

Factors influencing the functional capacity of soil micro-life.

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.

Overview of selected GO processes and related KEGG genes and their presence in the dataset.

Subject	Processes	Process in data	Genes KEGG/KO (those found in data marked red)	KEGG names in data
Nitrogen	"GO:0019740: [BP] nitrogen utilization"	Yes		
Nitrogen	"GO:0009399: [BP] nitrogen fixation"	Yes	nifH K02588 nifD K02586 nifK K02591 nifB K02585 nifX K02596 nifA K02584 nifC K05596	[1] "K02584: Nif-specific regulatory protein" [2] "K02586: nitrogenase molybdenum-iron protein alpha chain [EC:1.18.6.1]" [3] "K02591: nitrogenase molybdenum-iron protein beta chain [EC:1.18.6.1]" [4] "K05596: LysR family transcriptional regulator, chromosome initiation"
Nitrogen	GO:0009877: [BP] Nodulation	Yes	nodA K14658 nodB K14659 nodC K14666 nodD2 K14657 nodD K14648 nodI K14649 nodJ K14650	None

Nitrogen	GO:0019333: [BP] denitrification pathway	Yes	narG K00370 narH K00371 narI K00374 nirK K00368 nirS K15864 norB K04561 norC K02305 nosZ K00376 nosR K00376 fnr K03726 nirQ K15876	[1] "K00368: nitrite reductase (NO-forming) [EC:1.7.2.1]" [2] "K00370: nitrate reductase / nitrite oxidoreductase, alpha subunit [EC:1.7.5.1 1.7.99.4]" [3] "K00371: nitrate reductase / nitrite oxidoreductase, beta subunit [EC:1.7.5.1 1.7.99.4]" [4] "K00374: nitrate reductase gamma subunit [EC:1.7.5.1 1.7.99.4]" [5] "K00376: nitrous-oxide reductase [EC:1.7.2.4]" [6] "K02305: nitric oxide reductase subunit C" [7] "K03726: helicase" [8] "K04561: nitric oxide reductase subunit B [EC:1.7.2.5]" [9] "K15876: cytochrome c nitrite reductase small subunit"
Nitrogen	Nitrification GO:0019331: anaerobic respiration, using ammonium as electron donor	No	amoA - K10944 amoB - K10945 amoC - K10946 hao - K10535 nxA - K00370 nxB - K00371	[1] "K00370: nitrate reductase / nitrite oxidoreductase, alpha subunit [EC:1.7.5.1 1.7.99.4]" [2] "K00371: nitrate reductase / nitrite oxidoreductase, beta subunit [EC:1.7.5.1 1.7.99.4]" [3] "K10944: methane/ammonia monooxygenase subunit A [EC:1.14.18.3 1.14.99.39]" [4] "K10945: methane/ammonia

				monooxygenase subunit B" [5] "K10946: methane/ammonia monooxygenase subunit C"
phosphate	<p>phosphate release</p> <p>No single GO process captures phosphate release fully.</p> <p>There are two Enzyme-Specific Activities for Phosphate Release:</p> <p>GO:0016158 - Phytase activity</p> <ul style="list-style-type: none"> • Specific to enzymes that hydrolyze phytic acid to release phosphate. <p>GO:0016791 - Phosphatase activity</p> <ul style="list-style-type: none"> • Encompasses enzymes that release phosphate by hydrolyzing phosphate monoesters. 		<p>PhoA/PhoB K01077</p> <p>PhoD K01113</p> <p>pstS K02040</p> <p>BPP K15922</p> <p>phyA K15921</p>	

For each of these processes we have systemized a step-by-step process to evaluate how these associates with soil properties/farming practises and which of the different recorded variables are most important for dictating the level of the GO process.

First, we look if key variables of interest that group samples into max 2 groups (JB groups, organic farming and livestock manure) have significant and consistent different levels across the years of sampling. Then, we use a machine learning approach (random forest) to identify which variables are most important to predict the GO process level, first across all samples, and then within the key groups if any of these were significant.

We then address the question “which microbial genes are important for x,y,z soil property/management practise?” We first do a targeted evaluation of the selected genes involved in the nitrogen cycle and relate these to Rt. Then we take a more exploratory approach and include all detected KEGG genes allowing the machine learning model to select those most relevant for the variables being analysed.

Overview of sections

Nitrogen	5
Nitrogen fixation.....	5
Nitrogen utilization.....	8
Nodulation.....	11
Denitrification.....	14
Which KEGG genes are important for soil property or management practices?	16
Which of the nitrogen cycle related KEGG genes are important for pH (Rt) ?.....	16
Which KEGG genes or GO processes are important for JB groups ?	17
Which KEGG genes or GO processes are important for organic versus conventional farming?	18
For organic fields, which KEGG genes or GO processes are important for ‘kløvergræs min. 3 år’ versus ‘Enårige afgrøder’?	19
For the conventional fields, which KEGG genes or GO processes are important for ‘pløjefri min. 3 år’ versus ‘pløjet’?	20

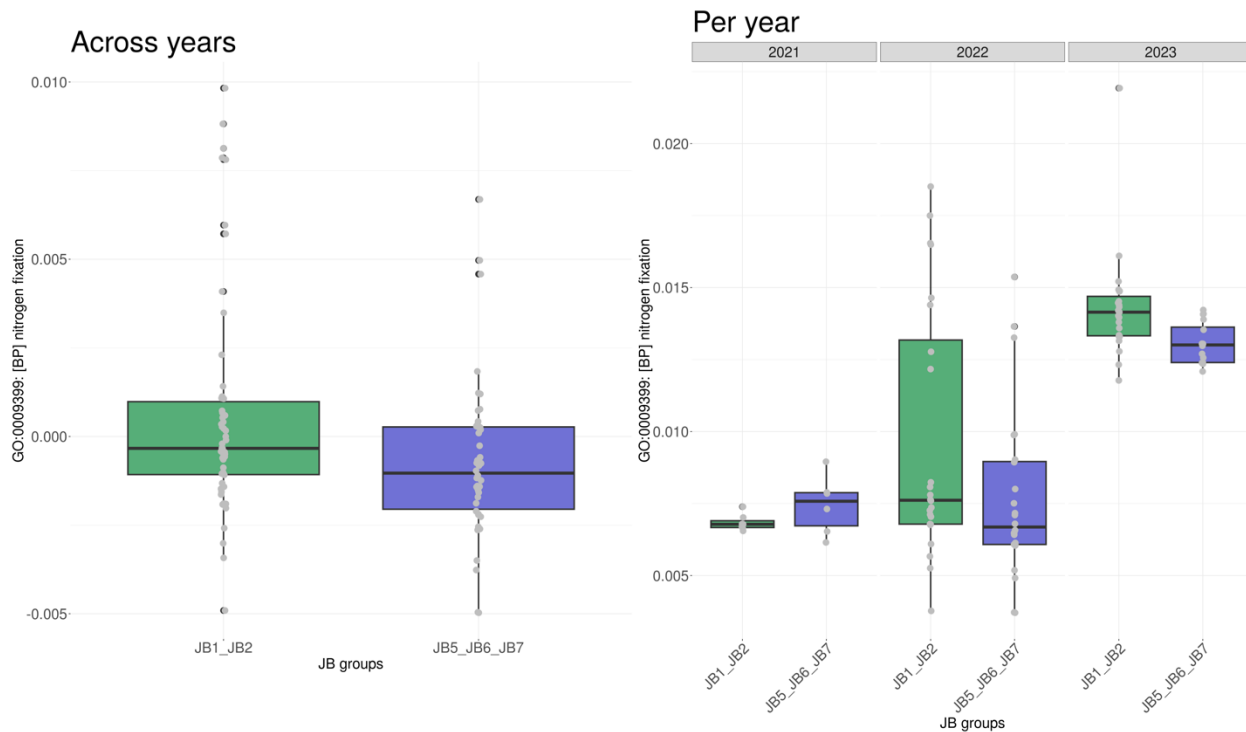
Nitrogen

Nitrogen fixation

Significant differences between key group variables?

Var		Est	p
:-----		-----	-----
JB_groups		-0.0013661	0.0162226
Organic_farm		0.0001116	0.8475328
Livestock_manure		0.0010027	0.1398968

We see that **nitrogen fixation is lower in the higher JB group** (JB 5-6-7, versus JB1-2). There is a trend for higher levels in samples treated with livestock manure, but it is not significant.



- Lower JB numbers (e.g., JB 1-3): Typically represent sandy soils with lighter textures.
- Higher JB numbers (e.g., JB 6-8): Represent clay-rich soils or heavier soils with higher water retention but poorer drainage.

Clay-rich soils in the higher JB groups tend to have poor aeration due to their fine texture and compact nature. Nitrogen-fixing bacteria, such as Rhizobium in legumes or free-living nitrogen fixers, require oxygen to function efficiently. Poor soil aeration in heavy, clay soils can create anaerobic conditions, which can inhibit nitrogen fixation. Nitrogen fixation relies heavily on the activity of symbiotic bacteria in legume crops or free-living microbes in the soil. If the microbial community is less active or diverse due to poor soil conditions, nitrogen fixation rates can decline.

Which variables predict the GO process?

First, we include all samples and most recorded variables.

We use %IncMSE to identify the most important variables. %IncMSE stands for the percentage increase in Mean Squared Error. It is a measure of how much the prediction error increases when a particular variable is permuted (i.e., its values are randomly shuffled) while all others are left unchanged.

Top predictors based on %IncMSE (most important in bottom):

	%IncMSE	IncNodePurity
Cast	1.001002	3.783424e-07
Livestock	1.324397	5.342630e-06
Potassium	1.406530	5.405899e-05
Cobber	1.659749	4.610826e-05
Crop_category	1.769027	4.073508e-05
No_plough	2.616501	7.888228e-06
Mulching_of_straw	2.657513	9.226233e-06
Degassed.fertilizer	3.027235	6.303558e-06
Rt	3.260664	7.933858e-05
Years_since_plowing	3.388055	3.267153e-05
Nitrogen_perc	3.532678	5.273530e-05
Organic_material_perc	5.491209	6.274347e-05
Magnesium	6.287489	8.600427e-05
Clay_perc	8.145284	1.160517e-04

We see that Clay% is the most important predictor, followed by magnesium and organic material. Rt is also important here.

Nitrogen% is also important but perhaps just as much correlated with nitrogen fixation because it is a consequence. Perhaps it should not be included in the analyses?

Now we select the most important variables (%IncMSE>1) and look how much of the variation in the GO process these variables explain:

% Variance explained: 0%

Are the important variables different between the two JB groups?

Below is a table of the variables selected as important in each JB group, with overlapping variables in bold text. And below the table is the % variance explained in each group.

We see that many variables are the same, but that more of the nitrogen utilization levels can be explained by evaluated variables in JB5-7 than in JB1-2 soil types. In JB5-7, organic farming is very important but does not appear to play a role for JB1-2.

JB12_predictors	JB567_predictors
NA	Magnesium
NA	ConservationAgriculture
NA	Nitrogen_perc
NA	Haveparkaffald
NA	Mulching_of_straw
NA	Vinasse
NA	No_plough
No_plough	Organic_material_perc
Clay_perc	Potassium
Magnesium	Years_since_plowing

% Variance explained for JB1+2: **6.5%**

% Variance explained for JB5-7: **0%**

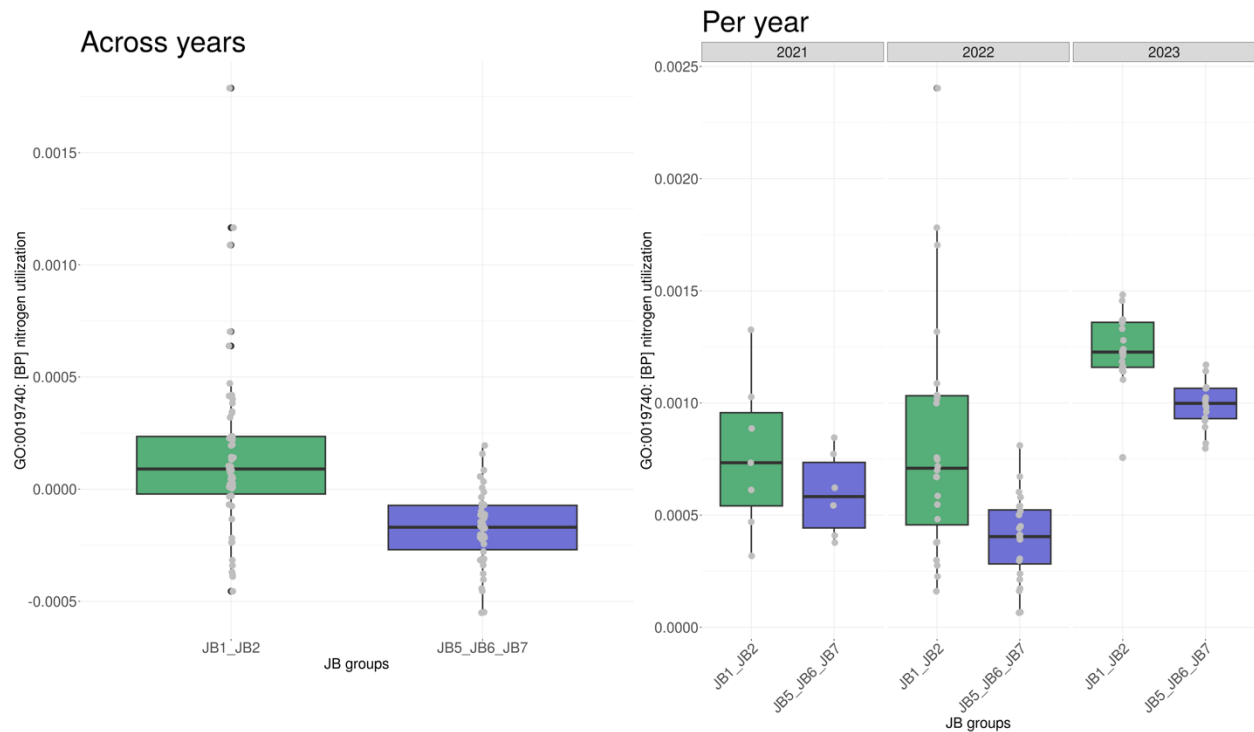
Theory notes: Magnesium levels in sandy soils can play a crucial role in nitrogen fixation, particularly because magnesium is an essential nutrient for plant metabolism and the functioning of nitrogen-fixing bacteria. Magnesium plays a role in **energy transfer** within plants, especially in the formation of ATP (adenosine triphosphate). ATP is needed in high amounts for the nitrogenase enzyme to fix atmospheric nitrogen (N₂) into ammonia (NH₃), which plants can use. Low magnesium can hinder ATP production, slowing down the nitrogen fixation process.

Nitrogen utilization

Significant differences between key group variables?

Var	Est	p
JB_groups	-0.0003255	0.0000009
Organic_farm	-0.0000303	0.6694074
Livestock_manure	0.0001518	0.0665372

We see that **nitrogen utilization is lower in the higher JB group** (JB 5-6-7, versus JB1-2). There is a trend for higher levels in samples treated with livestock manure, but it is not significant.



Which variables predict the GO process?

First we include all samples and most recorded variables.

We use %IncMSE to identify the most important variables. %IncMSE stands for the percentage increase in Mean Squared Error. It is a measure of how much the prediction error increases when a particular variable is permuted (i.e., its values are randomly shuffled) while all others are left unchanged.

Top predictors based on %IncMSE (most important in bottom):

	%IncMSE	IncNodePurity
OrganicMaterial_category	1.062575	6.841784e-08
Cast	1.208609	1.350048e-08
Livestock	1.281521	7.117355e-08
Magnesium	1.772825	8.554073e-07
Livestock_manure	2.666643	1.140522e-07
Years_since_plowing	2.816543	2.934125e-07
No_plough	3.135550	1.980305e-07
Commercial.fertilizer	3.151469	6.493701e-08
Organic_farm	3.697827	7.349624e-08
Potassium	3.711279	6.935349e-07
JB_groups	5.451393	7.213051e-07
Clay_perc	8.392563	1.663885e-06
Rt	8.616384	1.301289e-06

We see that soil pH is the most important predictor, followed by Clay% and JB groups (two interdependent variables). Organic farming is also found to be important for predicting the GO

process even as we did not detect significant consistent differences in between the groups in the above analysis indicating that organics farming is likely important in combination with some of the other variables.

Theory notes: Soil pH influences the availability of nitrogen in the soil, the activity of soil microorganisms involved in nitrogen cycling, and the forms of nitrogen present, all of which affect how efficiently plants can absorb and utilize nitrogen.

Now we select the most important variables (%IncMSE>1) and look how much of the variation in the GO process these variables explain:

% Variance explained: **18.8%**

Are the important variables different between the two JB groups?

Below is a table of the variables selected as important in each JB group, with overlapping variables in bold text. And below the table is the % variance explained in each group.

We see that many variables are the same, but that more of the nitrogen utilization levels can be explained by evaluated variables in JB5-7 than in JB1-2 soil types. In JB5-7, organic farming is very important but does not appear to play a role for JB1-2.

JB567_predictors	JB12_predictors
Organic_material_perc	NA
OrganicMaterial_category	NA
Cobber	NA
Livestock_manure	Organic_material_perc
Clovergrass_within_3_years	Livestock_manure
Crop_category	Magnesium
Years_since_plowing	No_plough
No_plough	Commercial.fertilizer
Rt	Clay_perc
Commercial.fertilizer	Years_since_plowing
Organic_farm	Rt

% Variance explained for JB1+2: 0%

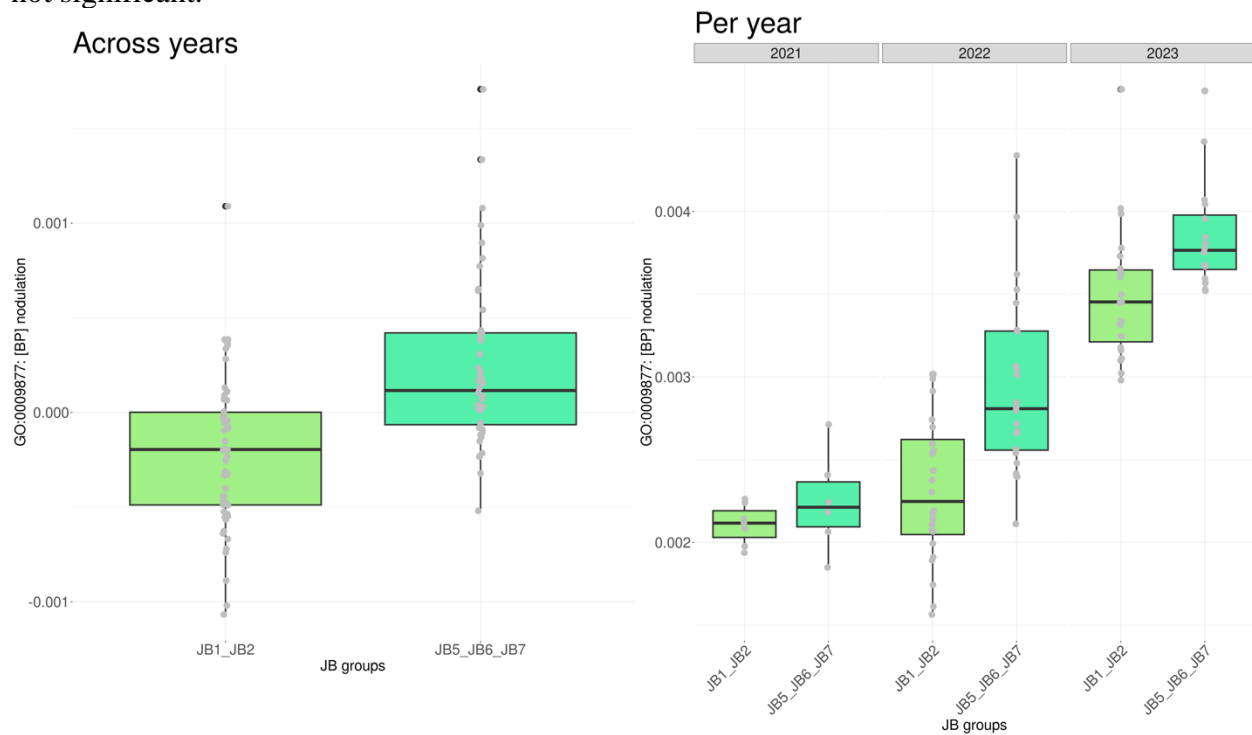
% Variance explained for JB5-7: **12.3%**

Nodulation

Significant differences between key group variables?

Var	Est	p
JB_groups	0.0004502	0.0000007
Organic_farm	-0.0001263	0.1918295
Livestock_manure	-0.0001736	0.1266196

We see that **nitrogen utilization is higher in the higher JB group** (JB 5-6-7, versus JB1-2). There is a trend for higher levels in samples treated with livestock manure and organic farming, but it is not significant.



Looking at 'Clovergrass_within_3_years' separate for organic and conventional fields, we see a trending positive association in organic (p=0.19) but no association in conventional.

Which variables predict the GO process?

First we include all samples and most recorded variables.

Top predictors based on %IncMSE (most important in bottom):

	%IncMSE
Cast	1.174673
Degassed.fertilizer	1.209275
Commercial.fertilizer	1.259265

Compact_soil	1.784173
No_plough	2.338121
Vinasse	2.791062
Years_since_plowing	3.659147
Organic_material_perc	3.692543
ConservationAgriculture	3.848921
Magnesium	4.590831
Nitrogen_perc	5.716781
Potassium	5.790658
JB_groups	7.881255
Clay_perc	9.864235
Rt	15.725176

We see that Rt stands out as an important predictor.

Theory notes: The pH of soil plays a critical role in nodulation, as it influences the survival, growth, and activity of both rhizobia and their host plants. Optimal nodulation typically occurs in soils with a near-neutral pH (around 6.0–7.0). Acidic soils (low pH) can inhibit rhizobial survival.

Now we select the most important variables ($\%IncMSE > 1$) and look how much of the variation in the GO process these variables explain:

% Variance explained: 39.7%

Are the important variables different between the two JB groups?

Below is a table of the variables selected as important in each JB group, with overlapping variables in bold text. And below the table is the % variance explained in each group.

JB12_predictors	JB567_predictors
Clovergrass_within_3_years	NA
Degassed.fertilizer	NA
field_well_drained	Magnesium
Vinasse	Compact_soil
Organic_farm	Livestock
Magnesium	No_plough
Organic_material_perc	ConservationAgriculture
Compact_soil	Organic_material_perc
Potassium	Potassium
Nitrogen_perc	Years_since_plowing
Livestock_manure	Nitrogen_perc
Rt	Rt

% Variance explained for JB1+2: 9.5%

% Variance explained for JB5-7: 39.2%

The samples are collected from bulk soil after crop harvest (for fields with 1-year crops) and thus are not collected to be representative of the root near rhizosphere. Therefore, the nodulation processes and genes analyzed here might be more representative of free-living nitrogen fixing bacteria.

Now we select the organic fields and look if the nodulation process differs between fields with one year crops versus min. 3 year clover grass. We see no difference in soil of type JB12 but a significant higher level of nodulation activity in the fields with min. 3 year clover grass for JB567 ($p=0.0074$).

Now we ask which of the remaining variables that are important for the nodulation activity within the JB groups (again in organic fields only). We see that more variables play a role in the JB567 soil than JB12, where only clay percentage is important.

In JB567, nodulation is higher with livestock and livestock manure, nitrogen and organic material, and higher with more years since plowing as well as positively correlated with Rt.

Var	Est. JB12	p-val JB12	Est. JB567	p-val JB567
Cast	-0.0001675	6.8194e-01	-0.0006079	2.3173e-01
Chalked	0.0004339	2.9391e-01	-0.0001587	6.1135e-01
Clay_perc	-0.0000879	4.9836e-02	0.0000347	2.7760e-01
Cobber	0.0000099	8.9984e-01	-0.0000512	7.1743e-01

Cold_soil		0.0001032		6.5116e-01		0.0000772		7.1387e-01	
Compact_soil		0.0002490		5.4967e-01		0.0002584		2.7559e-01	
Crop_category		0.0002517		6.2174e-01		0.0002686		5.7524e-01	
Degassed.fertilizer		-0.0002686		1.6843e-01		-0.0004184		8.1751e-02	
Earthworm_status		0.0001585		3.0580e-01		-0.0000276		9.1488e-01	
field_well_drained		-0.0000456		8.0416e-01		0.0001451		7.7857e-01	
Haveparkaffald		-0.0001182		7.7698e-01		-0.0006079		2.3173e-01	
Livestock		-0.0000908		5.7867e-01		0.0004634		2.4597e-02	
Livestock_manure		-0.0001400		5.1508e-01		0.0006429		3.0816e-02	
Magnesium		-0.0000045		9.2471e-01		0.0000537		1.1725e-01	
Mulching_of_straw		0.0000196		8.9883e-01		-0.0002940		1.6817e-01	
Nitrogen_perc		0.0007109		5.3390e-01		0.0072178		1.1102e-04	
Organic_material_perc		0.0000241		5.6924e-01		0.0003085		3.9083e-03	
OrganicMaterial_category		-0.0001876		3.5142e-01		-0.0003085		1.4918e-01	
Phosphorus		0.0000229		7.2134e-01		-0.0000570		6.1594e-01	
Potassium		-0.0000215		2.9727e-01		0.0000227		2.0733e-01	
Rt		-0.0000888		5.9606e-01		0.0004716		1.0868e-02	
Vinasse		-0.0003728		1.2003e-01		0.0000844		7.6311e-01	
Years_since_plowing		-0.0000040		9.2285e-01		0.0001375		1.9555e-02	

Denitrification

Significant differences between key group variables?

Var		Est		p
:-----		-----:		-----:
JB_groups		0.0004502		0.0000007
Organic_farm		-0.0001263		0.1918295
Livestock_manure		-0.0001736		0.1266196

Var		Est		p
:-----		-----:		-----:
JB_groups		-2.02e-05		0.6056840
Organic_farm		3.05e-05		0.4386522
Livestock_manure		8.65e-05		0.0596367

We see that **denitrification is trending higher in** livestock treated fields.

Looking at 'Clovergrass_within_3_years' separate for organic and conventional fields, we see a trending positive association in organic (p=0.19) but no association in conventional.

Which variables predict the GO process?

First we include all samples and most recorded variables.

Top predictors based on %IncMSE (most important in bottom):

	%IncMSE
Crop_category	1.125443
Nitrogen_perc	1.278718
Livestock_manure	1.499138
Organic_farm	1.810621
Clovergrass_within_3_years	2.198347
OrganicMaterial_category	2.88583
Rt	2.989402
Commercial.fertilizer	3.088263
JB_groups	3.507959
Cobber	3.714459
Clay_perc	4.071963
Magnesium	4.093623
Organic_material_perc	4.210996
Potassium	4.829112

Theory notes: There are more variables that appear to be evenly important.

Now we select the most important variables ($\%IncMSE > 1$) and look how much of the variation in the GO process these variables explain:

% Variance explained: 4.7%

Are the important variables different between the two JB groups?

Below is a table of the variables selected as important in each JB group, with overlapping variables in bold text. And below the table is the % variance explained in each group.

JB12_predictors	JB567_predictors
Livestock	NA
Cobber	NA
Organic_material_perc	NA
Nitrogen_perc	NA
Clay_perc	Degassed.fertilizer
Degassed.fertilizer	Nitrogen_perc
Magnesium	Organic_material_perc
Rt	Livestock_manure
Potassium	Cobber

% Variance explained for JB1+2: 7.1%
% Variance explained for JB5-7: 17.9%

Which KEGG genes are important for soil property or management practices?

Which of the nitrogen cycle related KEGG genes are important for pH (Rt) ?

I take all 16 KEGG genes from the nitrogen related processes and use random forest to identify those most important for Rt (after adjusting for year).

	%IncMSE	KEGG.annotation
1	6.563775	K00368: nitrite reductase (NO-forming) [EC:1.7.2.1]
2	2.367503	K00370: nitrate reductase / nitrite oxidoreductase, alpha subunit [EC:1.7.5.1 1.7.99.4]
3	8.323235	K00371: nitrate reductase / nitrite oxidoreductase, beta subunit [EC:1.7.5.1 1.7.99.4]
4	15.193928	K00374: nitrate reductase gamma subunit [EC:1.7.5.1 1.7.99.4]
5	9.061071	K00376: nitrous-oxide reductase [EC:1.7.2.4]
6	5.828677	K02305: nitric oxide reductase subunit C
7	4.940667	K02584: Nif-specific regulatory protein
8	2.452712	K02586: nitrogenase molybdenum-iron protein alpha chain [EC:1.18.6.1]
9	16.627319	K03726: helicase
10	7.553961	K04561: nitric oxide reductase subunit B [EC:1.7.2.5]
11	3.370954	K05596: LysR family transcriptional regulator, chromosome initiation
12	1.450148	K10944: methane/ammonia monooxygenase subunit A [EC:1.14.18.3 1.14.99.39]
13	4.744887	K10945: methane/ammonia monooxygenase subunit B
14	15.105992	K10946: methane/ammonia monooxygenase subunit C
15	12.920122	K15876: cytochrome c nitrite reductase small subunit

% Variance explained: **46.5%**

That is a high variation explained with leading genes being from the nitrification and denitrification processes.

Which KEGG genes are important for JB groups ?

From a technical point of view this analysis is slightly different from the above use of random forest as our “dependent variables” is not numeric (like pH) but groups with two JB groups. That means our analyses is now a classification and not a regression. However, the information we can extract from it is largely the same.

We first remove the effect of year from the KEGG genes detected across the 100 fields and then run the random forest analyses looking for microbial genes that differ in abundance between the JB groups and ask how well we can predict a soil samples JB type by just knowing the abundance of the most important genes.

We arbitrarily decide to identify the 20 genes that the model finds to be most important, and then ask for the classification accuracy when using these genes.

We get a classification accuracy of 94% meaning that the dataset indicate that we can predict the soil type of a soil sample with 94% accuracy by knowing the levels of these 20 genes in the sample.

We notice on the list of enzymes given below; several are relevant for the motility of the soil bacteria. Motility of soil bacteria is relevant for the soil's function in an agricultural context because it directly impacts processes that are critical for soil health, crop productivity, and ecosystem sustainability. We find the genes related to motility to be higher in the sandy soil with JB1 and JB2.

KEGG Names of 20 top predictors

K02074: zinc/manganese transport system ATP-binding protein
K02356: elongation factor EF-P
K02392: flagellar basal-body rod protein FlgG
K02400: flagellar biosynthesis protein FlhA
K02410: flagellar motor switch protein FliG
K02412: flagellum-specific ATP synthase [EC:3.6.3.14]
K02556: chemotaxis protein MotA
K03238: translation initiation factor 2 subunit 2
K03465: thymidylate synthase (FAD) [EC:2.1.1.148]
K04801: replication factor C small subunit
K05816: sn-glycerol 3-phosphate transport system ATP-binding protein [EC:3.6.3.20]
K07357: type 1 fimbriae regulatory protein FimB
K07726: putative transcriptional regulator
K08096: GTP cyclohydrolase IIa [EC:3.5.4.29]
K09128: hypothetical protein
K13798: DNA-directed RNA polymerase subunit B [EC:2.7.7.6]
K16306: fructose-bisphosphate aldolase / 2-amino-3,7-dideoxy-D-threo-hept-6-ulosonate synthase [EC:4.1.2.13 2.2.1.10]
K17285: NO_NAME
K18594: 3-hydroxypropionyl-CoA synthetase (ADP-forming) [EC:6.2.1.-]
K18855: NO_NAME

Which KEGG genes are important for organic versus conventional farming?

With the same logic as used above for JB groups, we here compare the organic and conventional fields that are two large groups of fields in the study.

We get a classification accuracy of 81%.

One gene stands out among the predictors namely “K05834: homoserine/homoserine lactone efflux protein” which is slightly higher in the organic fields. The *rhtB* gene encodes for a transporter protein involved in the efflux of homoserine and potentially homoserine lactones (HSLs), which are significant molecules in both microbial metabolism and intercellular communication.

Among the genes are also K19713 that was not annotated by the analyses software but by looking up in the KEGG database we find it named as a thiosulfate dehydrogenase involved in Sulfur metabolism.

Another gene in the list is K13796 named “cobZ, tcuA; tricarballylate dehydrogenase” that is increased in the organic fields. CobZ is involved in the biosynthesis of cobalamin (vitamin B12), a vital cofactor for many microbial enzymes. Microbial cobalamin production supports enzymes in carbon (e.g., methanogenesis) and nitrogen (e.g., nitrate reduction) cycling, which are critical for soil fertility.

KEGG Names of 20 top predictors

K05834: homoserine/homoserine lactone efflux protein

K04754: lipoprotein

K01250: pyrimidine-specific ribonucleoside hydrolase

K07303: isoquinoline 1-oxidoreductase, beta subunit

K11079: mannopine transport system permease protein

K03119: taurine dioxygenase [EC:1.14.11.17]

K22213: NO_NAME

K13796: NO_NAME

K14170: chorismate mutase / prephenate dehydratase [EC:5.4.99.5 4.2.1.51]

K19713: NO_NAME

K02020: molybdate transport system substrate-binding protein

K09992: hypothetical protein

K00596: 2,2-dialkylglycine decarboxylase (pyruvate)

K00356: NADH dehydrogenase [EC:1.6.99.3]

K03182: 4-hydroxy-3-polyprenylbenzoate decarboxylase [EC:4.1.1.98]

K01912: phenylacetate-CoA ligase [EC:6.2.1.30]

K02196: heme exporter protein D

K00948: ribose-phosphate pyrophosphokinase [EC:2.7.6.1]

K04759: ferrous iron transport protein B

K01752: L-serine dehydratase [EC:4.3.1.17]

For organic fields, which KEGG genes are important for ‘kløvergræs min. 3 år’ versus ‘Enårige afgrøder’?

For this analysis we first subset the fields to only look at organic farms.

We get a classification accuracy of 77%.

One of the main predictors is K03921 named as acyl-[acyl-carrier-protein] desaturase. The gene is notably higher in the fields with clover grass “min 3 years”. The article “Acyl-Acyl Carrier Protein

Desaturases and Plant Biotic Interactions” (DOI: [10.3390/cells10030674](https://doi.org/10.3390/cells10030674)) gives some interesting insight into the role of this enzyme in soil microbiology.

KEGG Names of 20 top predictors
K03921: acyl-[acyl-carrier-protein] desaturase [EC:1.14.19.2 1.14.19.11 1.14.19.26]
K14952: UDP-MurNAc hydroxylase
K02610: ring-1,2-phenylacetyl-CoA epoxidase subunit PaaB
K10956: protein transport protein SEC61 subunit alpha
K00411: ubiquinol-cytochrome c reductase iron-sulfur subunit [EC:1.10.2.2]
K21744: NO_NAME
K09957: hypothetical protein
K16566: NO_NAME
K03583: exodeoxyribonuclease V gamma subunit [EC:3.1.11.5]
K21140: NO_NAME
K01166: ribonuclease T2
K02385: flagellar protein FlbD
K15981: cholest-4-en-3-one 26-monooxygenase [EC:1.14.13.141]
K03892: ArsR family transcriptional regulator
K10005: glutamate transport system substrate-binding protein
K10747: DNA ligase 1 [EC:6.5.1.1 6.5.1.6 6.5.1.7]
K20265: glutamate:GABA antiporter
K01587: phosphoribosylaminoimidazole carboxylase / phosphoribosylaminoimidazole-succinocarboxamide synthase [EC:4.1.1.21 6.3.2.6]
K10221: 2-pyrone-4,6-dicarboxylate lactonase [EC:3.1.1.57]
K06303: spore germination protein PE

For the conventional fields, which KEGG genes are important for ‘pløjefri min. 3 år’ versus ‘pløjet’?

For this analysis we first subset the fields to only look at organic farms.

We get a classification accuracy of 88%.

Among the predictors is ‘3-ketosteroid 9 α -monooxygenase subunit B’ that is much higher in untilled soil.

The observation that 3-ketosteroid 9 α -monooxygenase subunit B is much higher in untilled soil is consistent with several ecological and microbial factors that distinguish untilled and tilled soils. This highlights the enzyme's role in organic matter degradation and its dependence on stable, less disturbed soil environments. Untilled soils often have higher levels of organic matter, including plant residues, root exudates, and animal-derived steroids from manure or natural inputs. These substrates provide a rich source of 3-ketosteroids, which are degraded by the 9 α -monooxygenase enzyme.

KEGG Names of 20 top predictors

K05565: multicomponent Na⁺:H⁺ antiporter subunit A_res

K15983: 3-ketosteroid 9alpha-monooxygenase subunit B [EC:1.14.13.142]_res

K09901: hypothetical protein_res

K02199: cytochrome c biogenesis protein CcmG, thiol:disulfide interchange_res

K03090: RNA polymerase sigma-B factor_res

K09925: hypothetical protein_res

K09164: hypothetical protein_res

K02498: HemY protein_res

K06918: NO_NAME_res

K07340: hypothetical protein_res

K08984: putative membrane protein_res

K00942: guanylate kinase [EC:2.7.4.8]_res

K09910: hypothetical protein_res

K02881: large subunit ribosomal protein L18_res

K04567: lysyl-tRNA synthetase, class II [EC:6.1.1.6]_res

K12979: NO_NAME_res

K07117: NO_NAME_res

K07645: two-component system, OmpR family, sensor histidine kinase QseC
[EC:2.7.13.3]_res

K01079: phosphoserine phosphatase [EC:3.1.3.3]_res

K03432: proteasome alpha subunit [EC:3.4.25.1]_res