water

MDPI

Supplementary Material

Ecological Study of Aquaponics Bacterial Microbiota over the Course of a Lettuce Growth Cycle

Mathilde Eck , Iris Szekely , Sébastien Massart and M. Haïssam Jijakli *

Integrated and Urban Plant Pathology Laboratory, Gembloux Agro–Bio Tech, University of Liège, Passage des Déportés 2, 5030 Gembloux, Belgium; mathilde.eck@uliege.be (M.E.); iris.szekely@uliege.be (I.S.); sebastien.massart@uliege.be (S.M.)

* Correspondence: mh.jijakli@uliege.be; Tel.: +32-(0)81-62-2431

† These authors contributed equally to this work.

Citation: Eck, M.; Szekely, I.; Massart, S.; Jijakli, M. H. Ecological Study of Aquaponics Bacterial Microbiota over the Course of a Lettuce Growth Cycle. *Water* **2021**, *13*, 2089. https://doi.org/10.3390/w13152089

Academic Editor: Abasiofiok Mark Ibekwe

Received: 24 June 2021
Accepted: 28 July 2021
Published: 30 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

Disease symptoms observed on the first group of lettuces during the fifth week of growth

Details on the hierarchical clustering and PCA procedures

1. Pre-treatment of the data

For each parameter, the average of the data was calculated for each day. Those mean values per day were then used for the rest of the analysis after being standardised, i.e. a mean of zero and a variance of 1, with the command scale().



a **b**

Figure S1. Disease symptoms observed on the shoot (a) and roots (b) parts of the first group of lettuces during the fifth week of growth.

2. Hierarchical clustering analysis

This part aimed to differentiate each date of the experiment depending on physico-chemical parameters and to create clusters of highly homogeneous dates [48]. To do so, the first step was to measure the Euclidian distances matric with the command dist(), which characterises the dissimilarity between two dates in the variable space [48].

The dates were then clustered with the hclust() command using the complete linkage method, which calculated the distance between the two most distant dates of two separate clusters, for all clusters. The resulting classification was visualised in the form of a dendrogram with the command plot(), which provides an image of the groups and their proximity (Figure S2) [48].

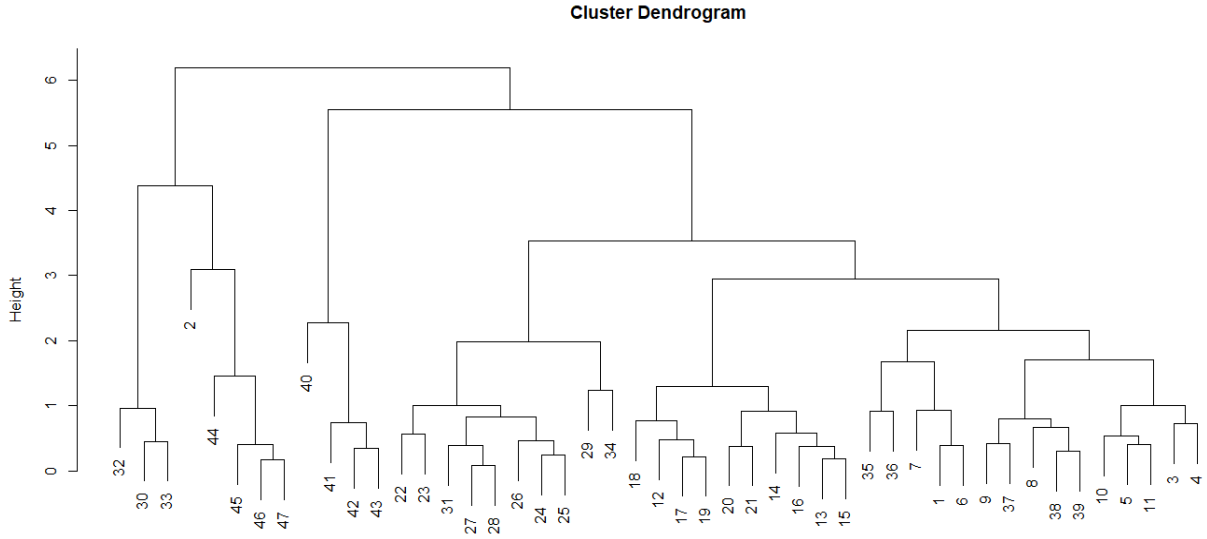


Figure S2. Cluster dendrogram generated by the command plot(), each number representing a day of the experiment (n° 1: March 8; n°48: May 13).

A partition then needed to be established, corresponding to the number of retained clusters. To do so, the R2 criteria provided by the command hclust.rsq() and the prsq plot() was used (FigureS3). R2 is the result of the sum of squared deviations between the groups divided by the sum of squared deviations of all parameters. The higher this value, the more initial information is retained after the clustering. Thus, the partition was chosen by looking at the successive clustering. The clustering that resulted in a strong R2 decrease and therefore, a big loss of information between the groups was spotted and the fusions were interrupted before this clustering. In this way, a partition made of 5 groups was retained.

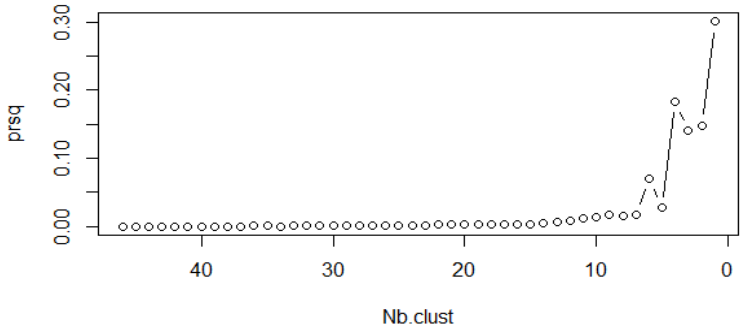


Figure S3. Graph generated by the command prsq plot(), representing the R2 decrease at each individuals fusion, i.e. dates fusion.

The robustness of the partition was checked using an alternative partition method, i.e. the centroid method. Hence, the gravity centre of each group was calculated with the command aggregate() using the standardised data. The centroid classification was then made with the command kmeans() using the calculated centroids, which aimed to transfer some individuals between groups in order to improve the classification. The two obtained classifications were then compared thanks to the command table() (FigureS4). Since the two classifications did not differ much, we concluded that the chosen partition of 5 clusters was robust and it was thus retained.

group	1	2	3	4	5
1	25	0	0	0	0
2	0	4	0	0	1
3	0	0	10	0	0
4	0	0	0	3	0
5	0	0	0	0	4

Figure S4. Table showing the comparison of the 5 clusters generated by the two classifications.

The chosen partition was visualised on the dendrogram with the command cutree() (FigureS5) and characterised in term of composition with the sort() function, in term of individuals with table() and in term of mean and standard deviation with aggregate() command.

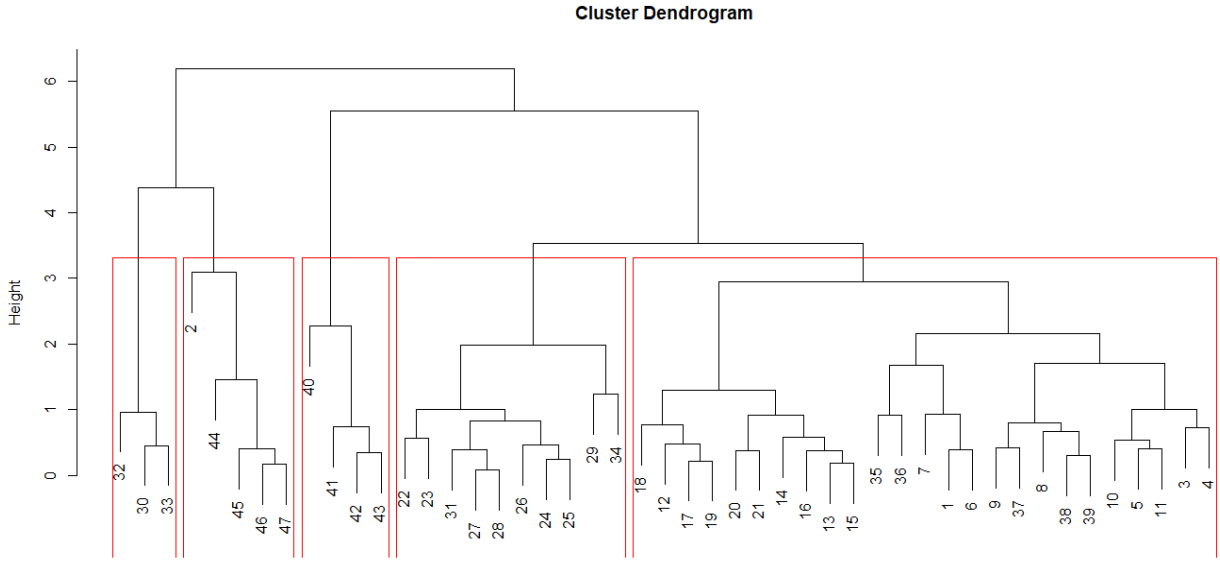


Figure S5. Cluster dendrogram showing the partition made of the 5 physicochemical groups.

3. Visualisation of the clusters via PCA

In order to describe the identified physicochemical groups in a multivariate space, the data were subjected to a principal component analysis (PCA). This enabled to reduce the number of variables, i.e. parameters, and to visualise the clusters in the factorial plans retained by the PCA. To do so, the FactoMineR package was loaded and the PCA() command was carried out on standardised data. In order to determine the number of retained factorial plans, the eigen value graph was generated with the command plot() (FigureS6).

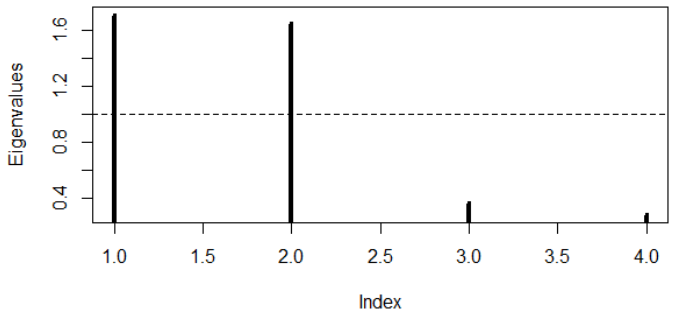


Figure S6. Eigen values graph.

According to this graph, only the first factorial plan made of the principal components 1 and 2 was retained, as both of them were highly superior to the mean, i.e. 1 since the data were standardised, which was not the case for the components 3 and 4.

The variable factor map showing the correlations between each parameter to the two axes, i.e. principal components 1 and 2, was generated thanks to the plot() command (Figure S7).

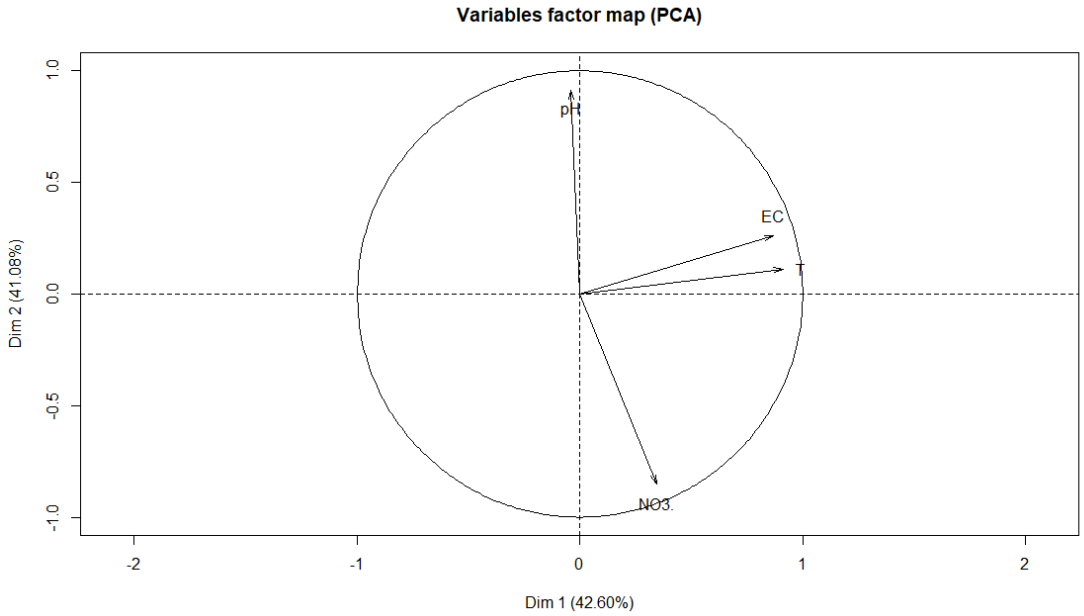


Figure S7. Variables factor map of the first factorial plan.

The physicochemical clusters were then visualised on the first factorial plan using the command rainbow(), that assigns a colour to each cluster, and the plot() function. The centroids of the groups were also added to the plan by using the function PCA() on another data set made of the clusters' identifiers and the standardised data.

In order to interpret the data in the factorial map and to understand the two principal components, the correlations of each parameter to the 2 new axes were extracted from the PCA object. The command dimdesc() facilitate the understanding by sorting the correlations in descendent order for each principal component. The quality of the representation of the parameters in the factorial map was also extracted from the PCA object.

Q-values of pairwise Kruskal-Wallis test performed to compare the diversity indices (Shannon and observed-otus) of the four compartments

- Shannon

Table S1. Q-values of pairwise Kruskal-Wallis test performed on the Shannon indices of the four compartments.

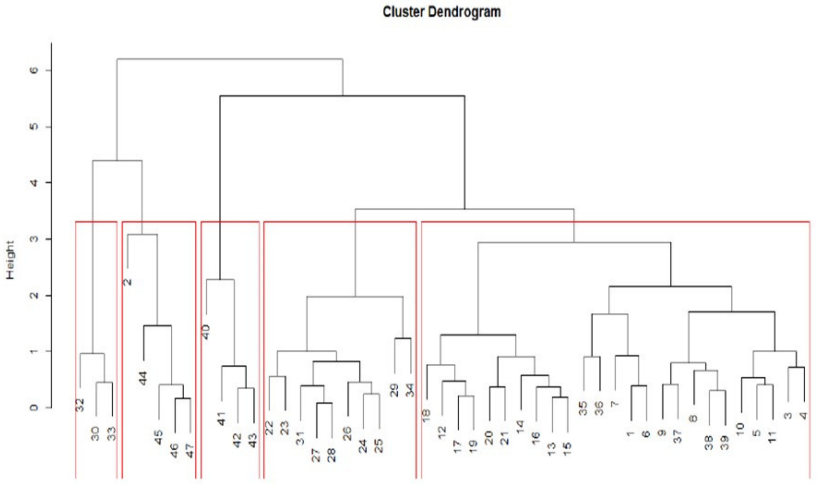
q-values	Biofilter	Sump	Rhizoplane	Root microbiota
Biofilter	/	0.0005	0.0007	0.4288
Sump	/	/	0.0005	0.0007
Rhizoplane	/	/	/	0.0008
Root microbiota	/	/	/	/

- Observed_otus

Table S2. Q-values of pairwise Kruskal-Wallis test performed on the Observed-otus indices of the four compartments.

q-values	Biofilter	Sump	Rhizoplane	Root microbiota
Biofilter	/	0.0452	0.0152	0.0012
Sump	/	/	0.0012	0.5750
Rhizoplane	/	/	/	0.0012
Root microbiota	/	/	/	/

Hierarchical clustering of water parameters and details of the water parameters for each group



N°	Date	16	09.04	32	28.04
1	08.03	17	10.04	33	29.04
2	11.03	18	11.04	34	30.04
3	12.03	19	12.04	35	01.05
4	13.03	20	13.04	36	02.05
5	14.03	21	14.04	37	03.05
6	15.03	22	18.04	38	04.05
7	18.03	23	19.04	39	05.05
8	19.03	24	20.04	40	06.05
9	21.03	25	21.04	41	07.05
10	26.03	26	22.04	42	08.05
11	28.03	27	23.04	43	09.05
12	05.04	28	24.04	44	10.05
13	06.04	29	25.04	45	11.05
14	07.04	30	26.04	46	12.05
15	08.04	31	27.04	47	13.05

Figure S8. Cluster dendrogram showing the repartition in 5 groups (**a**) of sampling dates based on water parameters and correspondence between date and day number (**b**).

Table S3. Physicochemical groups' characteristics. $\bar{x}$: mean; σ: standard deviation; CV: coefficient of variation; Min: minimum; Max: maximum

.Group	Individuals		EC	pH	T	NO_3^-
Group A	25	$\bar{x}$	1317.83	7.88	24.29	208.83
		σ	40.02	0.08	0.61	11.63
		CV	0.03	0.01	0.03	0.06
		Min	1241.00	7.72	23.20	179.21
		Max	1389.46	8.04	25.15	234.62
Group B	5	$\bar{x}$	1168.10	7.68	24.15	200.35
		σ	34.27	0.21	0.19	14.04
		CV	0.03	0.03	0.01	0.07
		Min	1137.50	7.52	22.95	186.24
		Max	1227.00	8.07	23.40	219.20
Group C	3	$\bar{x}$	1223.23	8.09	20.80	124.49
		σ	10.72	0.02	0.52	1.32
		CV	0.01	0.002	0.03	0.01
		Min	1211	8.07	20.33	123.57
		Max	1231.00	8.11	21.35	126.00
Group D	10	$\bar{x}$	1317.47	8.06	24.41	134.82
		σ	9.92	0.06	0.66	11.37
		CV	0.01	0.01	0.03	0.08
		Min	1302.97	7.91	23.20	124.21
		Max	1330.24	8.13	25.06	155.58
Group E	4	$\bar{x}$	1304.12	7.45	23.32	228.96
		σ	11.99	0.21	0.12	5.84
		CV	0.01	0.03	0.01	0.03
		Min	1288.50	7.14	23.15	222.00
		Max	1317.50	7.60	23.40	235.66

References

48. Kazi, T.G., Arain, M.B., Jamali, M.K., Jalbani, N., Afridi, H.I., Sarfaz, R.A., Baig, J.A., and Shah, A.Q. (2009). Assessment of water quality of polluted lake using multivariate statistical techniques: a case study. Ecotoxicol. Environ. Saf. *72*, 301–309.