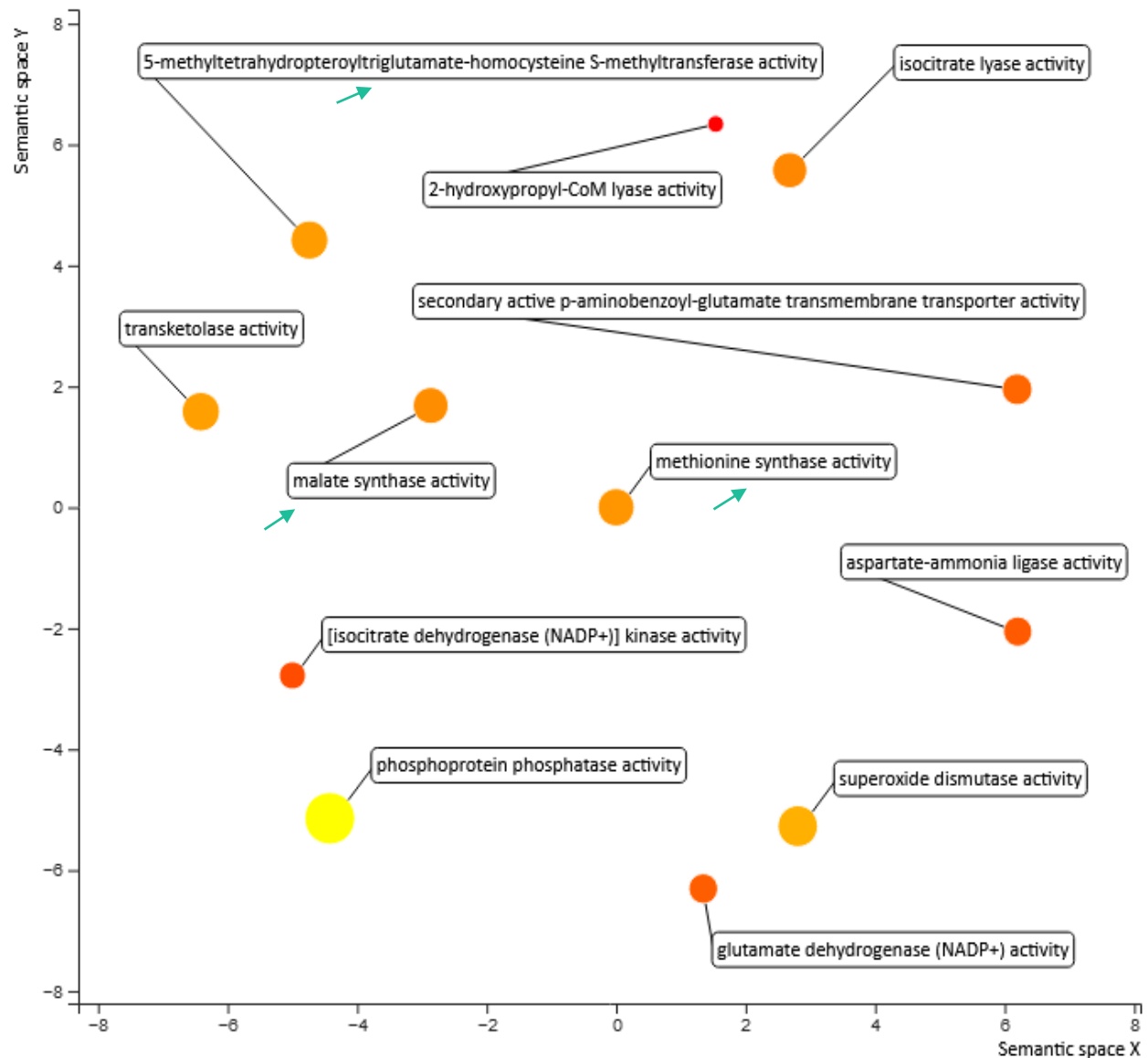
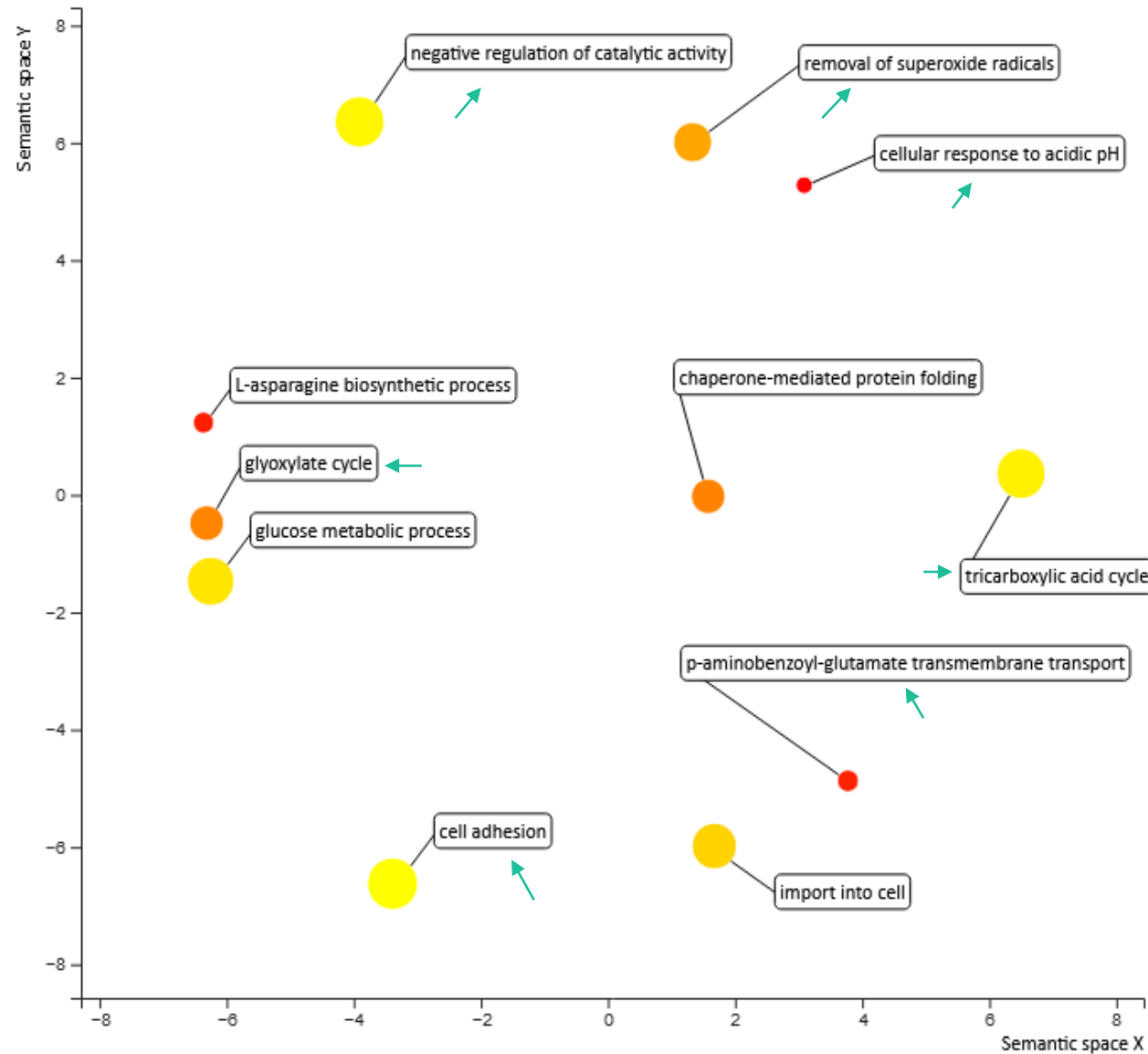
MF



WT-LowN vs WT-HighN (down regulated genes)

CC: Pilus
BP

MF

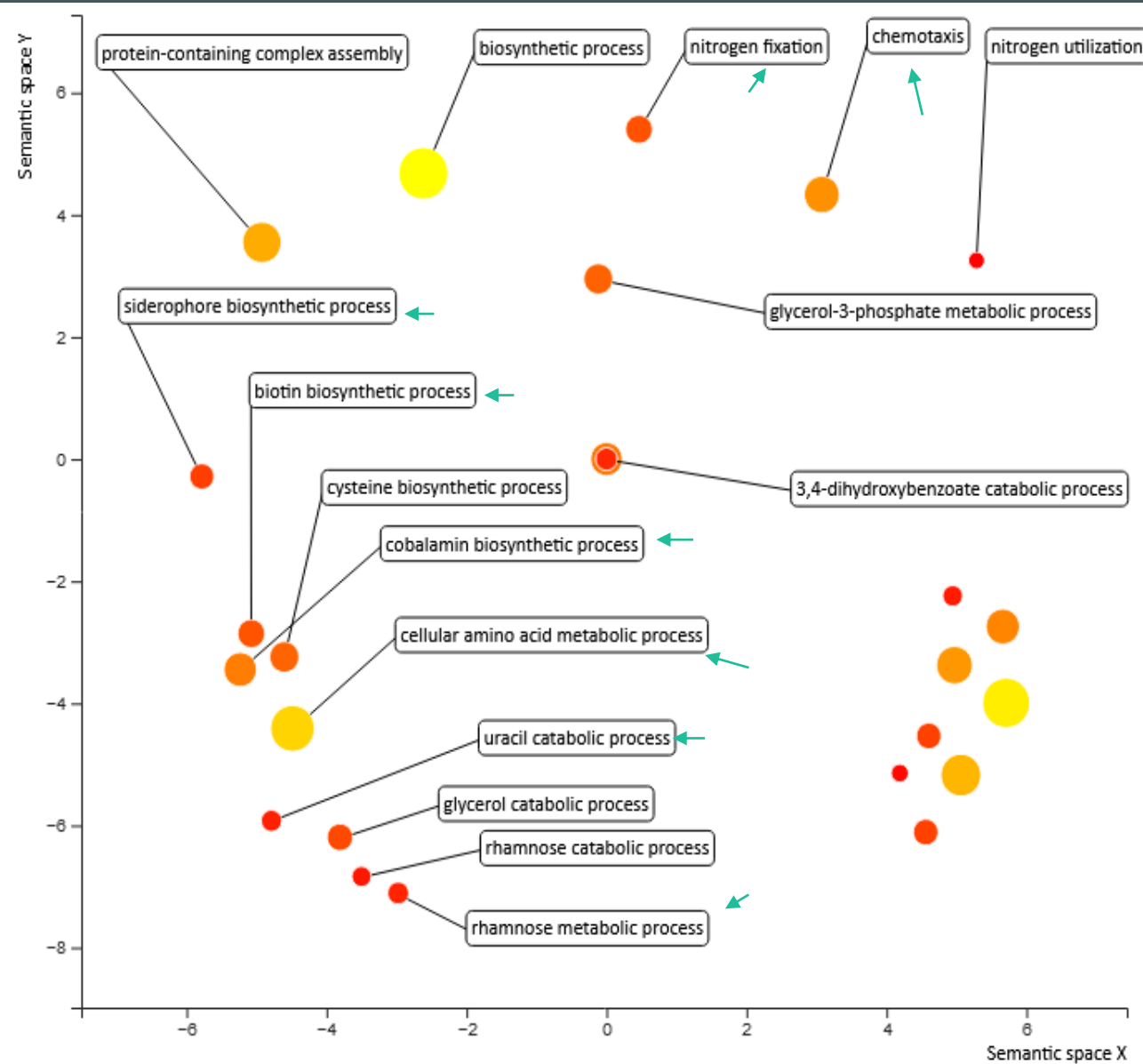
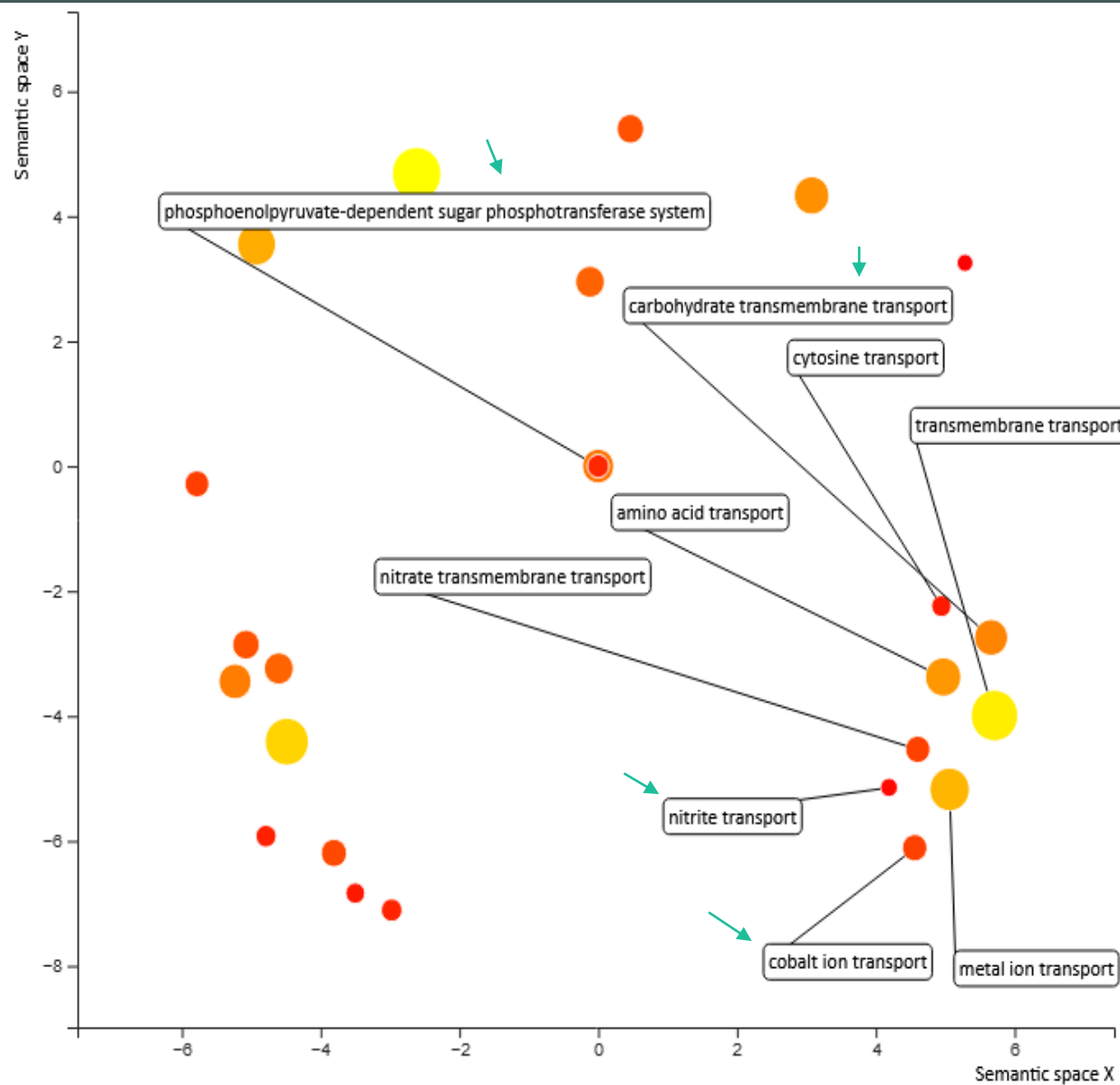


Nif-LowN vs WT-LowN (all genes)

CC: glycerol-3-phosphate dehydrogenase complex

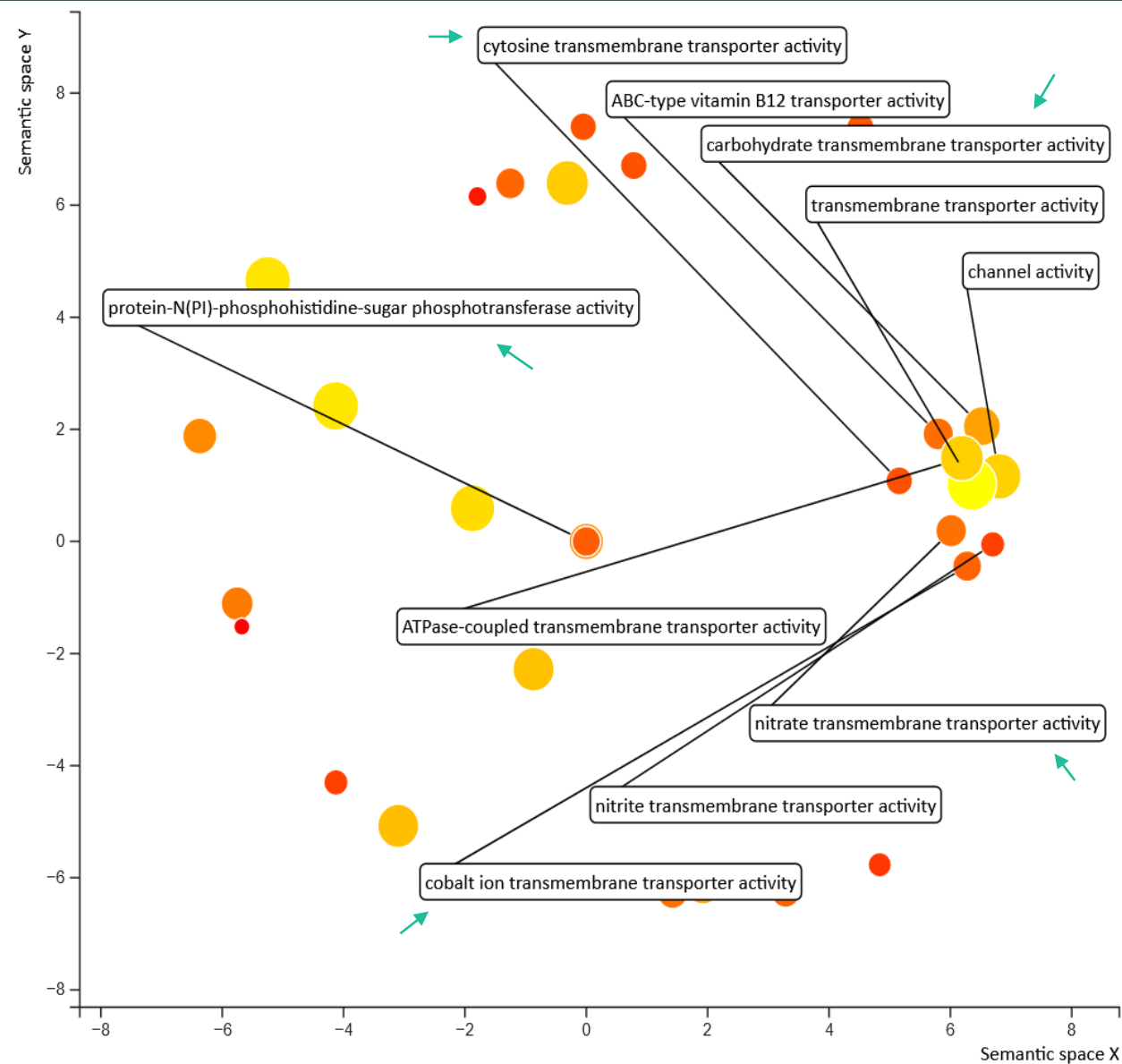
BP

BP

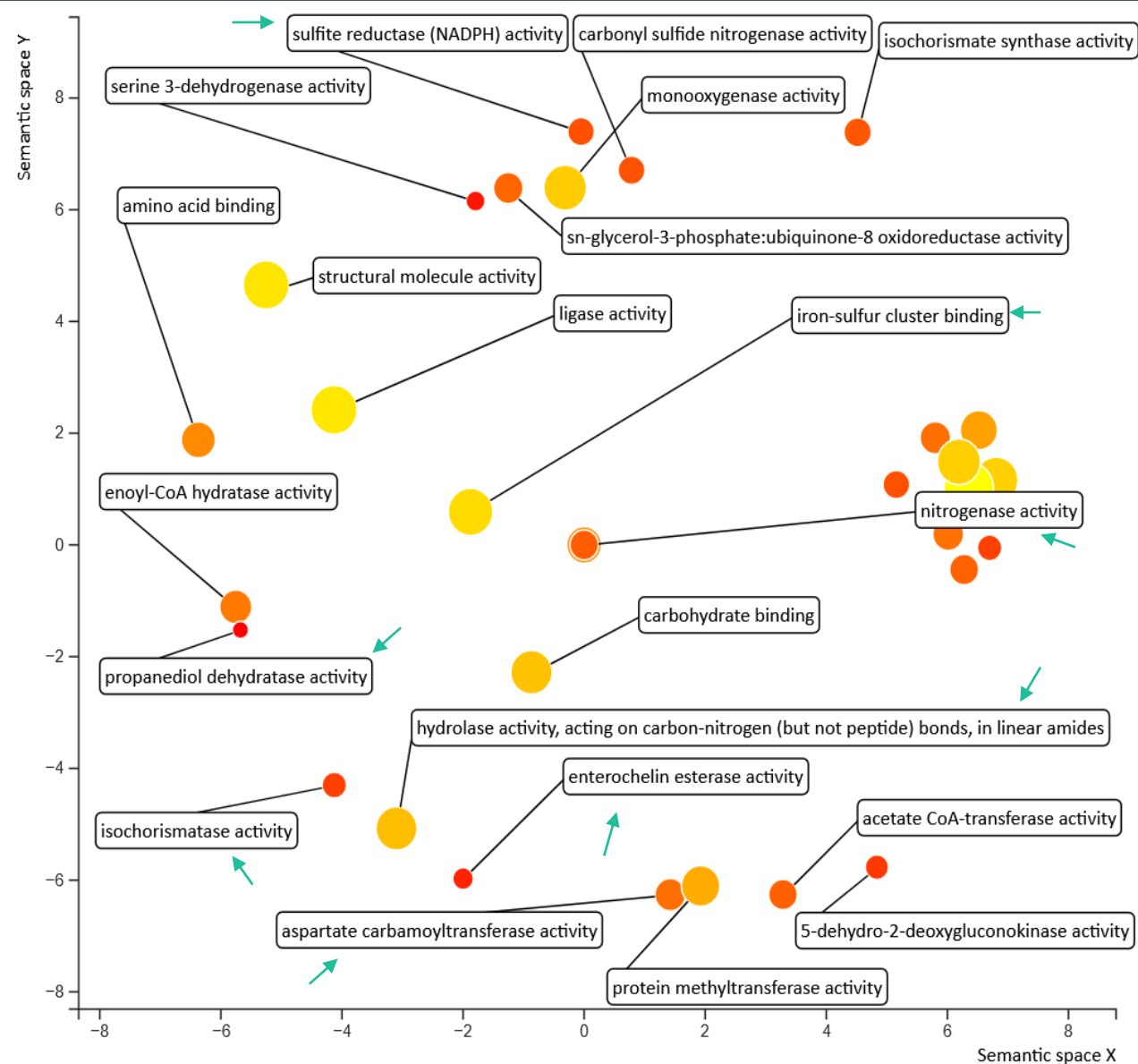


Nif-LowN vs WT-LowN (all genes)

MF

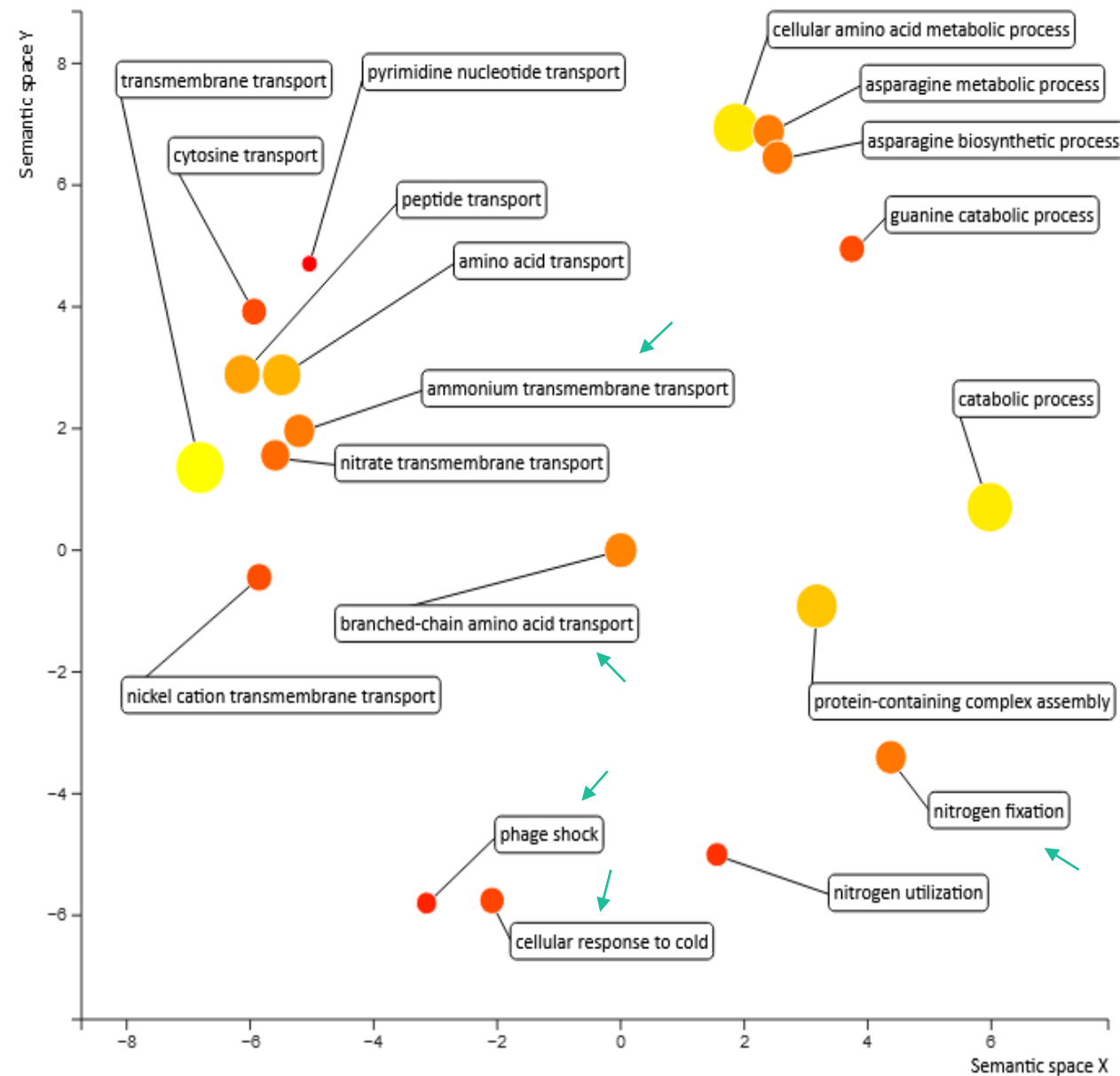


MF

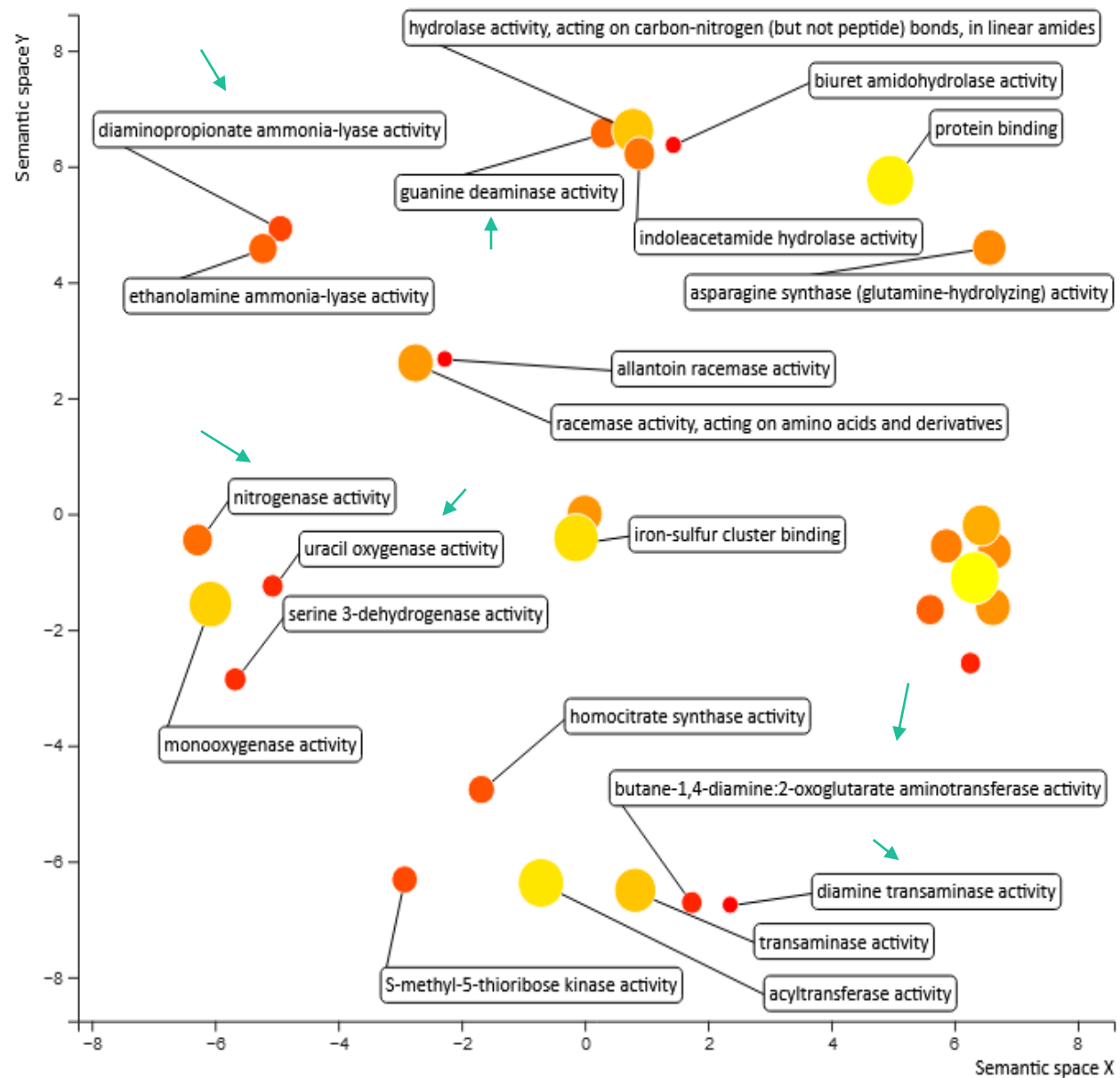


Nif-LowN vs WT-LowN (up regulated genes)

BP

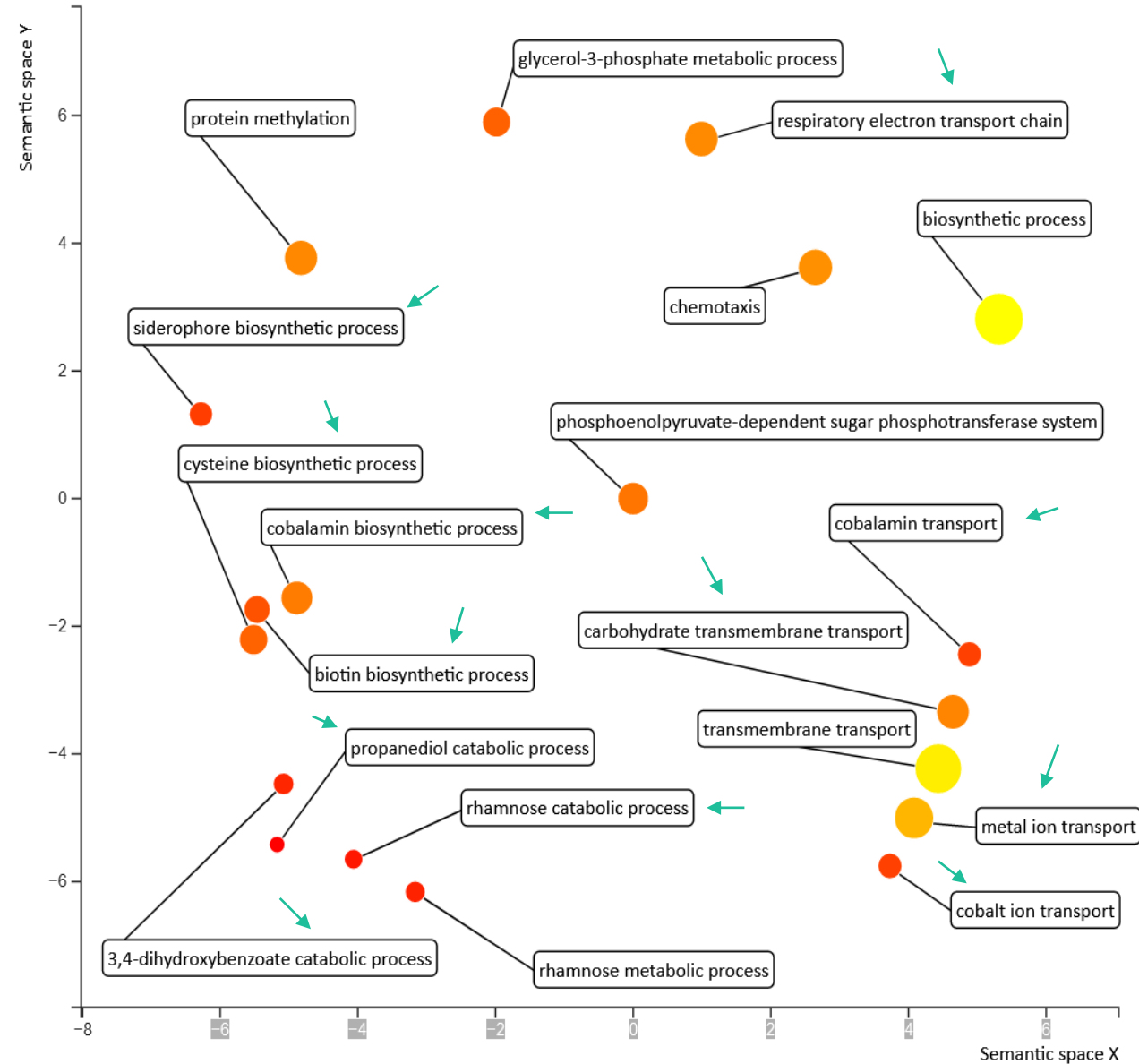


MF

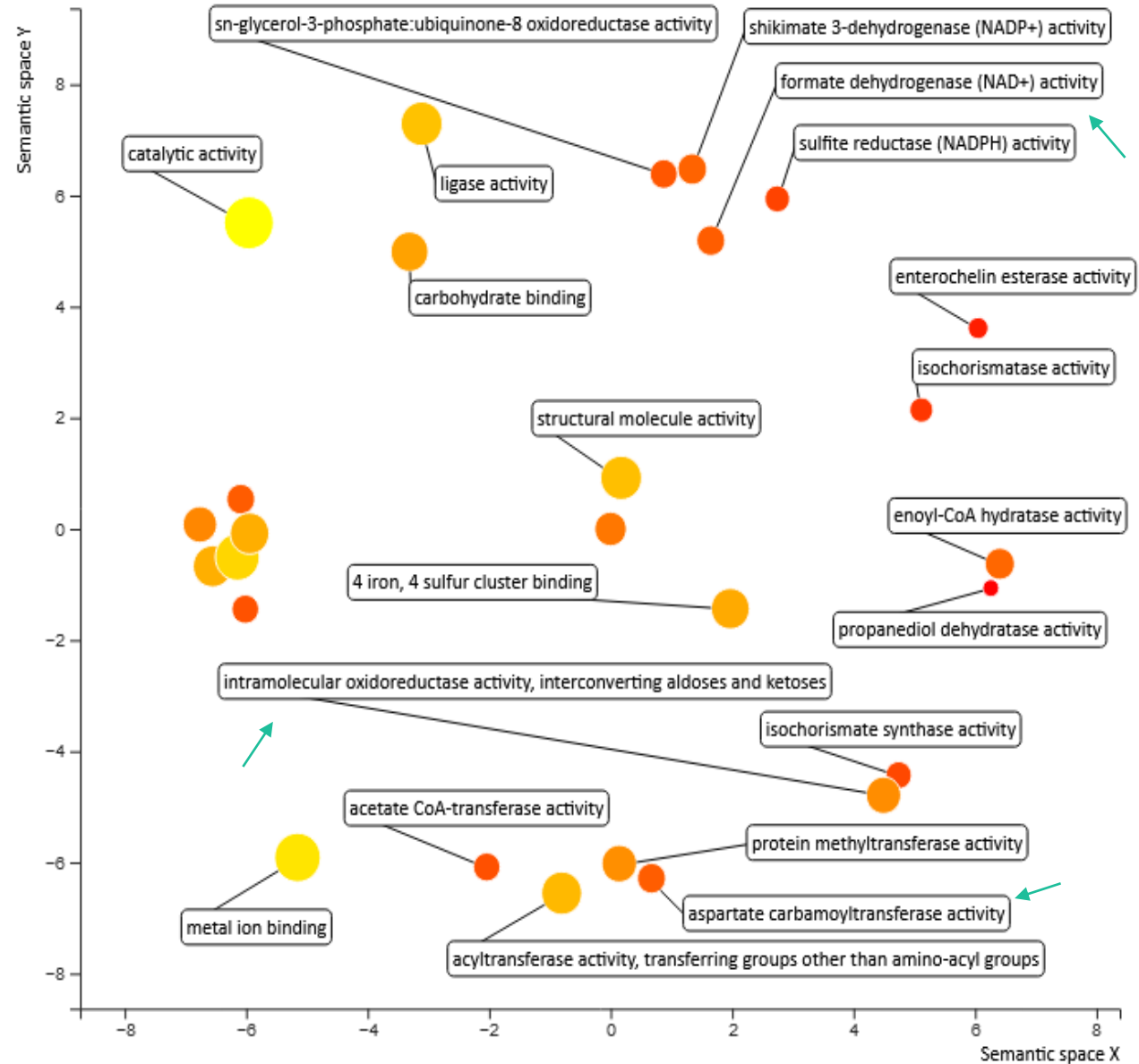


Nif-LowN vs WT-LowN (down regulated genes)

BP



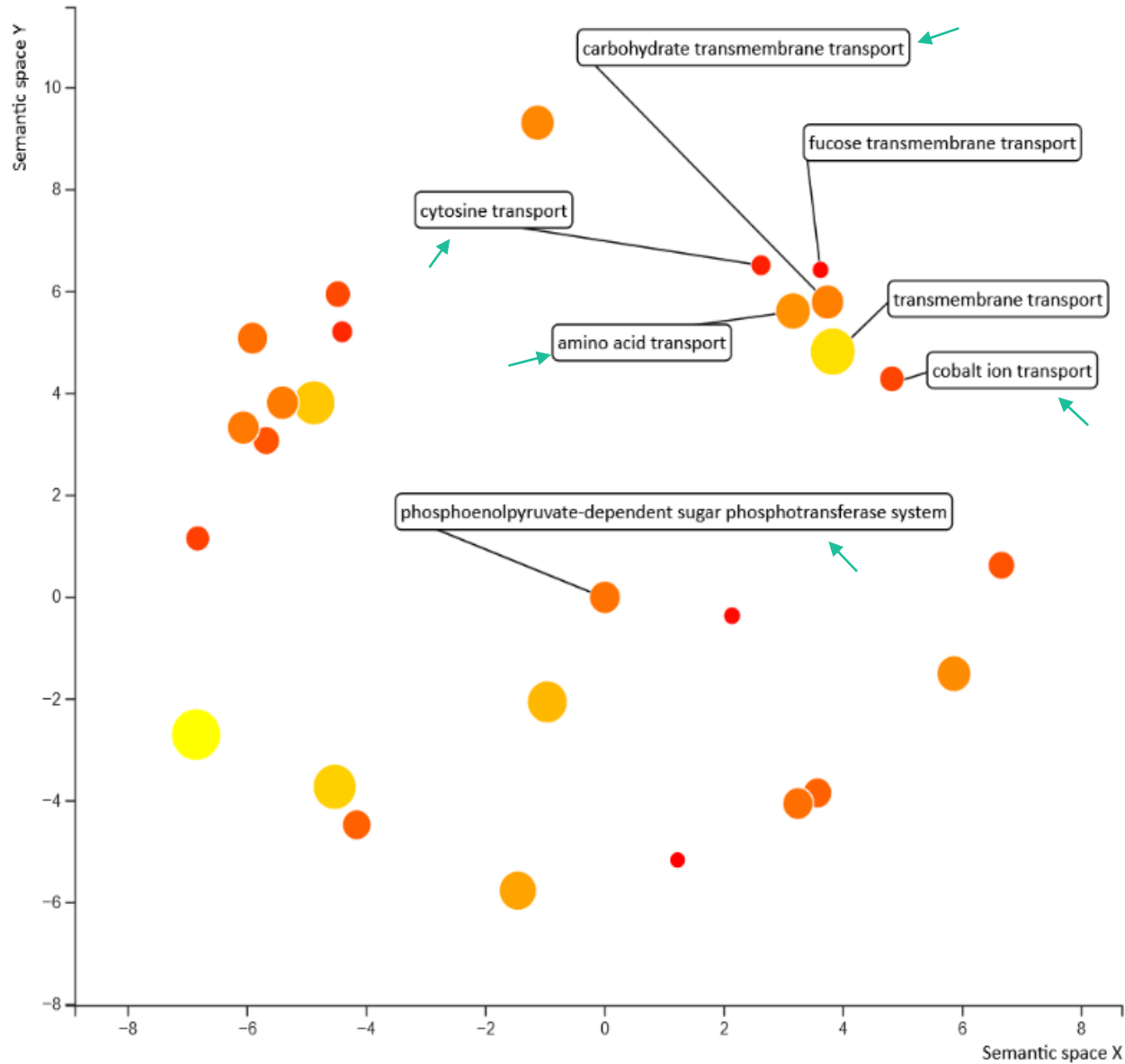
MF



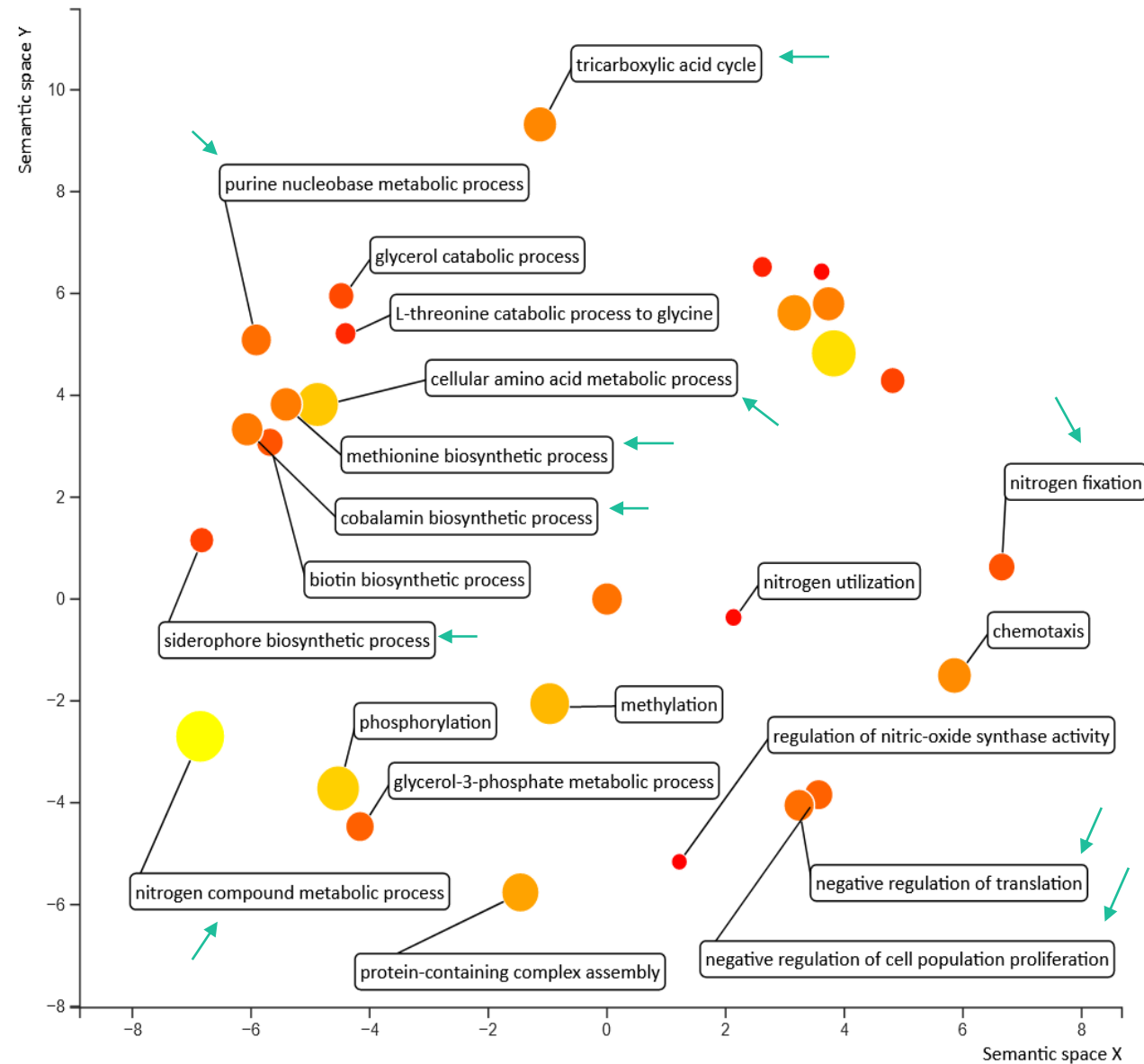
Nif-LowN vs WT-HighN (all genes)

CC: molybdenum-iron nitrogenase complex

BP

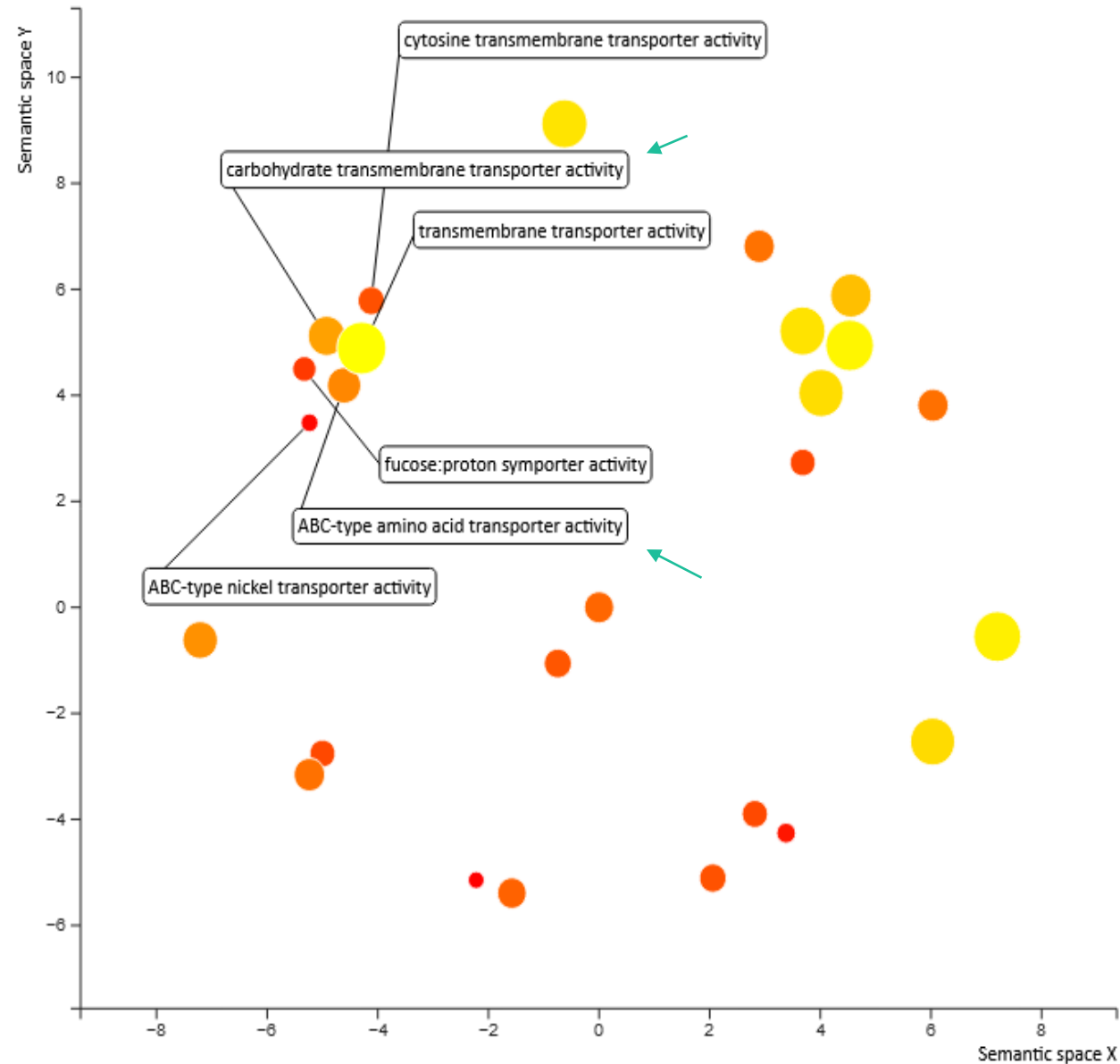


BP

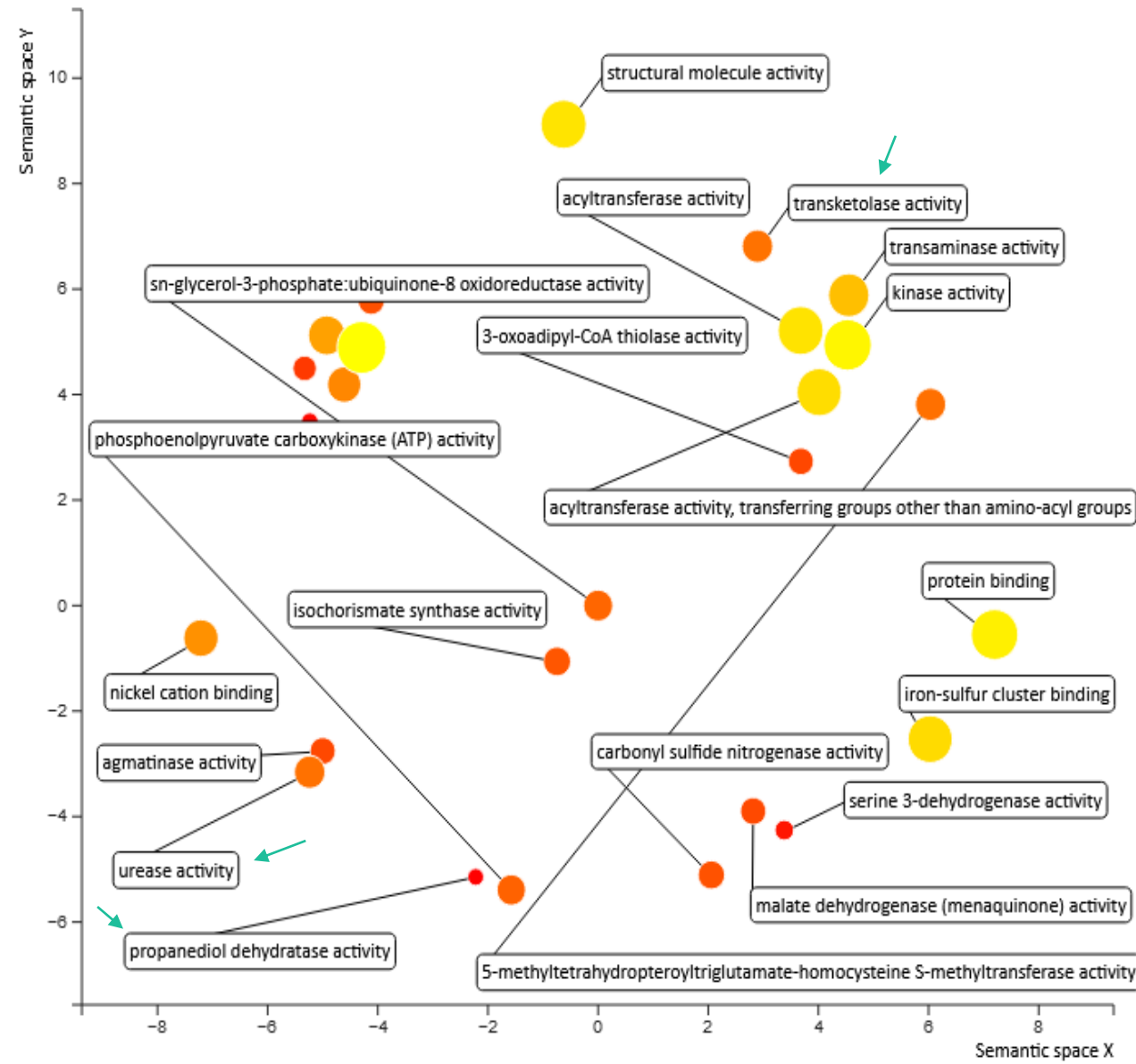


Nif-LowN vs WT-HighN (all genes)

MF



MF

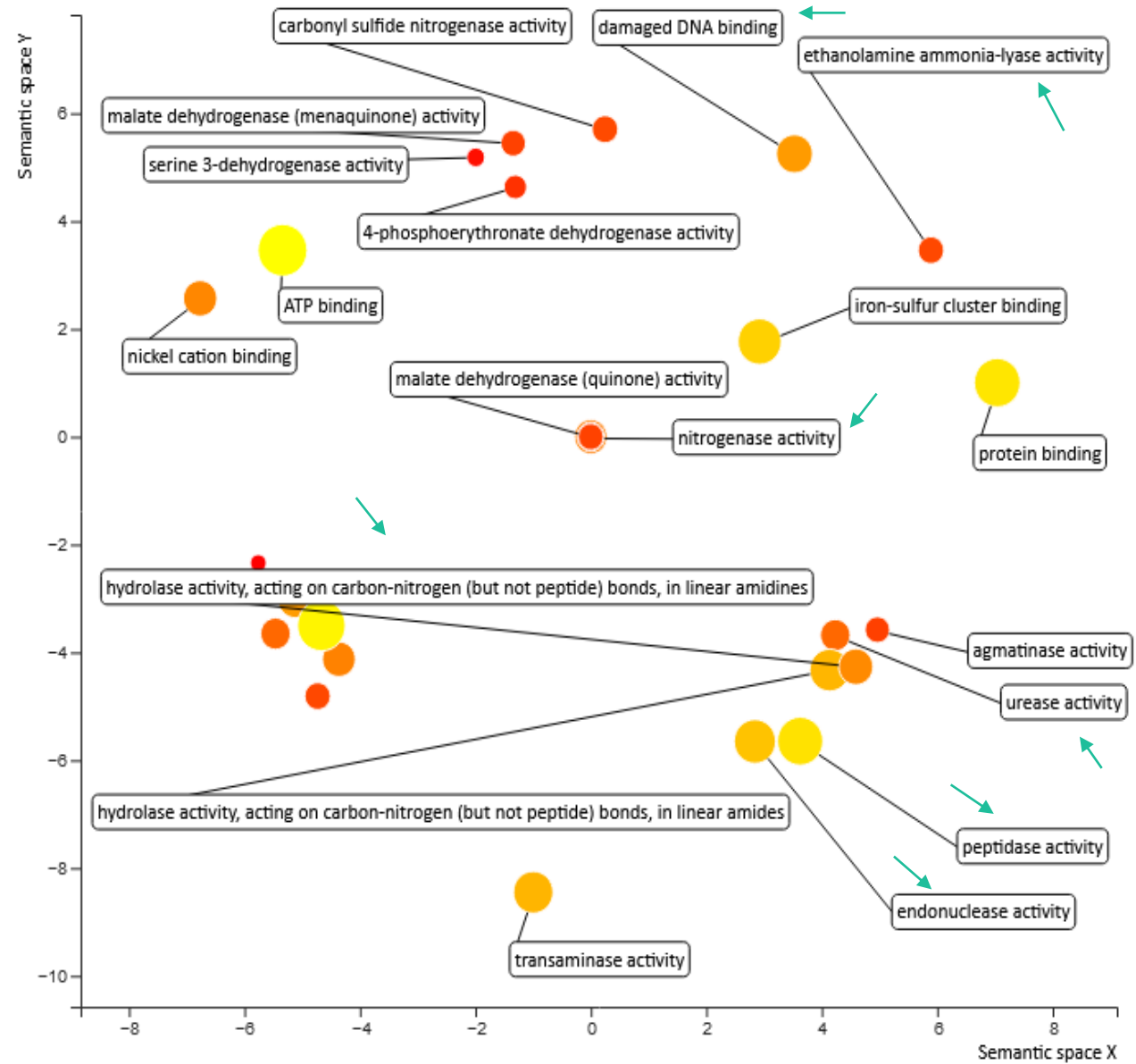
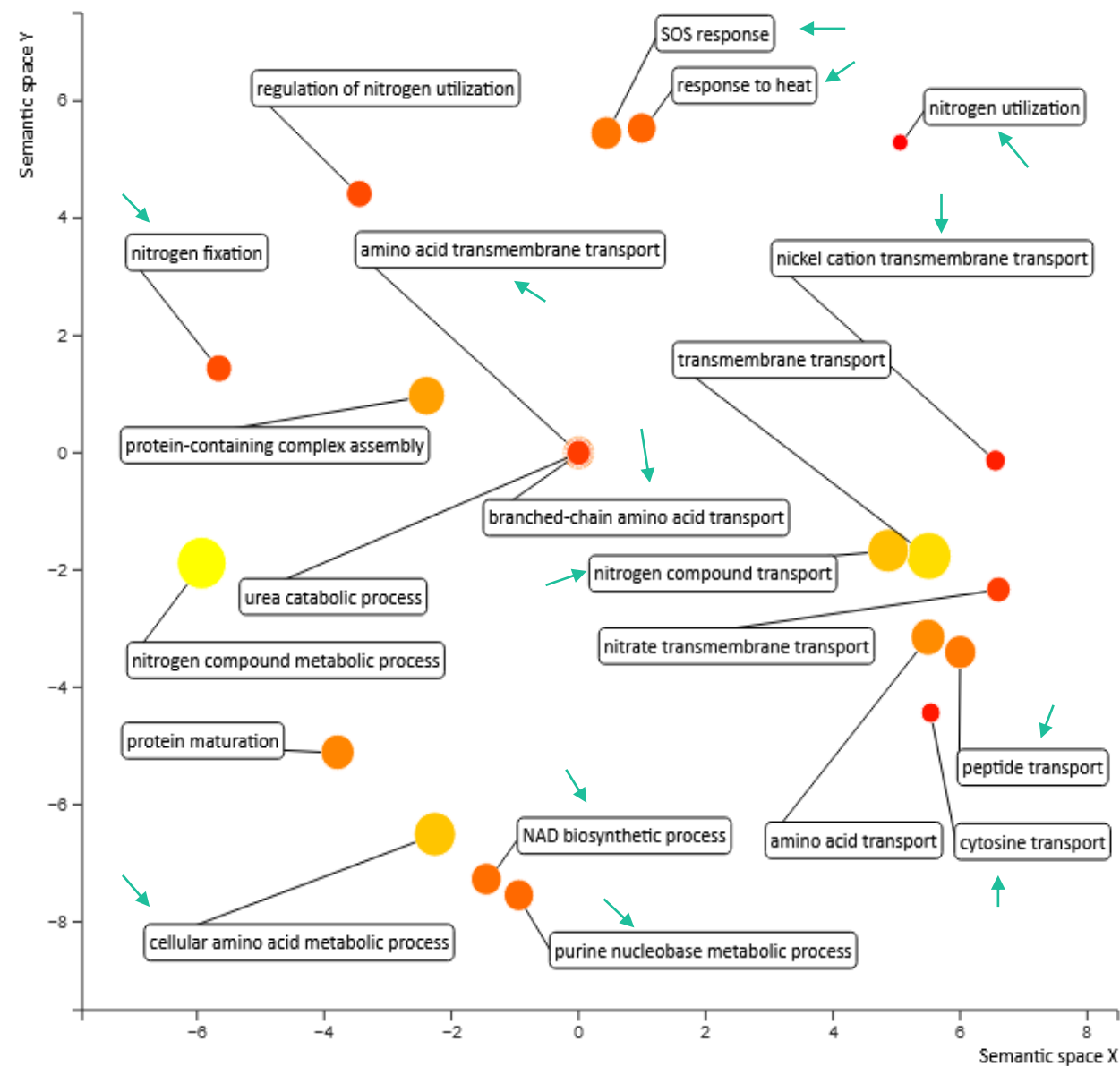


Nif-LowN vs WT-HighN (up regulated genes)

CC: molybdenum-iron nitrogenase complex

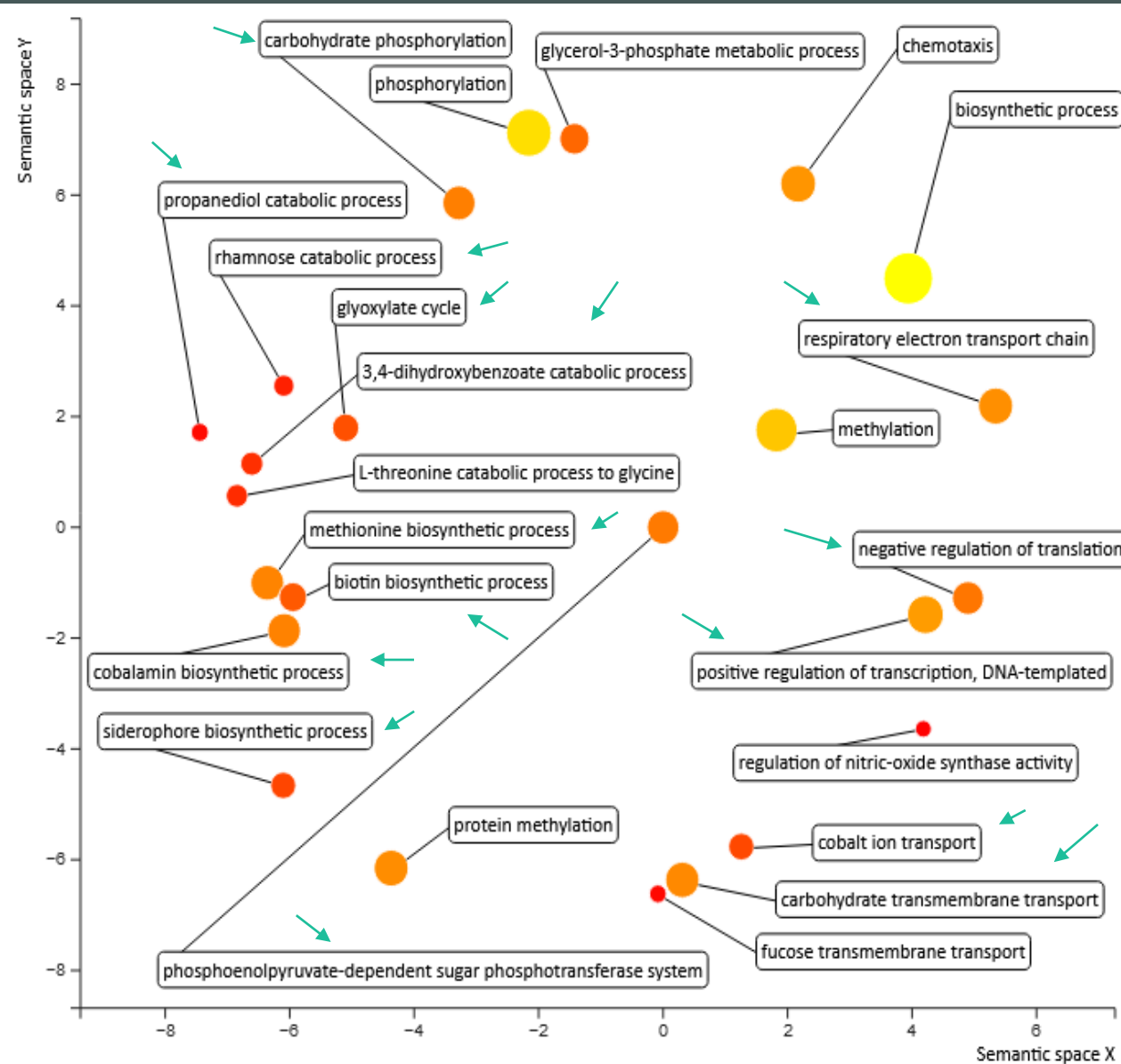
BP

MF



Nif-LowN vs WT-HighN (down regulated genes)

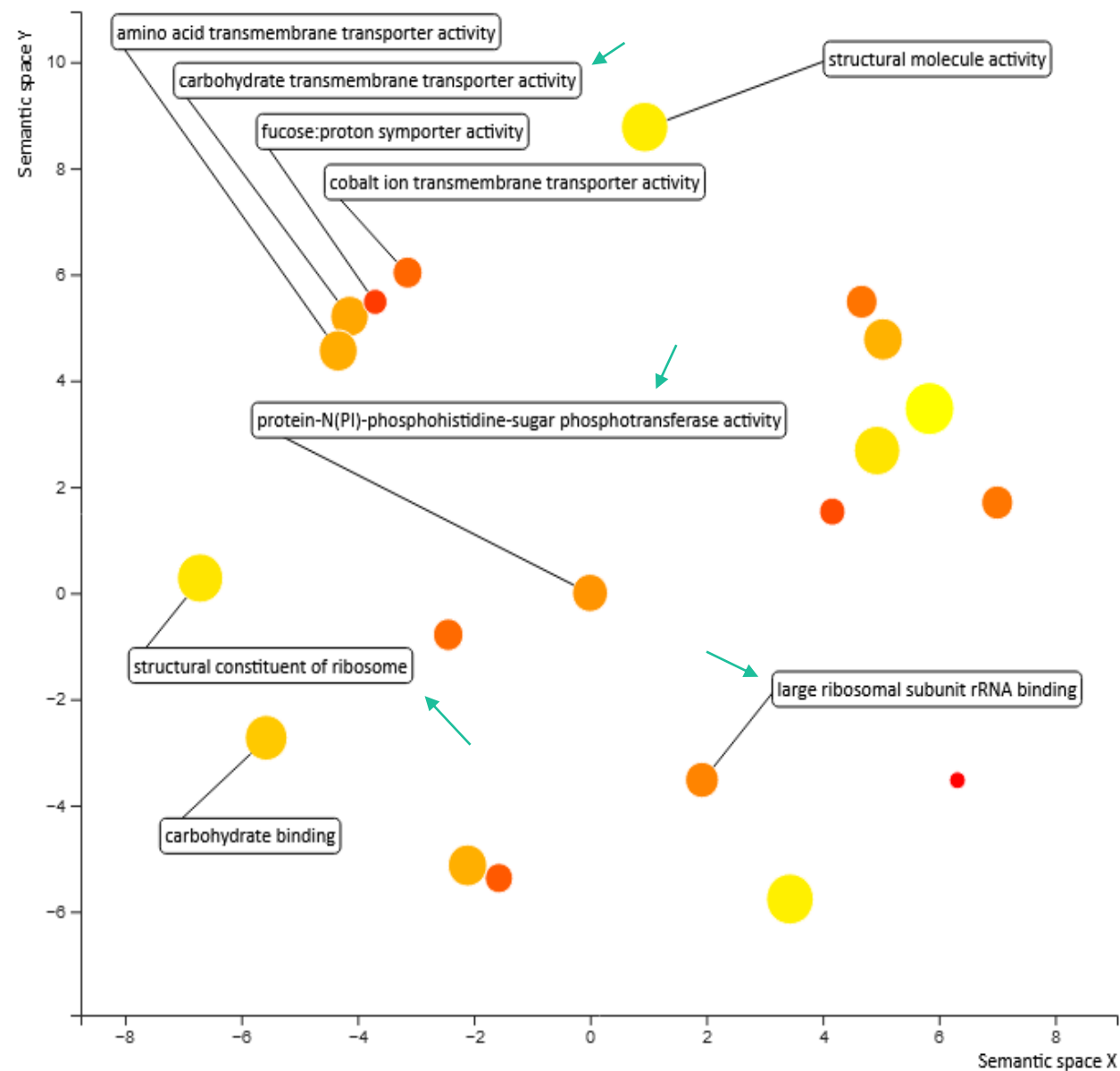
BP



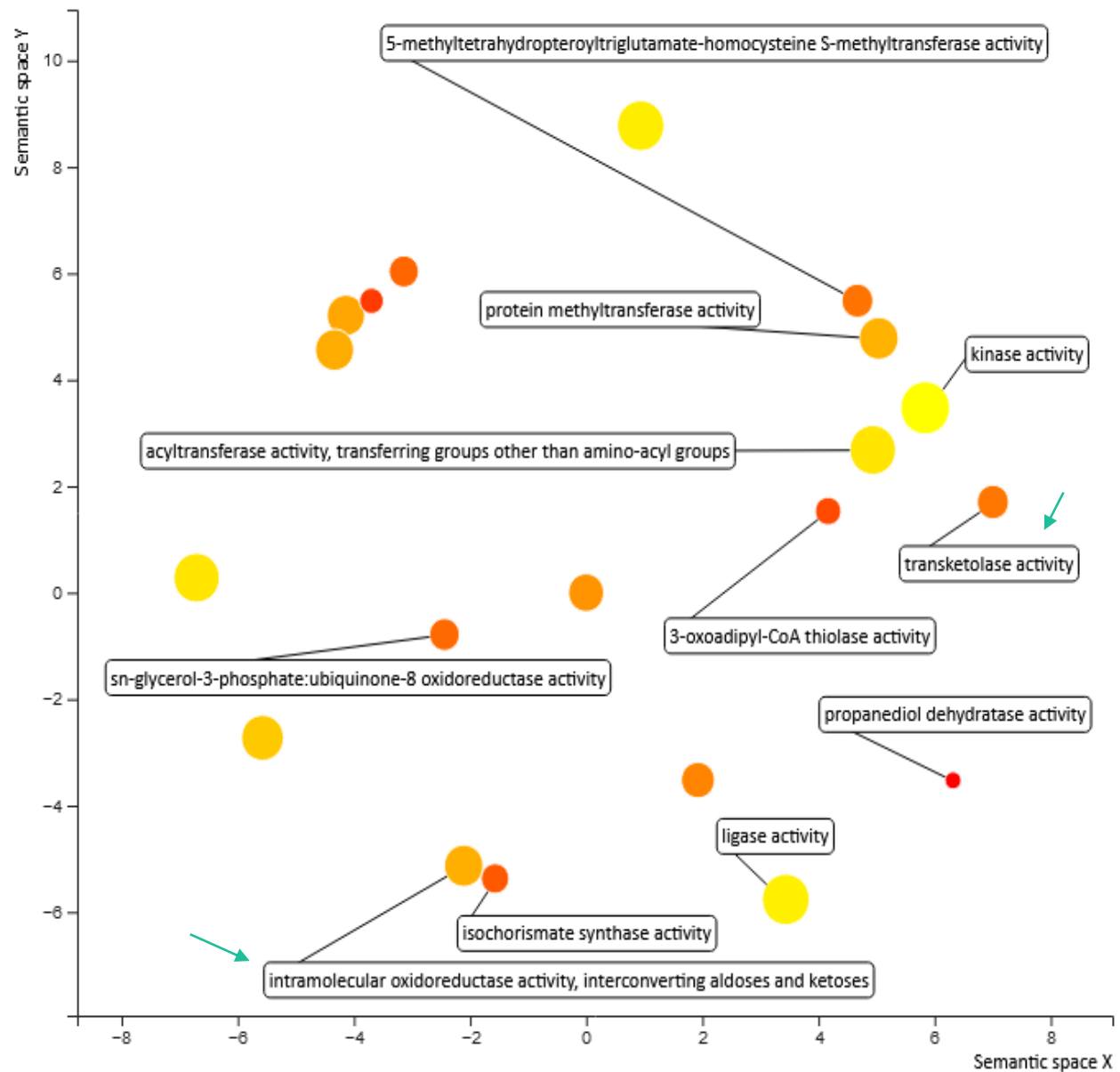
CC: cytosolic small ribosomal subunit, cytosolic large ribosomal subunit, bacterial microcompartment, propanediol degradation polyhedral organelle, and pore complex

Nif-LowN vs WT-HighN (down regulated genes)

MF



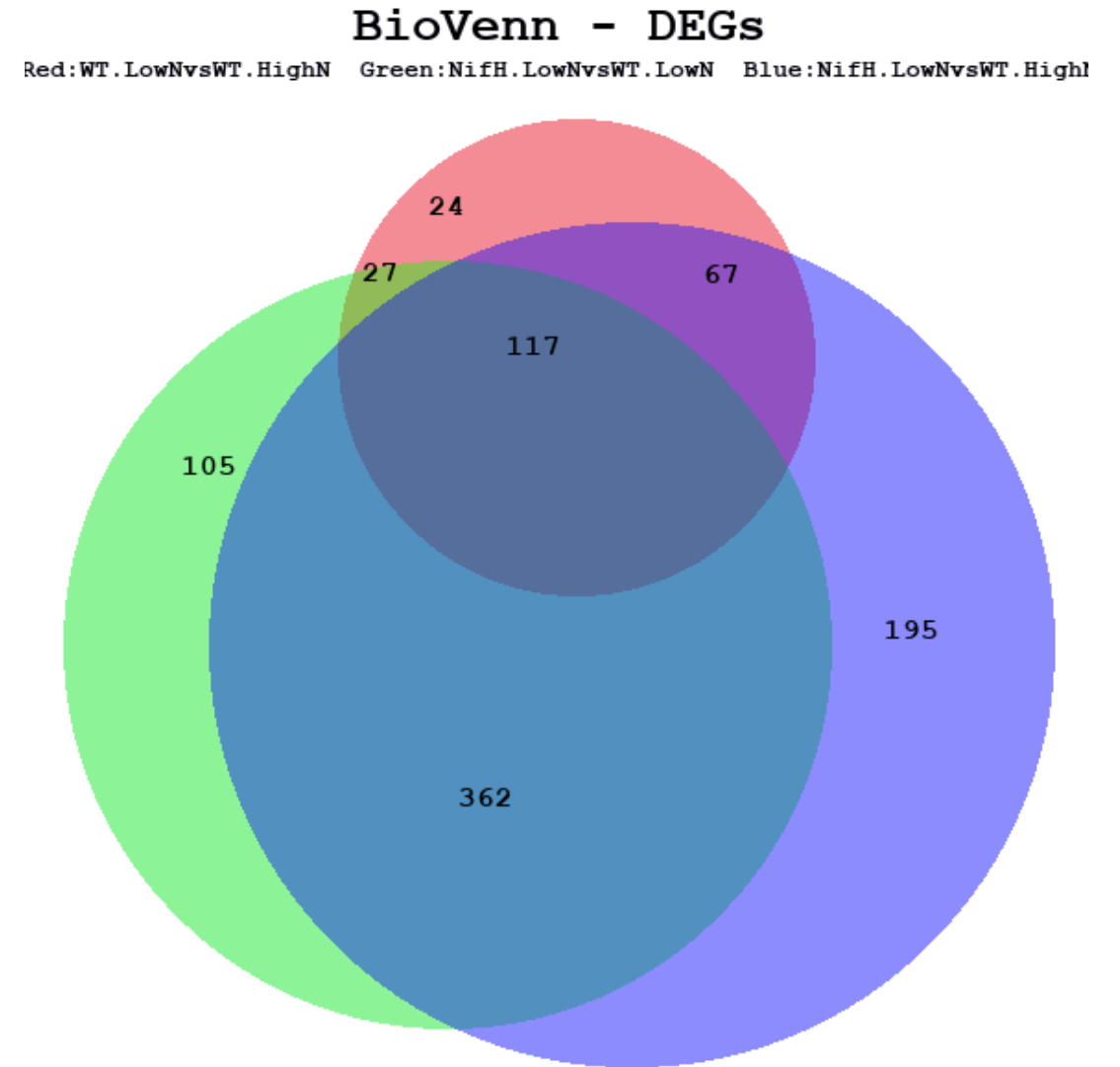
MF



Discussion

- Nitrogen fixation : costly in resources and space
- Diazotrophic conditions → affect large part of the genome
- Which genes and processes are activated/deactivated in diazotrophic conditions
- How *K. oxytoca* can withstand the metabolic stress

- Global gene expression changes based on nitrogen condition
- Expression of nitrogen related genes based on cytosolic levels of nitrogen
- Increase of DEGs as the comparisons become more extreme



Differential gene expression

- **WT-LowN vs WT-HighN**: ↑ nitrogen compensation, BCAAs transport, transcription regulation ↓ carbon metabolism
- **NifH-LowN vs WT-LowN**: ↑ nitrogen compensation, transcription regulation, **growth**, BCAAs transport, **stress** ↓ carbon metabolism, **PTS system**, stress, growth, transcription regulation
- **NifH-LowN vs WT-HighN**: ↑ nitrogen compensation, BCAAs transport, stress, **growth**, catabolism ↓ carbon metabolism, PTS system, transcription regulation, **stress**, growth

Enrichment of GO terms

- **WT-LowN vs WT-HighN**: ↑ nitrogen compensation, BCAAs transport, pyruvate-flavodoxin oxidoreductase ↓ carbon metabolism, translation, **oxidoreductase activity**
- **NifH-LowN vs WT-LowN**: ↑ nitrogen compensation, BCAAs transport, stress ↓ carbon metabolism, PTS system, growth, oxidoreductase activity
- **NifH-LowN vs WT-HighN**: ↑ nitrogen compensation, BCAAs transport, stress ↓ carbon metabolism, PTS system, **translation**, **transcription**, **ribosomes**, oxidoreductase activity

Nitrogen stress

1. Nitrogen fixation
2. Conservation of energy and reductants
3. Conservation of translational machinery
4. Stress responses
5. Reduction of growth

Potential regulators for adaptation to nitrogen stress

- Survival in hostile conditions -> stress responses
- Reduction of metabolism to the minimum
- Conservation of energy for surviving as long as possible
- *K. oxytoca* → conservation of energy and proteomic space for nitrogenase
- BCCAs levels regulate bacterial responses for adaptation to hostile environments (Kaiser & Heinrichs, 2018)
- PTS system regulates the cell's metabolism and adaptation to stress conditions (Galinier & Deutscher, 2017)

Conclusions

- ❑ Global gene changes based on nitrogen conditions
- ❑ Nitrogen compensating mechanisms respond to fixed nitrogen
- ❑ Up regulation of nitrogen compensating mechanism
- ❑ Down regulation of energy consuming mechanism
- ❑ Facilitation of the nitrogen fixation process
- ❑ Correlation between nitrogen stress and cell responses
- ❑ Possible system of regulation: transport of BCAAs and PTS system

Suggestions for future studies

- Confirmation with more biological replicates and more environmental conditions
- Confirmation of gene expression at mRNA level (qPCR) and protein level (mass spectrometry)
- Studies with gene deletions or gene over-expressions
- Computational tools and techniques for identification of candidate genes for future experiments

Thank you for your attention!

Special thanks: Dr Ioly Kotta-Loizou, Matthew Carrey, Dr Zoi Lygerou, Dr Sefernia Mavroudi

$$(6) K_{ij} \sim NB(\text{mean} = \mu_{ij}, \text{dispersion} = a_i)$$

$$(7) \mu_{ij} = s_j q_{ij}$$

$$(8) \log q_{ij} = \sum_r x_{jr} \beta_{ir}$$

i: gene

j: sample

q: quantity of cDNA fragments from a gene in a sample

x: design matrix element

β : coefficient

r: replicates

$$(9) \quad Sa = \lambda a$$

32

S: covariance matrix of the original dataset

a: eigenvectors

λ : eigenvalues

$$(10) \quad r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

r : Pearson correlation coefficient

x_i : x variable values

y_i : y variable values

\bar{x} : mean of x variable values

\bar{y} : mean of y variable values

$$(11) \quad z_i = \frac{x_i - \bar{x}}{s}$$

z_i : z score of a value

x_i : value number

\bar{x} : mean of a set of values

s : standard deviation of a set of values

$$p(x) = \frac{\binom{m}{x} \binom{n}{k-x}}{\binom{m+n}{k}}$$

x : the number of observed successes

m : the number of successes

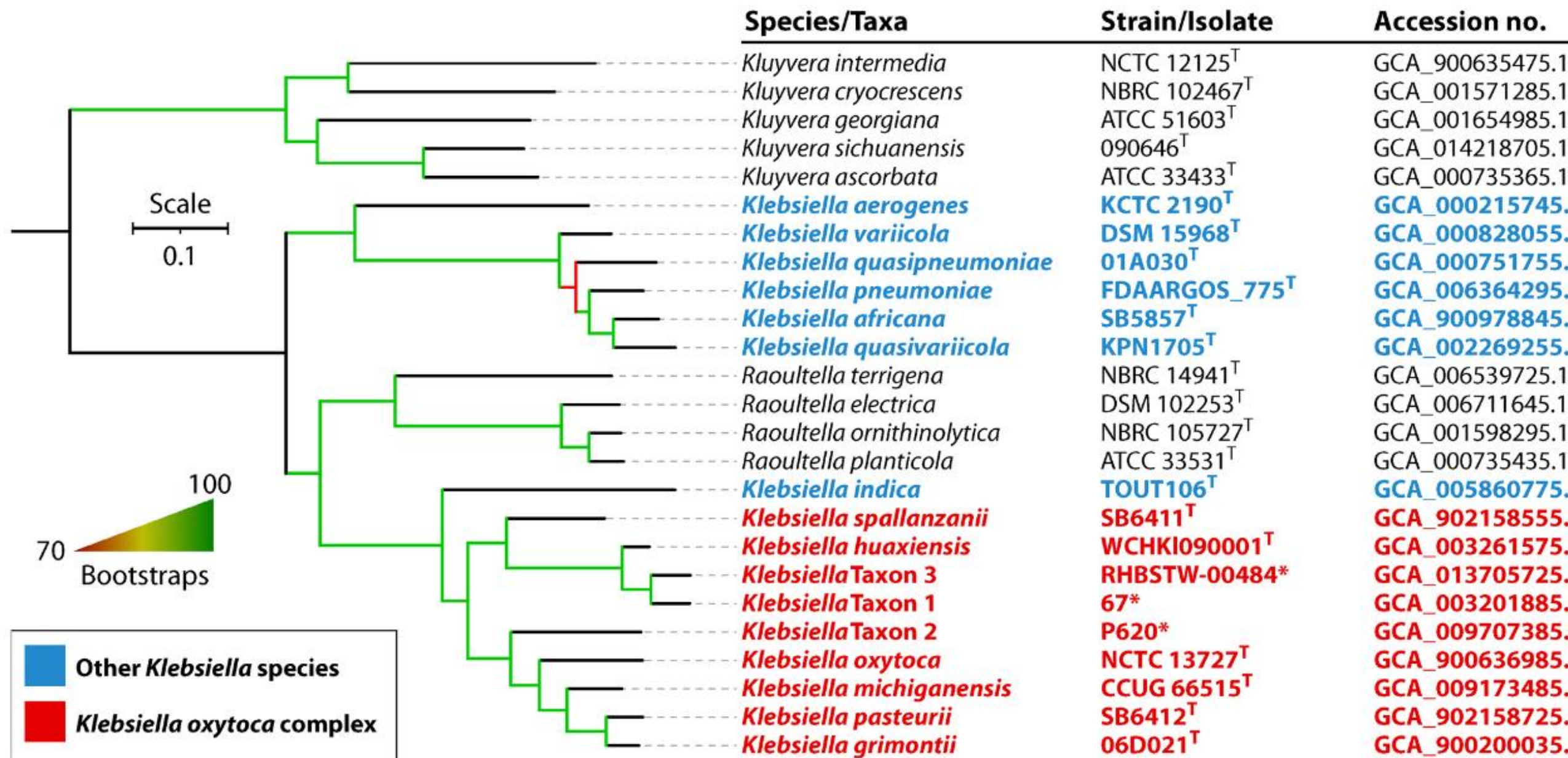
n : the number of failures

k : the number of draws

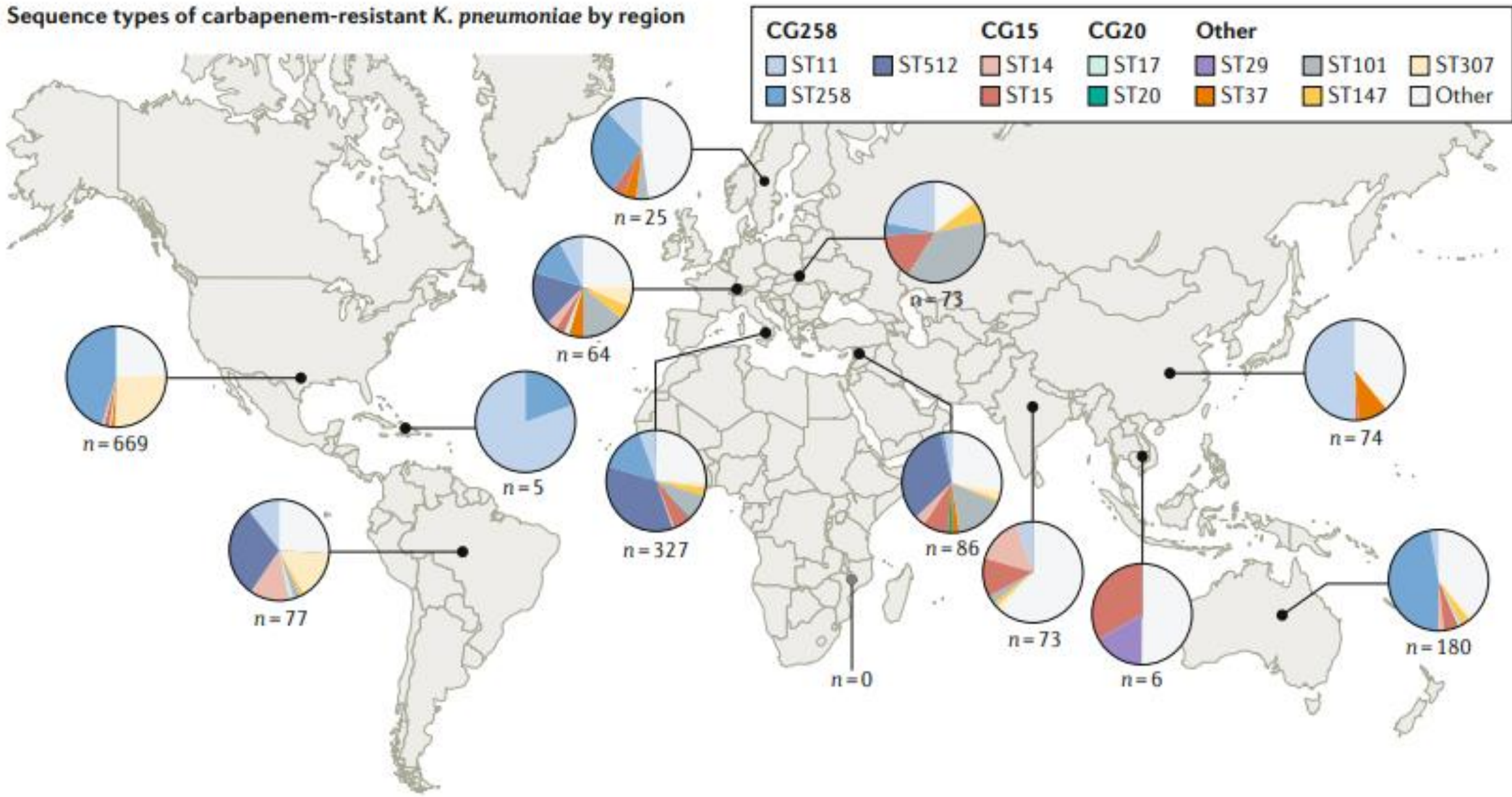
[22]

Modelling stress responses of *Klebsiella* bacteria

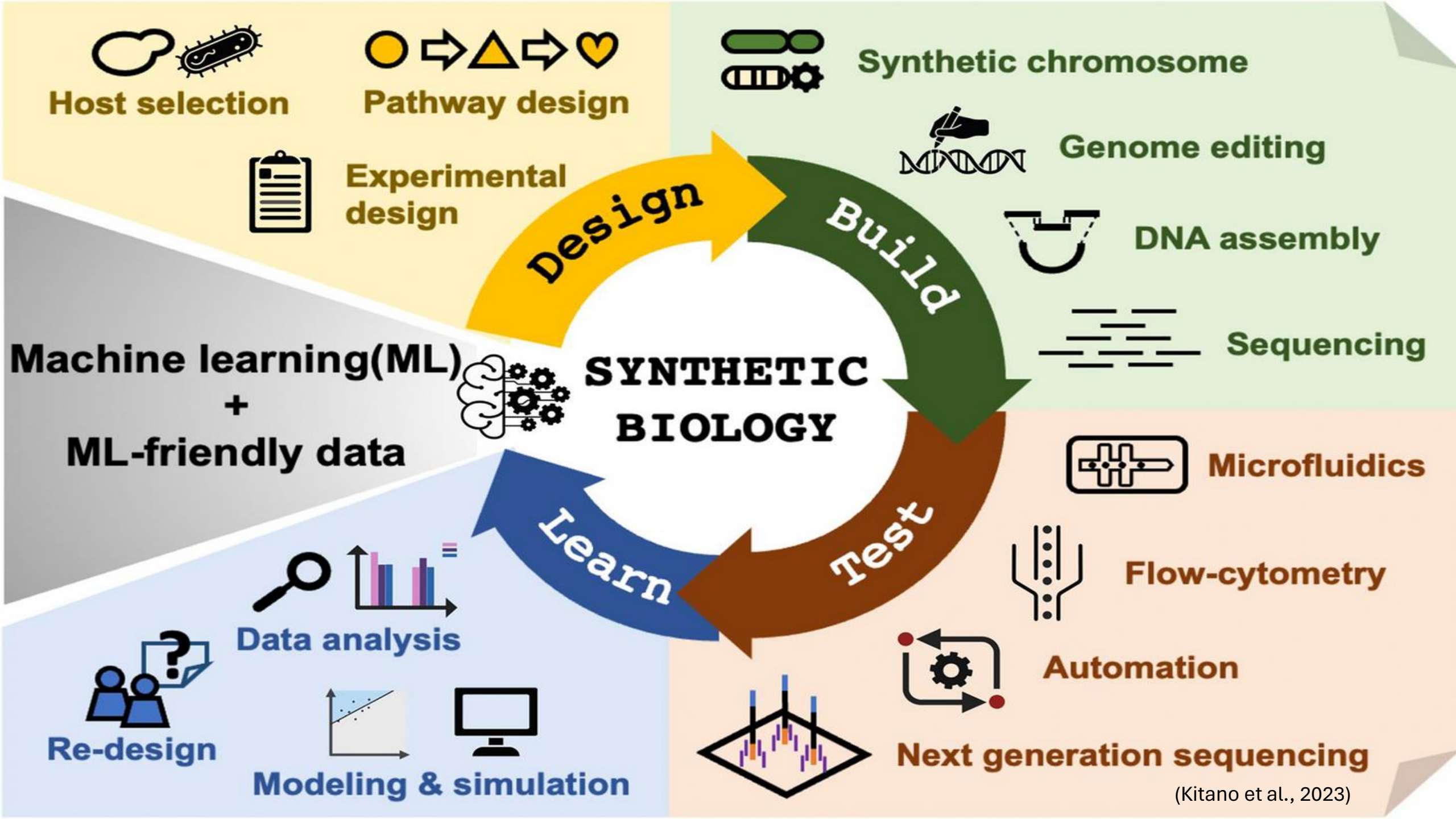
- *Klebsiella* → Biotechnological and medical interest
- Virulence and nitrogen fixation → metabolic stress → bacterial adaptation
- Model of stress responses → bacterial manipulation
- Models → require large amount of data → omics datasets
- PhD summary → data retrieval - DEA - PPINs- Annotation - GEMs

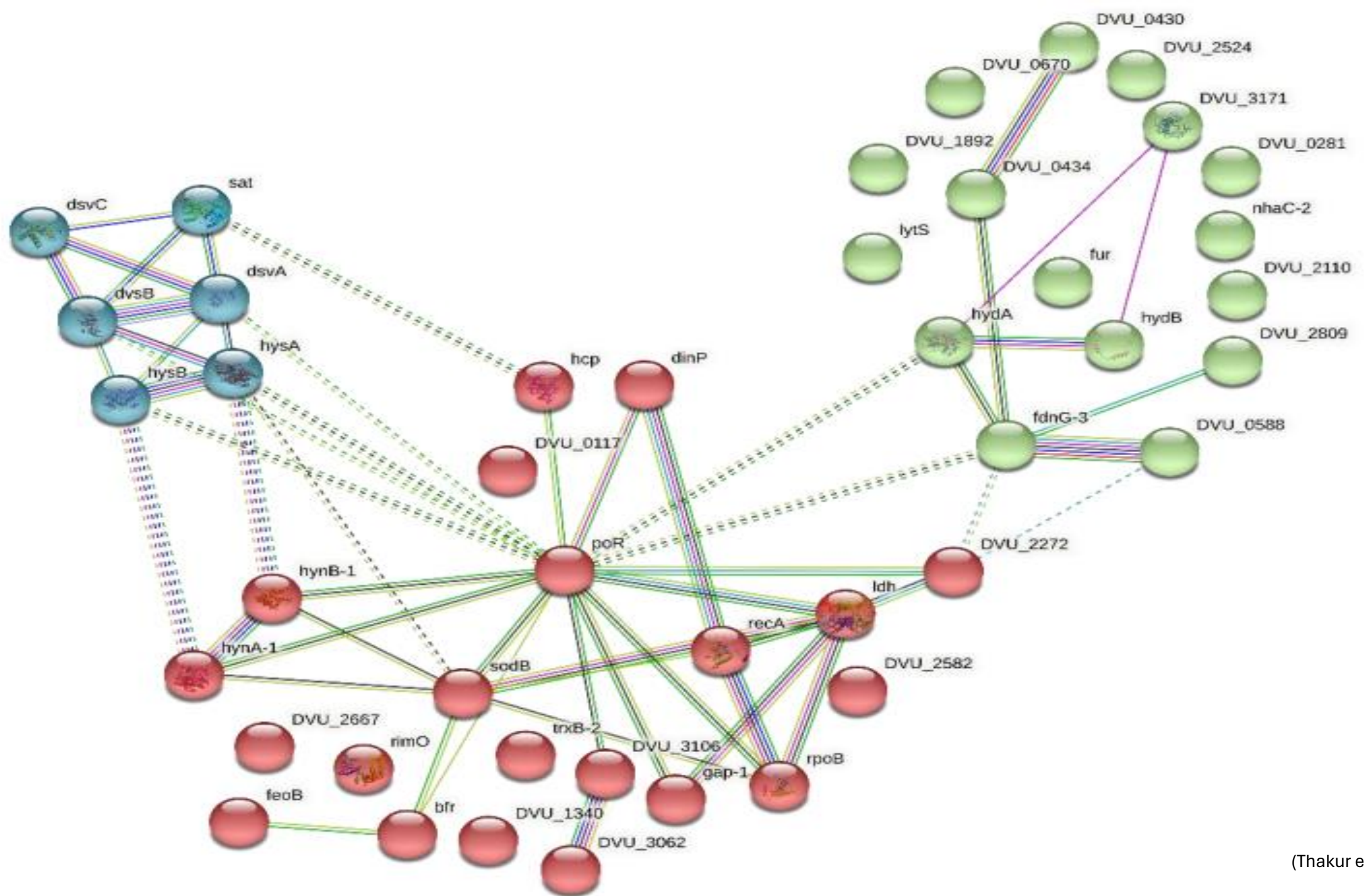


a Sequence types of carbapenem-resistant *K. pneumoniae* by region



(Wyres et al., 2020)





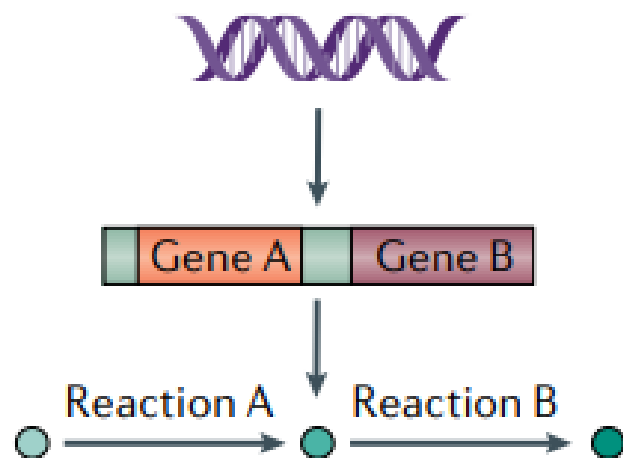
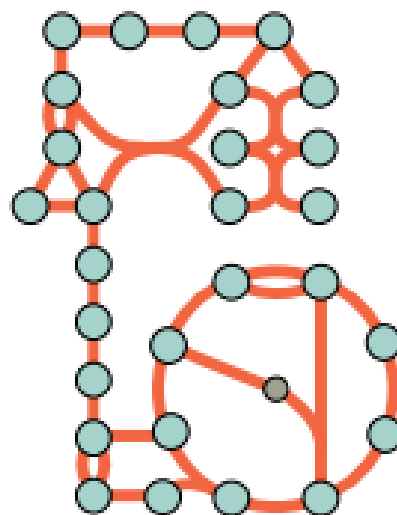
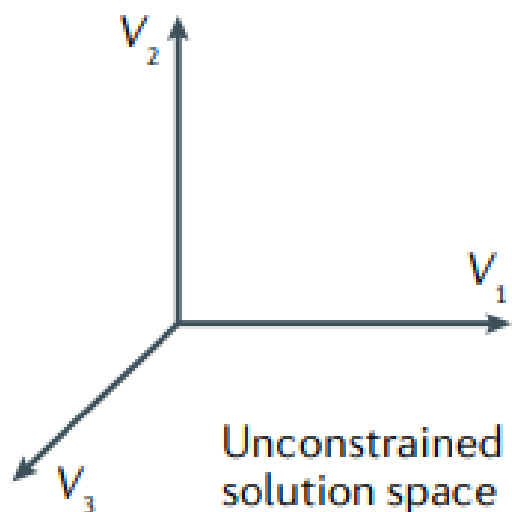
a Annotated genome**b Metabolic network****c S matrix**

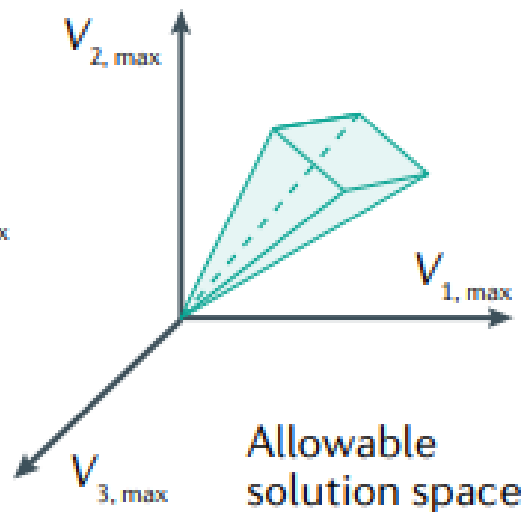
Diagram illustrating the S matrix. The matrix is a 7x7 grid with "Metabolites" on the vertical axis and "Reactions" on the horizontal axis. The matrix is:

0	1	0	0	0	-1	0
0	0	-1	1	0	0	0
0	0	1	0	0	0	1
1	0	0	0	1	0	0
-1	0	0	2	0	0	0
0	-1	0	-1	0	1	0
0	0	-1	0	0	0	-1

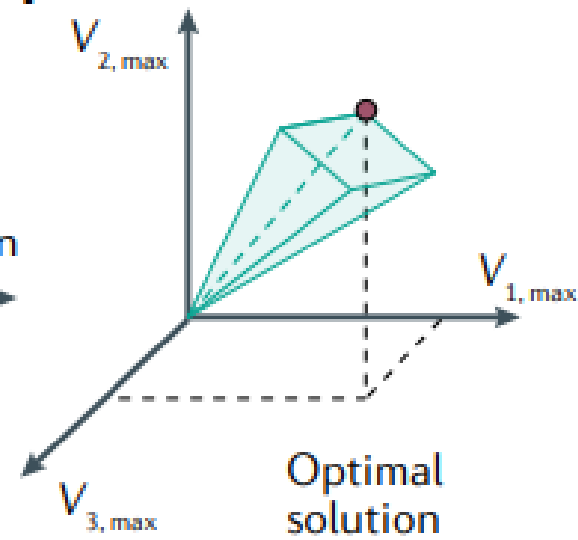
The matrix is labeled "Solve" on the right side.

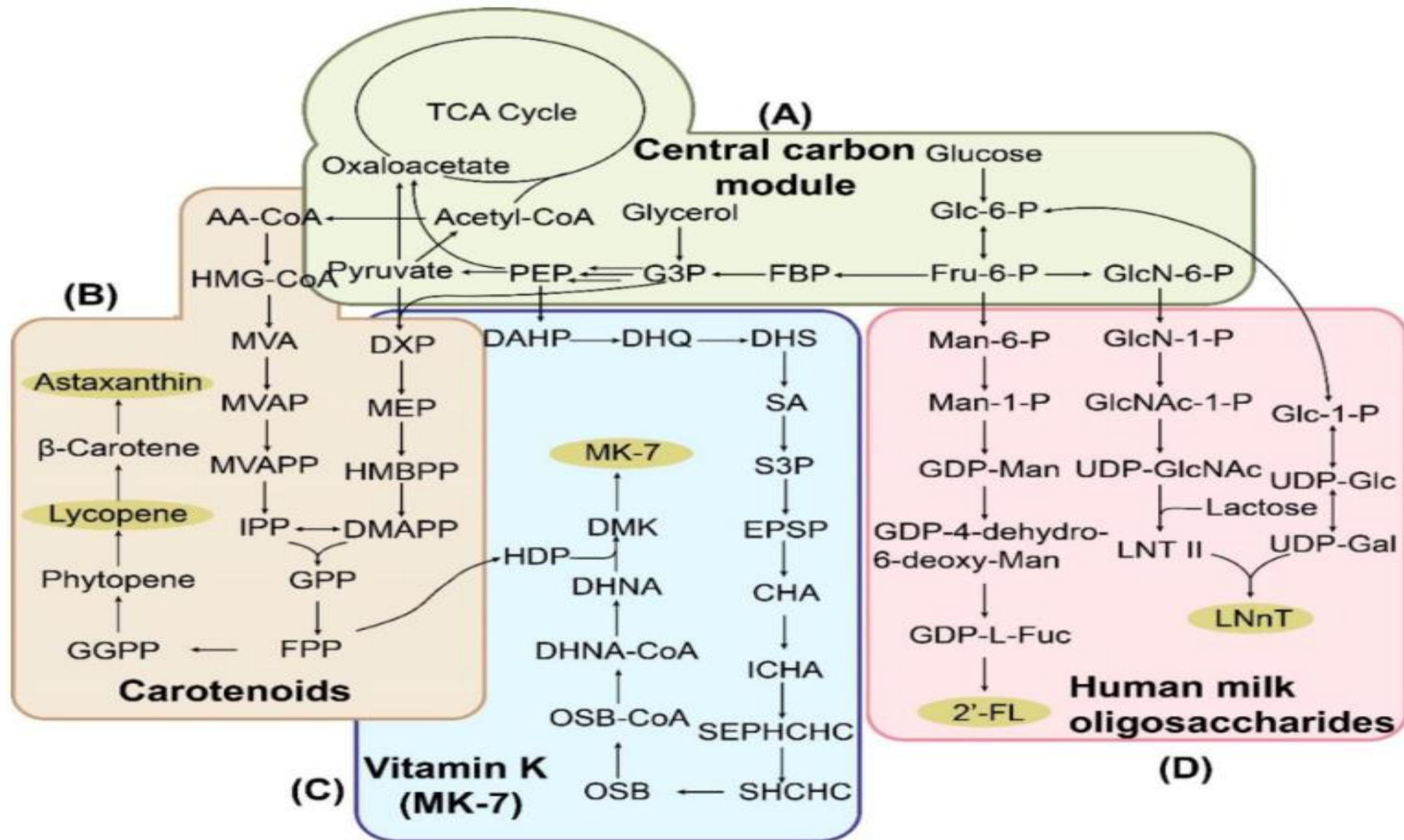
d

Constraints
 1. Mass balance
 2. $V_{i, \min} < V_i < V_{i, \max}$

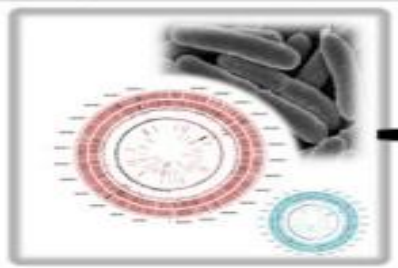
e

Optimization of objective function

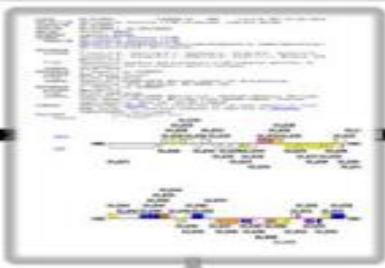
f



1. Full genome sequencing



2. Genome annotation



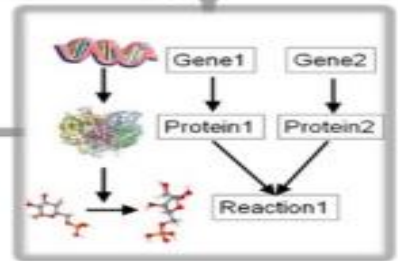
3. Collecting biochemical reactions (Literature, database, experiments)



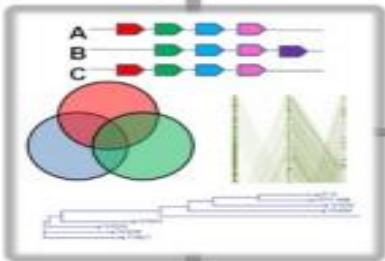
4. Manual curation (Literature, database, experiments)



6. Gene-protein-reaction correlation



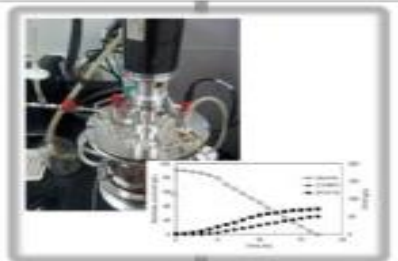
7. Comparative genome analysis



5. Model refinement (Gap filling, error correction, ...)



9. Comparison with experimental data (Experiments, biomass equation, ...)



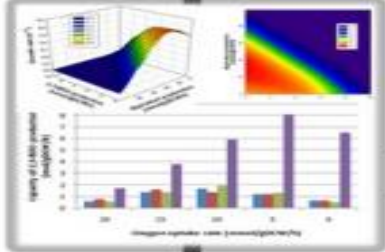
10. Model validation



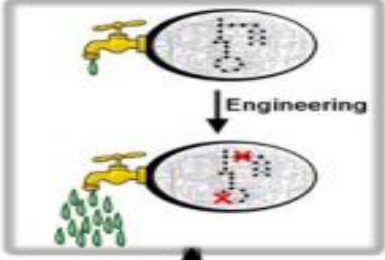
8. Static simulation



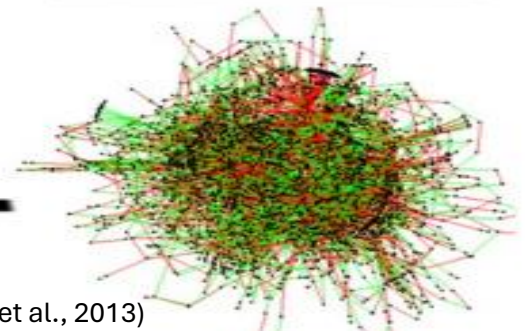
11. In silico simulation and analysis



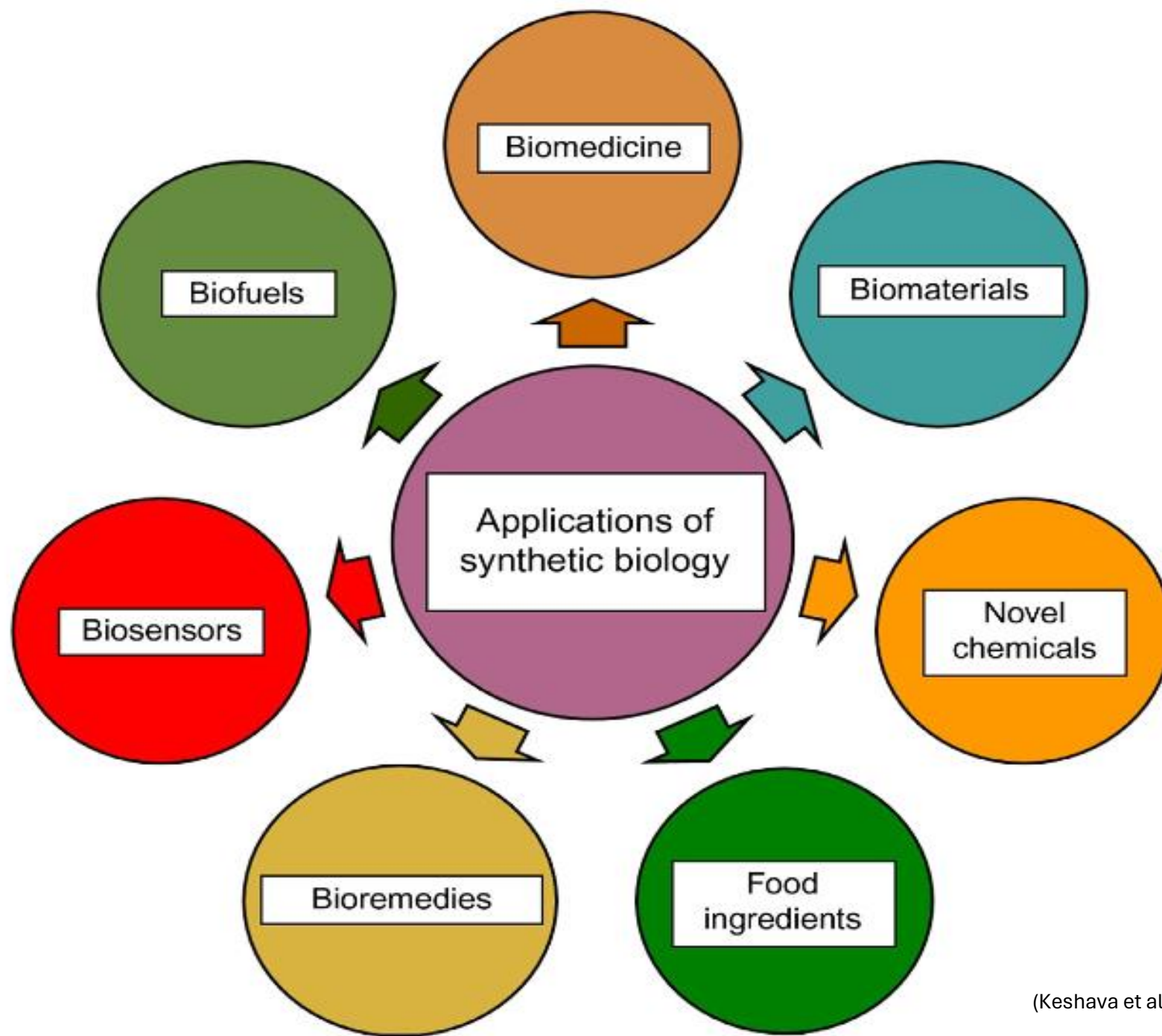
12. Application and strain improvement



In silico genome-scale metabolic model



(Park et al., 2013)



(Keshava et al., 2018)