Statistical Analysis Report

Functional profiling from shotgun metagenomes
Biomcare ApS

06/01/2025

Customer	Innovation Centre for Organic Farming, Tove Mariegaard Pedersen
Customer ID	DA00206-23
Project	Regenerativt landbrug til videreudvikling af den økologiske planteproduktion
Sample Type	Soil
Number of samples	14 samples
Type of data	shotgun metagenomics

Introduction to the report

Functional data presented

The soil microbiome harbors many thousand different genes, and we therefore need to categories and select genes or gene groups of interest before we can relate these to the many variables for soil properties and farming practises. Here we have taken the approach to select key processes of interest, and identifying their GO process ID and KEGG genes. Therefore, the report contain information on GO processes and thereto related KEGG genes for interests areas related to nitrogen cycling, phosphor processes, mycorrhiza and methane cycling.

The Project

For this project 2-3 samples were collected from 5 farms. Each sample is from one field and the fields differ on key aspects as listed in the below picture and table.

Image provided by Tove as project overview.

Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Bedrift_name	14						
… Agernæs Økologi, Fyn	2	14%					
… Anders Knudsen, Møn	3	21%					
… Niels Hansen, Sjælland	3	21%					
… Nyborggaard, Jylland	3	21%					
… Øm Klostergaard, Sjælland	3	21%					
Bedrift	14						
… P1	3	21%					
… P2	2	14%					
… P3	3	21%					
… P4	3	21%					
… P5	3	21%					

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Rt	14	6.3	0.68	5.2	5.8	6.8	7.2
Fosfor	14	2.7	1.2	0.5	1.9	3.6	4.1
Kalium	14	11	3.1	6.7	8	13	16
Magnesium	14	5.6	2.2	2.8	3.9	6.6	9.6
Kobber	14	2.7	1.1	0.8	2.1	3	4.9
Organisk_stof	14	2.7	0.66	1.6	2.3	3.1	3.9
Lerindhold_perc	14	13	4.9	7.9	10	16	25
C.N_forhold	14	11	1.2	9	10	11	13
P.afgrøde_lager	14	341	191	60	228	440	700
Ca.plante_tilgængelig	14	364	178	115	192	454	725
Total_Ca_jordlager	14	6209	2651	3695	4288	9021	10715
Ler.humus_.CEC.	14	102	39	64	72	140	174
Ombyttelig_CEC_.perc.	14	96	5.3	84	95	100	100
Mikrobiel_biomasse	14	265	58	160	229	297	381
Mikrobiel_aktivitet	14	44	14	24	36	54	73
Svampe.bakterie_forhold	14	0.79	0.16	0.5	0.7	0.9	1.1

Table 1: Summary table of the farms and fields and summary stats of key variables across the fields.

GO Processes

A list of 15 GO processes has been selected that capture much of the biological process of a set of nutrient cycles of interest. The first table below list the processes and provide key information on what biological process they represent. More can be read on each process by visiting the website: https://www.ebi.ac.uk/QuickGO (https://www.ebi.ac.uk/QuickGO).

Selected GO Processes

	Group	GO_ID	Process	Definition
1	Nitrogen	GO:0019331 (GO:0019331)	**Anaerobic respiration, using ammonium as electron donor (Nitrification)**	The oxidation of ammonium (NH4) to nitrogen (N2) in the absence of oxygen, using nitrite (NO2) as the electron acceptor. It is suggested that hydroxylamine and ammonium are combined to yield hydrazine, which is subsequently oxidized to N2.
2	Nitrogen	GO:0019740 (GO:0019740)	**Nitrogen utilization**	A series of processes that forms an integrated mechanism by which a cell or an organism detects the depletion of primary nitrogen source, usually ammonia, and then activates genes to scavenge the last traces of the primary nitrogen source and to transport and metabolize alternative nitrogen sources.
3	Nitrogen	GO:0009399 (GO:0009399)	**Nitrogen fixation**	The process in which nitrogen is taken from its relatively inert molecular form (N2) in the atmosphere and converted into nitrogen compounds useful for other chemical processes, such as ammonia, nitrate and nitrogen dioxide.
4	Mycorrhizal	GO:0009877 (GO:0009877)	**Nodulation**	The formation of nitrogen-fixing root nodules on plant roots.
5	Nitrogen	GO:0019333 (GO:0019333)	**Denitrification**	A process in which facultative anaerobic bacteria sequentially reduce nitrate (NO3-) to dinitrogen (N2) via nitrite (NO2-), nitric oxide (NO), and nitrous oxide (N2O).
6	Phosphor	GO:0016791 (GO:0016791)	**Phosphatase activity**	Encompasses enzymes that release phosphate by hydrolyzing phosphate monoesters.
7	Phosphor	GO:0016158 (GO:0016158)	**Phytase activity**	Specific to enzymes that hydrolyze phytic acid to release phosphate.
8	Mycorrhizal	GO:0036377 (GO:0036377)	**Arbuscular mycorrhizal association**	A form of mutualism between a fungus and the roots of a vascular plant, where hyphae of the fungus penetrate the plant cell wall and invaginate its cell membrane. Once inside, the fungus forms highly branched structures for nutrient exchange with the plant called arbuscules. Aids in the acquisition by the plant of nutrients such as phosphorus from the soil.

	Group	GO_ID	Process	Definition
9	Methan	GO:0015948 (GO:0015948)	**Methanogenesis**	The formation of methane by microbes, typically archaea, through anaerobic processes.
10	Methan	GO:0015977 (GO:0015977)	**Methanogenesis from Carbon Dioxide**	Methane production via the reduction of carbon dioxide using hydrogen or other reducing agents.
11	Methan	GO:0019385 (GO:0019385)	**Methanogenesis from Acetate**	The formation of methane from acetate, a common substrate in soil environments
12	Methan	GO:0019386 (GO:0019386)	**Methanogenesis from Methylated Compounds**	Methane production from methylamines, methanol, or dimethyl sulfide.
13	Methan	GO:0015949 (GO:0015949)	**Methanotrophy**	The biological oxidation of methane by microbes, primarily aerobic methanotrophic bacteria
14	Methan	GO:0009061 (GO:0009061)	**Anaerobic Methane Oxidation**	The anaerobic oxidation of methane, often coupled with sulfate reduction
15	Methan	GO:0008942 (GO:0008942)	**Methane Monooxygenase Activity**	Catalysis of the oxidation of methane to methanol.

Figure 1: Overview of selected GO processes.

Summary Table of key GO processes

Bedrift	Mark	GO:0008942 (GO:0008942): [MF] nitrite reductase [NAD(P)H] activity	GO:0009061 (GO:0009061): [BP] anaerobic respiration	GO:0009399 (GO:0009399): [BP] nitrogen fixation	GO:0009877 (GO:0009877): [BP] nodulation	GO:0015948 (GO:0015948): [BP] methanogenesis	GO:0015949 (GO:0015949): [BP] nucleobase-containing small molecule interconversion	GO:0015977 (GO:0015977): [BP] carbon fixation	GO:0016 (GO:001615 [MF phyt: activ
P1	R1	3.68183	1.38142	12.83434	3.65871	5.39068	0.36608	9.10587	0.19(
P1	R2	2.76528	1.31216	13.05687	3.07590	4.63237	0.37432	8.00007	0.06
P1	R3	3.55521	1.53924	13.59561	3.64399	5.23113	0.42578	8.50771	0.26:
P2	R4	4.53119	1.41985	9.48830	4.20827	5.59755	0.34304	8.97023	0.64:
P2	R5	4.99961	1.47283	10.25420	3.56865	5.79406	0.43821	7.93956	0.82(
P3	R6	6.42792	1.43123	11.32486	4.41219	6.25697	0.34194	8.56973	0.86∠
P3	R7	5.70092	1.51103	11.06802	4.29091	5.92309	0.39741	8.96306	0.83:
P3	R8	4.62572	1.71448	11.86542	3.90386	4.72182	0.27911	8.51991	0.48:
P4	R9	3.51369	1.50145	11.48255	3.37651	5.18692	0.34767	8.20848	0.18!
P4	R10	3.37955	1.52001	13.19299	4.27041	4.95585	0.32025	8.78746	0.17!
P4	R11	4.31906	1.26876	12.32823	3.75487	4.46414	0.41238	9.51261	0.09(
P5	R12	3.25281	1.94179	12.42589	3.93357	4.53407	0.33143	8.08750	0.22∠
P5	R13	3.93464	1.86195	10.99522	4.20273	5.23045	0.20055	7.99365	0.50(
P5	R14	3.67019	1.83856	13.72653	4.08678	5.02605	0.39795	7.70737	0.35

Figure 2: This table show the abundance of the selected GO processes in each field (rows). Each GO process column has been colored to support vidual evaluation using the concept familiar from excel with "conditional formatting". Thus, on a per-column bases, higher values have a darker background color while lower values have a lighter background color.

Note on the scale The data is generated using the software HUMANn that mapp reads to the big Uniref50 database of known genes. This parent-table has then been regrouped to cluster genes into GO processes, where the values are in reads per kilobase (RPK) which normalized fro gene lenght: RPK units reflect relative gene (or process) copy number in the community. RPK values were then further sum-normalized to adjust for differences in sequencing depth across samples resulting in relative abundance that we set to sum to 1e5. What it sums to is not very important but knowing that numbers are relative and can be compared across processes and samples is important.

KEGG genes

For the different GO process a number of key genes can be relevant to inspect specifically. For each 'group' of GO processes we have selected a set of key genes and these are shown in separate tables below.

Nitrogen

The first table below list the KEGG genes with KEGG ID and gene name. More can be read on each gene by visiting the website: https://www.genome.jp (https://www.genome.jp).

Selected KEGG genes

	Group	KEGG_ID	Enzyme
1	Nitrification	K10944	**Ammonia monooxygenase subunit A (amoA)**
2	Nitrification	K10945	**Ammonia monooxygenase subunit B (amoB)**
3	Nitrification	K10946	**Ammonia monooxygenase subunit C (amoC)**
4	Nitrification	K10535	**Hydroxylamine oxidoreductase (hao)**
5	Nitrification	K00370	**Nitrite reductase (NAD(P)H) large subunit (nxrA)**
6	Nitrification	K00371	**Nitrite reductase (NAD(P)H) small subunit (nxrB)**
7	Denitrification	K00370	**Nitrite reductase (NAD(P)H) large subunit (narG)**
8	Denitrification	K00371	**Nitrite reductase (NAD(P)H) small subunit (narH)**
9	Denitrification	K00374	**Nitrous oxide reductase (narI)**
10	Denitrification	K00368	**Nitrate reductase (nirK)**
11	Denitrification	K15864	**Periplasmic nitrate reductase (nirS)**
12	Denitrification	K04561	**Nitrate/nitrite transporter (norB)**
13	Denitrification	K02305	**Cytochrome cd1 nitrite reductase (norC)**
14	Denitrification	K00376	**Dissimilatory nitrate reductase (nosZ)**
15	Denitrification	K19339	**Assimilatory nitrate reductase (nosR)**
16	Denitrification	K03726	**Copper-containing nitrite reductase (fnr)**
17	Denitrification	K15876	**Nitrous oxide reductase (nirQ)**
18	Nitrogen Fixation	K02588	**Nitrogenase iron protein (nifH)**
19	Nitrogen Fixation	K02586	**Nitrogenase molybdenum-iron protein alpha chain (nifD)**
20	Nitrogen Fixation	K02591	**Nitrogenase molybdenum-iron protein beta chain (nifK)**
21	Nitrogen Fixation	K02585	**Nitrogen fixation protein NifB (nifB)**
22	Nitrogen Fixation	K02596	**Nitrogen fixation protein (nifX)**
23	Nitrogen Fixation	K02584	**Nif-specific regulatory protein (nifA)**

Figure 3: Overview of selected KEGG ensymes.

Summary Table of key KEGG enzymes

Bedrift	Mark	K00368: nitrite reductase (NO-forming) [EC:1.7.2.1]	K00370: nitrate reductase / nitrite oxidoreductase, alpha subunit [EC:1.7.5.1 1.7.99.4]	K00371: nitrate reductase / nitrite oxidoreductase, beta subunit [EC:1.7.5.1 1.7.99.4]	K00374: nitrate reductase gamma subunit [EC:1.7.5.1 1.7.99.4]	K00376: nitrous-oxide reductase [EC:1.7.2.4]	K02305: nitric oxide reductase subunit C	K02584: Nif-specific regulatory protein	K02585: nitrogen fixation protein NifB	K02586: nitrogenase molybdenum-iron protein alpha chain [EC:1.18.6.1]	m
P1	R1	0.43496	0.39417	0.13753	0.22951	0.25744	0.17196	0.12953	0.00000	0.00000	
P1	R2	0.59600	0.26434	0.04564	0.19358	0.19130	0.05159	0.05954	0.00000	0.00000	
P1	R3	0.46076	0.36764	0.11827	0.26190	0.28710	0.22896	0.13278	0.01731	0.00000	
P2	R4	0.43462	0.42910	0.21134	0.31925	0.32842	0.22055	0.15737	0.01123	0.00767	
P2	R5	0.45652	0.44307	0.13273	0.25918	0.31939	0.30810	0.17980	0.00647	0.00000	
P3	R6	0.47476	0.32218	0.15568	0.26897	0.40815	0.29487	0.15923	0.00000	0.00000	
P3	R7	0.40799	0.29877	0.04882	0.28630	0.35494	0.29135	0.15393	0.00000	0.01513	
P3	R8	0.42595	0.27148	0.17538	0.17139	0.26737	0.27596	0.09539	0.00000	0.00000	

Bedrift	Mark	K00368: nitrite reductase (NO-forming) [EC:1.7.2.1]	K00370: nitrate reductase / nitrite oxidoreductase, alpha subunit [EC:1.7.5.1 1.7.99.4]	K00371: nitrate reductase / nitrite oxidoreductase, beta subunit [EC:1.7.5.1 1.7.99.4]	K00374: nitrate reductase gamma subunit [EC:1.7.5.1 1.7.99.4]	K00376: nitrous-oxide reductase [EC:1.7.2.4]	K02305: nitric oxide reductase subunit C	K02584: Nif-specific regulatory protein	K02585: nitrogen fixation protein NifB	K02586: nitrogenase molybdenum-iron protein alpha chain [EC:1.18.6.1]	m
P4	R9	0.39519	0.35198	0.14240	0.16152	0.17463	0.17269	0.10475	0.00000	0.00000	
P4	R10	0.32783	0.29362	0.07858	0.19152	0.18842	0.18476	0.10142	0.00000	0.00000	
P4	R11	0.39679	0.33685	0.28987	0.27312	0.13124	0.08696	0.06320	0.00000	0.00000	
P5	R12	0.38056	0.29474	0.11699	0.21201	0.17558	0.12263	0.11376	0.00000	0.00000	
P5	R13	0.39548	0.26705	0.15763	0.31096	0.24021	0.26966	0.15148	0.00000	0.00000	
P5	R14	0.41918	0.38830	0.22514	0.18990	0.27521	0.24889	0.12833	0.01372	0.00000	

Figure 4: This table show the abundance in each field (rows) of the selected KEGG genes related to selected nitrogen GO processes.
Each KEGG column has been colored to support visual evaluation using the concept known from excel with "conditional formatting". Thus, on a per-column bases, higher values have a darker background color while lower values have a lighter background color.

Methane

The first table below list the KEGG genes with KEGG ID and gene name. More can be read on each gene by visiting the website: https://www.genome.jp (https://www.genome.jp).

Selected KEGG genes

	Group	KEGG_ID	Enzyme
1	Methane Cycle	K00399	**Particulate methane monooxygenase subunit A (pmoA)**
2	Methane Cycle	K10944	**Ammonia monooxygenase subunit A (amoA)**
3	Methane Cycle	K10945	**Ammonia monooxygenase subunit B (amoB)**
4	Methane Cycle	K10946	**Ammonia monooxygenase subunit C (amoC)**

Figure 5: Overview of selected KEGG genes.

Summary Table of key KEGG enzymes

Bedrift	Mark	K10944: methane/ammonia monooxygenase subunit A [EC:1.14.18.3 1.14.99.39]	K10945: methane/ammonia monooxygenase subunit B	K10946: methane/ammonia monooxygenase subunit C
P1	R1	0.00000	0.02845	0.63382
P1	R2	0.00954	0.00000	1.12008
P1	R3	0.02972	0.00000	0.51883
P2	R4	0.00895	0.00000	1.40343
P2	R5	0.02801	0.04538	0.98540
P3	R6	0.00000	0.16450	1.55836
P3	R7	0.00000	0.22249	1.52241
P3	R8	0.00000	0.20222	1.10894
P4	R9	0.00000	0.00000	0.89641
P4	R10	0.01242	0.00000	0.83572
P4	R11	0.00000	0.02793	0.48259
P5	R12	0.00000	0.00955	1.25588
P5	R13	0.00000	0.00000	1.57802
P5	R14	0.00377	0.00000	1.10898

Figure 6: This table show the abundance in each field (rows) of the selected KEGG genes related to selected methane GO processes.
Each KEGG column has been colored to support visual evaluation using the concept known from excel with "conditional formatting". Thus, on a per-column bases, higher values have a darker background color while lower values have a lighter background color.

Phosphor

The first table below list the KEGG genes with KEGG ID and gene name. More can be read on each gene by visiting the website: https://www.genome.jp (https://www.genome.jp).

Selected KEGG enzymes

	Group	KEGG_ID	Enzyme
1	Phosphatase activity	K01077	**Alkaline phosphatase (phoA)**
2	Phosphatase activity	K01113	**alkaline phosphatase D (phoD)**
3	Phosphatase activity	K02040	**Phosphate transport system substrate-binding protein (pstS)**
4	Phytase activity	K15921	**Phytase (phyA)**
5	Phytase activity	K15922	**Beta-propeller phytase (BPP)**

Figure 7: Overview of selected KEGG genes.

Summary Table of key KEGG enzymes

Bedrift	Mark	K01077: alkaline phosphatase [EC:3.1.3.1]	K01113: alkaline phosphatase D [EC:3.1.3.1]	K02040: phosphate transport system substrate-binding protein	K15922: alpha-glucosidase [EC:3.2.1.20]
P1	R1	0.12093	1.24422	2.49777	0.01408
P1	R2	0.07177	0.92907	2.63197	0.00000
P1	R3	0.11615	1.04639	2.55549	0.01261
P2	R4	0.24426	1.44097	2.85485	0.00000
P2	R5	0.22852	1.42369	2.82298	0.00000
P3	R6	0.14557	1.35313	3.15154	0.00000
P3	R7	0.15621	1.25112	3.33942	0.00000
P3	R8	0.18455	1.27863	2.68042	0.00000
P4	R9	0.09591	1.06301	2.21034	0.00000
P4	R10	0.13266	0.96113	2.52178	0.00000
P4	R11	0.07407	0.80824	2.51675	0.00000
P5	R12	0.13691	1.13030	2.43508	0.01571
P5	R13	0.17390	1.27934	2.95897	0.00000
P5	R14	0.12796	1.20502	2.78356	0.00000

Figure 8: This table show the abundance in each field (rows) of the selected KEGG genes related to selected Phosphor GO processes.
Each KEGG column has been colored to support visual evaluation using the concept known from excel with "conditional formatting". Thus, on a per-column bases, higher values have a darker background color while lower values have a lighter background color.

Mycorrhizal

Only few good gene candidates for Mycorrhizal processes could be found and the documentation int he KEGG database as well as linking to other databases was unspecific and unsystematic. Thus we decided to not provide gene details for this group of GO processes.

Version information

Table 2: List of used software including the used R-programming environment packages.

Package	Version	Package	Version
OS	Ubuntu 20.04.4 LTS	**ape**	5.8
R	4.3.3	**zip**	2.3.1
rstudioapi	0.16.0	**glue**	1.7.0
magrittr	2.0.3	**nlme**	3.1-165
TH.data	1.1-2	**rhdf5filters**	1.14.1
estimability	1.5.1	**Rtsne**	0.17
farver	2.1.2	**cluster**	2.1.6

Package	Version	Package	Version
nloptr	2.1.1	reshape2	1.4.4
rmarkdown	2.27	ade4	1.7-22
zlibbioc	1.48.2	generics	0.1.3
vctrs	0.6.5	gtable	0.3.5
multtest	2.58.0	tzdb	0.4.0
minqa	1.2.7	hms	1.1.3
RCurl	1.98-1.16	xml2	1.3.6
htmltools	0.5.8.1	utf8	1.2.4
S4Arrays	1.2.1	foreach	1.5.2
curl	5.2.1	pillar	1.9.0
Rhdf5lib	1.24.2	robustbase	0.99-3
SparseArray	1.2.4	splines	4.3.3
rhdf5	2.46.1	survival	3.7-0
sass	0.4.9	deldir	2.0-4
bslib	0.7.0	tidyselect	1.2.1
sandwich	3.1-0	V8	4.4.2
plyr	1.8.9	svglite	2.1.3
zoo	1.8-12	xfun	0.46
cachem	1.1.0	DEoptimR	1.1-3
igraph	2.0.3	stringi	1.8.4
lifecycle	1.0.4	yaml	2.3.9
iterators	1.0.14	boot	1.3-30
pkgconfig	2.0.3	evaluate	0.24.0
sjlabelled	1.2.0	codetools	0.2-20
R6	2.5.1	interp	1.1-6
fastmap	1.2.0	cli	3.6.3
GenomeInfoDbData	1.2.11	RcppParallel	5.1.8
numDeriv	2016.8-1.1	xtable	1.8-4
digest	0.6.36	systemfonts	1.1.0
colorspace	2.1-0	munsell	0.5.1
rematch2	2.1.2	jquerylib	0.1.4
prismatic	1.1.2	coda	0.19-4.1
hwriter	1.3.2.1	png	0.1-8
fansi	1.0.6	parallel	4.3.3
timechange	0.3.0	latticeExtra	0.6-30
abind	1.4-5	jpeg	0.1-10
mgcv	1.9-1	bitops	1.0-7
compiler	4.3.3	viridisLite	0.4.2
withr	3.0.0	mvtnorm	1.2-5
highr	0.11	insight	0.20.2
MASS	7.3-60.0.1	pcaPP	2.0-4
DelayedArray	0.28.0	crayon	1.5.3
biomformat	1.30.0	rlang	1.1.4
tools	4.3.3	multcomp	1.4-26
rrcov	1.7-5		