Data analysis

April 19th, 2023

Vitali Francesco, CREA

EXCALIBUR Training Series:

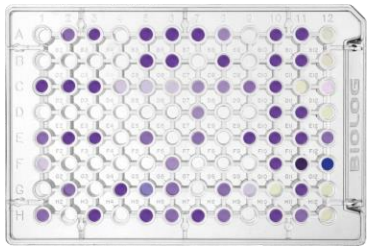
Addressing microbial metabolic profile by means of Phenotype Microarray technology (BIOLOG)

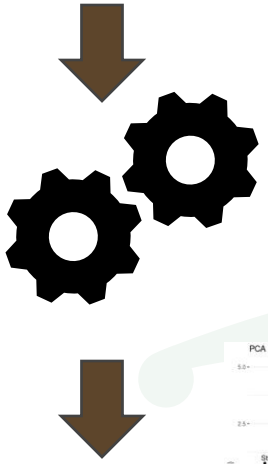
EXCALIBUR

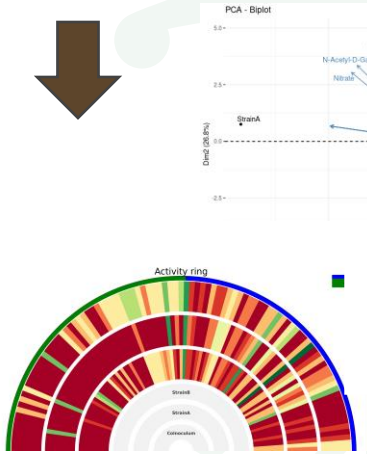
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 817946

Outline

- DuctApe installation and troubleshooting
- DuctApe workflow with example for co-inoculum experiment
- Getting data out of DuctApe, what to look for
- Using other tools for data analysis and visualization
  a) R
  b) PAST



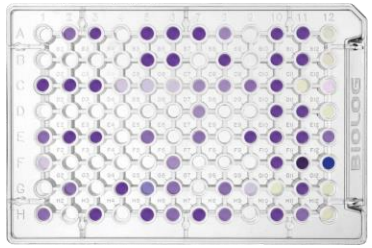




Workshop material, included the code to reproduce all analysis, is avalable in "Phenotype microarray workshop" folder at:

https://github.com/excaliburh2020/EXCALIBUR_training_series

Data analysis worflow



DuctApe

The missing link between genomics and phenomics

Analysis and visualization

AV data

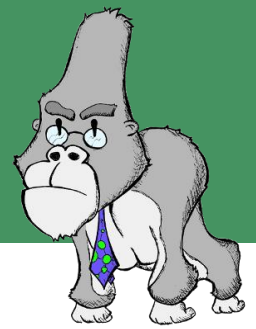
Import to other tool

Analysis and visualization

Code for reproducing the analysis in this workshop is available on GitHub at

https://github.com/excaliburh2020/EXCALIBUR_training_series

DuctApe installation and troubleshooting



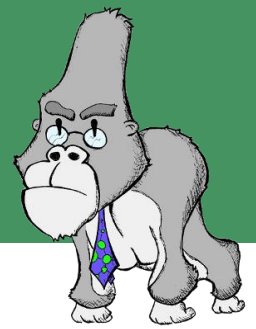
1. Install **conda**.
   Depending on your system, instruction can be found here https://docs.conda.io/en/latest/miniconda.html

   https://conda.io/projects/conda/en/latest/user-guide/getting-started.html <- general guide

   https://docs.anaconda.com/anaconda/install/ <- Installation guide

Conda is an open source package management system and environment management system that runs on Windows, macOS, and Linux. Conda quickly installs, runs and updates packages and their dependencies.

DuctApe installation and troubleshooting

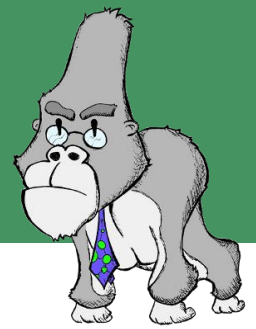


1. Install **conda**.
   Depending on your system, instruction can be found here https://docs.conda.io/en/latest/miniconda.html

2. Install **DuctApe**
   Follow instruction here https://combogenomics.github.io/DuctApe/howto.html

```
conda create -n ductape pip numpy scipy scikit-learn matplotlib biopython networkx blast

conda activate ductape

python -m pip install DuctApe
```

DuctApe installation and troubleshooting



1. Install **conda**.
   Depending on your system, instruction can be found here https://docs.conda.io/en/latest/miniconda.html

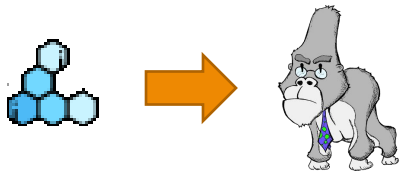
2. Install **DuctApe**
   Follow instruction here https://combogenomics.github.io/DuctApe/howto.html

3. **Troubleshooting**
   Resolve incompatibility of some packages version by downgrading the conda environment python version and the Biopython package

   1. Activate ductape conda environment with conda activate ductape
   2. Downgrade python installation inside the conda environment conda install python3.7
   3. Downgrade Biopython installation inside the conda environment conda install Biopython=1.77

Preparing biolog data for use with DuctApe

There are some compatibility issues between the current format of data export from Biolog Data Analysis and what DuctApe is expecting. Currently, the best option is to use the "old" structure of the file



StrainA.csv - LibreOffice Calc

	A	B
1	Data File	C:\Path
2	Set up Time	5/5/2022 2:53:36 PM
3	Position	3-B
4	Plate Type	PM 3
5	Strain Type	NOT APPLICABLE
6	Sample Number	StrainA
7	Strain Name	StrainA
8	Strain Number	StrainA
9	Other	

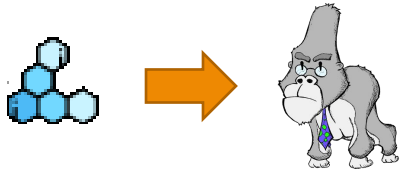
Plate Header information

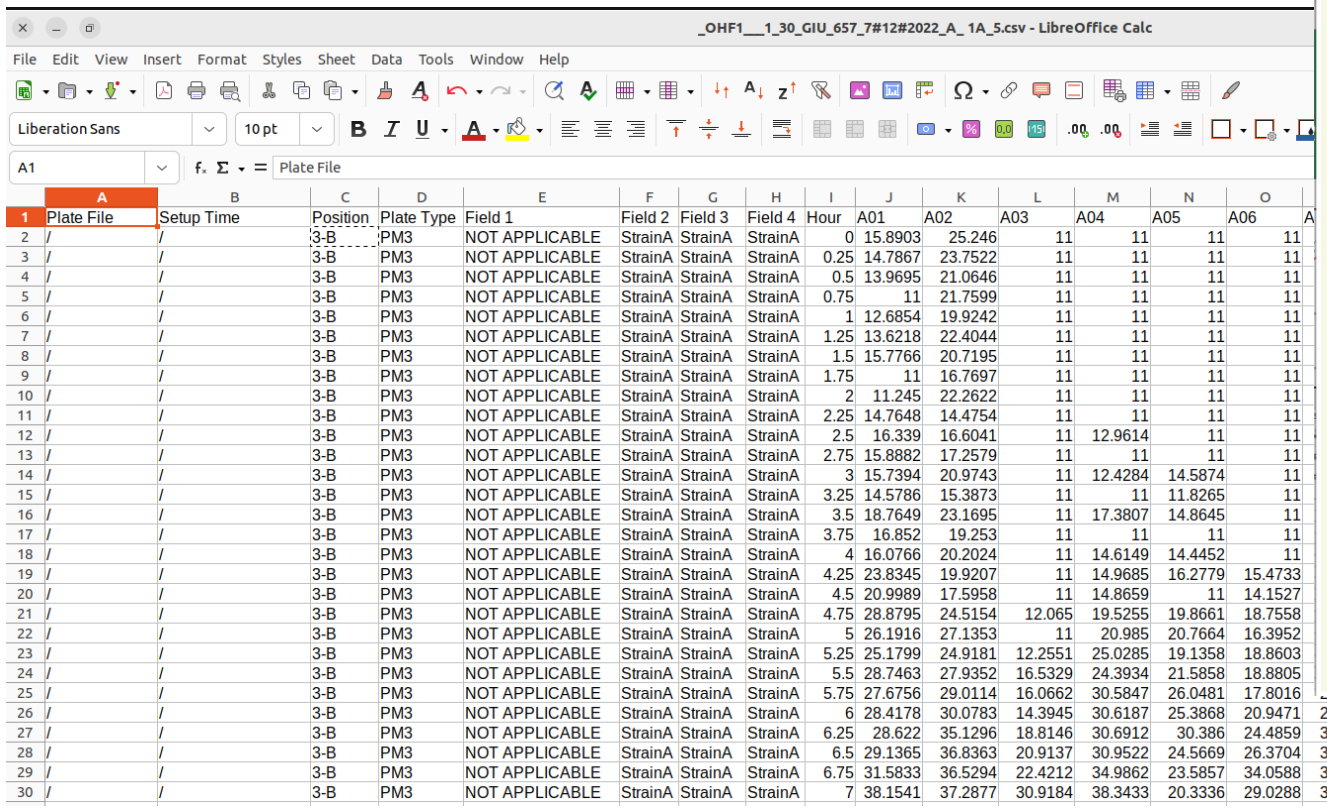
Plate readings data

Hour	A01	A02	A03	A04	A05	A06	A07	A08	A09	A10	A11	A12	B01	B02	B03	B04	B05	B06	B07	B08	B09	B1
0	53.8081	41.8799	44.6243	46.2106	30.8859	40.8293	36.6912	39.4762	39.3445	31.5718	40.9643	52.0935	55.7587	34.2062	21.6603	35.8076	24.8213	15.6654	24.9649	20.2152	75.8834	3
0.25	45.9529	37.0401	37.49	35.6212	23.8698	34.2158	25.5375	31.5377	35.3193	26.0633	36.1311	43.1558	47.3081	23.6286	15.6211	31.0659	18.6459	13.8971	20.3802	16.1387	69.5175	2
0.5	37.6619	31.5589	34.5123	26.5409	23.7762	27.634	24.1402	28.7832	29.8737	25.2924	34.1541	41.6294	37.6711	20.2976	11	22.0448	16.2713	12.0843	18.2921	16.9208	62.0042	2
0.75	40.7557	31.4265	33.0043	28.1854	23.5439	28.9789	24.5616	29.3019	30.4593	26.4728	35.239	39.9653	39.3726	21.108	11.6418	25.8016	18.2371	14.4965	16.2477	16.9558	63.0844	2
1	39.3866	30.1354	30.8191	26.5989	17.8946	25.8249	19.4798	25.8048	29.113	22.3128	30.0579	37.4498	37.4745	18.6723	11	23.1145	15.2338	11	14.701	12.1166	62.9696	2
1.25	49.4822	38.2955	41.2372	36.3842	27.5951	38.0109	33.3989	40.3709	37.7905	32.8905	41.2569	45.5438	48.9184	24.9893	16.7736	31.4063	22.3294	15.4673	23.7797	18.9378	68.0414	3
1.5	50.7411	41.9042	43.6232	40.4769	31.6982	38.6371	35.6152	44.4314	40.7817	31.3489	43.9519	45.5891	52.8066	28.5594	19.3993	31.3173	26.0612	16.2971	25.6847	18.1912	70.0571	3
1.75	47.0019	36.3349	38.4416	33.5149	28.5413	36.3309	31.2176	36.9435	37.2628	31.4597	39.668	45.8416	44.6549	23.2488	13.3766	27.1211	23.0982	15.7448	23.2936	20.908	65.6819	3
2	39.15	29.4591	29.4983	27.9941	23.5762	25.5026	23.6291	29.3399	31.7496	28.663	34.1285	42.3499	36.522	19.0961	11	22.3054	19.8391	15.0111	18.7653	17.8035	59.3538	2
2.25	40.7423	30.3565	30.7227	26.3324	21.8618	27.5154	24.8718	29.7277	30.5041	26.9277	33.3449	41.7365	36.1149	20.6202	11	23.1339	19.3028	14.3009	19.7601	15.6384	57.246	2
2.5	47.546	35.9537	39.4036	34.5675	26.7259	33.3088	29.8657	35.4886	36.0225	30.0634	40.5538	44.4513	44.4439	23.286	13.9031	29.5763	22.9266	14.8744	22.4188	19.631	63.212	2
2.75	38.5008	29.8473	28.2376	26.5457	20.6719	23.4547	21.4091	27.6362	31.3923	26.734	33.7758	40.6865	35.2591	18.6607	11	19.698	17.5264	11	18.2731	13.2876	54.2969	2
3	37.5053	27.2647	28.1638	26.444	22.7928	24.2864	22.7438	28.6136	28.2889	29.2104	34.588	41.943	32.557	18.832	11	20.6489	20.4138	11	18.3447	15.6117	52.7969	2
3.25	46.7463	35.7074	37.7842	32.9296	27.6489	34.4609	32.0056	36.6069	37.036	29.0101	38.6531	45.1784	41.6595	24.79	15.2619	26.593	23.7066	13.2945	21.0942	17.6141	59.5236	2
3.5	36.5442	28.3743	26.7532	23.0524	23.0925	24.9794	23.0883	26.2314	30.0207	28.3433	34.4238	42.9633	31.5335	16.7692	11	17.9018	20.0201	11.268	17.4066	15.313	51.2094	2
3.75	43.7534	34.2182	34.1588	30.6058	26.5275	32.3612	29.4836	35.169	36.1102	31.4846	37.1669	47.2965	41.1324	21.6519	11.8601	25.0722	24.3344	12.4157	22.9263	18.5027	55.7691	3
4	46.8162	37.4966	37.0931	32.6315	26.8407	32.7043	30.3084	36.2457	36.4942	33.1237	42.878	46.9585	42.9275	24.7297	12.9435	27.3974	25.8507	16.0119	22.4213	20.4268	57.7563	3
4.25	41.4017	30.0067	29.4695	27.6915	21.6519	27.6292	23.5389	29.946	27.6688	26.7032	33.6287	40.0982	36.7734	18.1774	11	18.4511	22.1391	11	16.2398	12.8808	51.8625	2
4.5	48.257	36.6499	38.2013	35.2087	27.9899	34.7952	29.5075	35.5012	36.8296	32.7003	39.7568	46.0026	44.5332	24.925	14.5308	26.6823	26.8153	17.7085	22.8643	17.5136	58.3334	3

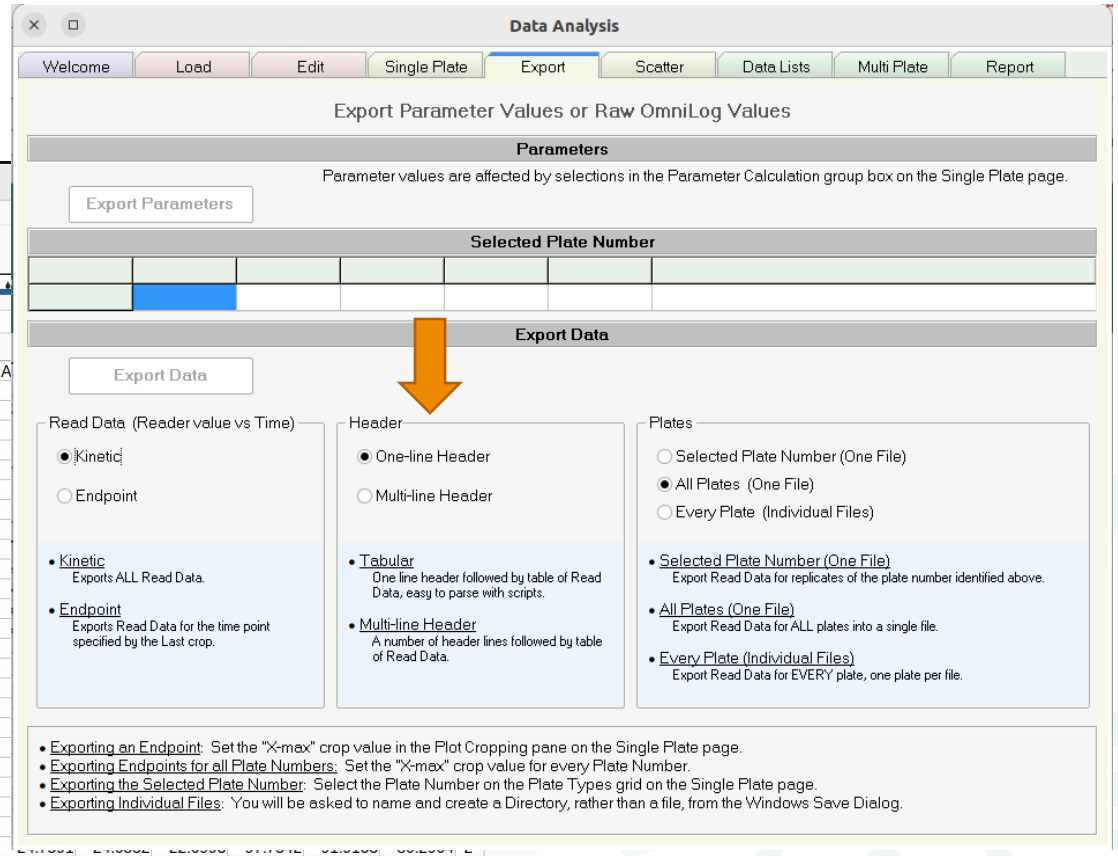
Preparing biolog data for use with DuctApe

There are some compatibility issues between the current format of data export from Biolog Data Analysis and what DuctApe is expecting. Currently, the best option is to use the "old" structure of the file



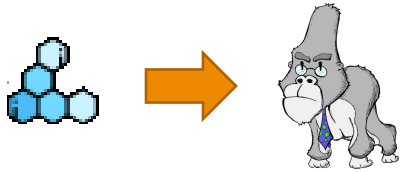


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Plate File	Setup Time	Position	Plate Type	Field 1	Field 2	Field 3	Field 4	Hour	A01	A02	A03	A04	A05	A06
2	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	0	15.8903	25.246	11	11	11	11
3	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	0.25	14.7867	23.7522	11	11	11	11
4	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	0.5	13.9695	21.0646	11	11	11	11
5	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	0.75	11	21.7599	11	11	11	11
6	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	1	12.6854	19.9242	11	11	11	11
7	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	1.25	13.6218	22.4044	11	11	11	11
8	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	1.5	15.7766	20.7195	11	11	11	11
9	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	1.75	11	16.7697	11	11	11	11
10	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	2	11.245	22.2622	11	11	11	11
11	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	2.25	14.7648	14.4754	11	11	11	11
12	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	2.5	16.339	16.6041	11	12.9614	11	11
13	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	2.75	15.8882	17.2579	11	11	11	11
14	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	3	15.7394	20.9743	11	12.4284	14.5874	11
15	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	3.25	14.5786	15.3873	11	11	11.8265	11
16	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	3.5	18.7649	23.1695	11	17.3807	14.8645	11
17	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	3.75	16.852	19.253	11	11	11	11
18	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	4	16.0766	20.2024	11	14.6149	14.4452	11
19	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	4.25	23.8345	19.9207	11	14.9685	16.2779	15.4733
20	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	4.5	20.9989	17.5958	11	14.8659	11	14.1527
21	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	4.75	28.8795	24.5154	12.065	19.5255	19.8661	18.7558
22	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	5	26.1916	27.1353	11	20.985	20.7664	16.3952
23	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	5.25	25.1799	24.9181	12.2551	25.0285	19.1358	18.8603
24	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	5.5	28.7463	27.9352	16.5329	24.3934	21.5858	18.8805
25	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	5.75	27.6756	29.0114	16.0662	30.5847	26.0481	17.8016
26	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	6	28.4178	30.0783	14.3945	30.6187	25.3868	20.9471
27	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	6.25	28.622	35.1296	18.8146	30.6912	30.386	24.4859
28	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	6.5	29.1365	36.8363	20.9137	30.9522	24.5669	26.3704
29	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	6.75	31.5833	36.5294	22.4212	34.9862	23.5857	34.0588
30	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	7	38.1541	37.2877	30.9184	38.3433	20.3336	29.0288

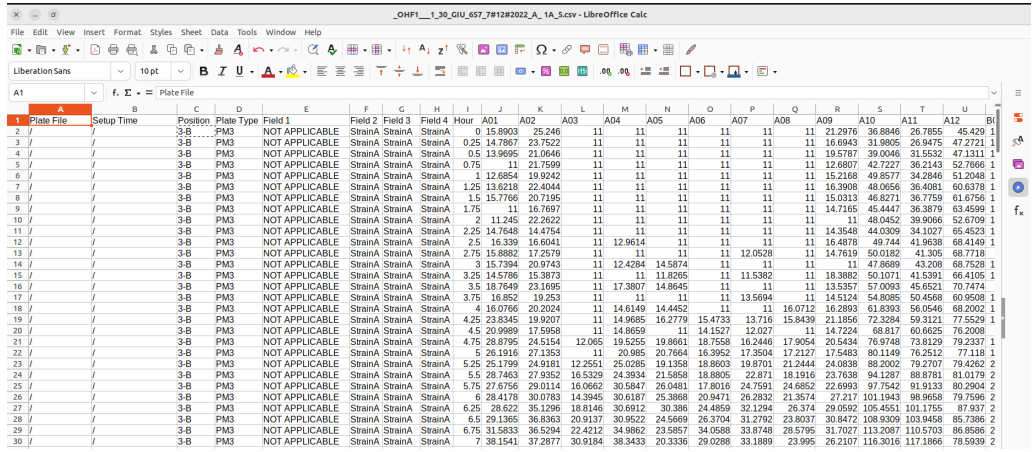


Preparing biolog data for use with DuctApe

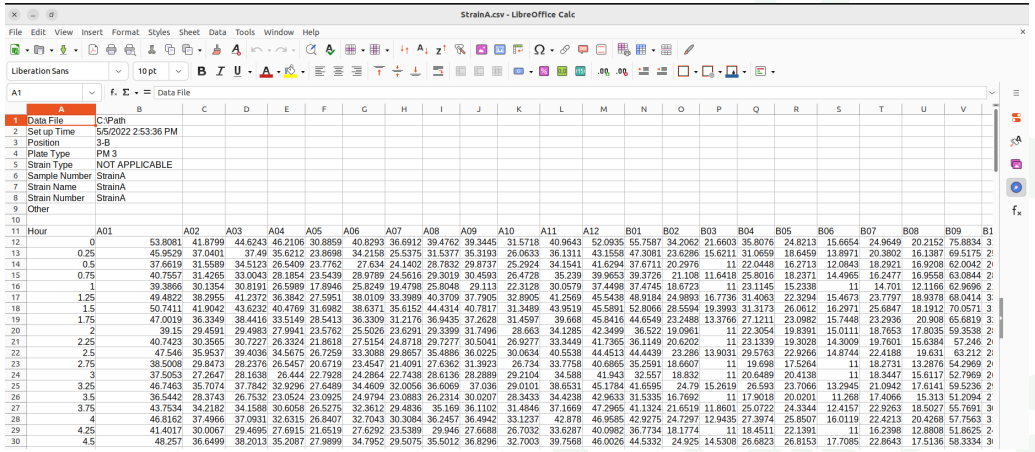
There are some compatibility issues between the current format of data export from Biolog Data Analysis and what DuctApe is expecting. Currently, the best option is to use the "old" structure of the file



We have developed an R script for converting "new" one-line header file in "old" multi-header file. Needs more testing and finalization (so expect a better version within next month) but is already available as GitHub material of this workshop.

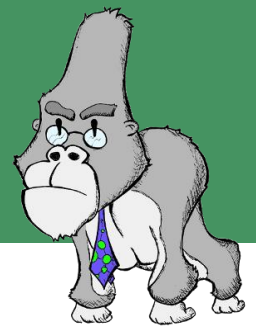


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Plate File	Setup Time	Position	Plate Type	Field 1	Field 2	Field 3	Field 4	Hour	A01	A02	A03	A04	A05	A06	A07	A08	A09	A10	A11	A12
2	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	0	15.8903	25.246	11	11	11	11	11	11	21.2976	36.8846	26.7855	45.429
3	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	0.25	14.7867	23.7522	11	11	11	11	11	11	16.6943	31.9805	26.9475	47.2721
4	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	0.5	13.9695	21.0646	11	11	11	11	11	11	19.5787	39.0046	31.5532	47.1311
5	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	0.75	11	21.7599	11	11	11	11	11	11	12.6807	42.7227	36.2143	52.7666
6	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	1	12.6854	19.9242	11	11	11	11	11	11	15.2168	49.8577	34.2846	51.2048
7	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	1.25	13.6218	22.4044	11	11	11	11	11	11	16.3908	48.0656	36.4081	60.6378
8	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	1.5	15.7766	20.7195	11	11	11	11	11	11	15.0313	46.8271	36.7759	61.6756
9	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	1.75	11	16.7697	11	11	11	11	11	11	14.7165	45.4447	36.3879	63.4599
10	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	2	11.245	22.2622	11	11	11	11	11	11	11	48.0452	39.9066	52.6709
11	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	2.25	14.7648	14.4754	11	11	11	11	11	11	14.3548	44.0309	34.1027	65.4523
12	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	2.5	16.339	16.6041	11	12.9614	11	11	11	11	16.4878	49.744	41.9638	68.4149
13	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	2.75	15.8882	17.2579	11	11	11	11	12.0528	11	14.7619	50.0182	41.305	68.7718
14	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	3	15.7394	20.9743	11	12.4284	14.5874	11	11	11	11	47.8689	43.208	68.7528
15	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	3.25	14.5786	15.3873	11	11	11.8265	11	11.5382	11	18.3882	50.1071	41.5391	66.4105
16	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	3.5	18.7649	23.1695	11	17.3807	14.8645	11	11	11	13.5357	57.0093	45.6521	70.7474
17	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	3.75	16.852	19.253	11	11	11	11	13.5694	11	14.5124	54.8085	50.4568	60.9508
18	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	4	16.0766	20.2024	11	14.6149	14.4452	11	11	16.0712	16.2893	61.8393	56.0546	68.2002
19	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	4.25	23.8345	19.9207	11	14.9685	16.2779	15.4733	13.716	15.8439	21.1856	72.3284	59.3121	77.5529
20	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	4.5	20.9989	17.5958	11	14.8659	11	14.1527	12.027	11	14.7224	68.817	60.6625	76.2008
21	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	4.75	28.8795	24.5154	12.065	19.5255	19.8661	18.7558	16.2446	17.9054	20.5434	76.9748	73.8129	79.2337
22	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	5	26.1916	27.1353	11	20.985	20.7664	16.3952	17.3504	17.2127	17.5483	80.1149	76.2512	77.118
23	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	5.25	25.1799	24.9181	12.2551	25.0285	19.1358	18.8603	19.8701	21.2444	24.0838	88.2002	79.2707	79.4262
24	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	5.5	28.7463	27.9352	16.5329	24.3934	21.5858	18.8805	22.871	18.1916	23.7638	94.1287	88.8781	81.0179
25	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	5.75	27.6756	29.0114	16.0662	30.5847	26.0481	17.8016	24.7591	24.6852	22.6993	97.7542	91.9133	80.2904
26	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	6	28.4178	30.0783	14.3945	30.6187	25.3868	20.9471	26.2832	21.3574	27.217	101.1943	98.9658	79.7596
27	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	6.25	28.622	35.1296	18.8146	30.6912	30.386	24.4859	32.1294	26.374	29.0592	105.4551	101.1755	87.937
28	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	6.5	29.1365	36.8363	20.9137	30.9522	24.5669	26.3704	31.2792	23.8037	30.8472	108.9309	103.9458	85.7386
29	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	6.75	31.5833	36.5294	22.4212	34.9862	23.5857	34.0588	33.8748	28.5795	31.7027	113.2087	110.5703	86.8586
30	/	/	3-B	PM3	NOT APPLICABLE	StrainA	StrainA	StrainA	7	38.1541	37.2877	30.9184	38.3433	20.3336	29.0288	33.1889	23.995	26.2107	116.3016	117.1866	78.5939



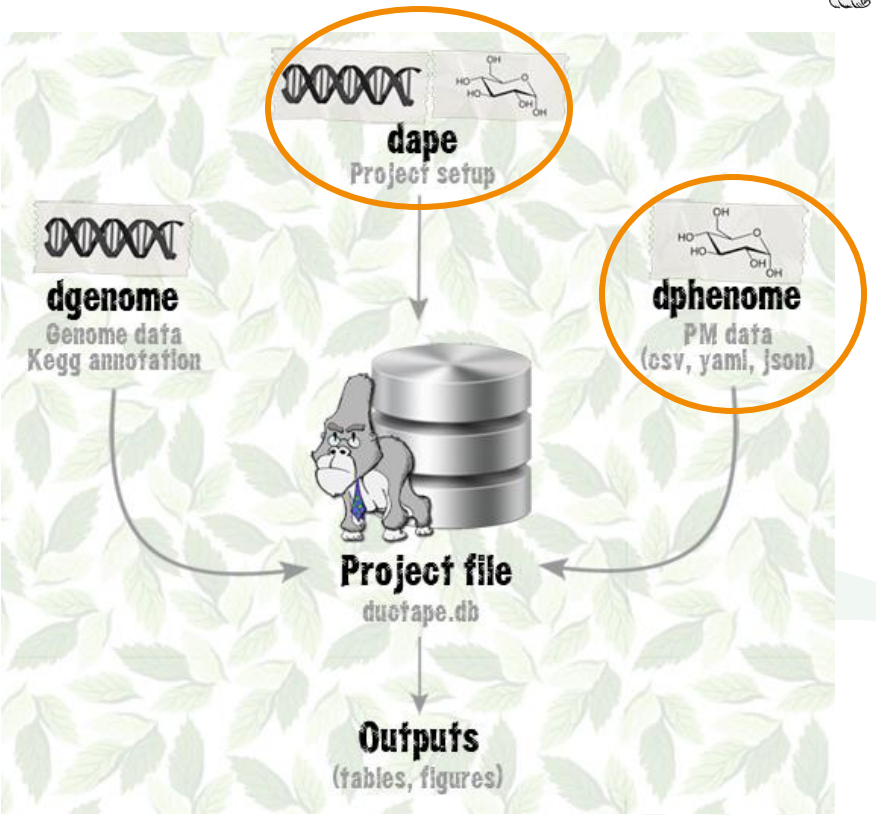
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	Data File	C:\Path																				
2	Set up Time	5/5/2022 2:53:36 PM																				
3	Position	3-B																				
4	Plate Type	PM 3																				
5	Strain Type	NOT APPLICABLE																				
6	Sample Number	StrainA																				
7	Strain Name	StrainA																				
8	Strain Number	StrainA																				
9	Other																					
10																						
11	Hour	A01	A02	A03	A04	A05	A06	A07	A08	A09	A10	A11	A12	B01	B02	B03	B04	B05	B06	B07	B08	B09
12	0	53.8081	41.8799	44.6243	46.2106	30.8859	40.8293	36.6912	39.4762	39.3445	31.5718	40.9643	52.0935	55.7587	34.2062	21.6603	35.8076	24.8213	15.6654	24.9649	20.2152	75.8834
13	0.25	45.9529	37.0401	37.49	35.6212	23.8698	34.2158	25.5375	31.5377	35.3193	26.0633	36.1311	43.1558	47.3081	23.6286	15.6211	31.0659	18.6459	13.8971	20.3802	16.1387	69.5175
14	0.5	37.6619	31.5589	34.5123	26.5409	23.7762	27.634	24.1402	28.7832	29.8737	25.2924	34.1541	41.6294	37.6711	20.2976	11	22.0448	16.2713	12.0843	18.2921	16.9208	62.0042
15	0.75	40.7557	31.4265	33.0043	28.1854	23.5439	28.9789	24.5616	29.3019	30.4593	26.4728	35.239	39.9653	39.3726	21.108	11.6418	25.8016	18.2371	14.4965	16.2477	16.9558	63.0844
16	1	39.3866	30.1354	30.8191	26.5989	17.8946	25.8249	19.4798	25.8048	29.113	22.3128	30.0579	37.4498	37.4745	18.6723	11	23.1145	15.2338	11	14.701	12.1166	62.9696
17	1.25	49.4822	38.2955	41.2372	36.3842	27.5951	38.0109	33.3989	40.3709	37.7905	32.8905	41.2569	45.5438	48.9184	24.9893	16.7736	31.4063	22.3294	15.4673	23.7797	18.9378	68.0414
18	1.5	50.7411	41.9042	43.6232	40.4769	31.6982	38.6371	35.6152	44.4314	40.7817	31.3489	43.9519	45.5891	52.8066	28.5594	19.3993	31.3173	26.0612	16.2971	25.6847	18.1912	70.0571
19	1.75	47.0019	36.3349	38.4416	33.5149	28.5413	36.3309	31.2176	36.9435	37.2628	31.4597	39.668	45.8416	44.6549	23.2488	13.3766	27.1211	23.0982	15.7448	23.2936	20.908	65.6819
20	2	39.15	29.4591	29.4983	27.9941	23.5762	25.5026	23.6291	29.3399	31.7496	28.663	34.1285	42.3499	36.522	19.0961	11	22.3054	19.8391	15.0111	18.7653	17.8035	59.3538
21	2.25	40.7423	30.3565	30.7227	26.3324	21.8618	27.5154	24.8718	29.7277	30.5041	26.9277	33.3449	41.7365	36.1149	20.6202	11	23.1339	19.3028	14.3009	19.7601	15.6384	57.246
22	2.5	47.546	35.9537	39.4036	34.5675	26.7259	33.3088	29.8657	35.4886	36.0225	30.0634	40.5538	44.4513	44.4439	23.286	13.9031	29.5763	22.9266	14.8744	22.4188	19.631	63.212
23	2.75	38.5008	29.8473	28.2376	26.5457	20.6719	23.4547	21.4091	27.6362	31.3923	26.734	33.7758	40.6865	35.2591	18.6607	11	19.698	17.5264	11	18.2731	13.2876	54.2969
24	3	37.5053	27.2647	28.1638	26.444	22.7928	24.2864	22.7438	28.6136	28.2889	29.2104	34.588	41.943	32.557	18.832	11	20.6489	20.4138	11	18.3447	15.6117	52.7969
25	3.25	46.7463	35.7074	37.7842	32.9296	27.6489	34.4609	32.0056	36.6069	37.036	29.0101	38.6531	45.1784	41.6595	24.79	15.2619	26.593	23.7066	13.2945	21.0942	17.6141	59.5236
26	3.5	36.5442	28.3743	26.7532	23.0524	23.0925	24.9794	23.0883	26.2314	30.0207	28.3433	34.4238	42.9633	31.5335	16.7692	11	17.9018	20.0201	11.268	17.4066	15.313	51.2094
27	3.75	43.7534	34.2182	34.1588	30.6058	26.5275	32.3612	29.4836	35.169	36.1102	31.4846	37.1669	47.2965	41.1324	21.6519	11.8601	25.0722	24.3344	12.4157	22.9263	18.5027	55.7691
28	4	46.8162	37.4966	37.0931	32.6315	26.8407	32.7043	30.3084	36.2457	36.4942	33.1237	42.878	46.9585	42.9275	24.7297	12.9435	27.3974	25.8507	16.0119	22.4213	20.4268	57.7563
29	4.25	41.4017	30.0067	29.4695	27.6915	21.6519	27.6292	23.5389	29.946	27.6688	26.7032	33.6287	40.0982	36.7734	18.1774	11	18.4511	22.1391	11	16.2398	12.8808	51.8625
30	4.5	48.257	36.6499	38.2013	35.2087	27.9899	34.7952	29.5075	35.5012	36.8296	32.7003	39.7568	46.0026	44.5332	24.925	14.5308	26.6823	26.8153	17.7085	22.8643	17.5136	58.3334

DuctApe workflow with example of co-inoculum experiment



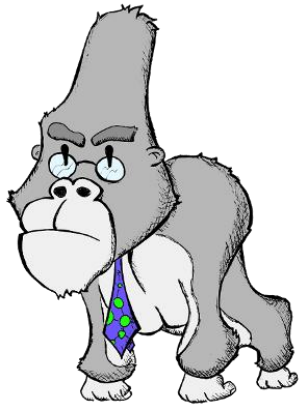
DuctApe is organized in modules of the analysis.
Syntax fo command is module option

- dape module import data and prepare a database
    - dape init initialize the database, is the first command
    - dape add add an organism (strain) entry to the database
- dphenome module performs phenomic data analysis
    - dphenome add-dir add phenomic data in a folder
    - dphenome zero performs zero/negative well substraction
    - dphenome start main command to start analysis
    - dphenome plot
    - dphenome rings
    - dphenome export



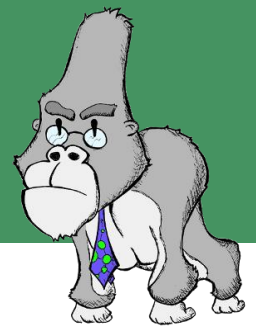
https://combogenomics.github.io/DuctApe/tutorial.html

DuctApe workflow with example of co-inoculum experiment

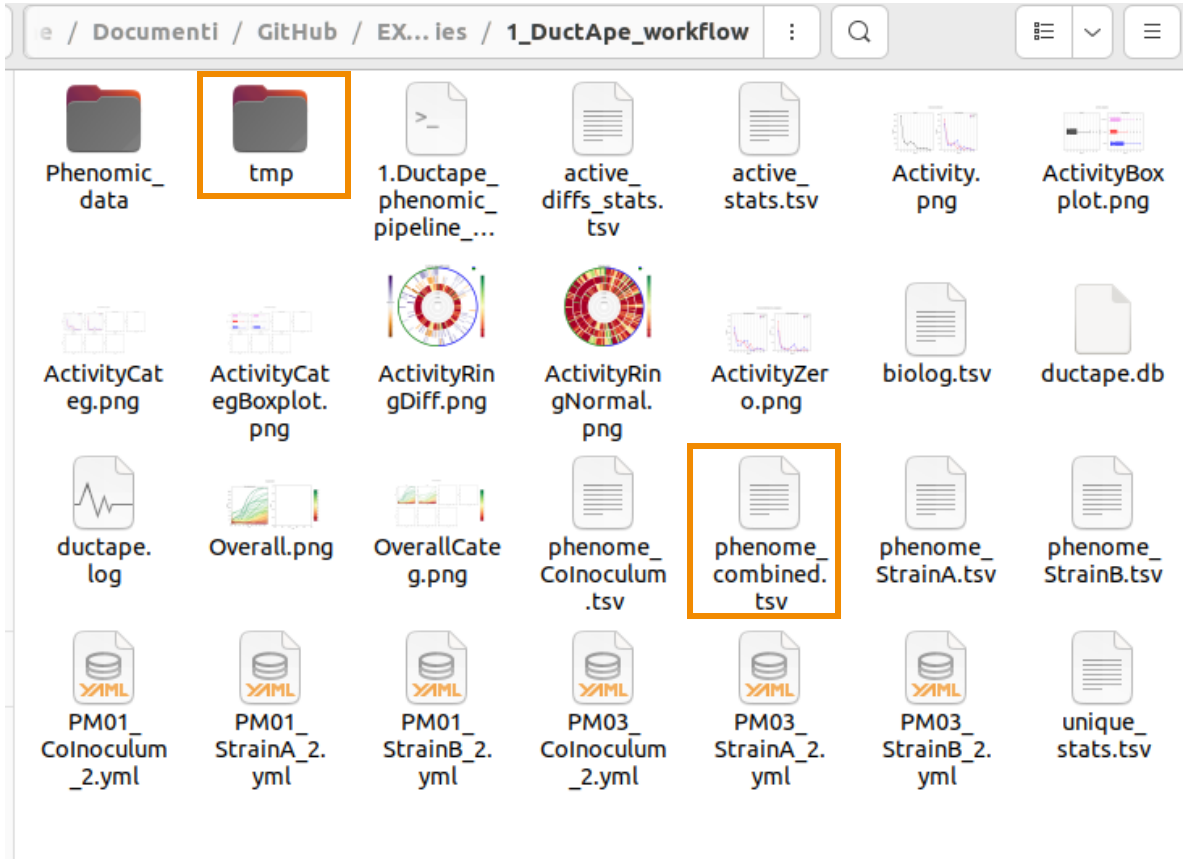


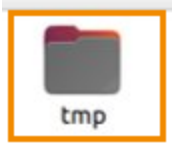
https://combogenomics.github.io/DuctApe/tutorial.html

Getting data out of DuctApe

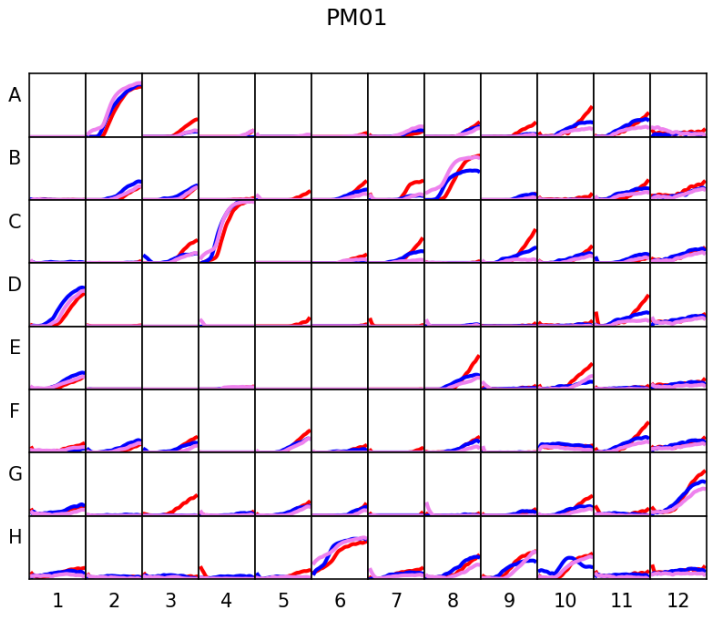


Those files and folder should have been created from the DuctApe pipeline



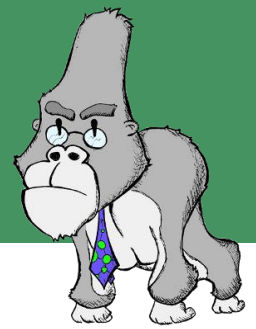


Contains plot of the plates and the curves for single compund

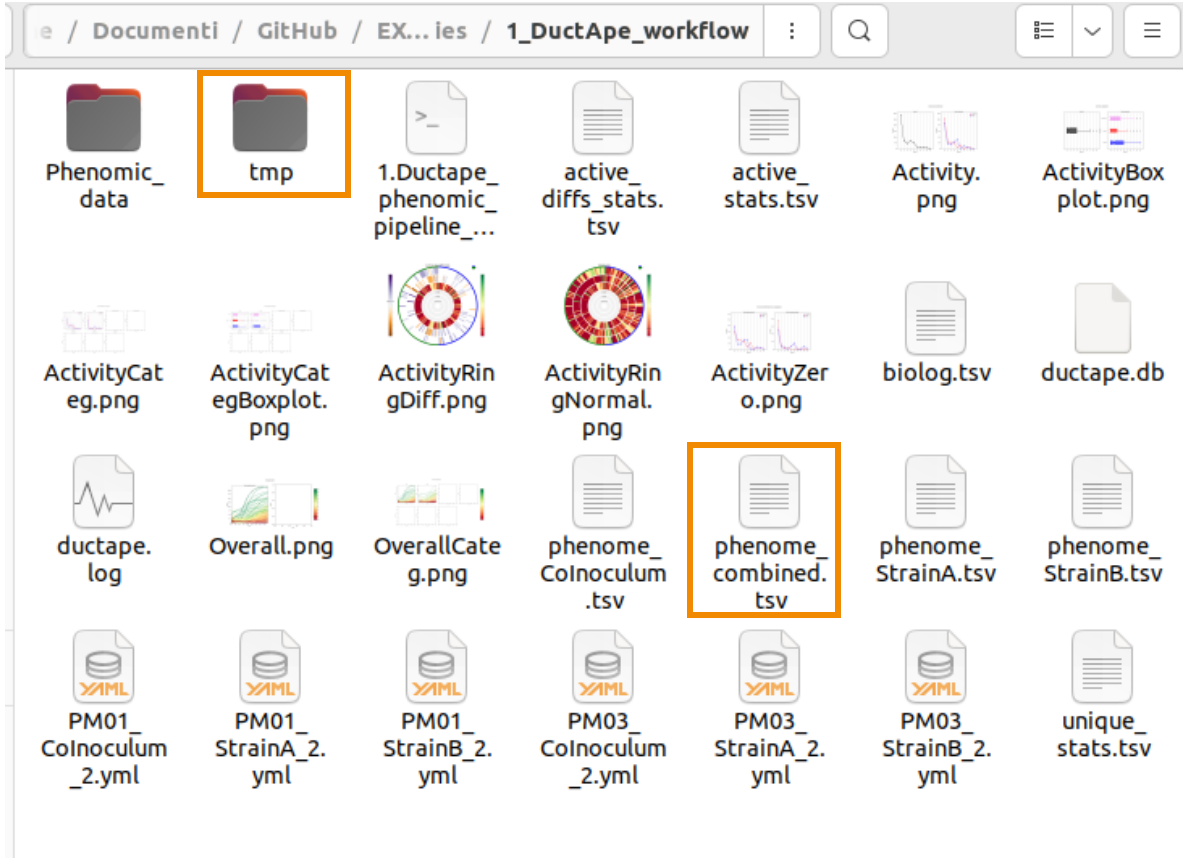


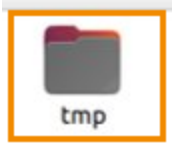
https://combogenomics.github.io/DuctApe/tutorial.html

Getting data out of DuctApe

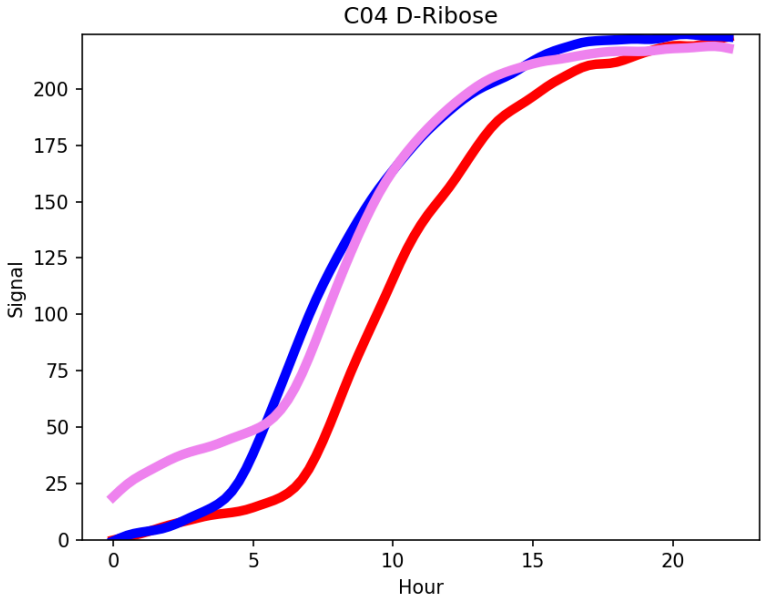


Those files and folder should have been created from the DuctApe pipeline



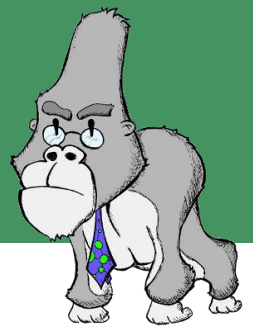


Contains plot of the plates and the curves for single compund

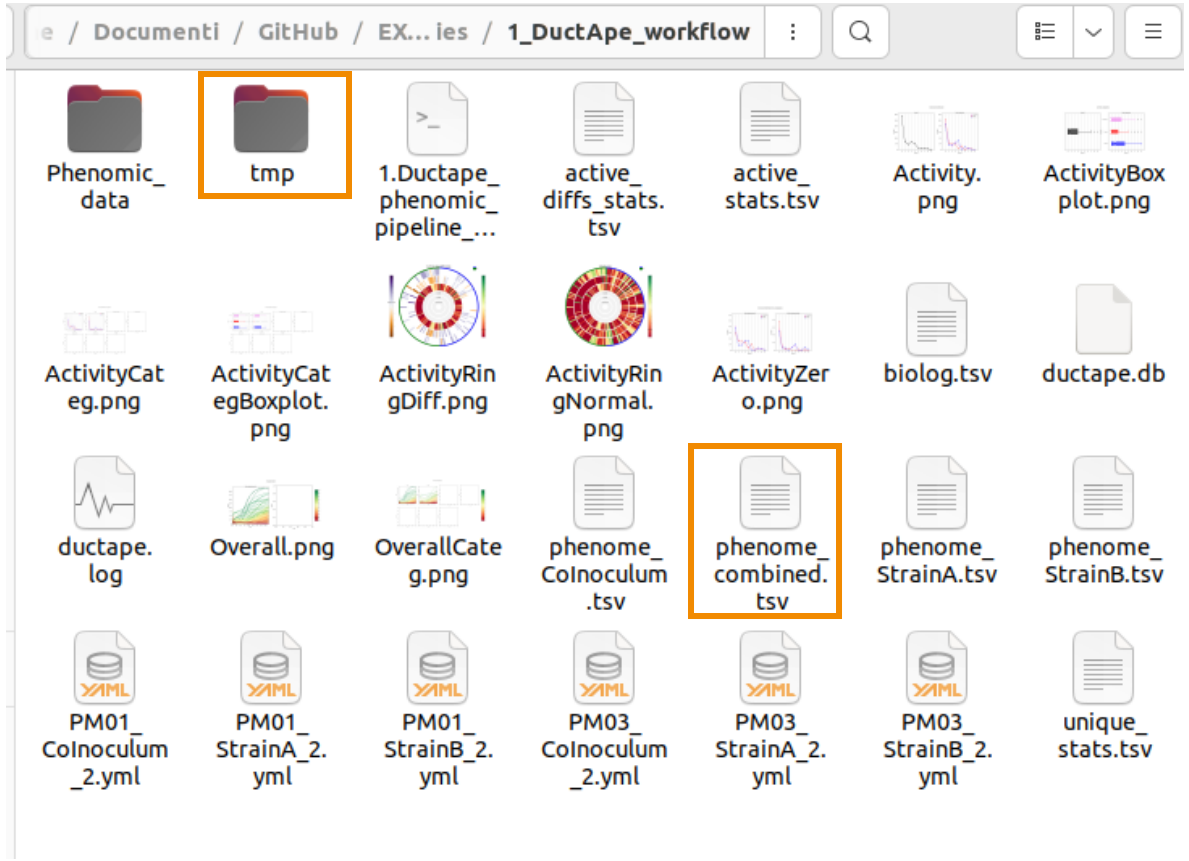


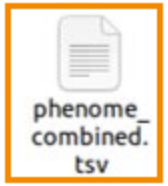
https://combogenomics.github.io/DuctApe/tutorial.html

Getting data out of DuctApe

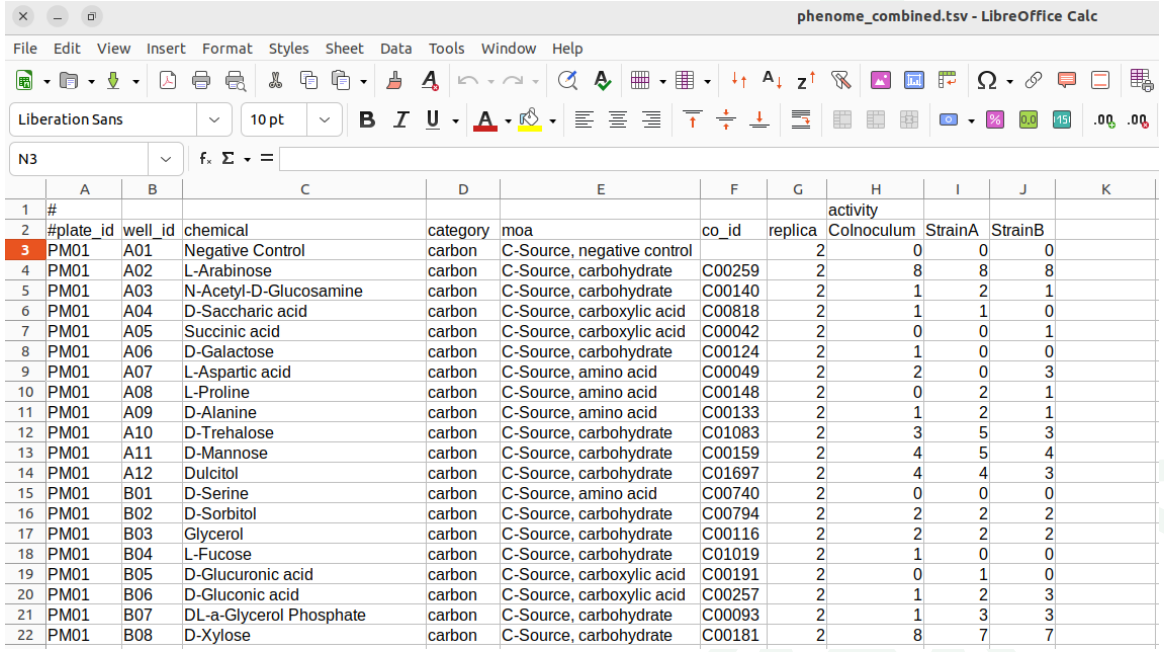


Those files and folder should have been created from the DuctApe pipeline





Contains the calculated AV data



	A	B	C	D	E	F	G	H	I	J	K
1	#							activity			
2	#plate_id	well_id	chemical	category	moa	co_id	replica	CoInoculum	StrainA	StrainB	
3	PM01	A01	Negative Control	carbon	C-Source, negative control		2	0	0	0	
4	PM01	A02	L-Arabinose	carbon	C-Source, carbohydrate	C00259	2	8	8	8	
5	PM01	A03	N-Acetyl-D-Glucosamine	carbon	C-Source, carbohydrate	C00140	2	1	2	1	
6	PM01	A04	D-Saccharic acid	carbon	C-Source, carboxylic acid	C00818	2	1	1	0	
7	PM01	A05	Succinic acid	carbon	C-Source, carboxylic acid	C00042	2	0	0	1	
8	PM01	A06	D-Galactose	carbon	C-Source, carbohydrate	C00124	2	1	0	0	
9	PM01	A07	L-Aspartic acid	carbon	C-Source, amino acid	C00049	2	2	0	3	
10	PM01	A08	L-Proline	carbon	C-Source, amino acid	C00148	2	0	2	1	
11	PM01	A09	D-Alanine	carbon	C-Source, amino acid	C00133	2	1	2	1	
12	PM01	A10	D-Trehalose	carbon	C-Source, carbohydrate	C01083	2	3	5	3	
13	PM01	A11	D-Mannose	carbon	C-Source, carbohydrate	C00159	2	4	5	4	
14	PM01	A12	Dulcitol	carbon	C-Source, carbohydrate	C01697	2	4	4	3	
15	PM01	B01	D-Serine	carbon	C-Source, amino acid	C00740	2	0	0	0	
16	PM01	B02	D-Sorbitol	carbon	C-Source, carbohydrate	C00794	2	2	2	2	
17	PM01	B03	Glycerol	carbon	C-Source, carbohydrate	C00116	2	2	2	2	
18	PM01	B04	L-Fucose	carbon	C-Source, carbohydrate	C01019	2	1	0	0	
19	PM01	B05	D-Glucuronic acid	carbon	C-Source, carboxylic acid	C00191	2	0	1	0	
20	PM01	B06	D-Gluconic acid	carbon	C-Source, carboxylic acid	C00257	2	1	2	3	
21	PM01	B07	DL-a-Glycerol Phosphate	carbon	C-Source, carbohydrate	C00093	2	1	3	3	
22	PM01	B08	D-Xylose	carbon	C-Source, carbohydrate	C00181	2	8	7	7	

https://combogenomics.github.io/DuctApe/tutorial.html

Phenomic data analysis in R + metabolic reconstruction

https://www.r-project.org/

https://posit.co/download/rstudio-desktop/

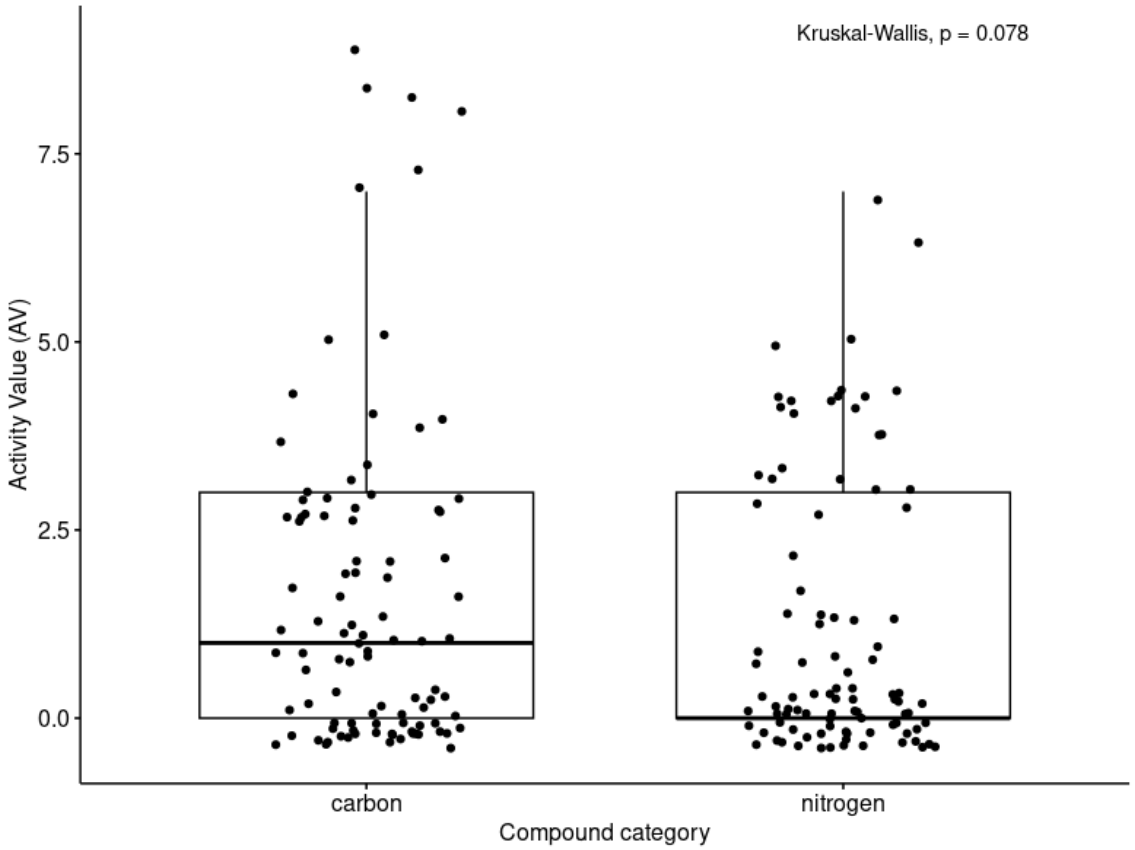
Phenomic data analysis in R + metabolic reconstruction

```r
# Loading needed libraries

library(tidyverse)
library(ggsci)
library(ggpubr)
library(reshape2)
library(ggside)
library(ggdist)
library(FactoMineR)
library(factoextra)
library(pathview)
```

```r
ductape_data <- read.table(file = "../1_DuctApe_workflow/phenome_combined.tsv",
                           header = F,
                           sep = "\t")

# Set column names

colnames(ductape_data) <- c("plate_id",
                            "well_id",
                            "chemical",
                            "category",
                            "moa",
                            "co_id",
                            "replica",
                            "CoInoculum",
                            "StrainA",
                            "StrainB")
```

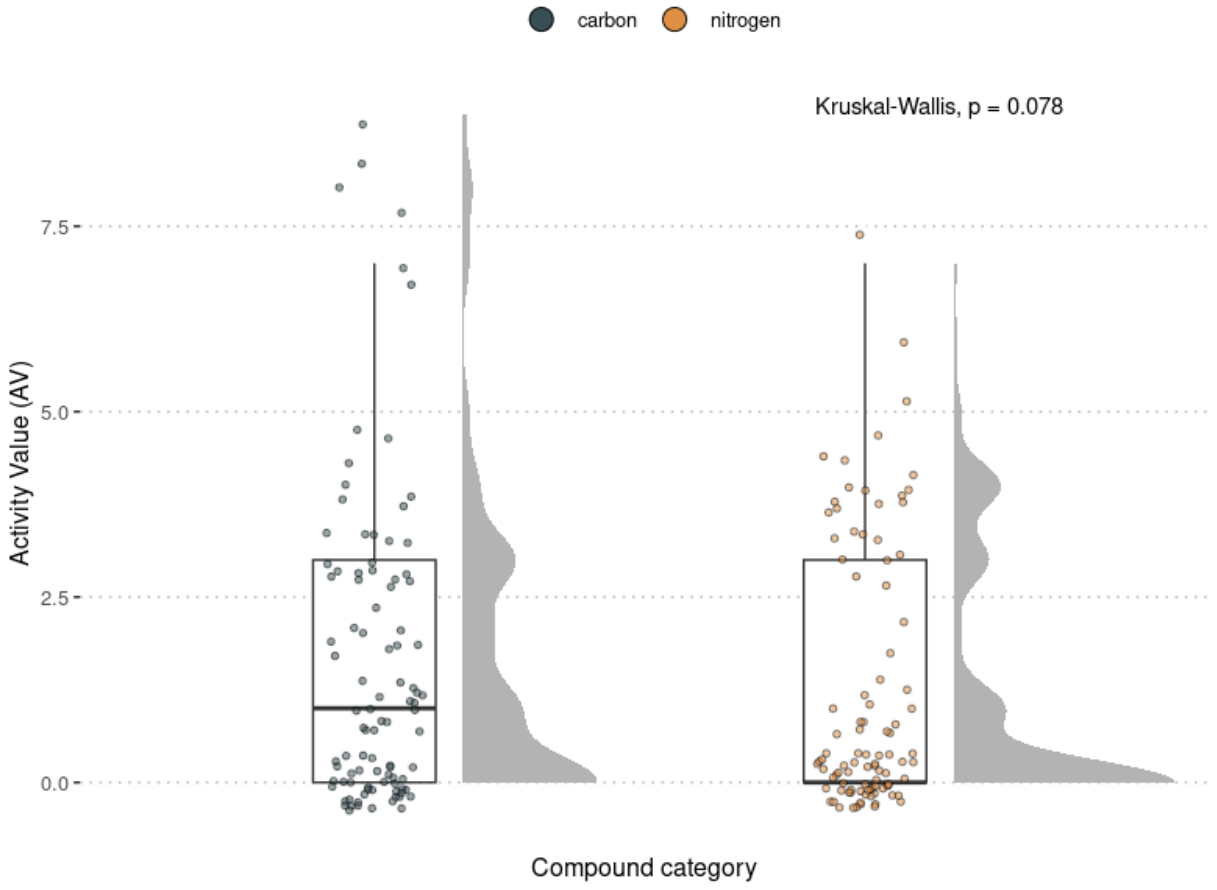
Phenomic data analysis in R + metabolic reconstruction

```
# Compare the obtained AV in the different compound category for the CoInoculum
# strain. The same plot could be obtained for the StrainA and StrainB by
# editing the "CoInoculum" part. We will also add a general Kruskal wallis test
# to know if AVs in the different compound categories are different for the
# selected strain

ductape_data %>%
  ggboxplot(x = "category", y = "CoInoculum", add = "jitter") +
  ylab("Activity Value (AV)") +
  xlab("Compound category") +
  stat_compare_means(method = "kruskal.test", label.x = 2)
```



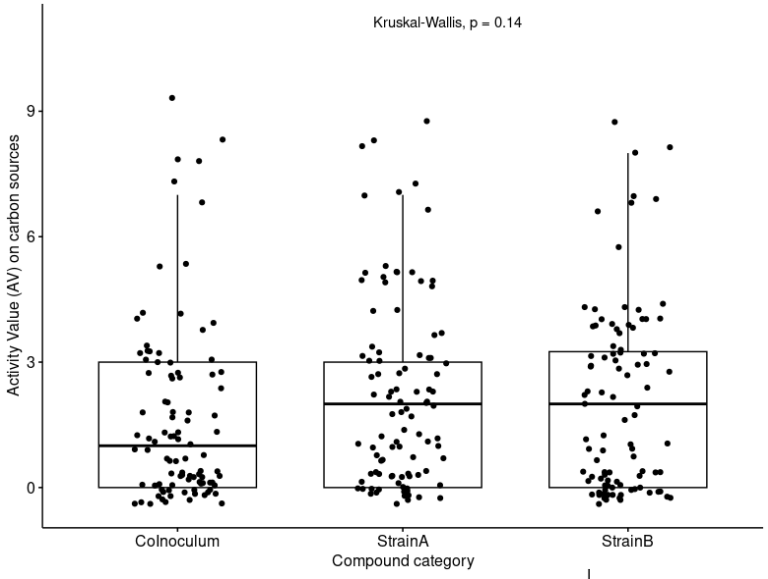
Phenomic data analysis in R + metabolic reconstruction

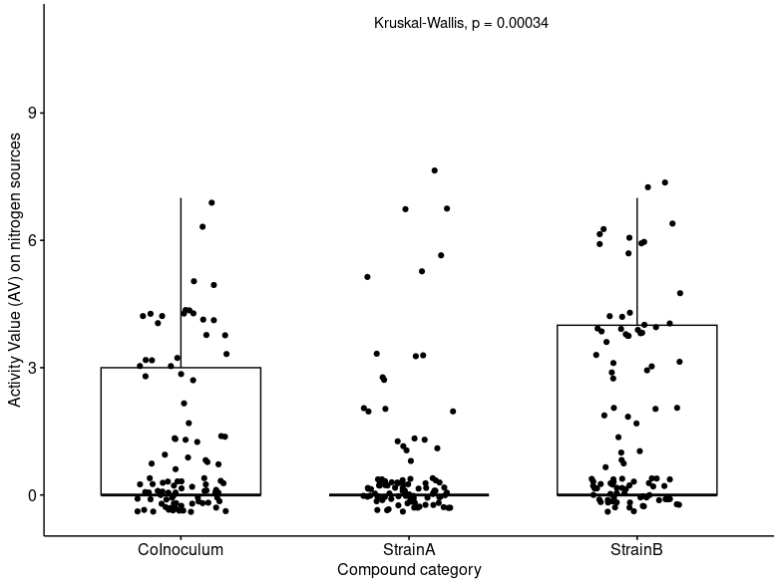
```r
ductape_data %>%
  ggplot(aes(x = category, y = CoInoculum)) +
  stat_halfeye(adjust = 0.5,
               width = 0.5,
               .width = 0,
               justification = -0.4,
               point_colour = NA,
               fill = "grey70") +
  geom_boxplot(width = .25,
               outlier.shape = NA,
               fill = "white") +
  geom_point(aes(fill = category),
             shape = 21,
             size = 1.3,
             alpha = .5,
             position = position_jitter(seed = 1, width = .1)) +
  theme_pubclean() +
  ylab("Activity Value (AV)") +
  xlab("Compound category") +
  scale_fill_jama() +
  stat_compare_means(method = "kruskal.test", label.x = 2) +
  theme(legend.position="top",
        legend.title = element_blank(),
        legend.box.background = element_blank(),
        legend.key = element_blank(),
        legend.key.size = unit(0.8, 'cm'),
        axis.ticks.x=element_blank(),
        axis.text.x=element_blank()) +
  guides(fill = guide_legend(override.aes = list(size = 5, alpha = 1)))
```



Phenomic data analysis in R + metabolic reconstruction

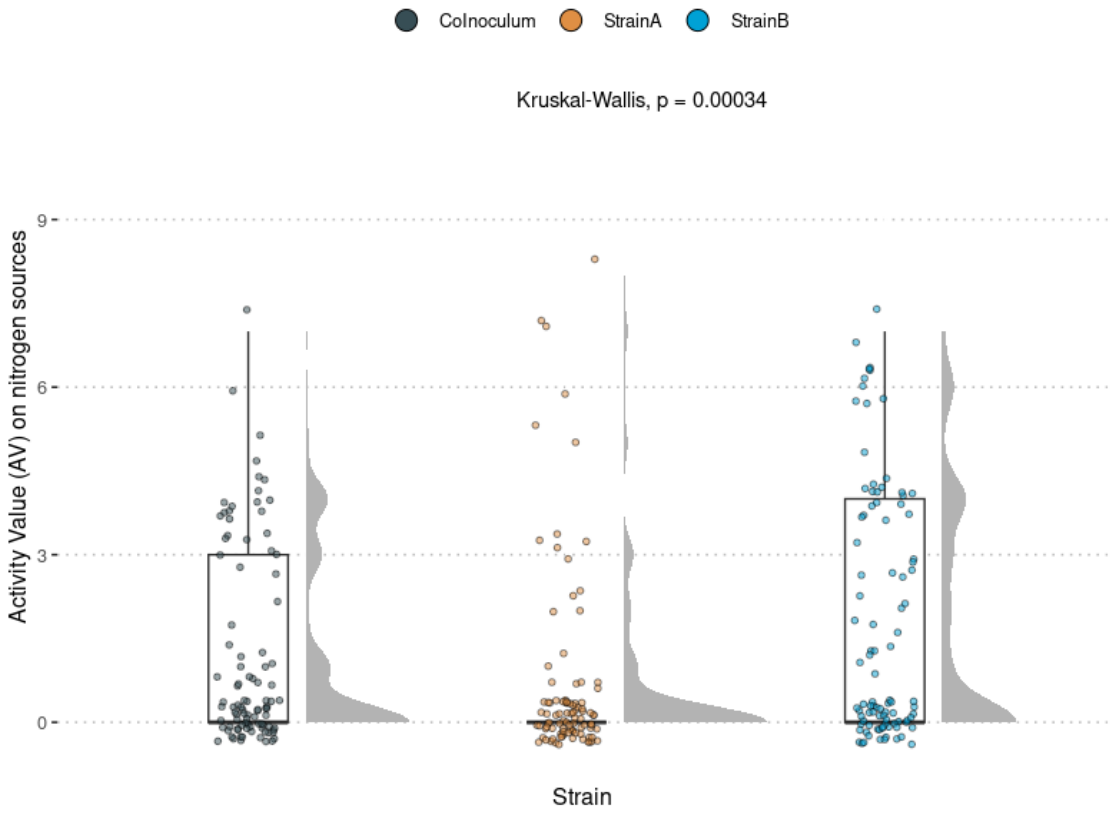
```r
# Here we test the Carbon sources
ductape_data %>%
  select(-replica) %>%
  melt() %>%
  filter(category == "carbon") %>%
  ggboxplot(x = "variable", y = "value", add = "jitter") +
  ylab("Activity Value (AV) on carbon sources") +
  xlab("Compound category") +
  stat_compare_means(method = "kruskal.test", label.x = 2, label.y = 11)

# Here we test the Nitrogen sources
ductape_data %>%
  select(-replica) %>%
  melt() %>%
  filter(category == "nitrogen") %>%
  ggboxplot(x = "variable", y = "value", add = "jitter") +
  ylab("Activity Value (AV) on nitrogen sources") +
  xlab("Compound category") +
  stat_compare_means(method = "kruskal.test", label.x = 2, label.y = 11)
```





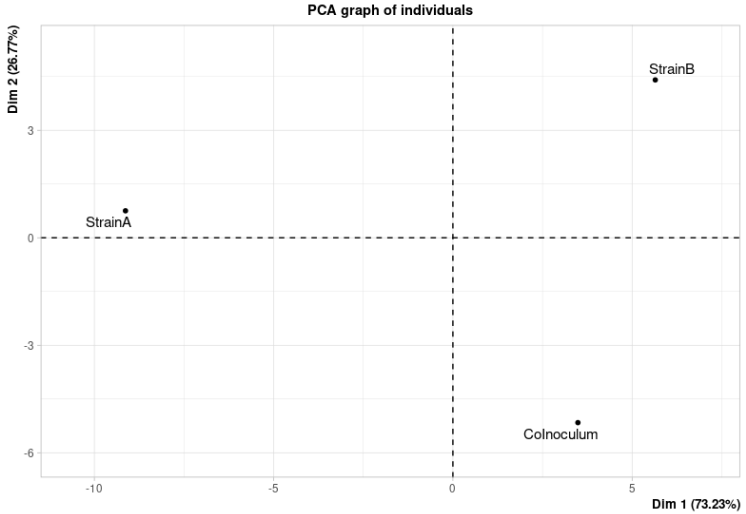
Phenomic data analysis in R + metabolic reconstruction

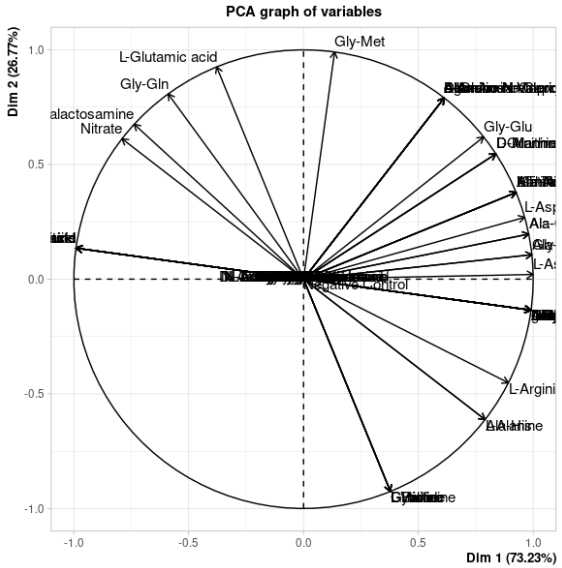
```r
ductape_data %>%
  select(-replica) %>%
  melt() %>%
  filter(category == "nitrogen") %>%
  ggplot(aes(x = variable, y = value)) +
  stat_halfeye(adjust = 0.5,
               width = 0.5,
               .width = 0,
               justification = -0.4,
               point_colour = NA,
               fill = "grey70") +
  geom_boxplot(width = .25,
               outlier.shape = NA,
               fill = "white") +
  geom_point(aes(fill = variable),
             shape = 21,
             size = 1.3,
             alpha = .5,
             position = position_jitter(seed = 1, width = .1)) +
  theme_pubclean() +
  ylab("Activity Value (AV) on nitrogen sources") +
  xlab("Strain") +
  scale_fill_jama() +
  stat_compare_means(method = "kruskal.test", label.x = 2, label.y = 11) +
  theme(legend.position="top",
        legend.title = element_blank(),
        legend.box.background = element_blank(),
        legend.key = element_blank(),
        legend.key.size = unit(0.8, 'cm'),
        axis.ticks.x=element_blank(),
        axis.text.x=element_blank()) +
  guides(fill = guide_legend(override.aes = list(size = 5, alpha = 1)))
```



Phenomic data analysis in R + metabolic reconstruction

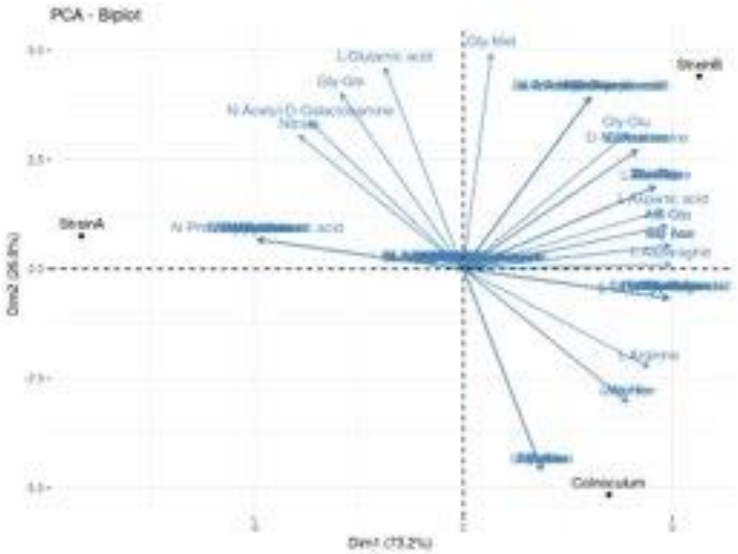
```
# Above analysis suggests no difference in how the two strains and the co-inoc
# experiment utilize Carbon sources, but a significant difference regarding the
# Nitrogen sources.
# We may futher explore the difference on Nitrogen sources with multivariate
# analysis. In this case, we use a Principal Component Analysis (PCA)

ductape_data %>%
  filter(category == "nitrogen") %>%
  select(c(3,8,9,10)) %>%
  column_to_rownames(var = "chemical") %>%
  t() %>%
  PCA() -> PCA_nitrogen
```

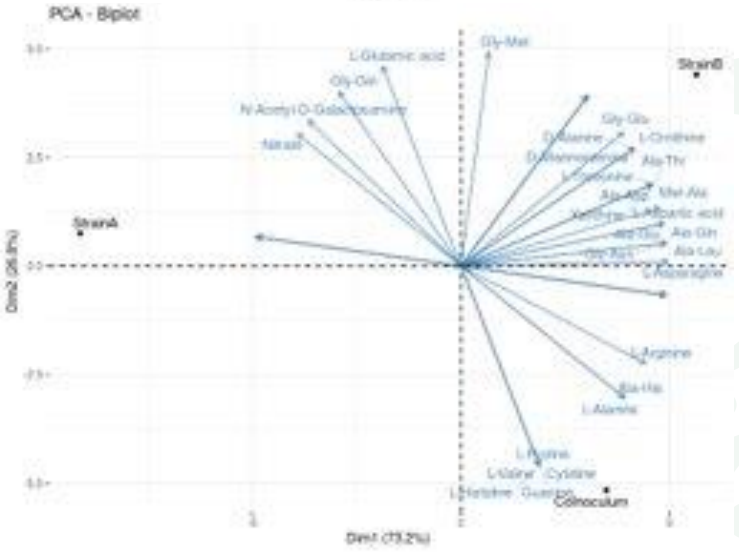




Phenomic data analysis in R + metabolic reconstruction

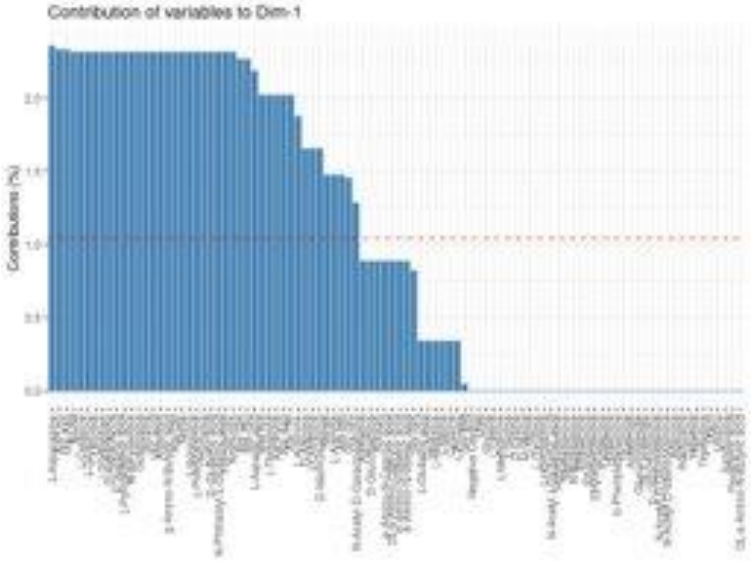
```
# The PCA function in the FactomineR package already produces plot. This can
# be avoided by adding graph = F in the PCA() command.
# Plotting is usually better done by using the "companion" package, which is
# factoextra. We will produce a biplot for the PCA ordination, which visualize
# in the same graphics both the samples (black points) and the variables (blue
# arrows). This is useful for interpretation of the results, as we can know
# which variable is more strongly contributing to the difference that we see in
# the samples. The arrows (which are usually named vectors) represent the
# direction in which each variable determine the placement of samples. Each
# variable "pull" the ordination in some direction

# Basic biplot
fviz_pca_biplot(X = PCA_nitrogen)

# The argument "repel = T" allows to avoid overlapping labels of variables, but
# some label are lost.
fviz_pca_biplot(X = PCA_nitrogen,
                repel = T)
```

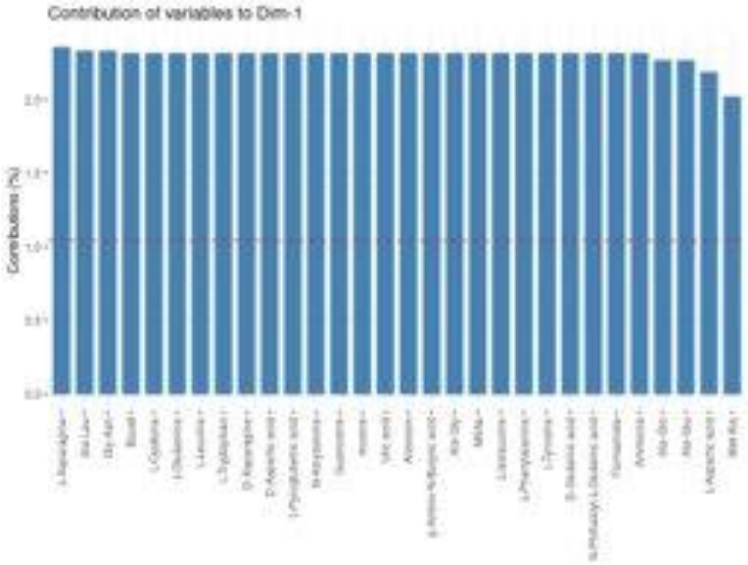




Phenomic data analysis in R + metabolic reconstruction

```r
# The fviz_contrib() fuction can be used to inspect variable contribution to
# each axis.
fviz_contrib(X = PCA_nitrogen,
             xtickslab.rt = 90,
             choice = "var",
             axes = 1) # change with 2 for second axis

fviz_contrib(X = PCA_nitrogen,
             xtickslab.rt = 90,
             choice = "var",
             top = 30, # set the number of variables to plot
             axes = 1) # change with 2 for second axis
```





Phenomic data analysis in R + metabolic reconstruction

```
# If genomic data are available for the strain, we may explore the genomic
# basis for the phenomic data observed by using Kegg map. This approach
# takes a specific pathway of interest; in this example the map00910 for the
# Nitrogen metabolism, as we have evidences of different use of nitrogen
# sources between the two strains. Compound from the PM plate which are included
# in the kegg pathway will be colored based on observed AV. Orthologs genes
# from the strain genome which are included in the kegg pathway will be colored
```

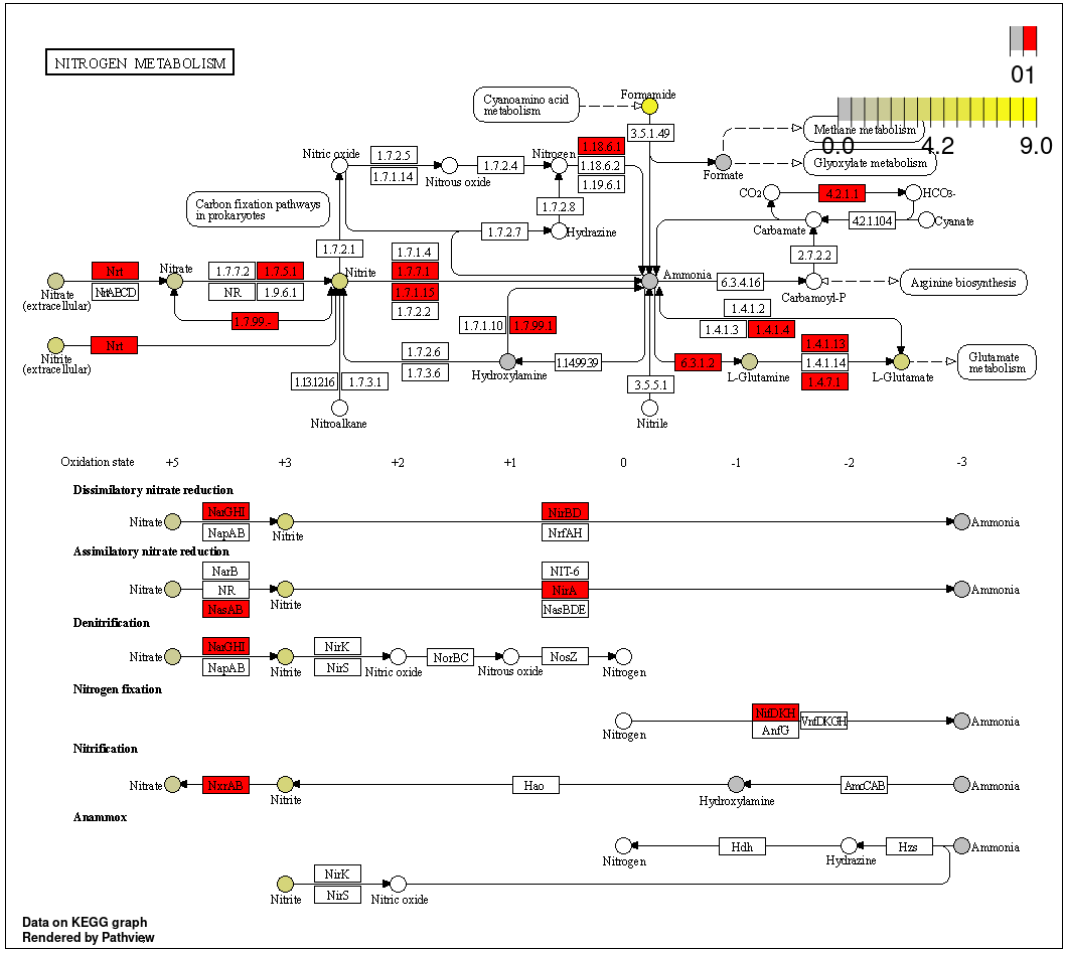
```
# Import gene annotation data. This contains mapping of protein sequences to
# the Kegg ortholog. Can be obtained using the KAAS service
# (https://www.genome.jp/kegg/kgas/) while protein sequences can be obtained
# by any pipeline for assembled genome annotation. As the function for plotting
# also allow to record the copy numbers for each orthologs in the genome, we
# include a column with all values = 1
StrainA_gene <- read.table(file = "./StrainA_KAAS.csv",
                           header = T,
                           sep = ",")

# we need to build a "named vector" for input in the function below
gene_data_strainA <- StrainA_gene$Presence
names(gene_data_strainA) <- StrainA_gene$Ortholog
```

```
# From the DuctApe output we can obtain the AV of each compound, and name it
# using the column "co_id" which is the compound code in Kegg
cpd_data_strainA <- ductape_data$StrainA
names(cpd_data_strainA) <- ductape_data$co_id
```

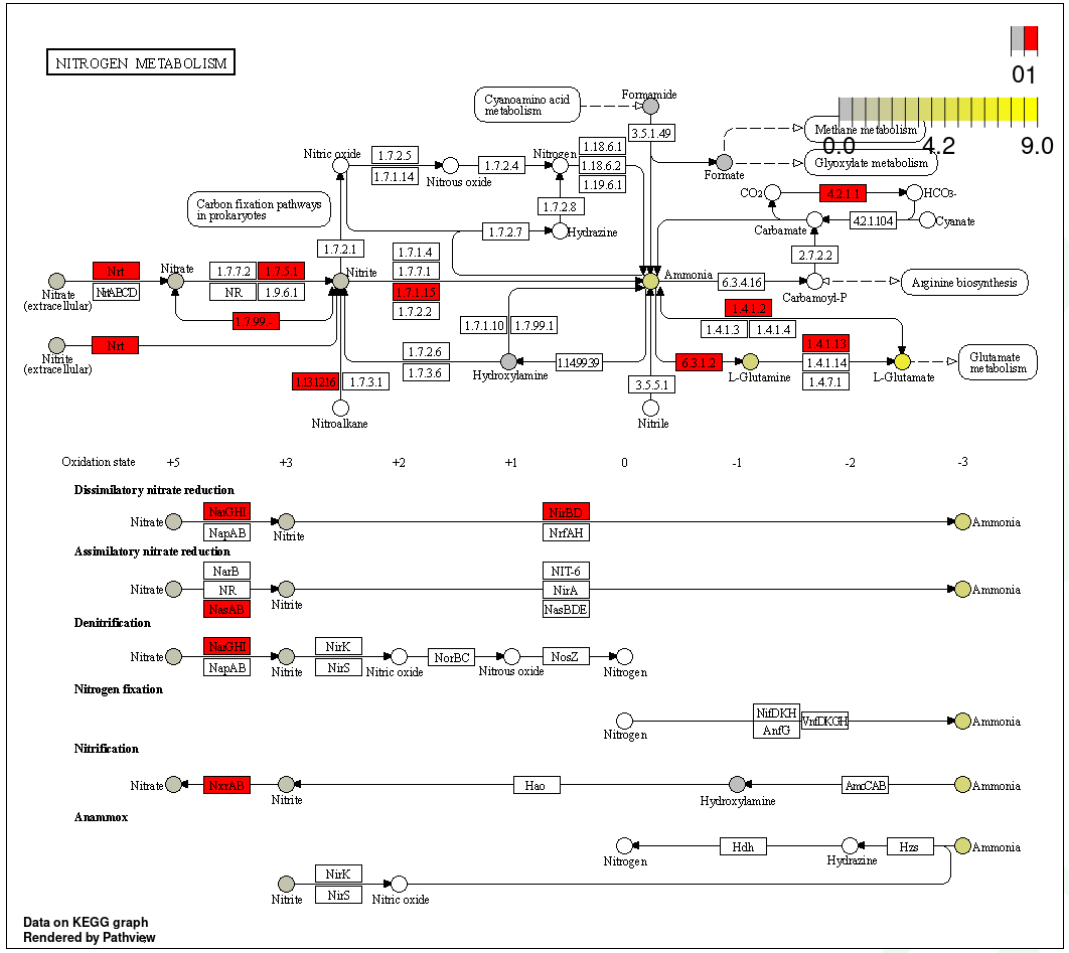
```
# Here we obtain the annotated Kegg map.
pv.out.N <- pathview(gene.data = gene_data_strainA,
                     cpd.data = cpd_data_strainA,
                     both.dirs = list(gene = FALSE, cpd = FALSE),
                     bins = list(gene = 1, cpd = 15),
                     discrete = list(gene = TRUE, cpd = FALSE),
                     limit = list(gene = 1, cpd = 9),
                     species = "ko",
                     cpd.idtype = "kegg",
                     gene.idtype = "KEGG",
                     pathway.id = "00910",
                     out.suffix = "strainA.N",
                     keys.align = "y",
                     kegg.native = T,
                     key.pos = "topright")
```

Phenomic data analysis in R + metabolic reconstruction

Strain A (20 genes)



Strain B (15 genes, but higher AV)

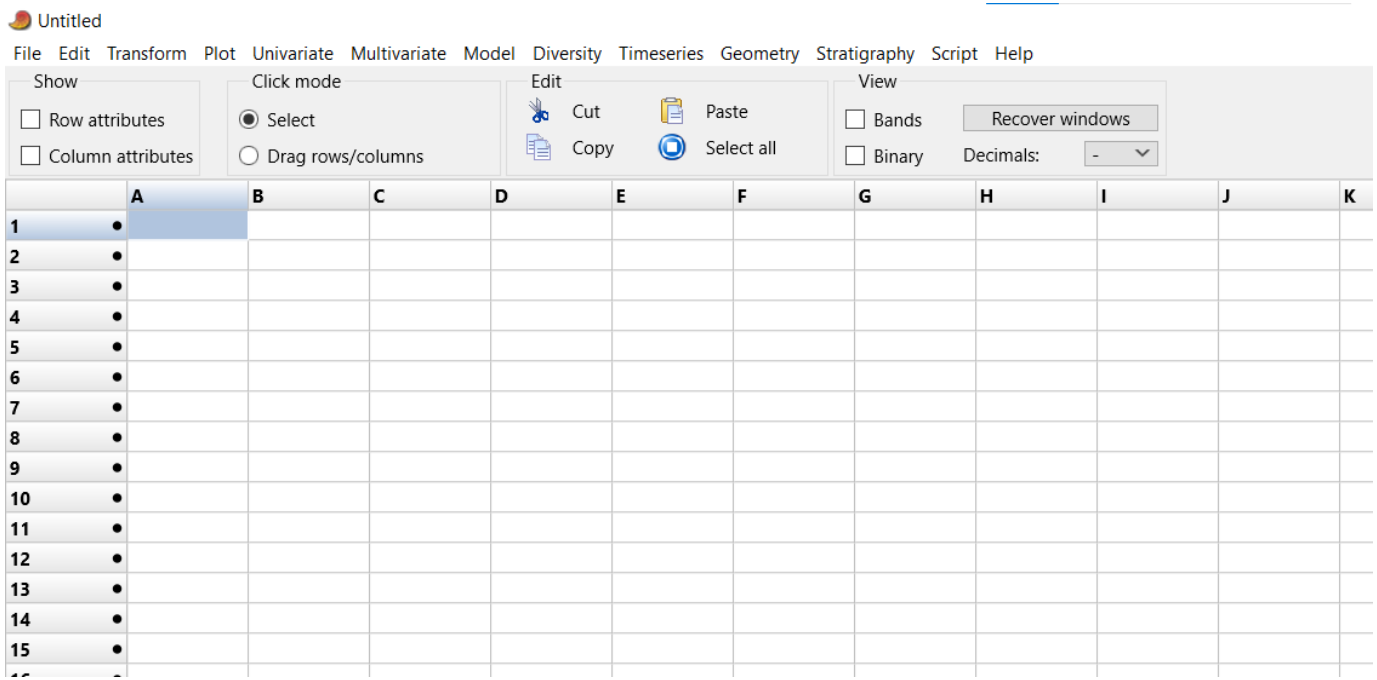


Phenomic data analysis in PAST

https://www.nhm.uio.no/english/research/resources/past/

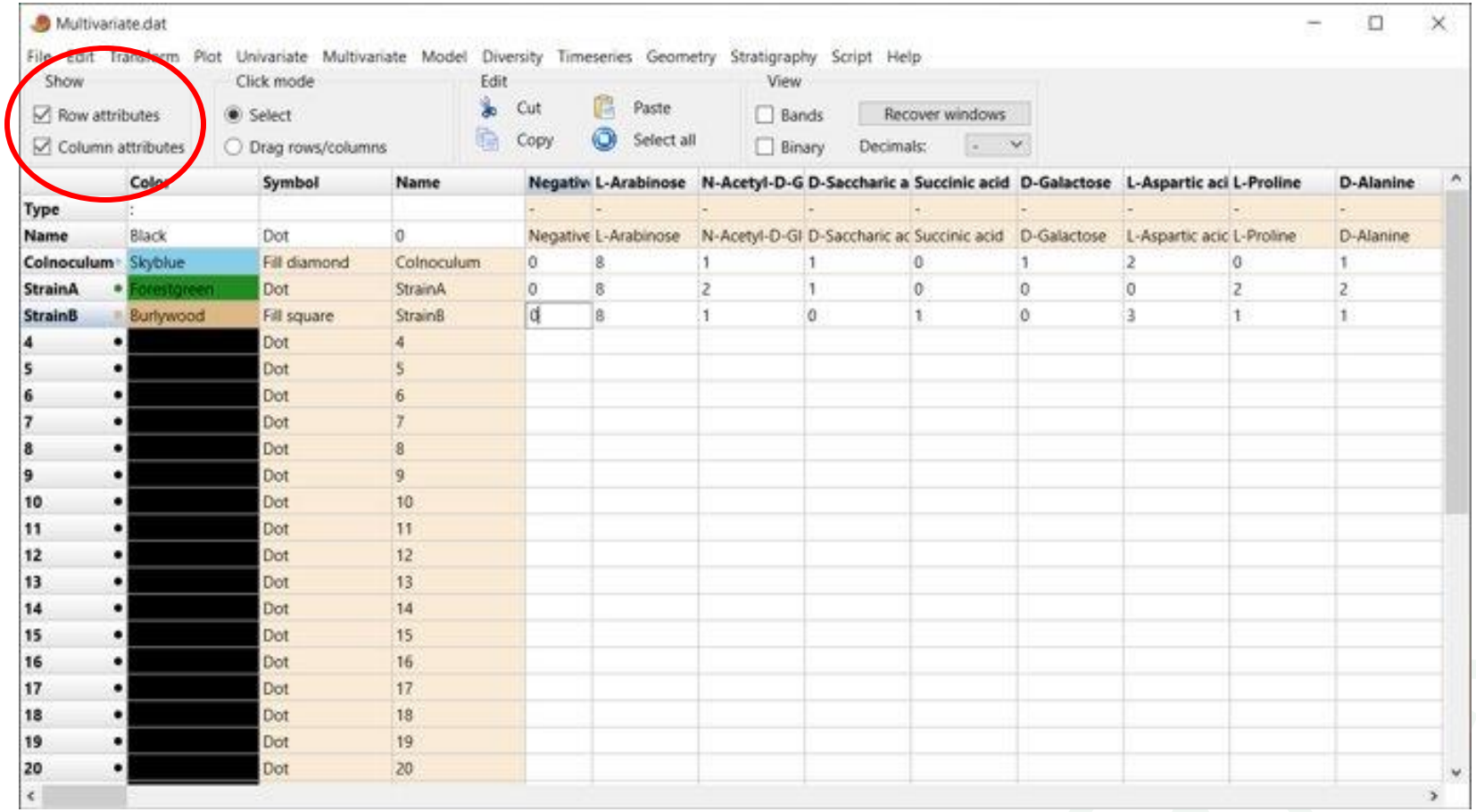
Phenomic data analysis in PAST

- Free, spreadsheet-like software for statistical analysis, tuned to ecological applications
- Runs in Windows and Mac (not so well on UNIX, tested with PalyOnLinux)



Phenomic data analysis in PAST

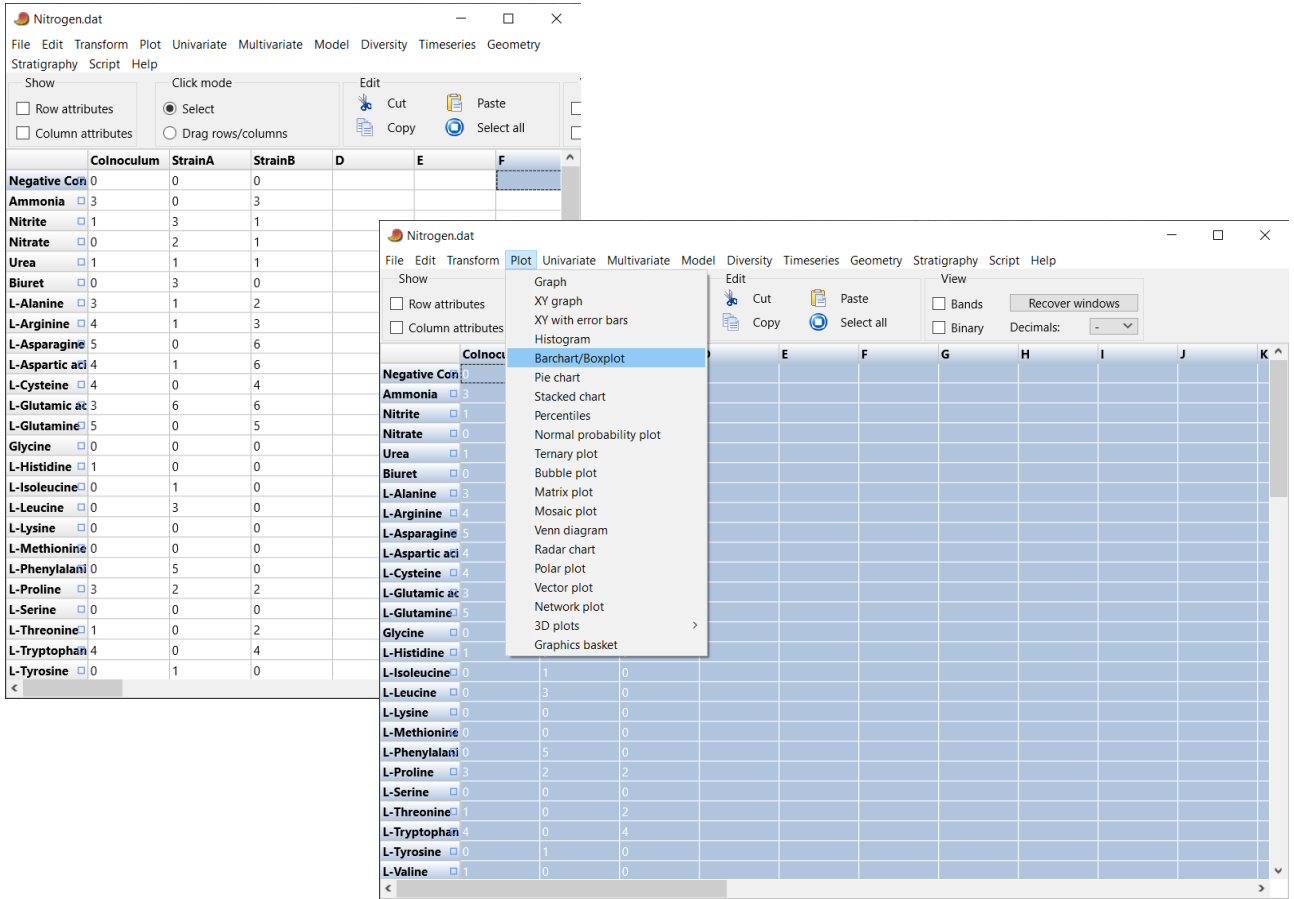
- Data can be copy-pasted from Excel (white cells are data)
- Meta-data can be included by checking the "Row attributes" and/or "Column attributes" mark. Yellow cells are used for variables and samples name, and to assign samples to groups

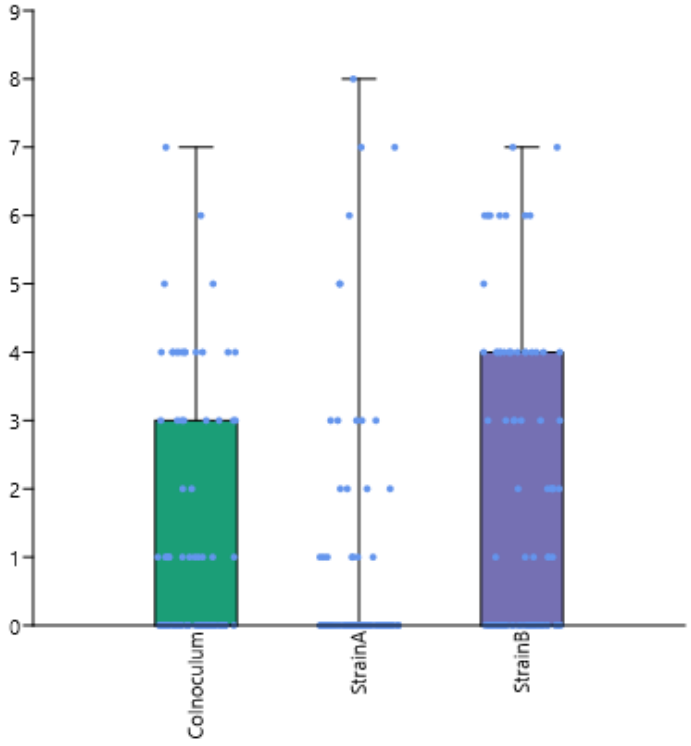


	Colo	Symbol	Name	Negativ	L-Arabinose	N-Acetyl-D-G	D-Saccharic a	Succinic acid	D-Galactose	L-Aspartic aci	L-Proline	D-Alanine
Type	:			-	-	-	-	-	-	-	-	-
Name	Black	Dot	0	Negative	L-Arabinose	N-Acetyl-D-Gl	D-Saccharic ac	Succinic acid	D-Galactose	L-Aspartic acic	L-Proline	D-Alanine
Colnoculum	Skyblue	Fill diamond	Colnoculum	0	8	1	1	0	1	2	0	1
StrainA	Forestgreen	Dot	StrainA	0	8	2	1	0	0	0	2	2
StrainB	Burlywood	Fill square	StrainB	0	8	1	0	1	0	3	1	1
4		Dot	4									
5		Dot	5									
6		Dot	6									
7		Dot	7									
8		Dot	8									
9		Dot	9									
10		Dot	10									
11		Dot	11									
12		Dot	12									
13		Dot	13									
14		Dot	14									
15		Dot	15									
16		Dot	16									
17		Dot	17									
18		Dot	18									
19		Dot	19									
20		Dot	20									

Phenomic data analysis in PAST

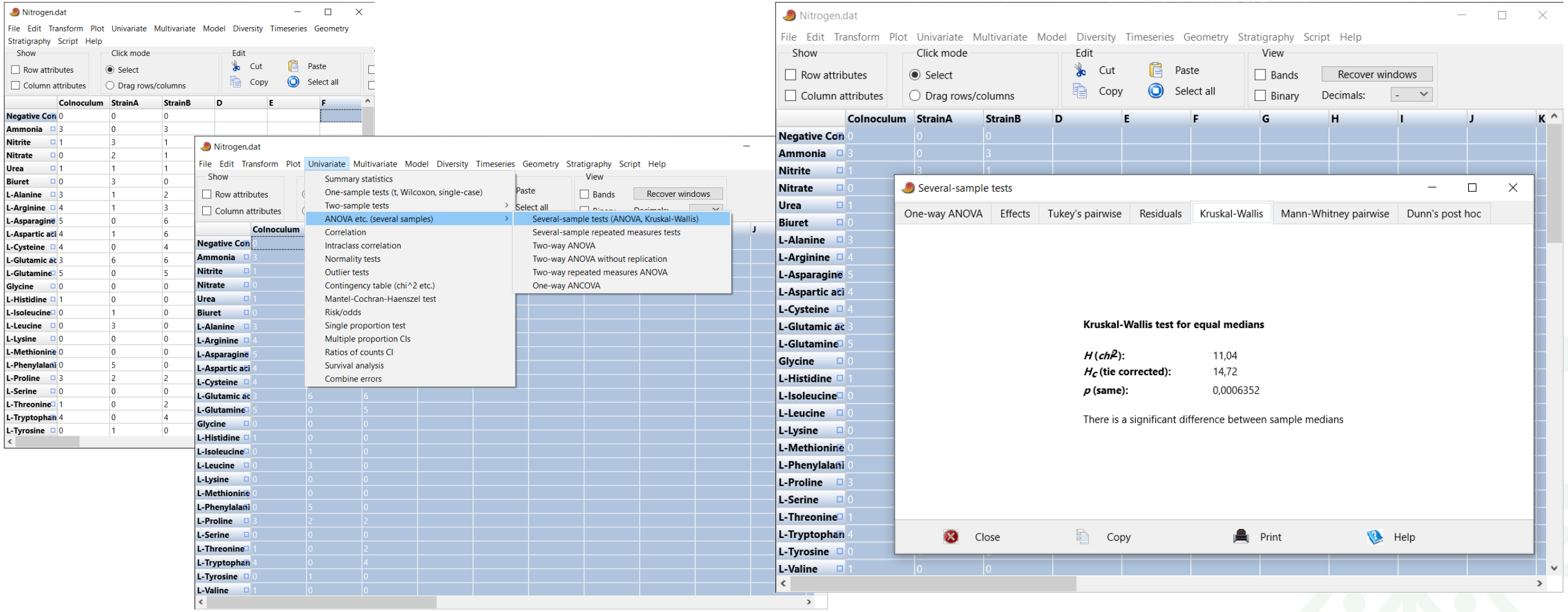
- Boxplot to visualize differences in AVs on the Nitrogen compounds between strains
- Attention on the orientation of data, for multivariate analysis we need need the "large" format (rows = samples; column = variables), for univariate analysis the "long" format (rows = columns; column = samples)





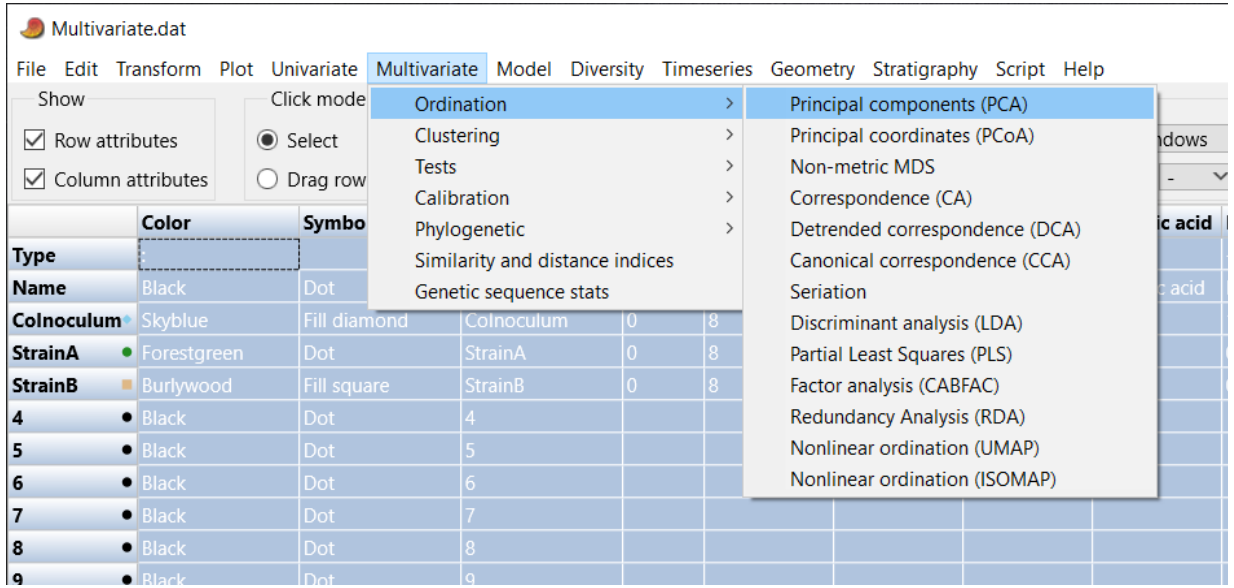
Phenomic data analysis in PAST

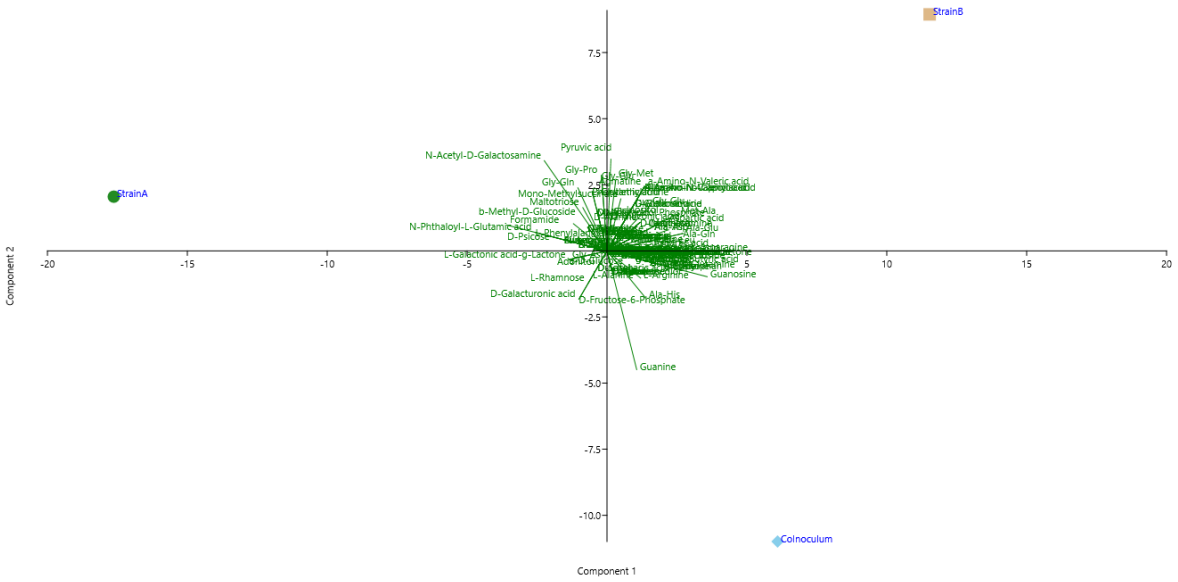
- Kruskal-Wallis test for the difference in AVs on the Nitrogen compound between strains



Phenomic data analysis in PAST

- PCA ordination analysis (or many other multivariate methods)
- In the same menu, also clustering analysis (i.e. UPGMA dendrograms)





EXCALIBUR

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 817946

Vitali Francesco
francesco.vitali@crea.gov.it
Github: FrancescoVit

Thanks for you attention