

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

COLONISATION DE SURFACES ROCHEUSES PAR DES MICROORGANISMES DANS UN AQUIFÈRE À
COVEY HILL

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

AU PROGRAMME DE MAÎTRISE EN BIOLOGIE

PAR

DIVYA PATEL

SEPTEMBRE 2023

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.12-2023). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Je veux d'abord donner un gros merci à ma directrice Cassandra LAZAR pour m'avoir donné l'opportunité de faire cette maîtrise et pour sa patience. Je veux aussi remercier tous ceux qui m'ont aidé durant l'échantillonnage: Benjamin Groult, Karine Villeneuve et Élise Lhoste. Je tiens aussi à remercier Marie Larocque et Sylvain Gagné pour m'avoir donné accès au puits à Covey Hill. Merci à Karine et Élise pour m'avoir aidé avec les scripts sur RStudio et un gros merci à Katel Hallaf d'avoir pris le temps de me montrer comment faire une dPCR. L'expérience était épuisante à certains moments, donc c'est toujours bien d'avoir des collègues de bureau qui sont là pour t'écouter et garder une atmosphère positive. Donc merci à Benjamin, Karine et Élise pour s'assurer que j'allais bien même si j'avais l'air d'un zombie.

Merci à mon frère qui n'hésitait pas à venir me chercher à l'UQAM à 23h-24h juste pour éviter que je prenne le métro. Et je remercie finalement mes amies qui m'ont encouragé à continuer et de persévérer.

Contents

REMERCIEMENTS	ii
LISTE DES FIGURES	vi
LISTE DES TEABLEAUX.....	viii
LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES	ix
RÉSUMÉ	x
ABSTRACT	xi
INTRODUCTION.....	12
1. La diversité des communautés microbiennes dans les aquifères	12
2. Changement de mode de vie planctonique à sessile	14
3. Colonisation et croissance de microorganismes sur les roches des aquifères	14
4. Objectifs et Hypothèses.....	16
Chapitre 1: Rock surface colonization by groundwater microorganisms in an aquifer system in Quebec, Canada	17
1.1 SUMMARY	17
1.2 INTRODUCTION	18
1.3 MATERIAL and METHODS	19
1.3.1 Study site	19
1.3.2 Rock chip preparation and properties measurements	21
1.3.3 Bioreactor incubation setup	22
1.3.4 Sample collection during the incubations in the bioreactors and DNA extractions	22
1.3.5 DNA sequencing and bioinformatics analysis.....	24
1.3.6 Digital PCR amplifications (dPCR)	24
1.3.7 Processing the 16S and 18S rRNA gene sequences	24

1.3.8	Statistical analyses	25
1.4	Results	26
1.4.1	Geochemical water and rock properties	26
1.4.2	Rock-attached bacterial absolute abundance and correlation with the rock properties.....	28
1.4.3	16S and 18S rRNA gene taxonomic composition.....	28
1.4.4	Alpha diversity and correlation with environmental parameters	30
1.4.5	Beta diversity ordination analyses and influence of environmental variables	32
1.4.6	Beta diversity correlation with the water geochemical properties and the rock mineralogy and roughness.....	34
1.4.7	Discriminative genera explaining differences between the planktonic and sessile communities.....	36
1.4.8	Microbial Source Tracking	38
1.5	Discussion.....	41
1.5.1	Evolution of planktonic communities over time during the incubations.....	41
1.5.2	Microbial colonization of the rock surfaces	45
1.5.3	Influence of the rock properties on surface-attached microbial communities.....	47
1.6	Conclusion	49
	DISCUSSION GÉNÉRALE	50
	Annexe A: Supplementary figures.....	53
	Annexe B: Supplementary tables	67
	Annexe C: PCR and dPCR primers	78
	Annexe D: PCR and dPCR conditions.....	79
	Annexe E: Scripts.....	80
	Annexe F: 3D Scanning methodology and chip isolation	105

Annexe G: Kit methodology	108
Références	110

LISTE DES FIGURES

Figure 1. Map of the aquifer sampling site in Quebec, Canada (using the Qgis software).	21
Figure 2. Average of the different geochemical parameter in each bioreactor	27
Figure 3. Linear regressions of the Roughness based on the bacterial abundance represented here by the number of gene copies in different samples.	28
Figure 4. Alpha-diversity (Shannon index) for the planktonic and sessile community for (A) the bacterial dataset and (B) eukaryotic dataset.....	31
Figure 5. (A) Bacterial and (B) eukaryotic β -diversity (using the Bray-Curtis distance matrix) represented in principal coordinate analysis. Different shapes were used to differentiate the samples based on the bioreactor (or field) it came from. Different colors were used to differentiate the type of the sample (Field for the initial planktonic community, Filter for the planktonic community and Rock for the sessile community). The Filter samples are numbered in order to know the day the sample were taken.	33
Figure 6. Visualization of the Distance-based redundancy analysis (dbRDA) based on (A) geochemical parameters of the water in the bioreactors for the bacterial planktonic community and (B) the mineralogy of the rock surfaces for the bacterial sessile community.	35
Figure 7. Visualization of the Distance-based redundancy analysis (dbRDA) based on (A) geochemical parameters of the water in the bioreactors for the eucaryotic planktonic community and (B) the mineralogy of the rock surfaces for the eucaryotic sessile community.....	36
Figure 8. Differential abundance of bacterial genera between the planktonic and sessile communities.	37
Figure 9. Differential abundance of eukaryotic genera between the planktonic and sessile communities.	38
Figure 10. Contribution of different bacterial sources to every sink. The sinks are divided into three groups based on which bioreactors the sink sample came from. The sources were also divided based on which bioreactors the sample was taken. However, the field data (P1 to P4) were sources for all three bioreactors. The planktonic sample are on the left side of each section while the rock samples are on the right side.....	39

Figure 11. Contribution of different eukaryotic sources to every sink. The sinks are divided into three groups based on which bioreactors the sink sample came from. The sources were also divided based on which bioreactors the sample was taken. However, the field data (P1 to P4) were sources for all three bioreactors. The planktonic sample are on the left side of each section while the rock samples are on the right side. 40

Figure 12. Contribution of different archaeal sources to every sink. The sinks are divided into three groups based on which bioreactors the sink sample came from. The sources were also divided based on which bioreactors the sample was taken. However, the field data (P1 to P4) were sources for all three bioreactors. In this case, there are no samples for the sessile community. 41

LISTE DES TEABLEAUX

Table 1. Component present in the 2 x AE buffer.	67
Table 2. Results of multiple Kruskal-Wallis test between geochemical properties as the rock properties between bioreactors.	67
Table 3. Results of the linear regression done on the bacterial abundance and the roughness of the rocks.	68
Table 4. Results of the Kruskal-Wallis test comparing the rock properties between bioreactors.	68
Table 5. Result of multiple Pearson correlation test between the mineralogy and the bacterial abundance.	68
Table 6. Results of the generalized linear model with the α -diversity. The p-values < 0.05 are in bold.....	69
Table 7. Result of multiple Pearson correlation test between the α -diversity and the rock properties.	69
Table 8. Results of the PERMANOVA analysis from the three domains. The p-values < 0.05 are in bold.....	71
Table 9. Results of the dbRDA analysis based on geochemical parameters of the water in the bioreactors and based on the mineralogy of the rock surfaces for the Bacteria and Eukaryote domain.....	73
Table 10. Results of the dbRDA analysis based on the mineralogy of the rock surfaces for the three domains. The p-values < 0.05 are in bold.....	74
Table 11. Bacterial abundance and roughness of each rock	75
Table 12. ASVs removed during the decontamination step 1.16 for each domain.	77
Table 13. ASVs found in the control samples.	77
Table 14. PCR primer pairs used for 16S/18S rRNA gene amplification and sequencing.	78
Table 15. Primer pairs and probe used for the dPCR assays.	78
Table 16. PCR conditions used for 16S/18S rRNA gene amplification and sequencing.....	79
Table 17. Conditions used for the dPCR assays.	79

LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES

ANOVA	Analysis of variance
ASV	Amplicon sequence variant
CERMO-FC	Centre d'Excellence en Recherche sur les Maladies Orphelines – Fondation Courtois
dbRDA	Distance-based redundancy analysis
DIC	Dissolved Inorganic Carbon
DNA	Deoxyribonucleic acid
DOC	Dissolved Organic Carbon
dPCR	Digital PCR
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethylene glycol tetraacetic acid
FEAST	Fast Expectation–mAximization microbial Source Tracking
LEfSe	Linear discriminant analysis Effect Size
NCBI	National Center for Biotechnology Information
PCoA	Principal Coordinate Analysis
PCR	Polymerase Chain Reaction
PERMANOVA	Permutational multivariate analysis of variance

RÉSUMÉ

La plupart des études sur les communautés microbiennes dans les aquifères se sont concentrées sur les communautés planctoniques. Cependant, depuis quelques années, des études montrent qu'il y a une grande diversité de microorganismes sessiles dans ces milieux. Maintenant qu'il y a un plus grand intérêt sur la diversité des microorganismes sessiles, cette étude vise à contribuer aux découvertes portant sur ces communautés en les comparant aux communautés planctoniques. Nous avons ainsi étudié l'impact de des caractéristiques abiotiques des roches (rugosité et composition minéralogique) sur la colonisation des surfaces des roches. Nous hypothétisons que les microorganismes présents à la surface rocheuse seraient des organismes planctoniques devenus sessiles. L'autre hypothèse est que les propriétés géologiques des roches influencent la structure des communautés microbiennes pouvant s'y installer. Pour ce projet, nous avons échantillonné de l'eau venant d'un aquifère à Covey Hill. Des pastilles stériles de roches venant du site ont été placées dans un système ouvert de bioréacteurs et l'eau souterraine contenant une communauté planctonique initiale y a été pompée. Nous avons ensuite identifié la taxonomie des microorganismes ayant évolué dans l'eau et ayant colonisé la surface des roches grâce au séquençage amplicon des gènes procaryote ARNr 16S et eucaryote ARNr 18S. Un scanner 3D a été utilisé pour mesurer la rugosité et la composition minéralogique a été mesurée avec Ilastik. On a trouvé qu'il y avait beaucoup de microorganismes dans les deux communautés et que la rugosité n'avait qu'un faible effet sur l'abondance de bactéries sur les surfaces. De plus, il n'y a que le quartz qui avait un effet significatif sur la composition de la communauté bactérienne. Cette étude a permis de mieux comprendre l'influence des microorganismes planctoniques sur la communauté sessile, ainsi que l'influence de la minéralogie sur les communautés sessiles.

Mots clés: communauté microbienne, communauté planctonique, communauté sessile, composition minéralogique, séquençage Amplicon

ABSTRACT

Most studies on microbial communities in aquifers have focused on planktonic communities. However, in recent years, studies have shown that there is a significant diversity of sessile microorganisms in these environments. With the growing interest in the diversity of sessile microorganisms, this study aims to contribute to discoveries regarding these communities by comparing them to planktonic communities. We have thus examined the impact of abiotic rock characteristics (roughness and mineral composition) on the colonization of rock surfaces. We hypothesize that the microorganisms present on the rocky surface are planktonic organisms that have become sessile. Another hypothesis is that the geological properties of the rocks influence the structure of microbial communities that can establish themselves there. For this project, we sampled water from an aquifer at Covey Hill. Sterile rock pellets from the site were placed in an open system of bioreactors, and groundwater containing an initial planktonic community was pumped into it. We then identified the taxonomy of microorganisms that evolved in the water and colonized the rock surfaces using amplicon sequencing of the 16S rRNA prokaryotic genes and 18S rRNA eukaryotic genes. A 3D scanner was used to measure roughness, and mineral composition was measured using Ilastik. We found that there were many microorganisms in both communities and that roughness had only a minimal effect on bacterial abundance on the surfaces. Furthermore, only quartz had a significant effect on the composition of the bacterial community. This study has provided a better understanding of the influence of planktonic microorganisms on sessile communities, as well as the influence of mineralogy on sessile communities.

Keywords: microbial community, planktonic community, sessile community, mineralogical composition, Amplicon sequencing

INTRODUCTION

1. La diversité des communautés microbiennes dans les aquifères

La subsurface terrestre constitue la partie allant de quelques mètres (Whitman et al., 1998), à une centaine de mètres sous la surface continentale. On retrouve de l'eau dans les zones saturées circulant entre les interstices de roches poreuses ou des particules solides, formant une région qui se nomme aquifère (Foster et Chilton, 2003). L'apport en nutriments dans cette région peut venir de différentes sources d'eaux de surface qui percole: les précipitations, l'irrigation de l'eau de surface ou la fonte de la neige (Foster et Chilton, 2003). Les conditions abiotiques dans les aquifères restent néanmoins peu propices à la vie. En effet, les aquifères sont souvent décrits comme étant un milieu oligotrophe (Amy et al., 1992; Griebler et al., 2002; Rempfert et al., 2017), il n'y a pas de lumière et souvent peu d'oxygène, de nutriments et sources d'énergie. Malgré cela, il existe une grande diversité de microorganismes dans ces régions (Flynn et al., 2013; Griebler et Lueders, 2009). En effet, selon Whitman et al. (1998), la majorité des procaryotes de la planète y seraient présents et selon McMahon et Parnell (2014), la biomasse de procaryotes dans la subsurface terrestre est similaire ou même plus grande que celle de la subsurface marine.

Par ailleurs, parmi les microorganismes habitant les aquifères, certaines espèces vont vivre dans l'eau tandis que d'autres vont se fixer à la surface des roches au travers desquelles l'eau passe pour y résider. Ainsi, on se retrouve avec deux communautés microbiennes: une communauté sessile (attachée) et une communauté planctonique (libre dans l'eau). Nous savons que certains de ces microorganismes ont la capacité de dégrader les surfaces rocheuses afin d'en ressortir des éléments essentiels à leur survie comme du phosphore et fer (Park et al., 2009; Stevens et McKinley, 1995; Bennett et al., 2001). D'autres vont pouvoir passer de la forme planctonique à sessile grâce à des processus physiologiques intracellulaires, lorsqu'il y a des surfaces rocheuses auxquelles s'accrocher ou lorsqu'ils sont en carence de nutriments (Bryce, 2015; Kjelleberg et al., 1982).

Il est important d'étudier les microorganismes vivant dans ces eaux souterraines, car ils jouent un rôle important dans les écosystèmes. En effet, étant assez nombreux dans la subsurface, ils jouent un rôle majeur dans les cycles biogéochimiques en transformant les composés

inorganiques et organiques (Kaur et al., 2022; Purkamo et al., 2017; Anantharaman et al., 2016). De plus, des recherches ont trouvé que ces microorganismes peuvent être utilisés dans la bioremédiation dans les aquifères contaminés (Ma et al., 2021; Kaur et al., 2022). Beaucoup d'études sur les communautés microbiennes de la subsurface se sont concentrées sur la communauté planctonique et moins sur la communauté sessile puisqu'il est plus difficile de récupérer des sédiments et surfaces rocheuses non-contaminés lors de forages que de récupérer l'eau souterraine par des puits forés (Alfreider et al., 1997). Ce n'est que depuis les deux dernières décennies que l'on s'est intéressé à cette communauté. Certaines études ont alors montré que la communauté sessile serait beaucoup plus diversifiée que la communauté planctonique, contrairement à ce qui était imaginé (Griebler et Lueders, 2009; Alfreider et al., 1997). En effet, les bactéries planctoniques ne représenteraient que 0,01 – 10% de la communauté bactérienne totale ou de la biomasse des aquifères tandis que le reste serait les bactéries sessiles qui sont fixées à des particules ou à des surfaces rocheuses (Griebler et al., 2002). Étant toujours un nouveau domaine de recherche, il reste encore plusieurs aspects peu étudiés chez ces communautés sessiles.

Les microorganismes réussissent à habiter ces milieux grâce à une pression sélective qui s'exerce en faveur des gènes essentiels à leur survie (Wackett et al., 2004), comme des gènes permettant aux microorganismes d'exploiter les minéraux dont les roches sont composées. En effet, une étude par Jones et Bennett (2014) montre que les microorganismes bénéficient de leur interaction avec les minéraux. Certains nutriments se libèrent des minéraux et peuvent être directement utilisés par les microorganismes tandis que d'autres forment des précipités qui doivent ensuite se solubiliser afin de devenir disponible pour eux (Uroz et al., 2015). Des exemples de cela mentionnés par Jones et Bennett (2014) incluent la présence d'accepteurs d'électrons minéraux, la dissolution de l'apatite pour libérer du phosphore et la dissolution du sulfure pour des cofacteurs. Ces nutriments peuvent se libérer de la surface minérale par des processus abiotiques ou par des activités biologiques des microorganismes présents à la surface des roches (Uroz et al., 2015). Ces nutriments dissous peuvent être nocifs ou bénéfiques selon la communauté microbienne. Puisque les minéraux sont répartis de manière hétérogène sur une roche, il y aura une sélection d'habitats qui s'opère, ce qui influencera la structure, la diversité et

la variabilité phylogénétique des communautés microbiennes dans les aquifères (Uroz et al., 2015; Jones et Bennett, 2017).

2. Changement de mode de vie planctonique à sessile

Dans les aquifères, un mode de vie sessile peut offrir un avantage sélectif par rapport à un mode de vie planctonique (Griebler et al., 2002). En effet, puisque des macromolécules et composés organiques ont tendance à s'accumuler sur les surfaces, les microorganismes sessiles auront un meilleur accès à ces nutriments (Griebler et al., 2002). De plus, les roches et minéraux servent de support physique pour que les microorganismes puissent s'y fixer et former des biofilms afin de se protéger de prédateurs et de toxines (Uroz et al., 2015; Griebler et al., 2002). À partir de cela, les individus sessiles peuvent former des communautés synergiques qui peuvent effectuer des processus combinés que les espèces individuelles ne peuvent pas faire (Beveridge et al., 1997) (Griebler et al., 2002). En effet, une étude par Breugelmans et al. (2008) a trouvé que trois espèces d'une communauté sessile dans un biofilm pouvaient dégrader un herbicide se trouvant dans l'eau beaucoup plus efficacement que si seulement une des espèces était présente. Un autre exemple serait une étude faite par Ross et al. (2001) où ils ont stimulé la formation de biofilm chez des microorganismes planctoniques vivant dans de l'eau souterraine. Ils ont trouvé que les microorganismes s'échangeaient des gènes, ce qui a mené à une augmentation de la diversité génétique des microorganismes. Cela va de pair avec les études de Uroz et al. (2015) et Escudero et al. (2018) qui expliquent que, lors de la formation de biofilm, une expression différentielle de gènes ou une modification dans la production de protéines est observée.

3. Colonisation et croissance de microorganismes sur les roches des aquifères

Plusieurs études ont visé à mieux comprendre la colonisation des microorganismes sur des surfaces solides, selon différents types géologiques de roches dans les aquifères. Ces recherches ont trouvé qu'une augmentation de la rugosité des surfaces augmentait la rétention des microorganismes sur une surface (Verran et Whitehead, 2005; Whitehead et Verran 2006; Dubois et al. 2010), ce qui leur permet de coloniser les surfaces avec plus d'aisance. D'autres, comme celle de Jones et Bennett (2014), ont montré qu'un ensemble de différentes espèces de microorganismes venant d'une grotte colonisait très peu les feldspaths (albite et microcline) et la

chaille, contrairement au calcaire et la dolomie. Dans l'étude de Rogers et Bennett (2004) qui se concentre plus sur la présence de nutriments à la surface rocheuse, il a été noté que les microorganismes préfèrent coloniser les silicates contenant du phosphore et fer, des nutriments limitants, que les silicates non nutritifs. De plus, la distribution des microorganismes et la formation de biofilm changent selon la roche (Jones et Bennett, 2014). Par exemple, sur les feldspaths, les microorganismes unicellulaires forment des amas discrets tandis que sur le quartz, ils sont répartis individuellement de manière homogène sur toute la surface (Jones et Bennett, 2014). En plus, il y a des accumulations épaisses, blanches et filamenteuses de biofilm sur les roches carbonatées et le basalte, tandis qu'il y en a très peu ou pas sur les silicates. Ces différents résultats sont dus aux différences dans la composition des roches (Rogers et Bennett, 2004; Rogers et al., 1998). En effet, il semble que les roches ayant une surface riche en phosphore et fer ont plus tendance à être colonisées par des microorganismes (Rogers et Bennett, 2004) et celles ayant une surface riche en carbone vont augmenter la quantité de biofilms produits sur celle-ci (Jones et Bennett, 2017). Dans l'expérience de Jones et Bennett (2017), l'ajout de carbone a même augmenté la diversité- α , c'est-à-dire le nombre d'espèces présent dans l'échantillon. De plus, les microorganismes apparentés, donc ayant le même genre ou famille, préféreraient coloniser les mêmes minéraux des roches (Jones et Bennett, 2017).

Un autre aspect à tenir en compte lorsque l'on discute la colonisation de roches est la charge de la surface rocheuse. Certaines classes de minéraux sont facilement altérées dans des conditions acides (Uroz et al., 2015). Une étude par Jones et Bennett (2014) et une par Rogers et Bennett (2004) expliquent que les microorganismes ont la capacité d'éroder les surfaces rocheuses grâce à des processus métaboliques, ce qui a pour effet de libérer des cations et ainsi modifier le pH du milieu ce qui va ultimement changer la charge à la surface rocheuse. À partir de cela, une surface rocheuse chargée positivement peut attirer des cellules bactériennes chargées négativement grâce à la protonation (Uroz et al., 2015; Rogers et Bennett, 2004; Harvey et al., 2007). De plus, la libération de certains éléments va attirer des organismes chimiotactiques vers la surface rocheuse (Rogers et Bennett, 2004; Lopez de Victoria, 1989).

4. Objectifs et Hypothèses

Sachant que les aquifères sont des milieux oligotrophes, il est sûrement avantageux pour les microorganismes planctoniques de coloniser les surfaces rocheuses afin d'avoir accès à plus de nutriments essentiels que s'ils se retrouvaient dans l'eau seulement; puisqu'une carence en nutriments peut déclencher des processus permettant la fixation d'un microorganisme sur une surface rocheuse. Ainsi, notre première hypothèse **[H1]** stipule que les organismes présents à la surface rocheuse des aquifères seraient des organismes planctoniques devenus sessiles (*i.e.* ayant changé de mode de vie).

Par ailleurs, les roches sont composées de minéraux dont certains peuvent être néfastes ou bénéfiques pour les microorganismes. Les microorganismes sont distribués de manière hétérogène à la surface afin de se retrouver sur un substrat adéquat pour eux. Ainsi, certaines zones des roches seront plus colonisées que d'autres. De plus, la rugosité des roches va changer la possibilité de colonisation par les microorganismes. On sait qu'il y a une plus grande abondance de microorganismes sur une surface plus rugueuse. Par conséquent, la deuxième hypothèse **[H2]** stipule que les propriétés physiques des surfaces rocheuses ainsi que ses propriétés minéralogiques influenceront la structure des communautés microbiennes pouvant s'y installer.

Ayant encore peu de connaissance sur la colonisation des surfaces rocheuses par des microorganismes dans les aquifères, ce projet de recherche vise à répondre aux objectifs suivants:

- (1) Comparer la communauté microbienne sessile à la communauté planctonique dans un système d'aquifère au Québec.
- (2) Déterminer l'influence de la composition minéralogique et la rugosité de la surface rocheuse sur la structure des communautés microbiennes colonisant les roches.

Les résultats obtenus pendant ce projet de recherche sont présentés par la suite sous forme d'un article scientifique qui sera soumis très prochainement.

Chapitre 1: Rock surface colonization by groundwater microorganisms in an aquifer system in Quebec, Canada

1.1 SUMMARY

Aquifers are rich in microbial diversity. They play an important part in geochemical cycles such as the carbon cycle. Yet, there's a lack of information about sessile communities and how they colonize surfaces in aquifers because of the difficulty of sampling sediments on which they would be found. This study's objective is to better understand the sessile community in an aquifer by comparing them to their planktonic counterparts. Additionally, the impact of rock roughness and mineral composition on colonization is explored. We hypothesized that the sessile community would consist of planktonic microorganisms that have become sessile, and that the geological properties of rocks will influence microbial colonization. Using a system of bioreactors, we recreated the environmental conditions of an aquifer in Covey Hill using water directly collected from the site as well as a sterilized rock sample. DNA extractions were performed as well as taxonomy identification through Amplicon sequencing, 3D scanning the rocks for roughness measurement, and mineral composition assessment using Ilastik. We found that many microorganisms are present in both planktonic and sessile communities. The main taxa in the groundwater were *Sulfurimonas*, unc. *Acetobacteraceae*, unc. *Ochrophya*, *Spumella* and unc. *Bathyaerchaia*, while on the rock surfaces they were unc. *Acetobacteraceae*, unc. *Ochrophya*, and *Spumella*. These lineages potentially interact both in the water and the biofilms (prey-predation), but future studies are needed to confirm this. The Roughness had a minimal impact on bacterial abundance on rock surfaces. A possible explanation would be that the range of the roughness was too small to see a correlation between the two parameters. Quartz was the only mineral significantly affecting bacterial community composition, which is consistent with other studies. Having this knowledge now, it will be possible to change certain parameters in the experiment such as having only one mineral in the bioreactors, the duration of the experiment or having a fix range of roughness to answer some of the questions this project brought to light.

1.2 INTRODUCTION

The subsurface contains up to 40% of the world's prokaryotic biomass (Whitman et al., 1998). These microorganisms, being quite abundant in the subsurface, play an important role in mediating biogeochemical cycles in groundwater such as the carbon cycle and the nitrogen cycle by transforming inorganic and organic compounds (Kaur et al., 2022; Purkamo and al., 2017; Anantharaman and al., 2016; Hug and al., 2011). This water is critical for essential human activities [e.g. for human consumption, irrigation, agriculture, and much more (Smith and al., 2018)] which is why it is so important to study these environment: we can better understand how they impact the geochemical cycles, how the microbial communities change in these environment and how to protect aquifers. Aquifers, which are regions in the subsurface where groundwater circulates between rocks and other solid matter, are an important reservoir of fresh water and are used by many people around the world (Foster and Chilton, 2003). These dark and nutrient-poor habitats contain microorganisms with two distinct lifestyles: planktonic and sessile (Smith and al., 2018). The planktonic microorganisms are free-living, while the sessile microorganisms are attached to surfaces. Sessile microbes have better access to minerals on rock surfaces that some species can exploit (Rogers and Bennett, 2004), as well as better access to more organic matter that tends to sediment on surfaces (Grösbacher and al., 2016). Microorganisms have different ways to colonize a surface. For instance, some studies found that the roughness of rock surfaces plays a significant role in the colonization and that microorganisms tend to colonize rough surfaces more effectively (Verran and Whitehead, 2005; Whitehead and Verran, 2006; Dubois et al., 2010). Researchers also found that different rock types show varying levels of colonization. For example, limestone and dolomite were more colonized by microorganisms than feldspar and shale (Jones and Bennett, 2014). Furthermore, microorganisms showed a preference for rocks containing important nutrients for microorganism's survival like phosphorus and iron (Rogers and Bennett, 2004). Another important factor to consider is the charge of the rock's surface. Positively charged rock surfaces can attract the negatively charged bacterial cell wall (Uroz et al., 2015; Rogers and Bennett, 2004; Harvey et al., 2007). Metabolic processes by the microorganisms can alter the rock's surface and thus release certain elements

that can attract chemotactic organisms to the rock (Rogers and Bennett, 2004; Lopez de Victoria, 1989).

Unfortunately, the difficulty in obtaining uncontaminated rock samples from aquifers challenges studies on sessile communities found there, and thus most studies will focus on the planktonic community (Lazar and al., 2019; Anantharaman and al., 2016; Alfreider and al., 1997). However, it is important to study the sessile communities too since they are more abundant compared to the planktonic community (Flynn and al., 2013; Griebler and Lueders, 2009; Alfreider and al., 1997). Thus, to better understand both communities present in aquifers, we used 16S rRNA amplicon sequencing to compare the planktonic community to the sessile community. We also studied the influence of the roughness of rocks and their mineralogical composition on the colonization of rock surfaces. We hypothesized that the microorganisms present on the rock surfaces would be planktonic organisms that changed their lifestyle to become sessile. We also hypothesized that the roughness and mineralogy of rock surfaces will influence the structure of the microbial communities that can settle there.

For this experiment, we used a system of three bioreactors (replicates) to recreate the environmental conditions and settings of the aquifer ecosystem in the laboratory. These bioreactors allowed us to circumvent the difficulties of sampling in-situ sessile populations in the subsurface.

1.3 MATERIAL and METHODS

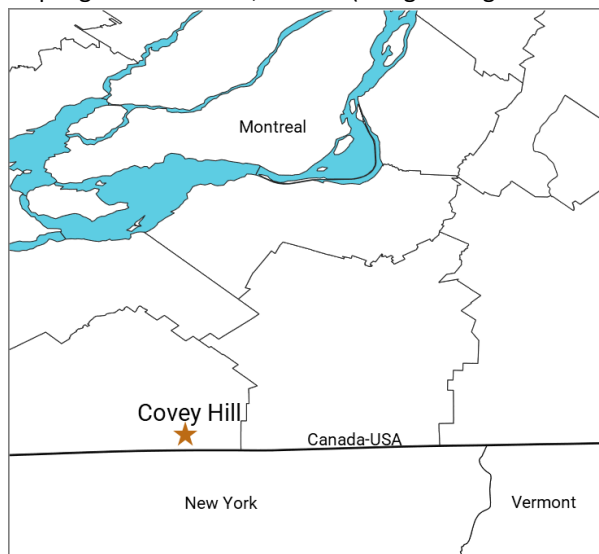
1.3.1 Study site

Rock and water samples were collected from a 1.5m well on Covey Hill which connects to an aquifer (Figure 1), located in the Basses-Terres of Saint-Laurent. The sample site was chosen because of the geology of the area. The different underground layers are mostly made of felspar (Globensky, 1987; McGerrigle, 1950), and they outcrop up to the surface. Thus, the rock samples at the surface have a similar mineralogy to those found in the aquifer ecosystem. Water samples were collected from the well after pumping out three times the volume to remove stagnant water. Then, 30 L was collected for the bioreactor incubations (10 L per bottle per bioreactor), and 1 L was taken four times (before collecting each 30 L) to determine the initial planktonic community

in the groundwater. An extra liter was taken from a peat bog that is located near the aquifer. Indeed, it is likely that the peatbog is a source of recharge for the groundwater. To maintain a similar environment in the bioreactors as found *in-situ*, we measured the temperature, pH, and percentage of dissolved O₂ in the groundwater using a multiparameter probe (OAKTON PD 450, Cole-Parmer, United Kingdom).

Furthermore, the following geochemical parameters were measured in the initial groundwater: DIC, DOC, NH₄, NO₂⁻ and NO₃⁻. In order to do so, aliquots of water were filtered into scintillation bottles during the sampling using a 0.45 µm filter. For the DIC and DOC, the water was collected in gas-free glass bottles and for the NH₄ samples, we used 0.2 µm filters. All the samples were analyzed at the laboratory of the Interuniversity Research Group in Limnology (GRIL). For DIC and DOC, samples were analyzed with an OI Analytical Aurora 1030W TOC Analyzer by using a persulfate oxidation method. For the NO₂⁻ and NO₃⁻ measurements, the samples were analyzed with a continuous flow analyzer (OI Analytical Flow Solution 3100) using an alkaline persulfate digestion method coupled with a cadmium reactor. Ammonium measurements were done with a Flow Solution 3100 autosampler by using a chloramine reaction with salicylate to form indophenol blue dye (EPA Method 350.1). These analyses were conducted at the Université du Québec à Montréal (UQAM)-GRIL analytical laboratory. These geochemical parameters were also measured every two days from day 6 of the experiment to day 24.

Figure 1. Map of the aquifer sampling site in Quebec, Canada (using the Qgis software).



1.3.2 Rock chip preparation and properties measurements

An outcropping rock slab was collected prior to the groundwater sampling and was first cleaned with soap and water. Seventy-two rock chips were then made by drilling (18" nova voyager dvr, King Industrial, Canada) and sawing (DREMEL 3000, DREMEL, USA) the rock slab. They were all approximately $\frac{1}{2}$ inch in diameter and $\frac{3}{32}$ inch large. The chips were further cleaned by vortexing them separately in a HCl 10% solution for 2 min and by sonicating them for 1min. The pellets were then autoclaved, and 24 chips were inserted in each bioreactor (Annexe A, Figure A).

To determine the mineralogy on the rock surface, pictures of both flat sides of the chips were taken. We used a stereo binocular microscope that had a camera attached to it for this part. Afterwards, the pictures were analyzed using two programs: starting with Ilastik on the Pixel Classification setting and then ImageJ. It is important to note that the identification of the minerals on the rock surface was first done manually in order for the program to then be able to identify the rest of the surface. Thus, the identification process was faster and more reliable than identifying ourselves. For the roughness measurements, a 3D scanner (Polyga HDI Compact C210, Burnaby, Canada) paired with the Flexscan3D software was used to take pictures of the surface of multiple rocks at once. Then, using the CloudCompare software, each rock chip was isolated, and the coordinates of each point present when scanning the surfaces were obtained. Finally, with these coordinates, the roughness was calculated with MathLab (Annexe E). The detailed instructions are provided in the Annexe F.

1.3.3 Bioreactor incubation setup

In order to recreate an environment similar to the aquifer in the laboratory, we used a bioreactor system (CBR 90 Standard CDC Biofilm Reactor, BioSurface Technologies, USA). We had three separate bioreactors acting as triplicates to check the results reproducibility (Annexe A, Figure B). The 30 L of aquifer water collected in three 10 L glass bottles were placed on magnetic stirring plates. In order to avoid microorganisms attaching to the glass walls of the bottles, the water was agitated during the incubations. The groundwater was pumped at a rate of 0.289 mL/min (IPC, Ismatec, Germany) from the 10 L bottles into the bioreactors for 24 days. The bioreactors were also on stirrer plates for the same reasons as mentioned above. The excess water in the bioreactor was collected in other 10 L containers and never recirculated in the bioreactors to mimic aquifer systems. Simultaneously, three common gases of the atmosphere (CO_2 , N and O_2) were injected at 200 sccm into the bioreactor using the MCQ GB100 gas mixer (Monkey Industrial Supply, USA). The pH of the water and % dissolved O_2 in the bioreactors were monitored and adjusted daily for the duration of the incubation to maintain a similar pH and % dissolved O_2 to the data measured in the groundwater during the sampling. The whole system of bioreactors (Annexe A, Figure B) was located in an incubation chamber, in which the temperature was set according to that measured in the groundwater during sampling (13°C).

1.3.4 Sample collection during the incubations in the bioreactors and DNA extractions

During the 24 days of incubation, 300 mL of water was filtered every two days from the three collection bottles to analyze the planktonic community. The water was filtered with sterile polyethersulfone membrane filters with 0.2 μm diameter pores (Sartorius, Germany). The filtration was done with a membrane vacuum pump (Welch 2019B-01, Welch, USA) in asepsis under a laminar flow hood. Each filter obtained was stored at -80°C. The excess water was used to measure the same physio-chemical parameters as those taken on the day of sampling (DIC, DOC, NH_4 , NO_2^- and NO_3^-). At the end of the experiment, all the rock pellets were taken from each of the 3 bioreactors to study the colonization of rock surfaces and the sessile community, and frozen at -20°C. DNA was extracted from the rock surfaces of each rock pellet separately, as well as from the filters obtained after the filtration of groundwater.

The DNA of the sessile microorganisms was extracted from each pellet following the protocols of Lazar and al. (2019) under asepsis conditions with sterile tubes. First, a 2x AE elution buffer was prepared (Annexe B, Table 1). The rock chip was then placed in a 50 mL Falcon tube with 1 g of autoclaved 0.1mm zircon/silica beads and 2 mL of 2x AE buffer, and vortexed for 30 seconds. Then, 500 µL of 20 % SDS was added, and the solution was vortexed for 1min at maximum speed. The solution was then centrifuged for 5 min at 8000 xg and the rock chip was put into a 15mL Falcon tube. Two mL of phenol/chloroform/isoamyl alcohol (25:24:1) was added, and the solution was vortexed at maximum speed for 1min and then centrifuged for 5 min at 8000 xg. Two liquid phases were present in the tube: an aqueous phase and a phenolic phase. The aqueous phase, which is above the organic phase, was transferred to a new 50 mL tube and 2 mL of chloroform was added. Then, the new tube was vortex for 1min and then centrifuged for 10 min at 8000 xg. The supernatant was then again transferred into a new 50 mL tube.

In order to precipitate the DNA present in the sample, 2 µL of glycogen (20 mg/mL), 200 µL of sodium acetate (3M) and 4 mL of 100 % ethanol were added to the aqueous phase. The tube was inverted and incubated overnight at 4°C. The next day, the tube was centrifuged for 30 min at 8000 xg. The obtained pellet was dried out and then rinsed with 0.5 mL of 70 % ethanol. The tube was then centrifuged for 10 min at 8000 xg and the ethanol was completely removed. The pellet was then re-suspended in 100 µL of 1x Tris EDTA (TE) buffer.

DNA was extracted from the filters using the RNeasy PowerSoil Total RNA kit (QIAGEN, Germany), followed by the RNeasy PowerSoil DNA elution kit (QIAGEN, Germany), following the instructions provided by the manufacturer (Annexe G). Initially, it was planned to also examine the RNA expressed in each sample, but that was not done since there was already a lot of data analysed in this paper. In order to make sure there was no contamination while doing the experiment, negative controls were done. For this, the bioreactor incubation experiment was done again in the same conditions, using sterile sterile miliQ water instead of groundwater, for the same duration. Plus, all the tubes and bioreactors were sterilized beforehand and new sterile rock chips were added into the bioreactors. After 24 days, 300 mL were filtered from each bioreactor, and the rock pellets were taken as well. The DNA was extracted following the same protocols as

detailed above. In Rstudio, `isContaminant()` was used to removed the ASV who were contaminants.

1.3.5 DNA sequencing and bioinformatics analysis

Bacterial and archaeal 16S rRNA, as well as eukaryotic 18S rRNA genes from both the rock chips and the filters were amplified using the polymerase Phusion Hot Start HF (ThermoFisher, Waltham, USA). Primer pair B341F – B785R (CCTACGGGAGGCAGCAG and GACTACHVGGGTATCTAATCC) was used for the Bacteria, A340F – A915R (CCCTACGGGGYGCASCAG and GTGCTCCCCGCCAATTCCT) for the Archaea, and E960F – E1438R (GGCTTAATTTGACTCAACRCG and GGGCATCACAGACCTGTTAT) for the Eukaryotes. PCR was carried out following the conditions described in the Annexes C and D. Sequencing was carried out using the Illumina Miseq platform of CERMO-FC (Centre d'Excellence en Recherche sur les Maladies Orphelines – Fondation Courtois), with the Miseq Reagent v3 600 cycles kit (Illumina, USA), and a paired reading of 300 bp. Negative PCR controls, as well as the control samples for water and rock samples, were also sequenced for all three domains. Despite all our efforts, we did not manage to amplify the archaeal 16S rRNA genes for the sessile community. Thus, only the planktonic archaeal samples were sequenced and not the sessile community.

1.3.6 Digital PCR amplifications (dPCR)

The digital PCR was only done for the bacterial sessile data. To determine the absolute abundance of bacterial 16S rRNA genes on the rock surfaces, we performed dPCR amplifications of the sessile community. We used the QuantStudio 3D Digital PCR (ThermoFisher, Waltham, MA, USA) instrument as well as the QuantStudio 3D PCR Master Mix v2 (ThermoFisher, USA). The primers, probes, and cycle specifications are available in the supplementary material (Annexes C and D).

1.3.7 Processing the 16S and 18S rRNA gene sequences

The DADA2 package (Callahan et al., 2016) in RStudio version 4.2.2 (R Core Team, 2023) was used to process the sequences and produce amplicon sequence variant (ASV) tables (Annexe E). Sequences present in the controls were considered contaminants and were removed from the datasets. For each domain (Bacteria, Archaea and Eukaryotes), we rarefied to the lowest ASV number by sample: 2300 for Bacteria, 4000 for Archaea and 3000 for Eukaryotes. We chose those

limits to be able to maintain as many samples as possible (Annexe A, figures C to E). Taxonomy was assigned using the Silva 138.1 reference database, and the SILVA v128 prokaryotic train set was used to get a more precise taxonomy (going as far as the species name when possible).

1.3.8 Statistical analyses

All the statistical analysis were performed with RStudio version 4.2.2 except for the calculation of the rock roughness (Ra), which was done on Matlab version R2022a. All the scripts used are available in the Annexe E. Unless written otherwise, each analysis was done for all three domains. The different geochemical properties were compared between the bioreactors to check whether the conditions in the three bioreactors were significantly similar, using the Kruskal-Wallis test with the `kruskal.test()` function. Shared genera between the planktonic and sessile communities were determined using the `VennDiagram` package (version 1.7.3). This step was not done for the Archaea.

For the dPCR, the results were expressed as gene copies/ μ l. After calculating the number of gene copies for each sample, we performed a linear regression with the `lm()` function and calculated the Pearson correlation coefficient using the `cor()` function.

The alpha diversity (Shannon index) was calculated using the `diversity()` function from the `vegan` package (version 2.6.4). To compare the alpha diversity between both lifestyles (sessile and planktonic) for Eukaryotes and Bacteria, we used a Welch's t-test. To determine the correlation between the geochemical properties of the water and the α -diversity, we ran multiple generalized linear models and kept the one with the best AIC (Akaike Information Criterion) score. Beta diversity was calculated using the Bray-Curtis dissimilarity matrix with the `vegdist()` function in the `vegan` package. We used PCoA (Principal Coordinate Analysis) to observe the distribution and clustering of the different water and rock samples, and the significance of the clustering was checked using PERMANOVA (Permutational multivariate analysis of variance). The planktonic communities had to be divided into smaller groups for this analysis. For the bacteria dataset, the days were grouped as such: days 1 to 4, days 5 to 6, and 7 to 12. In the case of the eukaryotic data, the groups were day 1 to 3, day 5 to day 6, and day 7 to 12. Lastly, the groups for the archaeal data were days 1–6 and days 7–10. Db-RDA (Distance-based redundancy analysis) was performed

to determine correlations between β -diversity of the planktonic community and the water geochemical properties, and between β -diversity of the sessile community and the rock mineralogy and roughness.

For each domain, a Linear discriminant analysis Effect Size (LEfSe) was performed in order to identify which microbial taxa are primarily responsible for differences between different groups. Not only did the LEfSe determined the taxa that were statistically significant, but it also determined the ones that are biologically meaningful in terms of their effect size. These analyses were performed on the LEfSe-Galaxy application found on the following site: <https://huttenhower.sph.harvard.edu/galaxy/>

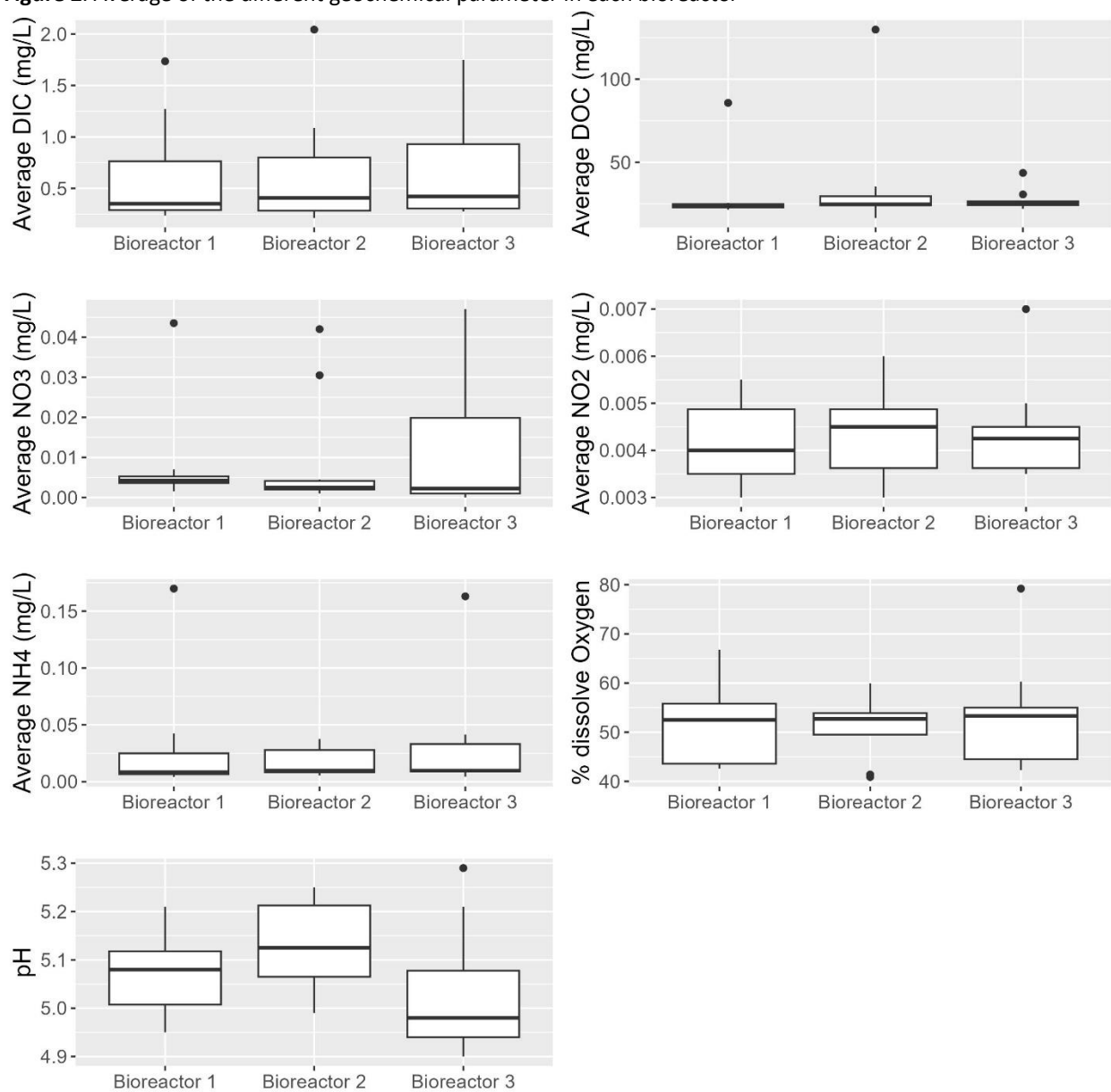
Finally, in order to determine the proportion of the planktonic community (sources) that contributed to the planktonic community at further time points (sinks) or that contributed to the sessile community (sinks), we ran multiple Fast Expectation-mAximization microbial Source Tracking (FEAST) (Shenhav et al., 2019).

1.4 Results

1.4.1 Geochemical water and rock properties

To check whether conditions within the bioreactors were reproducible, we compared the average DIC, DOC, NO_2^- , NO_3^- , NH_4 , % dissolved O_2 and pH values between each bioreactor (Figure 2) using a Kruskal-Wallis test. The figures showing fluctuations every two days in each bioreactor are in the supplementary materials (Annexe A, Figures F to L). We observed no significant difference between the bioreactors for all measured parameters (Annexe B, Table 2). Overall, for all three bioreactors, DIC concentrations decreased drastically after 6 days of incubation, while DOC concentrations remained stable throughout the experiment, with peaks at 12, 14 and 18 days, depending on the bioreactor. Ammonia concentrations decreased during the experiment, with a peak after 10 days of incubation. Nitrate concentrations were low throughout the experiment, with an increase (peak) after 10 days of incubation and a second increase at the end of the incubations after 22 days.

Figure 2. Average of the different geochemical parameter in each bioreactor



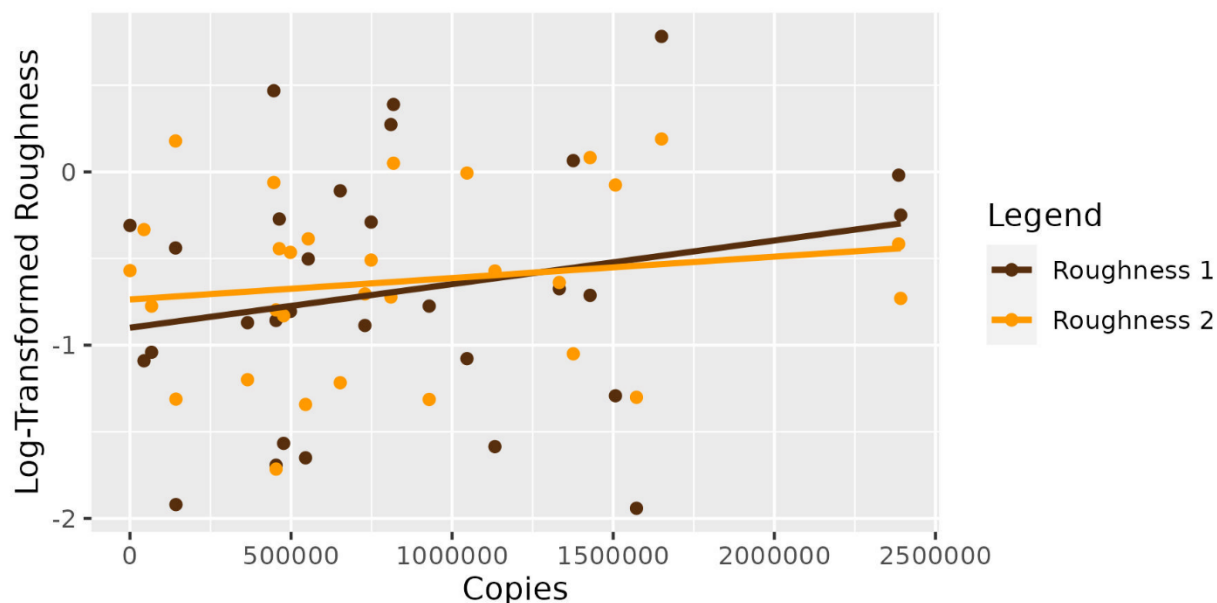
Looking at the rock properties, the values of the roughness vary from a minimum roughness of 0.14353 to a maximum of 2.18455 (Annexe A, Figure M). As for the mineralogy, four minerals were identified: mica, quartz, red feldspar and white feldspar (Annexe A, Figure N). There was also some iron oxide present on the surfaces. The parts that did not correspond to the previous minerals were marked as "undetermined". We noticed that quartz was the most abundant mineral in all three bioreactors. Its percentage on the surface varied from 19.41 to 98.14%. Mica was the least abundant mineral, varying from 0.00% to 41.00%. Neither the roughness nor

mineral composition of the rocks were significantly different between the three bioreactors (Annexe B, Table 4).

1.4.2 Rock-attached bacterial absolute abundance and correlation with the rock properties

The bacterial absolute abundance numbers per rock chip varied from 6.71×10^4 to 2.39×10^6 for bioreactor 1, 1.94×10^2 to 6.53×10^5 for bioreactor 2, and 1.43×10^5 to 2.39×10^6 for bioreactor 3 (Annexe B, Table 11). The linear regressions (Figure 3) showed a weak positive relationship between the roughness measurements (Roughness 1 and Roughness 2) and the bacterial absolute abundance, represented by the number of gene copies found in the rock samples (Figure 3). After doing a Pearson correlation coefficient, we found a weak positive correlation between the Roughness 1 and the Abundance (0.22) as well as between the Roughness 2 and the Abundance (0.15). There were no significant correlations between the absolute abundance and the mineralogy (Annexe B, Table 5).

Figure 3. Linear regressions of the Roughness based on the bacterial abundance represented here by the number of gene copies in different samples.



1.4.3 16S and 18S rRNA gene taxonomic composition

For the Bacteria domain, the peat bog was dominated by *Acidocella* (23.1 % of the total reads), *Granulicella* (14.8 %), and unc. Caulobacteraceae (14.4 %). The initial groundwater was composed of *Sulfurimonas* (52.4 %) and unc. *Thermodesulfobacteria* (9.6 %) for the P1 sample; *Sulfurimonas* (36.6 %), unc. *Thermodesulfobacteria* (10.1 %), and *candidatus Solibacter* (8.5 %) for P2;

Sulfurimonas (24.7 %), unc. *Thermodesulfovibrionia* (15.4 %), and unc. *Acetobacteraceae* (9.8 %) for P3; and *Sulfurimonas* (40.2 %), unc. *Thermodesulfovibrionia* (8.5 %), *cand. Solibacter* (6.8 %), and unc. *Acetobacteraceae* (6.7 %) for P4. In the planktonic community, we observed a similar temporal pattern in all three bioreactors (Annexe A, Figure O). *Sulfurimonas* was the most abundant genus in the beginning from days 2 to 12 and started to decrease after 2 days of incubation (bioreactor 1: relative abundance 79.3-60.7 %, bioreactor 2: 58.4-61.6 %, bioreactor 3: 65.7-69.0 %). Meanwhile, the unclassified (unc.) *Acetobacteraceae* became increasingly abundant, becoming the dominant genus on the 14th day of the experiment (bioreactor 1: 37.4%, bioreactor 2: 45.08% and bioreactor 3: 33.8%). This unc. *Acetobacteraceae* remained the most abundant genus until the end of the experiment. Looking at the sessile community, we see a similar community pattern among all the rock samples, regardless of the bioreactor (Annexe A, Figure P). Since the rock samples were taken on the last day of the experiment only, there's no temporal variation that can be discussed. In this community too, we see that the unc. *Acetobacteraceae* is the most abundant taxon (bioreactor 1: 39.5%, bioreactor 2: 40.9%, bioreactor 3: 32.3%).

With the Venn diagram, we were able to determine which genera were present in both communities as well as which ones were present in one community only. For the bacteria, 88.51% of the genera were present in both communities (Annexe A, Figure T). As for the rest, 8.05% were exclusively in the planktonic community and 3.45% were in the sessile community. It is important to note that the dominant genera in both communities are *Sulfurimonas* and unc. *Acetobacteraceae*.

For the Eukaryote domain, the peat bog was dominated by unc. *Charophyta* (22.6 %), *Trachelius* (9.5 %), and *Cyclopoida* (8.8 %). The initial groundwater was composed of unc. *Fungi* (33.8 %), *Ochromonas* (17.8 %), unc. *Ochrophyta* (12.9 %), and unc. *Cryptomonas* (9.6 %) for the P1 sample; unc. *Fungi* (39.8 %), *Ochromonas* (11.6 %), unc. *Ochrophyta* (14.9 %) for P2; unc. *Fungi* (22.1 %), *Stenostomidae* (22.1 %), *Ochromonas* (11.4 %), unc. *Ochrophyta* (14.1 %) for P3; and unc. *Fungi* (39.7 and unc. *Ochrophyta* (13.3 %) for P4. In the planktonic community, we observed a temporal pattern as well (Annexe A, Figure Q). In this case, there was a high abundance of unc. *Fungi*, *Ochromonas* and unc. *Ochrophyta* on the second day of the experiment. Then, the abundance of

Fungi and *Ochromonas* decreased until the end of the experiment. *Ochrophyta*'s abundance peaked on the 10th day (bioreactor 1: 74.9 %, bioreactor 2: 82.0% and bioreactor 3: 78.2 %). Afterwards, the *Ochrophyta* decreased while the *P34.45* (also known as *Spumella*) increased (Lefranc et al., 2005). At the end of the experiment, *P34.45* reached 59.3% for the first bioreactor, 64.8% for the second bioreactor and 66.6% for the third bioreactor. For the rock samples, we observed 2 major genera: *Spumella* (bioreactor 1: 44 %, bioreactor 2: 56.7 %, bioreactor 3: 47.3 %) and *Ochrophyta* (bioreactor 1: 39.5 %, bioreactor 2: 25 %, bioreactor 3: 23.1 %) (Annexe A, Figure R).

In the Venn diagram for the Eukaryotes, we see that both communities have 64.52% of all genera, 29.03% in the planktonic community only, and 6.45% in the sessile community (Annexe A, Figure U). In this case, the dominant genera are *P34.45*, *Ochromonas* and unc. *Ochrophyta*.

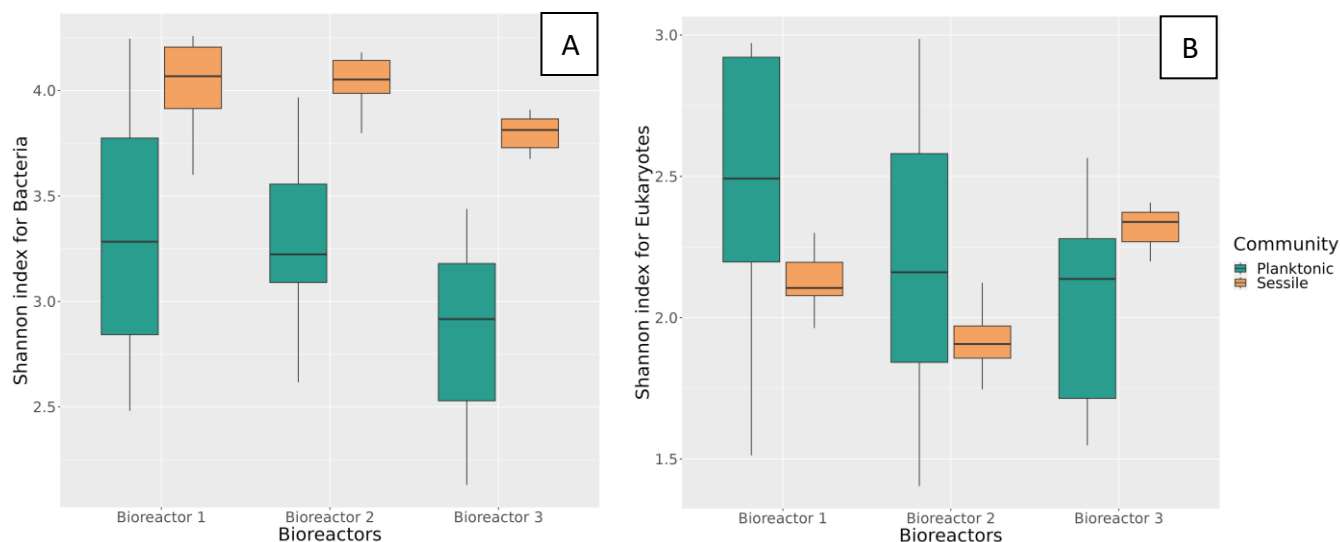
For the Archaea domain, due to the inability to sequence any of the DNA extracted from the rock pellets, we only analyzed the planktonic community. The peat bog was dominated by *Crenarchaeota* group 1.1c (32.1 %), *Nanoarchaeota* SCGC_AAA011-D5 (18.2 %), unc. *Bathyaerchaiea* (9.4 %), and *Methanoregula* (8.8 %). The initial groundwater was composed of unc. *Bathyarchaeia* (68.9 %) and *Crenarchaeota* group 1.1c (9.9 %) for the P1 sample; unc. *Bathyarchaeia* (53.1 %) and unc. *Crenarchaeota* (13.9 %) for P2; unc. *Bathyarchaeia* (47.1 %) and unc. *Crenarchaeota* (13.9 %) for P3; and unc. *Bathyarchaeia* (51.1 %) and unc. *Crenarchaeota* (18.1 %) for P4. Unlike the two other domains, we didn't observe major temporal patterns at the genus level in the planktonic community. The most abundant genera were the unc. *Bathyarachaiea* (bioreactor 1: 45.5 %, bioreactor 2: 40.9%, bioreactor 3: 47.6%) and the unc. *Crenarcheota* (bioreactor 1: 28.9 %, bioreactor 2: 29% and bioreactor 3: 27.9%) (Annexe A, Figure S).

1.4.4 Alpha diversity and correlation with environmental parameters

For the Bacteria domain, the α -diversity of the planktonic community varied from 2.130 to 4.246 while the ones for the sessile community varied from 3.601 to 4.259 (Figure 4A). There was a significant difference in the diversity indices between both lifestyles, with indices for the sessile community being higher (Kruskal-Wallis chi-squared = 31.729, p-value = 1.772×10^{-8}) (Annexe E).

To check which parameter was correlated to the alpha-diversity of the planktonic community in the bioreactors, we used multiple generalized linear models with the following parameters: DIC, DOC, NH_4 , NO_3^- and NO_2^- . The model with the best AIC score contained DIC, DOC, NH_4 , and none of the parameters were significantly correlated to the α -diversity (Annexe B, Table 6). For the sessile community, we only found that only mica was negatively correlated to the α -diversity (-27.32% with a p-value of 0.0347) (annexe B, Table 7).

Figure 4. Alpha-diversity (Shannon index) for the planktonic and sessile community for (A) the bacterial dataset and (B) eukaryotic dataset.



For the Eucaryote domain, the α -diversity indices varied from 1.404 to 2.899 for the planktonic community, and 1.746 to 2.407 for the sessile community. Only the third bioreactor had a higher α -diversity for the sessile community than the planktonic community (Figure 4B). However, there was no statistically significant difference between the two lifestyles (Kruskal-Wallis chi-squared = 0.31005, p-value = 0.5777). For the multiple generalized linear models, the better model also had DIC, DOC, NH_4 as the parameters. Here, DIC was the only parameter that significantly correlated with the α -diversity of the planktonic community (p-value = 0.00068) (Annexe B, Table 6). In this case, neither the minerals nor roughness were correlated to the α -diversity (Annexe B, Table 6).

For the Archaea domain, the α -diversity indices varied from 4.391 to 4.989 for the planktonic community. (Annexe A, Figure V). The best model had the same parameters as the other domains, and all the parameters were statistically correlated with the α -diversity (Annexe B, Table 6).

1.4.5 Beta diversity ordination analyses and influence of environmental variables

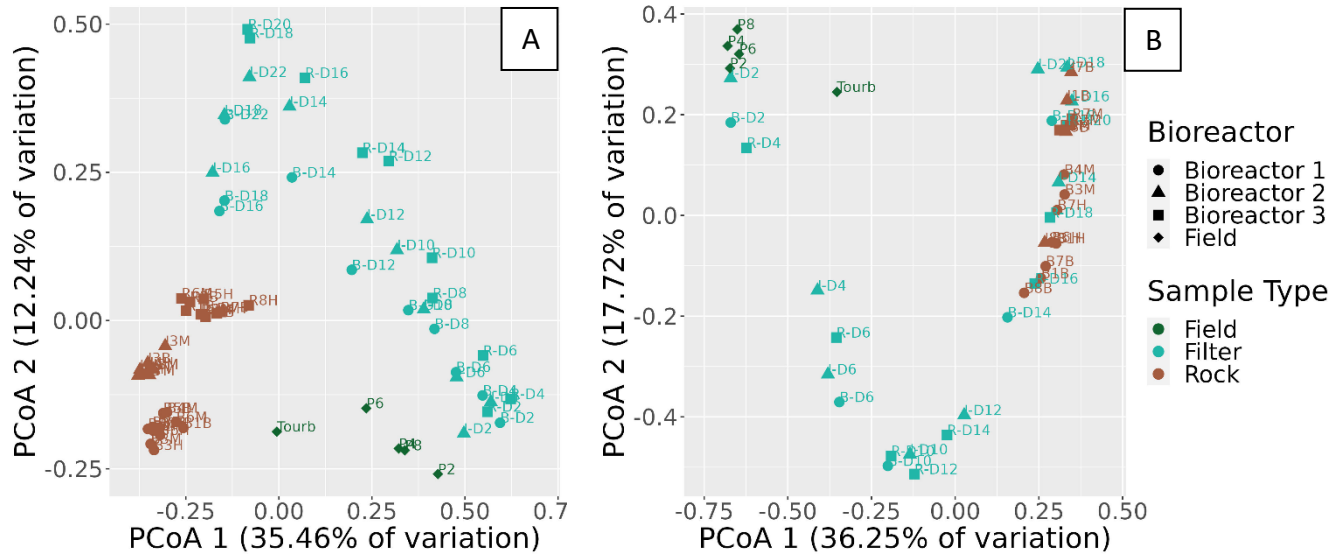
A PCoA was run for each domain in order to view the similarities between samples. For this analysis, we included the samples collected in the field on the day of sampling (the different groundwater samples and the peat bog water sample).

For the bacterial domain, the rock samples (sessile community) are all grouped based on which bioreactor they were in (Figure 5A). The filter samples (planktonic community) were grouped based on the day they were collected: days 2 to 8, 10 and 15, and 14 to 22. For the field data, the *in-situ* groundwater samples are closest to the water from days 2-8.

For the eukaryotic communities, the filter samples can be clustered into three groups: D2-D6, D10-D12 and D14-D22 (Figure 5B). As opposed to the bacterial community, we observe that the filter samples (D14-D22) are grouped with all the rock samples. Also, the *in-situ* samples taken from the well are closer to the D2-D4 samples. The peat bog data is isolated from the rest of the clusters.

For the archaeal community, only the field and filter samples are shown (Annexe A, Figure W). We observe two major groups with samples further apart from the rest. The first cluster contains most of the field *in-situ* water samples (P2-P4) and most of the filter samples (D2-D12). The second cluster contains the last days of the filter sample (D14-D18). Finally, one of the initial planktonic community samples (P1) as well as the peat bog sample are located far from the other groups.

Figure 5. (A) Bacterial and (B) eukaryotic β -diversity (using the Bray-Curtis distance matrix) represented in principal coordinate analysis. Different shapes were used to differentiate the samples based on the bioreactor (or field) it came from. Different colors were used to differentiate the type of the sample (Field for the initial planktonic community, Filter for the planktonic community and Rock for the sessile community). The Filter samples are numbered in order to know the day the sample were taken.



For the first PERMANOVA analysis, the filter and rock samples were separated and the field samples were excluded. The PERMANOVAs were done to determine the correlation between the β -diversity indices and the bioreactor replicates, the day the sample was taken (for the filter samples only). For the filter samples collected on days 2-8, the incubation time was significantly correlated with the β -diversity explaining 33.3 % of the community variance, while the bioreactor was not significantly correlated (Annexe B, Table 8). In contrast, for the filter samples collected days 10-12, both environmental parameters were significantly correlated with the β -diversity indices, with the bioreactor explaining 52.25 % of the variance, while incubation time explained 33.8 %. For the filter samples collected days 14-22, similarly to the previous cluster, the bioreactor explained 31.5 % of the variance while incubation time explained 27.8 %. As for the rock samples, the bioreactor was the only significant parameter, explaining 45.5 % of the variance. A second PERMANOVA was done using the filter and rock samples. The lifestyle was significantly correlated with the bacterial community and explained 32.3 % of the variance with a p-value of 0.001 (Annexe B, Table 8).

The results for the eukaryotic PERMANOVA are quite different than the ones found for the bacterial ones (Annexe B, Table 8). In this case, for the planktonic dataset, only the incubation

time parameter (DAY) was statistically significant, and for two clusters only ($R^2 = 56.31\%$ and $p\text{-value} = 0.002$ for the D2-D6 cluster and $R^2 = 35.47\%$ and $p\text{-value} = 0.006$ for the D14-D22 cluster). For the sessile dataset, only the bioreactor parameter was statistically significant ($p\text{-value} = 0.003$). Going through the different clusters, the incubation time explained more of the variations of the β -diversity (56.31 %) than the bioreactor (25.11 %) for the first group. We also see a similar case with the third cluster: 35.47 % for the day and 29.67 % for the Bioreactor parameter. In the second cluster, it's the opposite: the day only explains 26.413 % of the variation, while the bioreactor explains almost double the variation (47.02 %).

As for the sessile community, the bioreactor was the only significant parameter, explaining 46.5 % of the variance. In this case, the Bioreactor parameter was also statistically significant ($p\text{-value} = 0.003$). Furthermore, when comparing both lifestyles, we found that the planktonic and sessile communities were statistically different ($p\text{-value} = 0.001$). The second PERMANOVA showed that lifestyle (sessile/planktonic) was significantly correlated with the eukaryotic community and explained 18.8 % of the variance.

For the archaeal PERMANOVA, only the incubation time parameter was statistically different in the first cluster with a $p\text{-value}$ of 0.023 (Annexe B, Table 8), explaining 32.33 % of the variation.

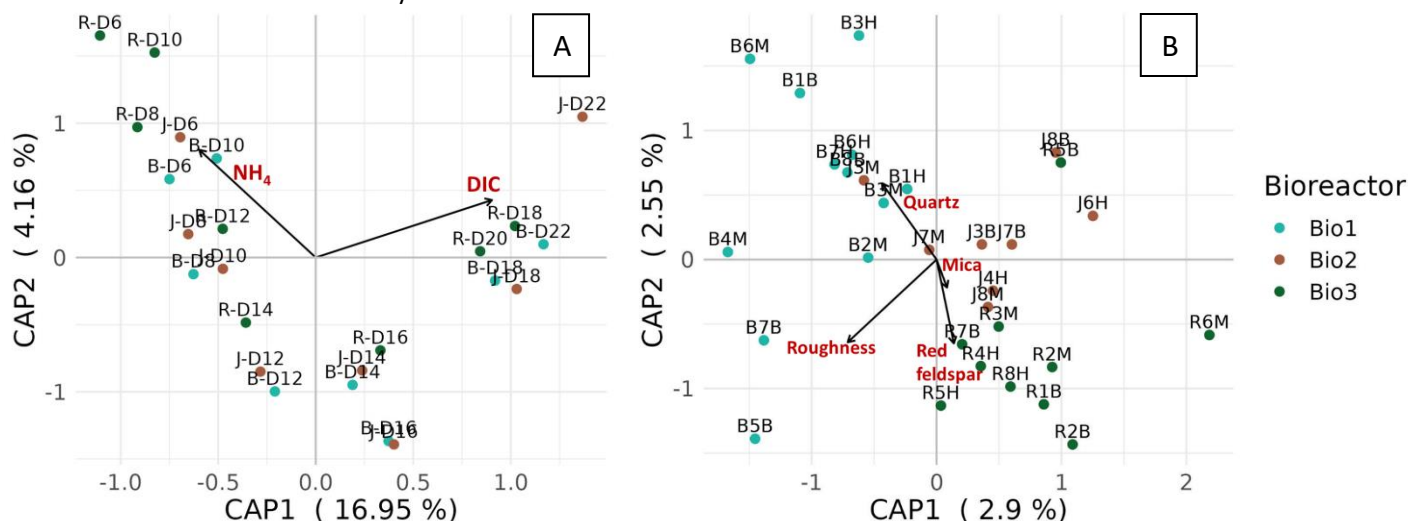
1.4.6 Beta diversity correlation with the water geochemical properties and the rock mineralogy and roughness

We used the same β -diversity dataset as in the previous analyses. However, the db-RDA did not include any of the field data. Since the parameters that were to be analyzed were different based on lifestyle, the planktonic dataset was analyzed separately from the sessile dataset. For the planktonic community, the geochemical parameters were analyzed, while for the sessile community, the rock properties were analyzed. Only the parameters present in the simplest model will be shown in the figures.

For the bacterial planktonic community, the simplest model had DIC and NH_4 , and the db-RDA showed that only the DIC was significantly correlated ($p\text{-value} = 0.003$) (Annexe B, Table 9). The db-RDA graph (Figure 6A) displayed different clusters based on incubation time: D2-D16 and D18-D22. NH_4 influences the community more at the beginning of the experiment (D2-D10) and DIC

has a higher influence at the end of the experiment (D18-D22). For the bacterial sessile community, the simplest model almost had all the parameters considered. However, there was collinearity between many minerals. Thus, we only used the following parameters: quartz, red feldspar, mica and roughness. We found that none of the parameters were significantly different parameter (Annexe B, Table 5). The db-RDA graph (Figure 6A) displayed clusters mainly based on the bioreactor, just like what we saw in the PcoA (Figure 5A). The bioreactor 1 looks more influenced by the quartz, while the second and third bioreactors are more influenced by the red feldspar.

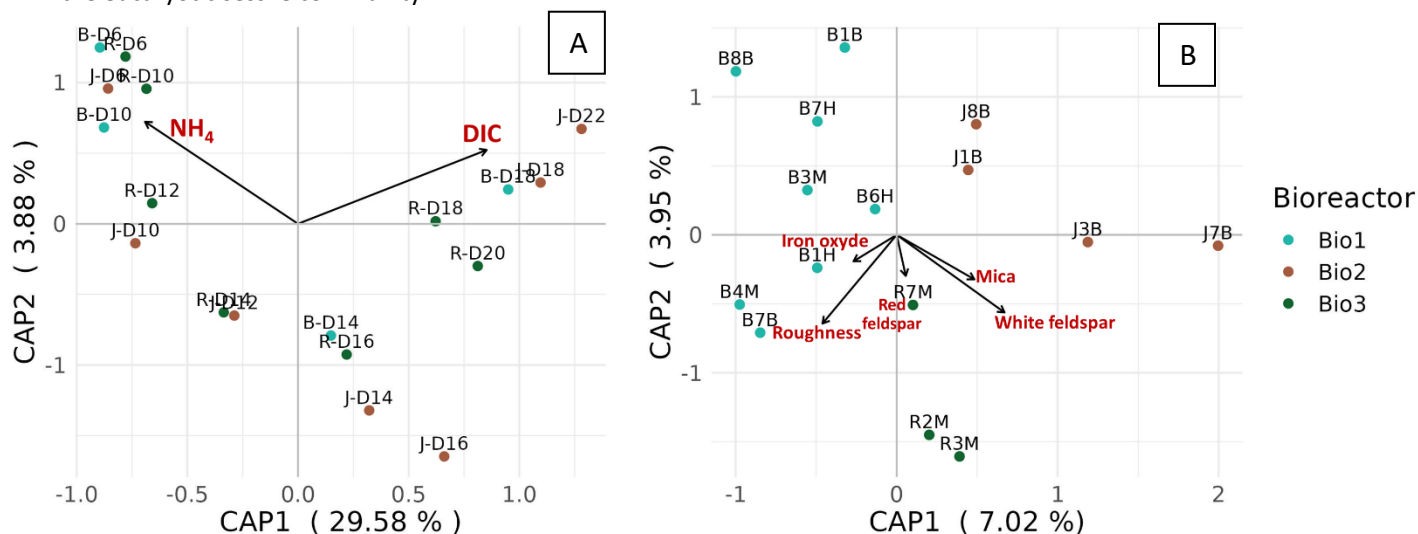
Figure 6. Visualization of the Distance-based redundancy analysis (dbRDA) based on (A) geochemical parameters of the water in the bioreactors for the bacterial planktonic community and (B) the mineralogy of the rock surfaces for the bacterial sessile community.



For the eucaryotic planktonic community, the simplest model had DIC and NH_4 , and only the DIC was significantly correlated with a p-value of 0.006 (Annexe B, Table 9). The db-RDA graph (Figure 7A) displayed clusters based on incubation time: D6-D10, D12-D16 and D9-D22. As for the Bacteria, even though it is not significantly correlated with the β -diversity, NH_4 influences the community more at the beginning of the experiment (days 6 to 10) while DIC influences the community at the end of the experiment (days 18 to 22). For the sessile community, the simplest model had none of the parameters except the undetermined minerals. Thus, we took the model in which there was no collinearity between the parameters. We ended up with a model containing the following parameters: iron oxide, red feldspar, white feldspar, mica and roughness. We found that none of the parameters were significantly correlated (Annexe B, Table 10). The db-RDA graph

displayed clusters based on the bioreactor, just like what we saw in the PCoA (Figure 5B). Even if the minerals are not statistically significant, it looks like the iron oxide and roughness influence more the first bioreactor, while the red feldspar influences more the third bioreactor (Figure 7B).

Figure 7. Visualization of the Distance-based redundancy analysis (dbRDA) based on (A) geochemical parameters of the water in the bioreactors for the eucaryotic planktonic community and (B) the mineralogy of the rock surfaces for the eucaryotic sessile community.



For the archaeal planktonic community, the simplest model had DIC and DOC, and both were significantly correlated (p-value of 0.001 and 0.002) (Annexe B, Table 10). The db-RDA graph (Annexe A, Figure X) displayed clusters based on incubation time: days 18 and 20 were far apart compared to the other samples and seemed to be explained by both DOC and DIC.

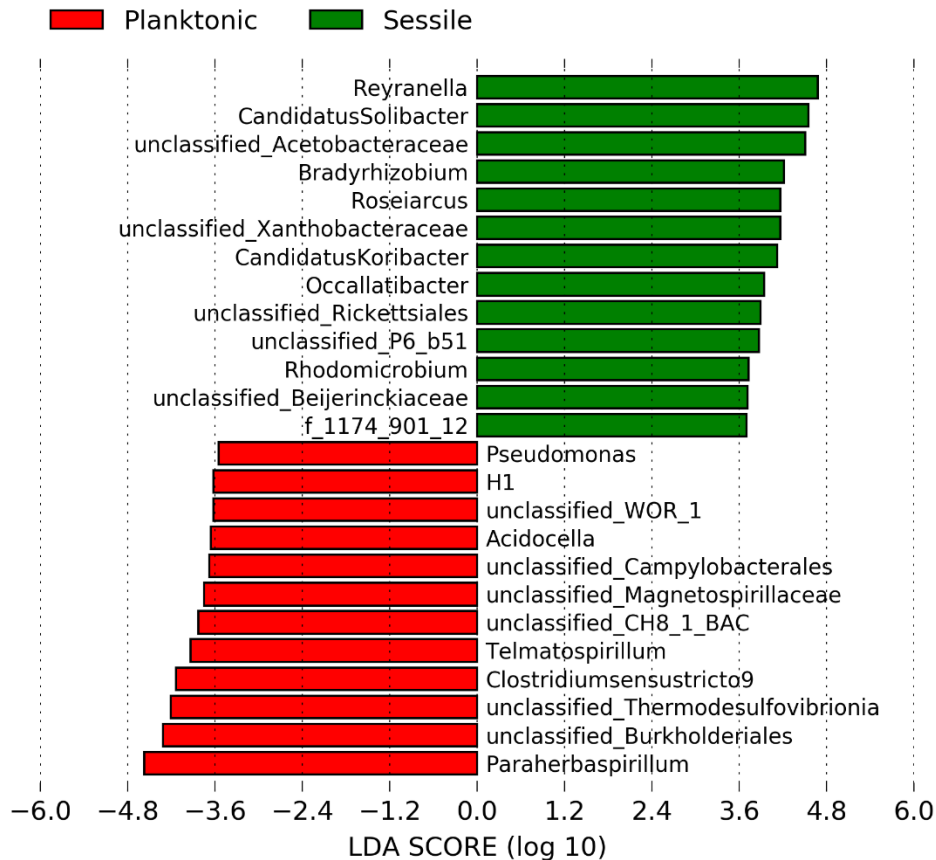
1.4.7 Discriminative genera explaining differences between the planktonic and sessile communities

In the LefSE analyses, we looked at which genera for each domain were significantly different in relative abundance for both different lifestyles. The LefSe scores shown in the figures 6 and 7 can be interpreted as the degree of consistent difference in relative abundance between taxa in the two lifestyles. There was no separation done based on the bioreactors.

For the sessile bacterial community (Figure 8), the genera with the highest LDA scores are: *Reyranelia*, *Candidatus Solibacter* and *Bradyrhizobium*. There is also the *unclassified* (unc.) *Acetobacteraceae* among them, but it has only been identified to the family. Looking at the

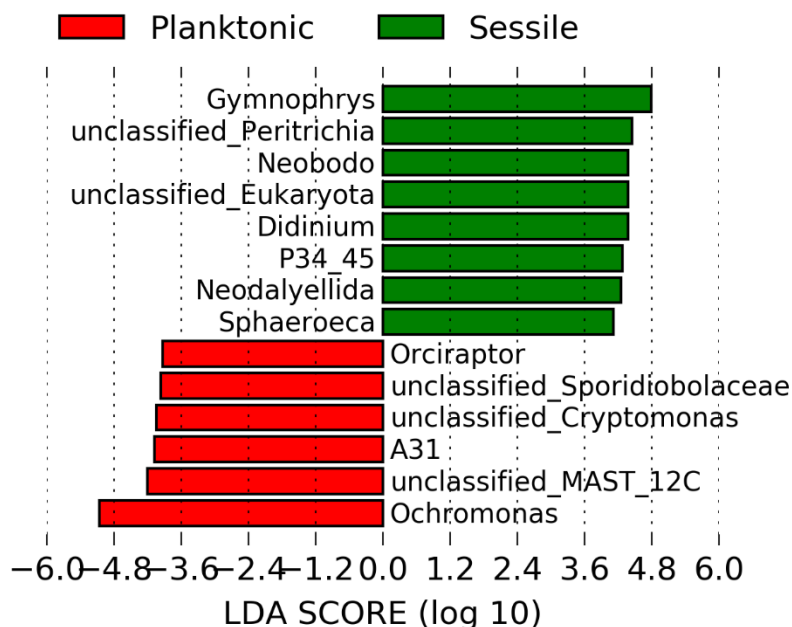
planktonic community, we notice fewer genera (5 genera in contrast to 9 genera in the sessile community). Here, the two genera with the highest score are the *Paraherbaspirillum* and *unclassified (unc.) Burkholderiales*. It is important to point out that these genera weren't necessarily the most abundant when looking at the relative abundance plots (Annexe A, Figures O and P).

Figure 8. Differential abundance of bacterial genera between the planktonic and sessile communities.



For the eukaryotic community, we notice there are fewer genera in the planktonic category than in the previous domain (Figure 9). Indeed, in this case, we only have two genera: *Ochromonas* with the highest score and A31. There are a lot more in the sessile category like in the bacterial community. The LDA scores are quite close, varying between 4 and 5. *Gymnophrys* has the highest score, while *Spumella* has the lowest.

Figure 9. Differential abundance of eukaryotic genera between the planktonic and sessile communities.



Since the goal of this analysis was to determine which genera for each domain were significantly different in abundance between the sessile and planktonic communities, the analysis was not done for the Archaea dataset.

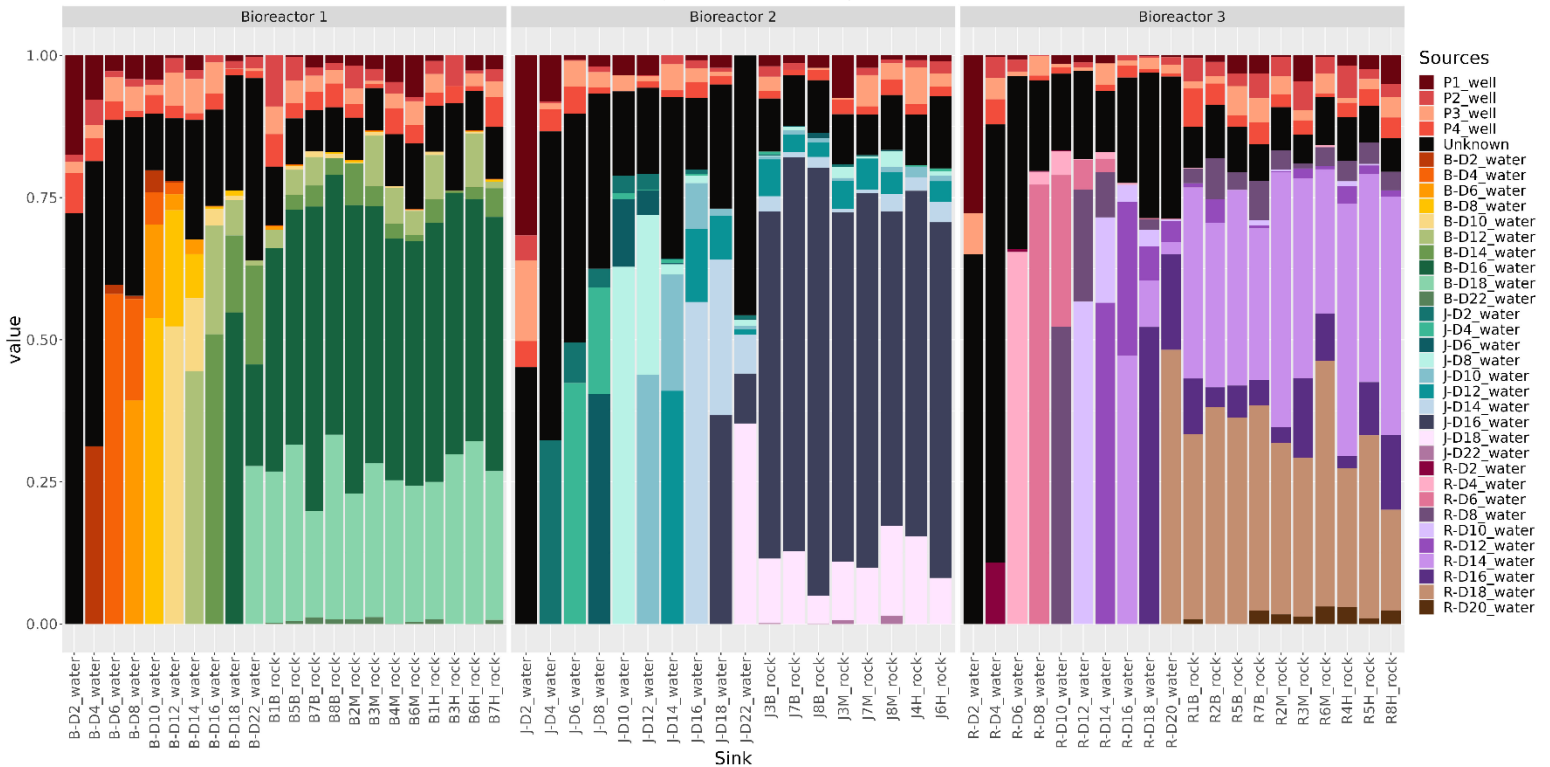
1.4.8 Microbial Source Tracking

Microbial source tracking analyses were done for each domain. For the water samples, we used the peat bog, the in-situ groundwater, and the previous time point water samples as potential sources. For the rock-attached samples, we used the peat bog, the *in-situ* groundwater, and the water samples from the previous time points as potential sources.

For the Bacteria domain, the peat bog was a huge contributor to the initial groundwater samples (data not shown, 49.1 to 62.3 %). For the planktonic community, we can see that there are mostly unknown sources for the first 4 days of incubation (Figure 10). However, afterward, we see that the previous water time points became the sources for the later incubation times. We see this trend in all three bioreactors. We also observe that the *in-situ* groundwater contributes more at the beginning than near the end of the experiment. Moreover, among the well sources, it's mostly the first liter of water that was pumped that is influencing the planktonic community. For the sessile community, the water samples from the last couple of days are the biggest contributors.

For bioreactor 1, it was the 16th and 18th day of the experiment. For the bioreactor 2, D16 was again the biggest contributor. The D18 sample was less important than for the previous bioreactor. Oddly enough, for these two bioreactors, the D22 sample, which was the 22nd day of the experiment, did not contribute as much to the sessile community. Lastly, for the bioreactor 3, D14 and D18 influenced the sessile community the most.

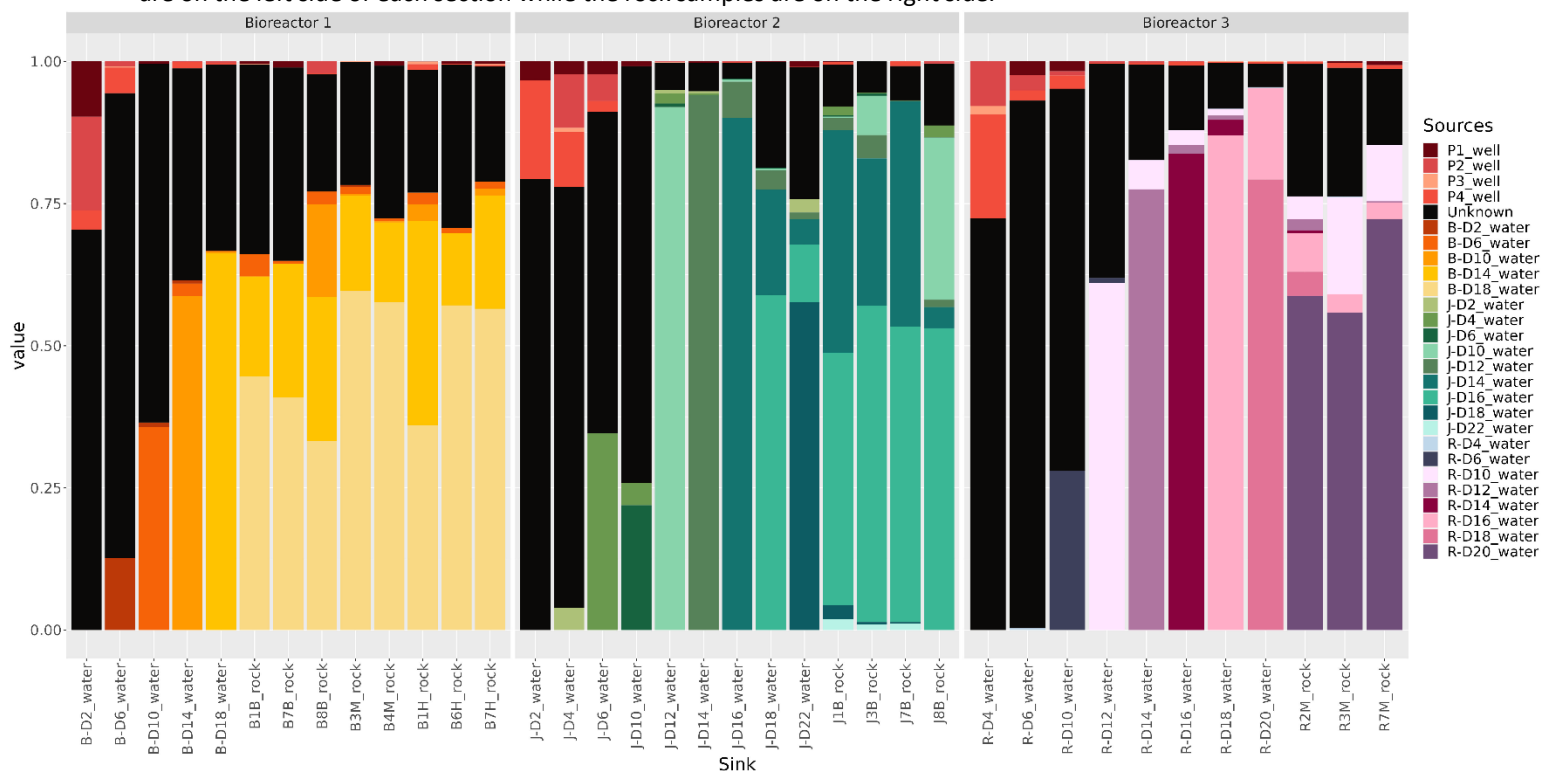
Figure 10. Contribution of different bacterial sources to every sink. The sinks are divided into three groups based on which bioreactors the sink sample came from. The sources were also divided based on which bioreactors the sample was taken. However, the field data (P1 to P4) were sources for all three bioreactors. The planktonic sample are on the left side of each section while the rock samples are on the right side.



For the Eukaryote domain, the peat bog was also a huge contributor to the initial groundwater samples (data not shown, 42 to 53.2 %). Unfortunately, we had fewer samples, whether for the planktonic community or the sessile. We do, however, notice a similar pattern to what we see with the bacterial community. Indeed, in this case too, we see that there are mostly unknown sources in the first water samples (Figure 11). After the 10th day of the experiment, the unknown sources became less important, and the previous water time points became the sources for the later incubation times. These trends are also visible in all three bioreactors. We also notice the groundwater data contributing a lot at the beginning of the experiment, especially the fourth

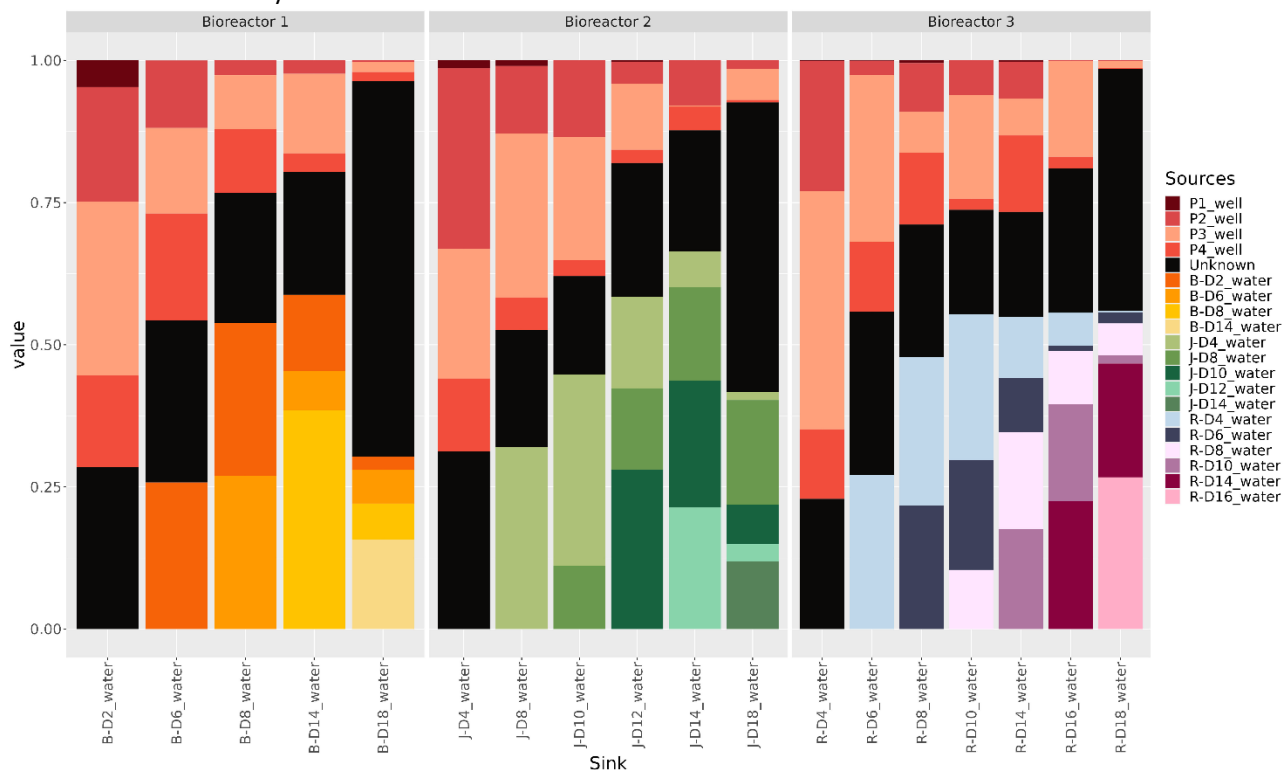
sample. However, they almost don't contribute at all after D10. In the sessile community, again, the last couple of days of our water samples are the biggest contributors. For bioreactor 1, D18 was the predominant contributor. For the second bioreactor, D14 and D16 were the biggest contributors. Lastly, in the final bioreactor, D20 influenced the sessile community the most.

Figure 11. Contribution of different eukaryotic sources to every sink. The sinks are divided into three groups based on which bioreactors the sink sample came from. The sources were also divided based on which bioreactors the sample was taken. However, the field data (P1 to P4) were sources for all three bioreactors. The planktonic sample are on the left side of each section while the rock samples are on the right side.



For the Archaea domain, the peat bog was also a huge contributor to the initial groundwater samples (data not shown, 35.1 to 41.8 %). Even though we have fewer data than the rest, we can still see some trends. In the beginning, the groundwater samples were the main contributors to the planktonic community (Figure 12), with the third one contributing more. As the day goes on, the previous water time points become important sources for the later samples. Yet, unlike what we saw in the bacterial and eukaryotic communities, there are more unknown sources in the last day (D18).

Figure 12. Contribution of different archaeal sources to every sink. The sinks are divided into three groups based on which bioreactors the sink sample came from. The sources were also divided based on which bioreactors the sample was taken. However, the field data (P1 to P4) were sources for all three bioreactors. In this case, there are no samples for the sessile community.



1.5 Discussion

1.5.1 Evolution of planktonic communities over time during the incubations

In the bacterial community, *Sulfurimonas* was prevalent in the water at the beginning of the experiment and was also the dominant taxon in the initial groundwater used for the incubations, but its relative abundance decreased until the end of the experiment. *Sulfurimonas* is a genus mostly found in sulfidic habitats such as hydrothermal vents, marine sediments, and a terrestrial oil field (Han & Perner, 2015). These bacteria use sulfur to generate energy (Inagaki and al., 2003). This genus was detected in the nearby surface peat bog acting as the main recharge zone for the aquifer ecosystem, where it was not a dominant taxon. Since the bacteria in the peat bog were shown to be major contributors to the groundwater community, it is likely this bacterium originates from the peat bog. There was likely a good amount of sulfur in the well (based on the smell of sulfur that could be detected while sampling), therefore this bacteria would have been able to thrive. Now, one reason that might explain the decline of this genus in the experiment

would be a reduction of sulfur or any sulfur derivative in the water. Since the only source of sulfur is the 10 L collected at the beginning, it could mean that there was not enough sulfur for all the *Sulfurimonas* to thrive. Unfortunately, the concentration of sulfite (SO_4) was not measured. For the eukaryotic planktonic community, there was *Ochromonas*, whose relative abundance increased but then decreased as *Spumella* increased. *Ochromonas*, are algae that can be found in very acidic lakes (Lessmann and al., 2000). They are flagellates and common dwellers in freshwater habitats (Wollmann and al., 2000). The increase of its abundance was not expected as there was little light in the incubator chamber in which the bioreactors were during the whole experiment. Indeed, the bioreactors were purposefully covered to avoid creating a new variable: the light intensity to which the water was exposed to. The moments when the microorganisms could have gotten some light would have been during the filtration during which the water was put into semi-opaque sterile bottles and during the small sample collections made to verify the pH and dissolved O_2 levels. However, such small exposure time could have not been the cause of such an increase. Therefore, it is possible that some change in the environment or community structure favored their proliferation. *Spumella* is a heterotrophic flagellate (Rothhaupt, 1996). A study conducted by Rothhaupt (1996) looked into the competition for nutrients between *Ochromonas* and *Spumella*. The study found that *Spumella* were better at acquiring food for themselves than *Ochromonas* in a dark environment. It's also mentioned that *Ochromonas* has difficulty growing in the dark and that it needs a sufficient concentration of bacteria to feed on. These two points might explain the observed variation in the relative abundances of *Ochromonas* and *Spumella*. For the Archaea, the taxa could only be identified at the phylum level: unc. *Bathyarchaeota* and the unc. *Crenarcheota*. Thus, it is difficult to gain specific information on the Archaea present in the aquifer.

When we looked to see if any of the geochemical water properties were correlated to the α -diversity, we only found that the DIC was significantly correlated to the α -diversity of the Eukaryotes. This could be explained by the high relative abundance of *Ochromonas*, a mixotrophic genus that uses dissolved inorganic carbon to develop (Katechakis and Stibor, 2006). However, it is important to note that they will use dissolved inorganic carbon when the bacterial abundance is low (Terrado and al., 2017). If not, they will mostly acquire carbon by consuming bacteria. In

our case, we have the data to corroborate this statement, but future studies could look into this potential interaction between DIC and *Ochromonas*'s acquisition of carbon. The other parameters (DOC and NH_4) were not, which could mean that we should look into other parameters such as the concentration of sulfur derivatives since we observed that some microorganisms in this aquifer metabolize sulfur.

When looking at the relationships between the community composition (β -diversity) and the incubation time, and the bioreactor factor, we found that for the bacteria, the incubation time factor had a significant effect on the variation of the planktonic community composition, explaining the highest variance in the community composition compared to the bioreactor factor. This shows that the community changed during the 24 days of the experiment. We also saw this pattern with the dominant taxa relative abundance fluctuations. The bioreactor parameter was almost always statistically significant apart from the first days of the experiment, suggesting that other factors not considered in this study had a big influence on the water's geochemical properties. Also, it could mean that even a small variation in the geochemical properties can affect the community composition. A study by Chandler and al. (2012) found that an increase in ions such as Na^+ would alter prokaryote communities. Other studies have also found that the variations in various ions, dissolved oxygen concentration, and pH could lead to a change in the composition and structure of the microbial community (Amalfitano et al., 2014; Sang et al., 2018; Chakraborty et al., 2022). It is important to note that in some of these studies, the variations of ions were statistically significant, which is not what we have here. It is also important to note that not all the parameters of the aquifer water were measured and reproduced in the bioreactor such as the pressure inside the aquifer. This, and random effects, could explain why the bioreactor parameter is significant.

For the planktonic community in the Eukaryote dataset, the incubation time significantly influenced two out of the three clusters (days 2 to 6, and days 14 to 24). This could mean that the Bacteria are more susceptible to geochemical changes in the water than Eukaryotes. However, further analysis should be done in order to validate this statement.

For the bacterial and eukaryotic planktonic communities, the db-RDA showed that only the DIC was significantly correlated. DIC influenced more the communities at the end of the experiment and that NH_4 , even though it was not statistically significant, influenced more the communities at the beginning of the experiment. This could be caused by a shift in the dominant bacteria and eukaryotes in the planktonic communities. As mentioned before, the bacterial planktonic community was mostly dominated by the *Sulfurimonas* genus. A study by Sievert and al. (2008) found that these bacteria are dependent on ammonia uptake from the environment to assimilate nitrogen. This could explain why NH_4 was influencing the community only at the beginning. It could have been turned into ammonia and then used by all the *Sulfurimonas* bacteria. As time went on, *Sulfurimonas* became less and less abundant, and unc. *Acetobacteraceae* became the most abundant bacteria. This shift could explain why suddenly another parameter started influencing the community, but it is not possible to link the high abundance of unc. *Acetobacteraceae* to DIC. For the eukaryotic planktonic community, we also saw a shift in the dominant genera. In the beginning, it was mostly unclassified genera, which makes it hard to associate a certain bacteria with a metabolic process involving ammonium or ammonia. Later in the experiment, *Spumella* becomes the most abundant. However, *Spumella* doesn't use dissolved organic carbon, as it is a heterotroph incapable of fixing carbon (Kim and al., 2020). This means that there must be another reason why DIC influenced more of the community at the end of the experiment. Maybe a look into the interactions between bacteria and eukaryotes could help us understand this change in parameter.

For the planktonic community in the Eukaryote dataset, the incubation time significantly influenced two out of the three clusters (days 2 to 6, and days 14 to 24). This could mean that the Bacteria are more susceptible to geochemical changes in the water than Eukaryotes. However, further analysis should be done in order to validate this statement.

Looking at the FEAST results for the archaeal community, we see that, even towards the end of the incubations, there were still a lot of unknown sources contributing to the planktonic communities. Indeed, the unknown sources are even the dominant source for the last sample in each bioreactor at day 24. They are also the dominant source in the first couple of samples of the planktonic community for the Bacteria and Eukaryotes too. This could suggest that the planktonic

community underwent many structural and diversity changes during the first days of the incubations. There might also be some contamination that happened during the experiment, such as from the gas mix machine or when the bottles were opened during the sampling. As for the sessile communities, for Bacteria and Eukaryotes, we found that it was the planktonic samples collected in the final days of the experiment that were the biggest source of their microbial communities. This could be due to the changes of relative abundance of the microorganisms in the bioreactors. As mentioned before, when looking at the relative abundance of planktonic communities, there was a temporal variation. However, at the end of the experiment, the most abundant microorganisms in the planktonic community were also the most abundant in the sessile community. Therefore, since the planktonic samples near the end had as similar distribution of dominant microorganism to those found in the sessile community, the analysis would have found them as being a bigger source than the other samples.

1.5.2 Microbial colonization of the rock surfaces

Looking at the bacterial sessile community, we observed that the unc. *Acetobacteraceae* was the most abundant taxon colonizing the rock surfaces after 24 days of incubation. This family of bacteria lives in acidic environments ranging from a pH of 4 to 7 (Rolf & Ahnert, 1999). Since our natural aquifer system is acidic (pH = 5.21) it would explain why they thrive. There are previous studies that also found this family in their aquifers, such as Villeneuve and al. (2022) and Lehosmaa and al. (2021). Some species in this family have the ability to form flocs in the water (Knotek-Smith and al., 2020), but there weren't studies that found them in groundwater biofilms. For the Eukaryotes, *Spumella* (P34.45) was one of the discriminative taxa in the sessile community. They were found in other aquifers (Loquay and al., 2009) and in biofilms (Böhme and al., 2009).

The Venn diagrams showed us that there were a lot of genera shared between the planktonic and sessile communities (88.51% for the Bacteria and 64.52% for the Eukaryotes), suggesting that a lot of microorganisms in the aquifer are able to switch lifestyles. As for the Archaea domain, it was not possible to retrieve sequences from the rock samples. In other studies, researchers were able to find archaeal species in their biofilm samples, but the environmental conditions of the aquifers were different from ours (Smith and al. 2019). In one case, the Archaea were associated

with sulfate-reducing bacteria. Having found bacteria such as *Sulfurimonas* who are able to metabolize sulfur and can be found in biofilm, it would be expected to find some archaeal species in the rock samples as well (Benedek and al., 2016; Probst and al., 2013). This could possibly mean that their abundance was too low or that the DNA extracted was too degraded to be able to be sequenced.

Furthermore, our study showed that bacterial α -diversity indices were significantly different between both lifestyles with sessile indices being higher, while eukaryotic diversity indices did not significantly differ. This could mean that the sessile lifestyle is preferred by the bacteria rather than the planktonic lifestyle. Indeed, the colonization and development on rock surfaces is quite common (Scholl and al., 1990; Nuppenen-Puputti and al., 2023; Vishnyakova and al., 2022), and the colonization of minerals on the surfaces can offer a physical protection against the water flow (Dar and al., 2019). Also, since 88.5% of the genera were shared between both communities, it might mean that most of the bacteria present in the aquifer have the ability to switch lifestyles.

The β -diversity indices were also significantly different between both lifestyles for both the Bacteria and Eukaryote domains. For the Bacteria, we found that the genera that were significantly higher in the sessile community have different characteristics. *Reyranella* can reduce NO_3 (Pagnier and al., 2011), *Candidatus solibacter* are acidophilic (Dedysh and al., 2017), and *Bradyrhizobium* can fix nitrogen (Ormeño-Orrillo and Martínez-Romero, 2019). Some genera that were significantly higher in the planktonic community (*Telmatospirillum*, *Acidocella*, and *Pseudomonas*) are acidophilic (Dobrovolskaya and al., 2020; Hiraishi, 2015; Dejsirilert and al., 1991). So, the planktonic community seems to contain more acidophilic microorganisms than the sessile community. This could be because the sessile organisms could have formed biofilms on the rock surfaces, which would protect them from environmental stressors like an acidic environment (Yin and al., 2019).

For the Eukaryote domain, some of the genera that were significantly higher in the sessile community were *Gymnophrys*, *Neobodo*, *Didinium*, and *P34.45* (*Spumella*). We were not able to find literature about the presence of *Gymnophrys* in biofilm, but we did find a study mentioning their presence in aquifers in Quebec (Groult and al., 2023). *Neobodo* is a genus commonly found

in stream biofilms (Böhme and al., 2009). They are heterotrophic like *Spumella* (P34.45) and *Didinium* (Böhme and al., 2009; Corliss and Esser, 1974). The presence of these heterotrophs in the sessile community could be explained by a higher density of bacteria (prey for the heterotrophs) on the rocks than in the water. As for the genera present in the planktonic community, we have *Ochromonas*, *A31*, and *Orciraptor*. As mentioned previously, *Ochromonas* are flagellate algae that can be found in very acidic lakes (Lessmann and al., 2000; Wollmann and al., 2000), which would not only explain their presence in the aquifer but also their significant abundance in the planktonic community. *Orciraptor* also happens to be a flagellate, and they can be found in acidic bogs or stagnate water (Hess and Melkonian, 2013). This could mean that *Orciraptor* came from the bog that is connected to the aquifer and managed to survive in the new environment. Not much information can be found on the *A31* genus, but if we look at the family (*Perkinsidae*), we find that this family interacts highly with other microorganisms. In the study, they found that *Perkinsidae* were involved in many predator-prey interactions. In a future study, it could be interesting to look more into the predator-prey interactions between bacteria and eukaryotes.

For sessile in both the Bacteria and Eukaryote datasets, we mostly see that the dominant sources of each sample are the planktonic samples from the last couple of days of the experiment. However, we see that the samples from previous days (the beginning of the experiment and even the well samples) also contribute to the sessile communities. This could give us an idea of when the organisms started colonizing the surface. However, competition among microorganisms for resources and space on the surfaces (Fierer and al., 2010) could explain why we see mostly samples taken in the last days as dominant sources for the sessile communities. The microorganisms that had colonized the surface at the beginning could have then been replaced by others (Abed and al., 2019; Walker and del Moral, 2011; Lee and al., 2008), since we did notice changes in relative abundance in the planktonic community over time.

1.5.3 Influence of the rock properties on surface-attached microbial communities

We found that the roughness had little effect on the bacterial absolute abundance. Indeed, we found that the correlation between the abundance and the roughness was only 0.22 (Roughness 1) and 0.15 (Roughness 2). Usually, the roughness would allow bacteria, and microorganisms in

general, to colonize a surface with greater ease (Viles, 1995; Taylor and al., 1998). It is possible that the values of roughness we had here (going from $R_a = 0.14353$ to $R_a = 2.18455$) were too high or too low to see a higher correlation (0.22 and 0.15). In a study conducted by Li and Logan (2004), they looked at bacterial adhesion on 11 different surfaces and found that there was more bacterial adhesion on rougher surfaces. They thus suggested that there may be ranges of roughness where bacterial adhesion is correlated with roughness, while other ranges would not be correlated with roughness. A previous study also found that the surface charge of the microorganism's cell wall helps them adhere to surfaces (Johnson & Logan, 1996; Li & Logan, 2004). In Johnson & Logan's study (1996). The authors found that surfaces coated with iron and quartz (Fe-quartz) would have more bacterial adhesion than uncoated surfaces. Metal-oxide, like what we found as a form of iron oxide, can also increase the adhesion of bacteria. However, in our experiment, there was a low percentage of iron oxide. Thus, the low percentage of iron oxide could explain the little effect of abundance and the roughness.

The bioreactors had a significant influence on both the bacterial and eukaryotic sessile communities. There were no significant differences in either roughness or mineralogy measurements between bioreactors, suggesting other factors may have influenced the differences in community structure between the three bioreactors. However, although not significant, the db-RDA graph suggested that different minerals were correlated to each bioreactor sessile community. Seepage of groundwater and microbial activities can modify pH and water geochemistry, which could have altered the mineral composition of the rocks over the course of the experiment. Local differences for each rock chip composing each bioreactor could have contributed to the observed influence of the bioreactor on the sessile community composition.

Only the quartz had a statistically significant effect on the β -diversity of the bacterial sessile community. This finding is corroborated by previous studies that found that some bacteria could colonize quartz (Roberts, 2004; Smith and al., 2000). Although another study conducted on soil did find that feldspar (more precisely potassium feldspar) also had an effect on the microbial community (Zhang and al., 2020). Since the other minerals did not significantly influence the bacterial sessile community, it might mean that the microorganisms in the aquifer prefer

colonizing the quartz over the other minerals. Another study with mineral chips instead of rock chips could help answer this supposition.

1.6 Conclusion

In summary, many microorganisms were found in both communities and their relative abundance varied a lot during the 24 days experiment. For example, *Sulfurimonas*, a sulfur-utilizing bacteria initially dominated the bacterial community but experienced a decline by the end of the experiment. A possible explanation of the fluctuation of relative abundance could have been predator-prey relationships. For instance, the eukaryotic planktonic was first dominated by the *Ochromonas* genus. *Ochromonas* was likely in competition with *Spumella*, a heterotrophic flagellate. At the end of the experiment, *Spumella* was the most abundant genera. The study also found that the α -diversity of eukaryotes was significantly correlated with dissolved inorganic carbon (DIC), suggesting a role for *Ochromonas*, who utilizes DIC for development. Examining beta diversity and community composition, the incubation time had a substantial impact on bacterial and eukaryotic communities, indicating changes throughout the experiment.

In the FEAST analyses, we discovered that the attached (sessile) communities in both bacteria and eukaryotes are mostly affected by the planktonic samples collected in the final days of the experiment. Although samples from earlier days had a slight influence, it implies a gradual colonization of surfaces over time. Also, despite the general expectation that rougher surfaces enhance bacterial colonization, this study only found a weak correlation. The range of roughness used might have been too or not extreme enough to demonstrate a stronger correlation.

For the bacterial sessile community, only quartz had a statistically significant effect on the β -diversity, which is consistent with previous studies. This might mean that there is a potential preference of microorganisms in this aquifer for colonizing quartz over other minerals. It would be interesting to verify this hypothesis with another study in which mineral chips would be used instead of rock chips.

DISCUSSION GÉNÉRALE

Les objectifs de cette étude étaient de comparer la communauté microbienne sessile à la communauté planctonique dans un système d'aquifère au Québec ainsi que de déterminer l'influence de la composition minéralogique et la rugosité de la surface rocheuse sur la structure des communautés microbiennes colonisant les roches.

Nous avons trouvé que la rugosité des surfaces rocheuses avait peu d'effet sur l'abondance des bactéries, expliquant seulement 4,8 % et 2,3 % de la variation de l'abondance. Ces résultats sont surprenants, car la rugosité est généralement associée à une meilleure colonisation d'une surface par les bactéries (Viles, 1995 ; Taylor et al., 1998). Nous pensons qu'il serait intéressant d'essayer des surfaces plus rugueuse et moins rugueuse afin de voir si ce n'était pas juste les valeurs de rugosité qui n'étaient pas optimales.

Nos résultats sur la taxonomie des microorganismes nous ont révélé que les genres les plus abondants dans la communauté planctonique à la fin de l'expérience sont aussi les plus abondants dans la communauté sessile. Nous avons aussi noté une variation temporelle pour les communautés planctoniques des bactéries et eucaryotes. Cela pourrait être dû aux variations des paramètres physico-chimiques ou d'une déplétion de nutriments. En effet, puisque les bioréacteurs ne sont que connectés à une bouteille contenant de l'eau de l'aquifère (elle-même peuplée de microorganismes), il se peut que vers les derniers jours, il y ait eu un déficit de nutriments que les microorganismes avaient besoin. Par ailleurs, il se peut aussi que les communautés microbiennes dans ces bouteilles sources aient également changé au cours du temps.

Par exemple, dans la communauté planctonique des bactéries, *Sulfurimonas* était le genre ayant l'abondance relative la plus haute au début, mais son abondance a diminué petit à petit (Annexe A, Figure O). *Sulfurimonas* est souvent trouvé dans des habitats sulfuriques (Han & Perner, 2015) et ils utilisent le soufre pour produire de l'énergie, d'où l'importance de mesurer la concentration de dérivés soufrés. La diminution de ce genre durant l'expérience pourrait s'expliquer par une réduction du soufre ou de ses dérivés dans l'eau.

Quant à l'analyse LEfSe sur les bactéries, elle nous a permis de déterminer quels genres étaient discriminatifs dans les communautés sessiles et planctoniques et nous avons remarqué que la communauté planctonique contenait davantage de microorganismes acidophiles que la communauté sessile. De plus, les diagrammes de Venn ont montré qu'il y avait de nombreux genres présents dans les deux communautés, ce qui nous laisse croire que plusieurs microorganismes dans cet aquifère peuvent changer de mode de vie. Par contre, n'ayant pas fait de métagénomique, il n'est pas possible de le confirmer.

Pour ce qui est des résultats sur la diversité- α , nous avons déterminé que celle-ci différait significativement entre les modes de vie, avec une préférence pour le mode sessile (p -valeur = $1,772 \times 10^{-8}$). Par contre, chez les eucaryotes, aucune différence significative n'a été observée (p -valeur = 0,5777). Les communautés bactériennes sessiles présentaient une diversité- α plus élevée que les communautés planctoniques, suggérant une préférence des bactéries pour le mode de vie sessile. En effet, la colonisation et le développement sur les surfaces rocheuses sont assez courants (Scholl et al., 1990; Nuppenen-Puputti et al., 2023; Vishnyakova et al., 2022), et la colonisation de minéraux à la surface peut offrir une protection physique contre le courant d'eau (Dar et al., 2019). Lorsque nous avons examiné si l'une des propriétés physico-chimiques de l'eau était corrélée à la diversité- α , nous avons trouvé que le carbone inorganique dissout (DIC) était le seul paramètre physico-chimique qui était significativement corrélé à la diversité- α des eucaryotes (p -valeur = 0,00068). Vu que les autres paramètres (DOC, NH_4) ne l'étaient pas, il serait bien d'examiner d'autres paramètres tels que la concentration des dérivés de soufre, comme il l'a été mentionné plus tôt.

L'étude a révélé que la composition des communautés bactériennes (diversité- β) a évolué de manière significative sur une période de 24 jours. Il est à noter que les communautés bactériennes sessiles étaient influencées par le bioréacteur dans lequel elles se trouvaient, même si les caractéristiques physico-chimiques de l'eau utilisée dans l'expérience étaient similaires entre les bioréacteurs.

Dans le cas des eucaryotes, il est intéressant de noter que la communauté sessile était moins influencée de manière significative par le bioréacteur dans lequel elle se trouvait. Cependant, la

communauté planctonique a changé de manière significative sur une période de 24 jours. Cela pourrait suggérer que les bactéries sont plus sensibles aux changements que les eucaryotes. Cependant, des analyses plus approfondies sont nécessaires pour valider cette hypothèse.

En ce qui concerne les minéraux et la rugosité, seul le quartz avait un impact significatif sur la diversité- β des bactéries sessiles, suggérant que ces micro-organismes ont une préférence pour le quartz par rapport à d'autres minéraux. En ce qui concerne la communauté eucaryote, aucun des minéraux l'a influencé. Une autre façon de faire ce projet impliquerait d'avoir des morceaux de minéraux au lieu de roche afin de mieux voir l'impact de chaque minéral sur les communautés.

Pour ce qui est des analyses FEAST, nous avons trouvé que dans les communautés des bactéries et eucaryotes, les communautés sessiles sont principalement influencées par les échantillons planctoniques prélevés aux derniers jours de l'expérience. Cependant, les échantillons pris plus tôt ont aussi contribué faiblement. Cela suggère une colonisation progressive des surfaces. De plus, la compétition entre les microorganismes pour les ressources et l'espace sur les surfaces (Fierer et al., 2010) pourrait expliquer pourquoi nous observons principalement des échantillons prélevés les derniers jours en tant que sources dominantes pour les communautés sessiles. Les micro-organismes qui avaient colonisé la surface au début auraient aussi pu être remplacés par d'autres (Abed et al., 2019 ; Walker et del Moral, 2011 ; Lee et al., 2008).

Pour la communauté archéenne, malgré une expérience de 24 jours, des sources inconnues demeurent importantes, même dans les derniers échantillons. Cela n'aurait pas dû se produire puisqu'il n'y a pas de sources externes qui pourraient expliquer ces sources inconnues. Des recherches supplémentaires seraient nécessaires pour identifier ces sources inconnues.

Annexe A: Supplementary figures

Figure A. Diagram of rock pellets and a bioreactor containing columns with the pellets. Note that the dimensions of the rock pellets are approximate and will vary depending on the pellet. Additionally, the bioreactor has room for 8 columns, unlike the 3 shown here. On the right is an example of pellets in a column.

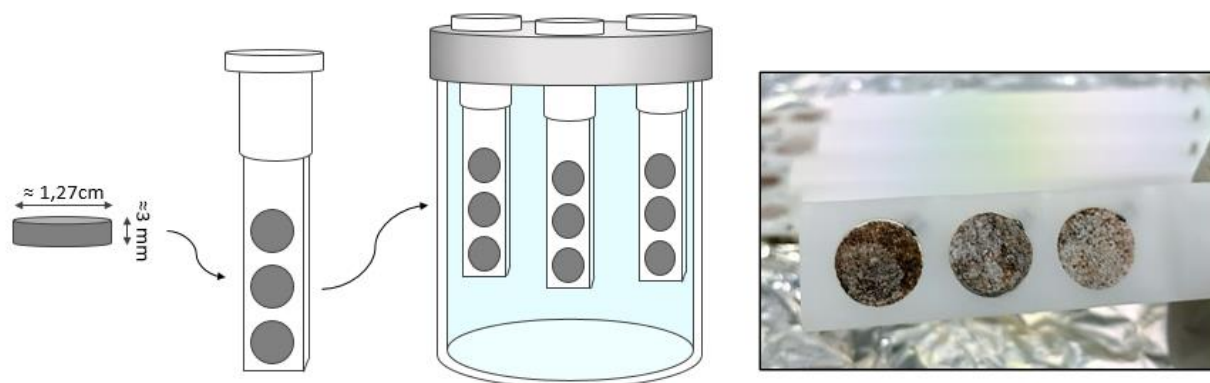


Figure B. Diagram of the assembly of the bioreactors. The arrows indicate the direction of movement of the water (in blue and gray) and the gas mixture (in black). It should be noted that the gas canisters are not illustrated and only the gas molecules are indicated.

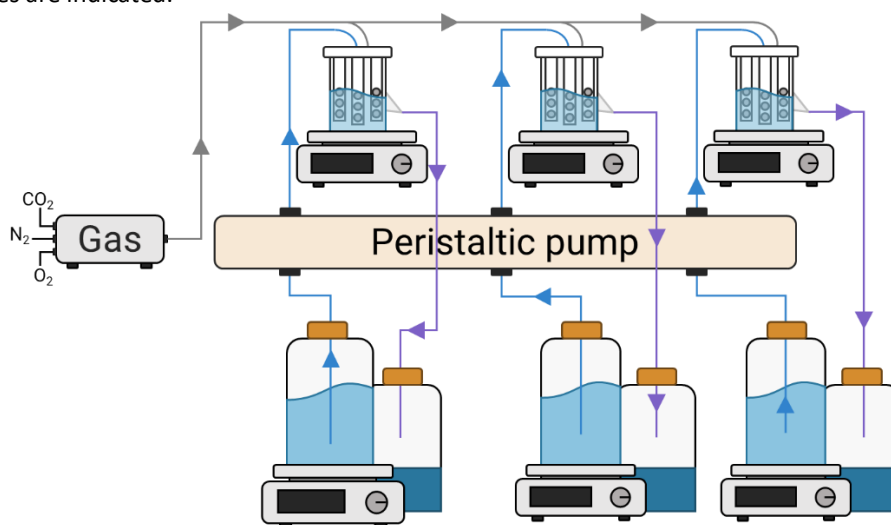


Figure C. Rarefaction curve of the Bacteria domain displaying the species richness against the number of sequences before rarefaction.

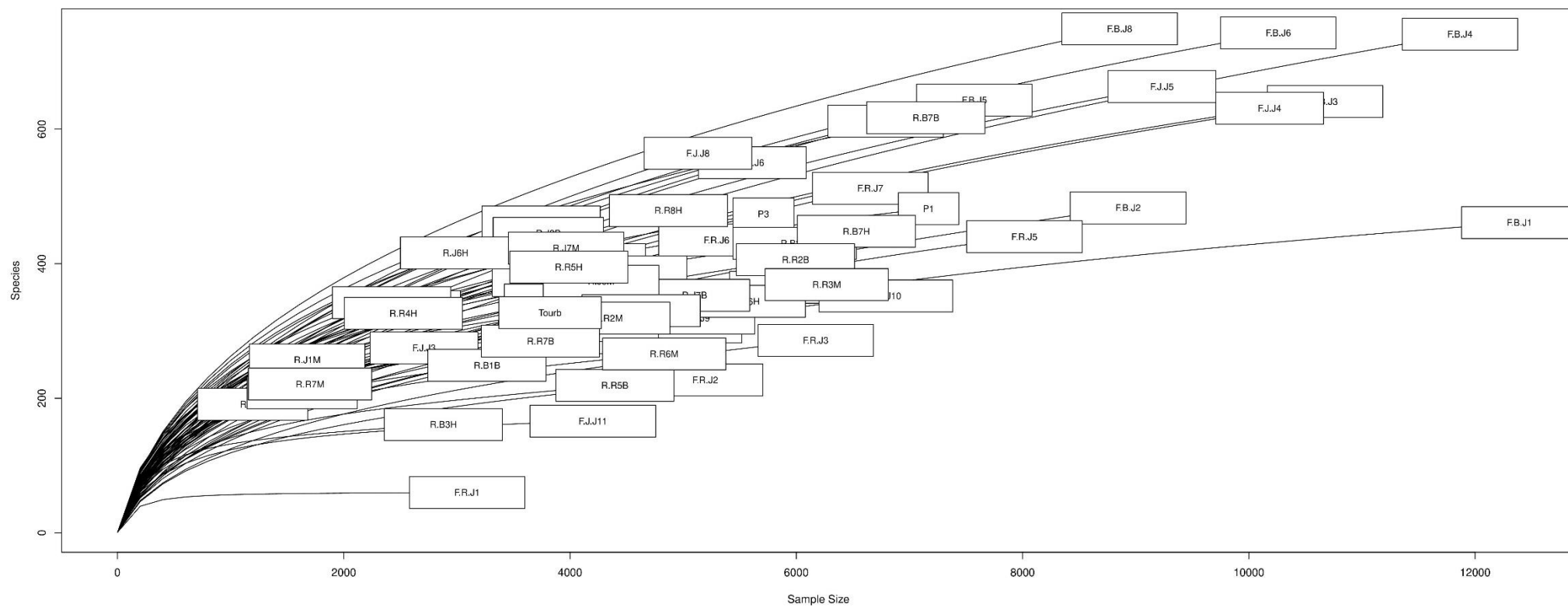


Figure D. Rarefaction curve of the Eukaryote domain displaying the species richness against the number of sequences before rarefaction.

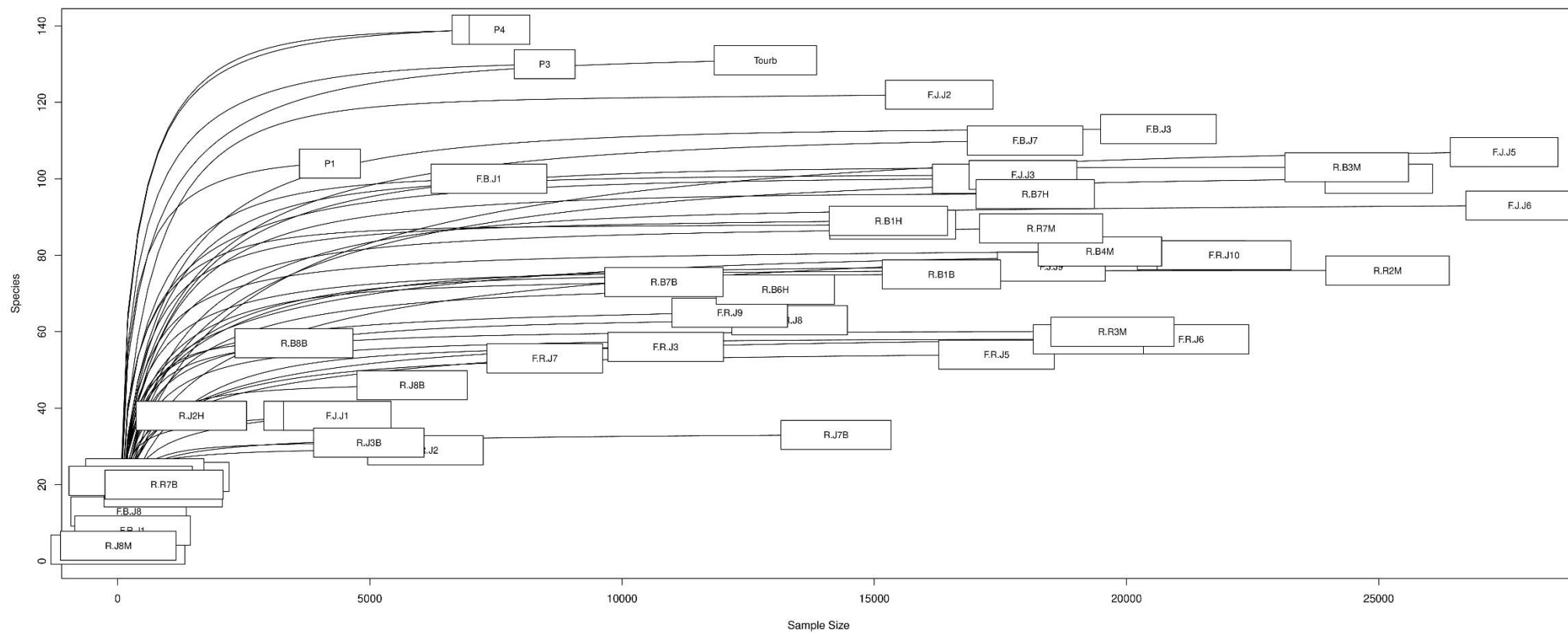


Figure E. Rarefaction curve of the Archaea domain displaying the species richness against the number of sequences before rarefaction.

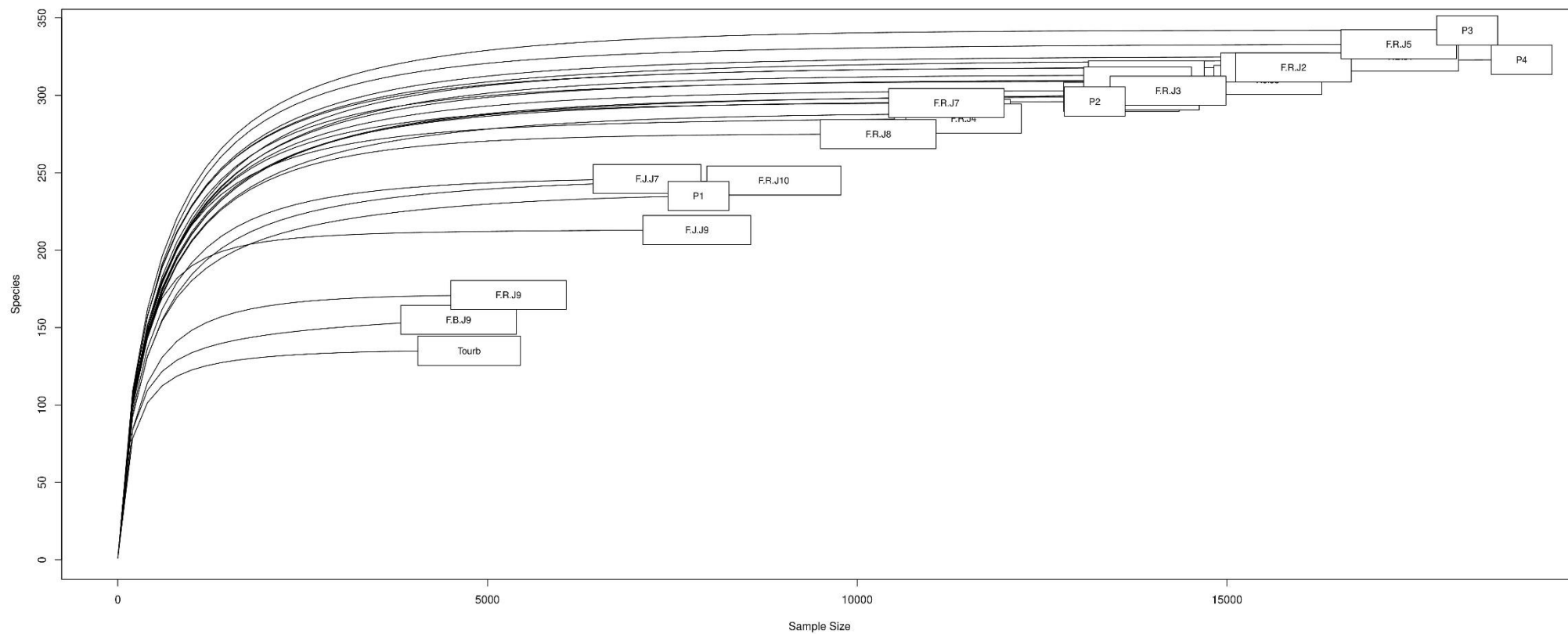


Figure F. Temporal changes of the average concentration (in mg/L) of dissolved inorganic carbon (DIC) during the 24 days of the experiment. The bioreactors are represented in different colors. The day 0 represent the data collected on the field.

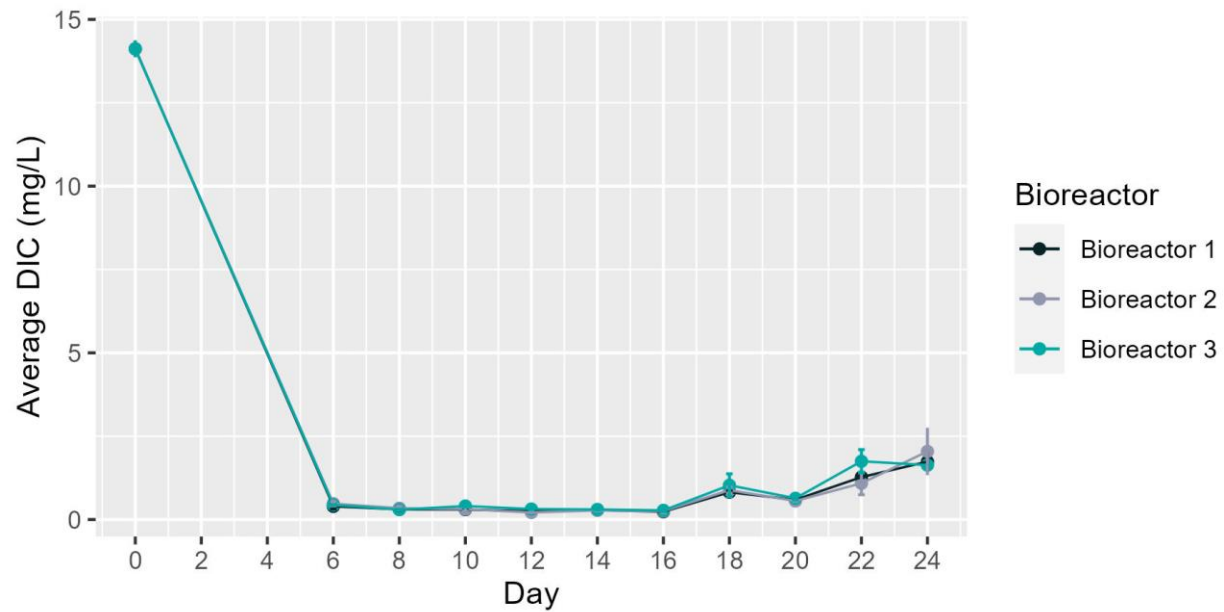


Figure G. Temporal changes of the average concentration (in mg/L) of dissolved organic carbon (DOC) during the 24 days of the experiment. The bioreactors are represented in different colors. The day 0 represent the data collected on the field.

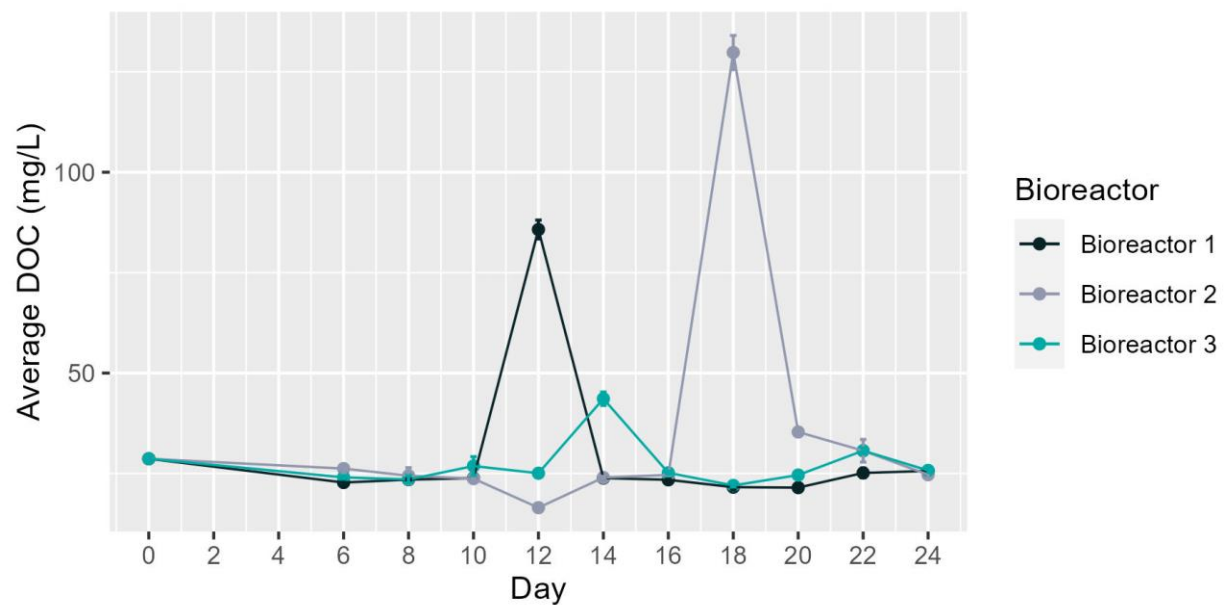


Figure H. Temporal changes of the average concentration (in mg/L) of NO_2^- during the 24 days of the experiment. The bioreactors are represented in different colors. The day 0 represent the data collected on the field.

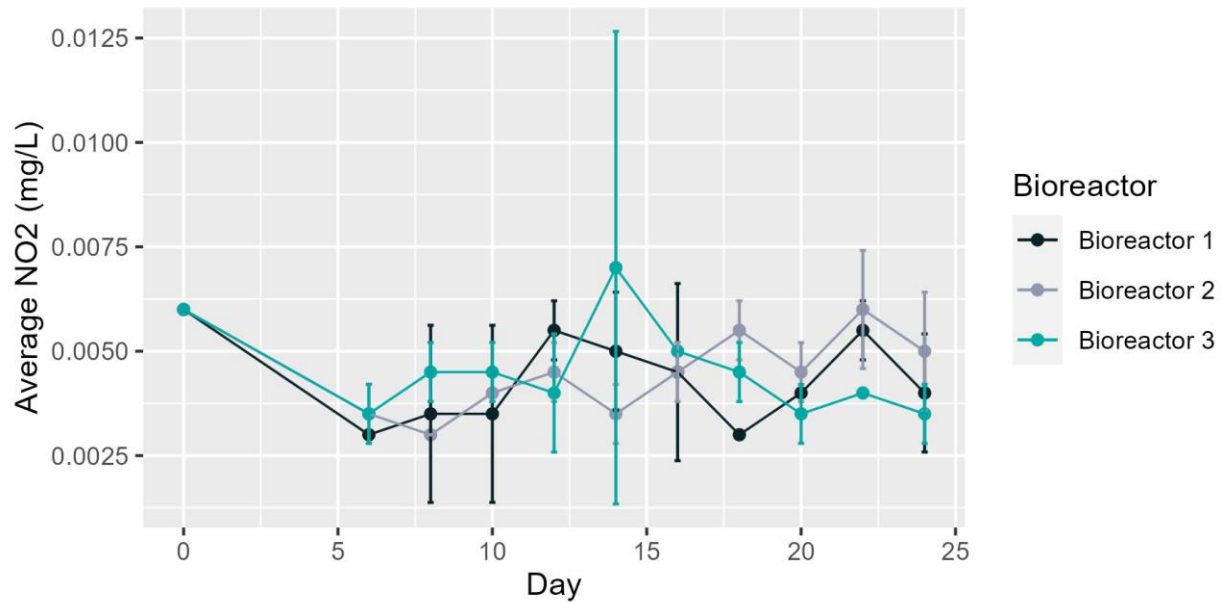


Figure I. Temporal changes of the average concentration (in mg/L) of NO_3^- during the 24 days of the experiment. The bioreactors are represented in different colors. The day 0 represent the data collected on the field.

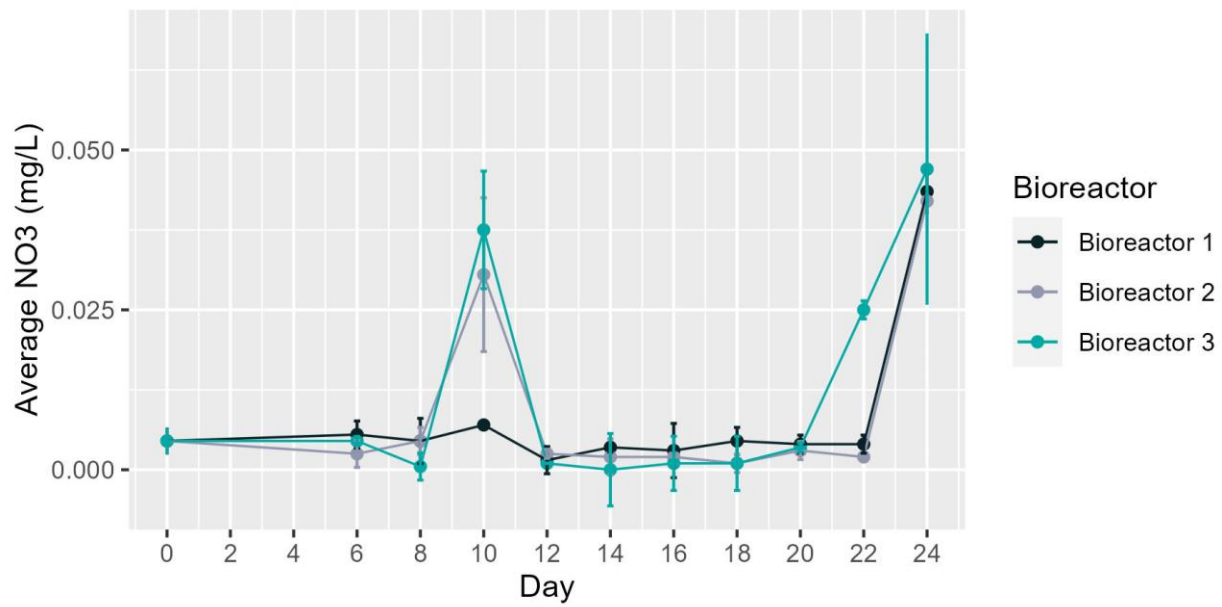


Figure J. Temporal changes of the average concentration (in mg/L) of NH_4 during the 24 days of the experiment. The bioreactors are represented in different colors. The day 0 represent the data collected on the field.

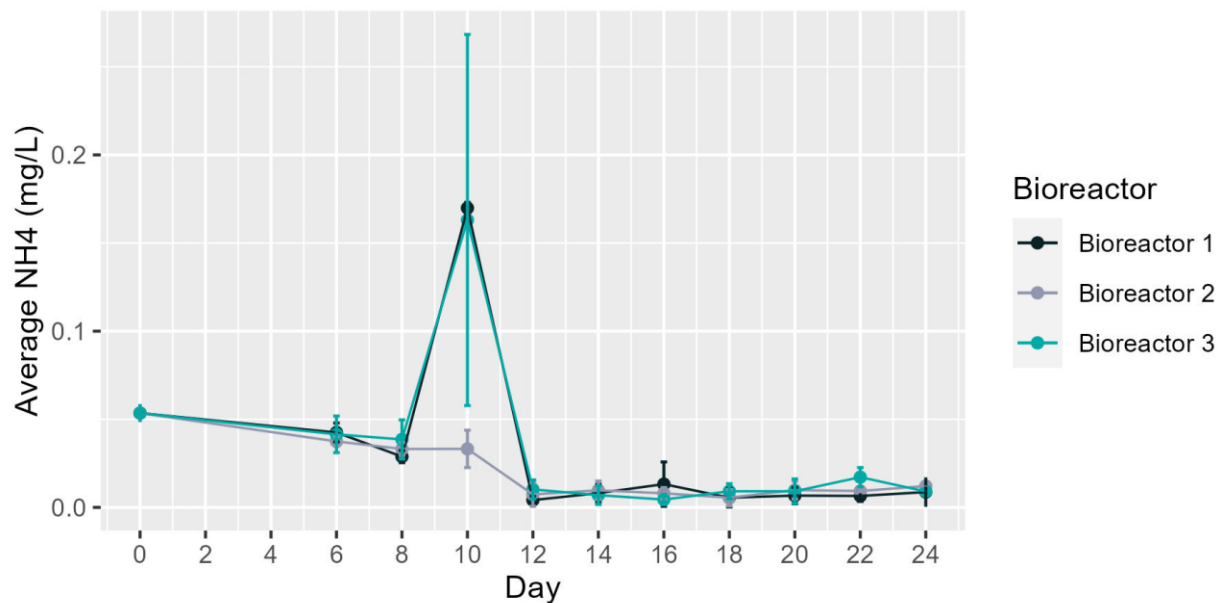


Figure K. Temporal changes of the average concentration (in mg/L) of dissolved O_2 during the 24 days of the experiment. The bioreactors are represented in different colors. The day 0 represent the data collected on the field.

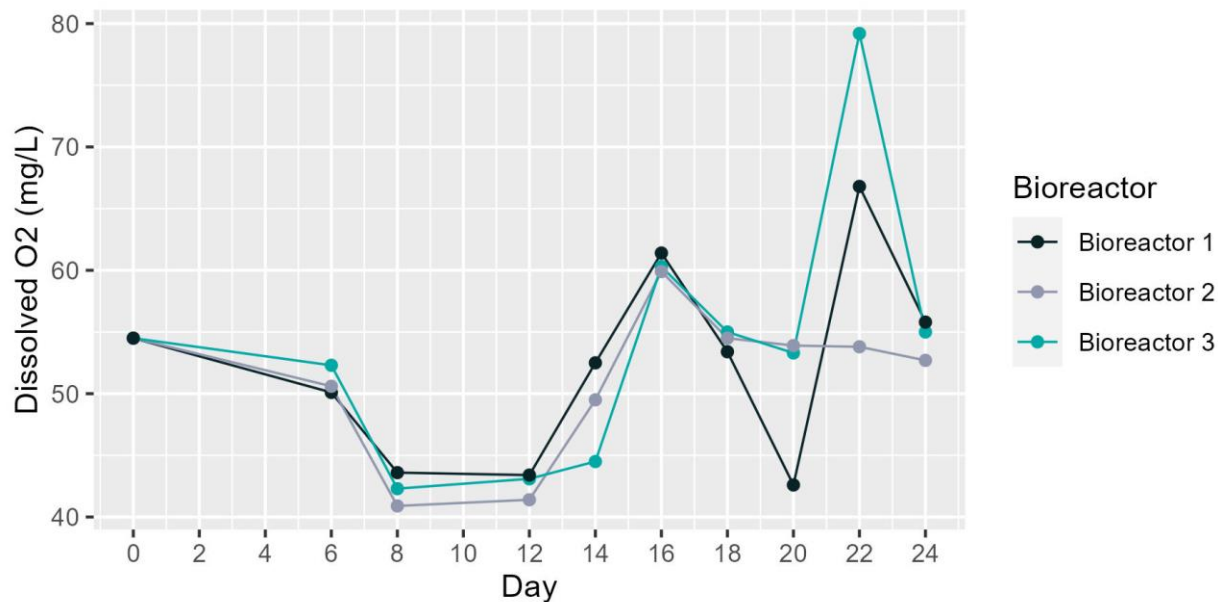


Figure L. Temporal changes of the pH during the 24 days of the experiment. The bioreactors are represented in different colors. The day 0 represent the data collected on the field.

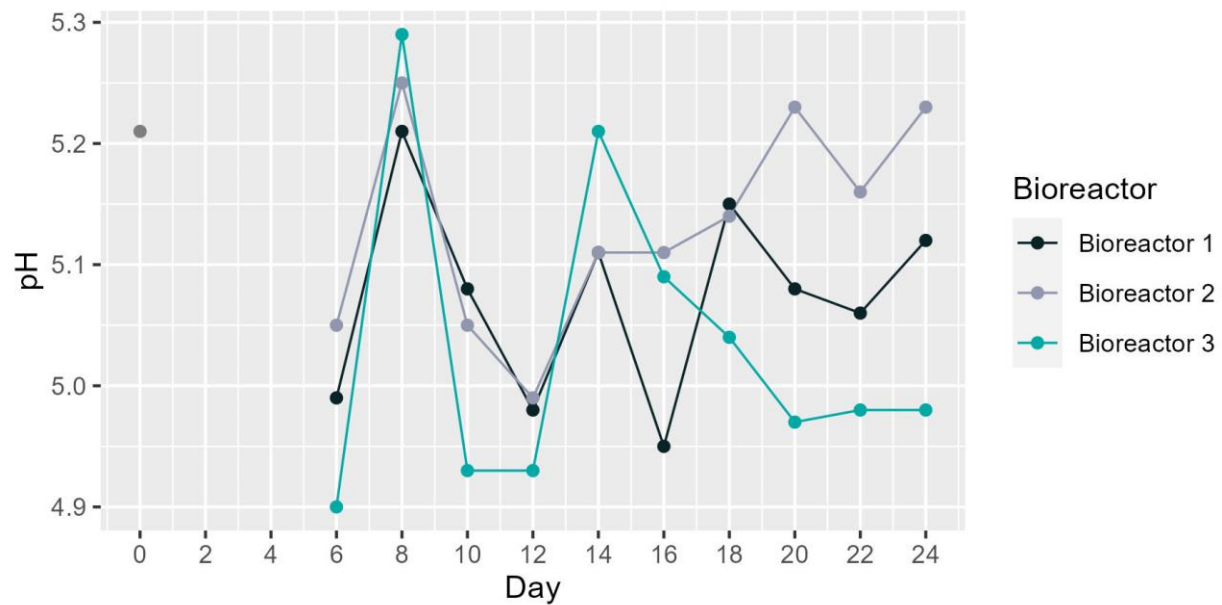


Figure M. Variation of the rock's roughness in each bioreactor. The values have been separated based on the bioreactor.

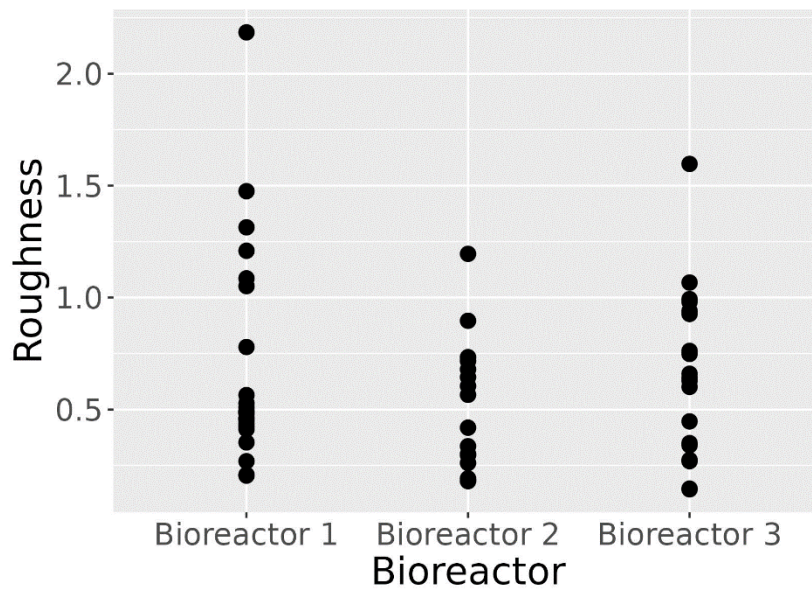


Figure N. Linear regressions of the Roughness based on the bacterial abundance represented here by the number of copies in different samples.

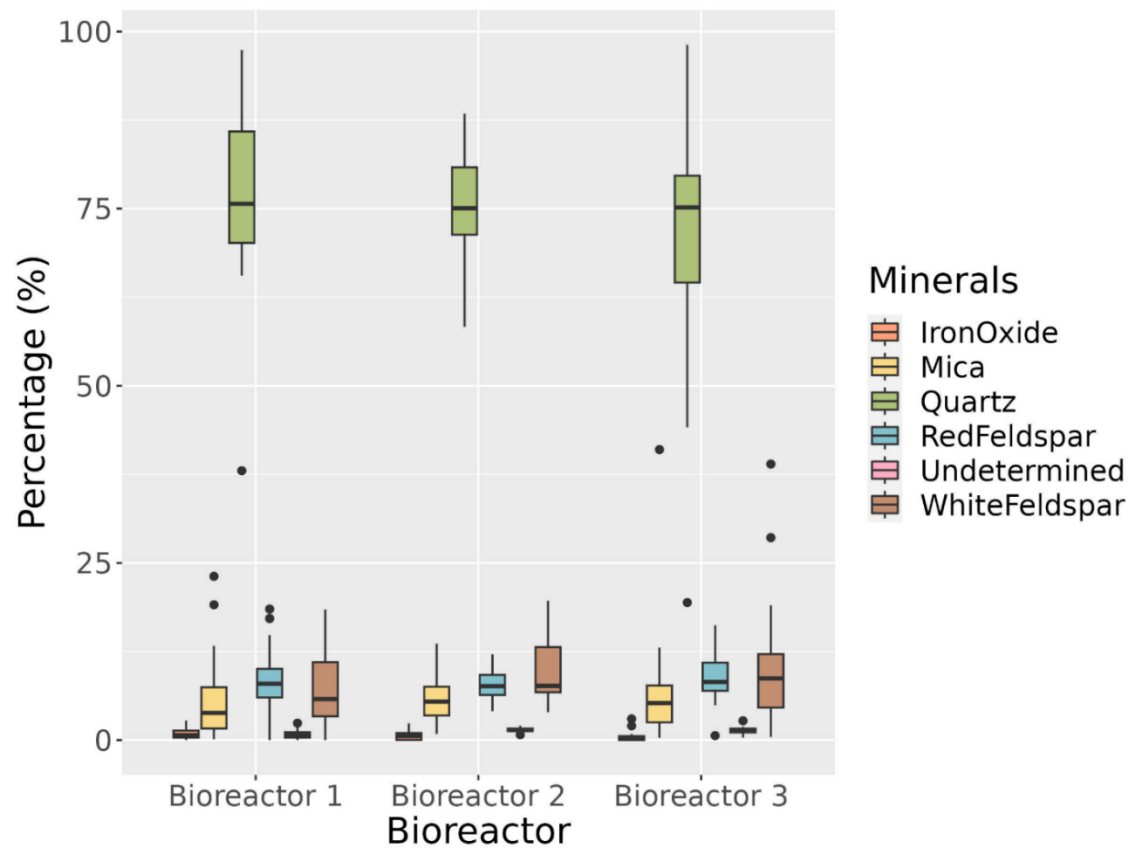


Figure O. Temporal changes of relative abundance (%) of bacterial genera in our planktonic community. Genera with an average relative abundance of less than 0.5% were categorized as “Other”. Going from left to right, the histogram is divided into 4 groups: the first bioreactor, the second one, the third one and the field data.

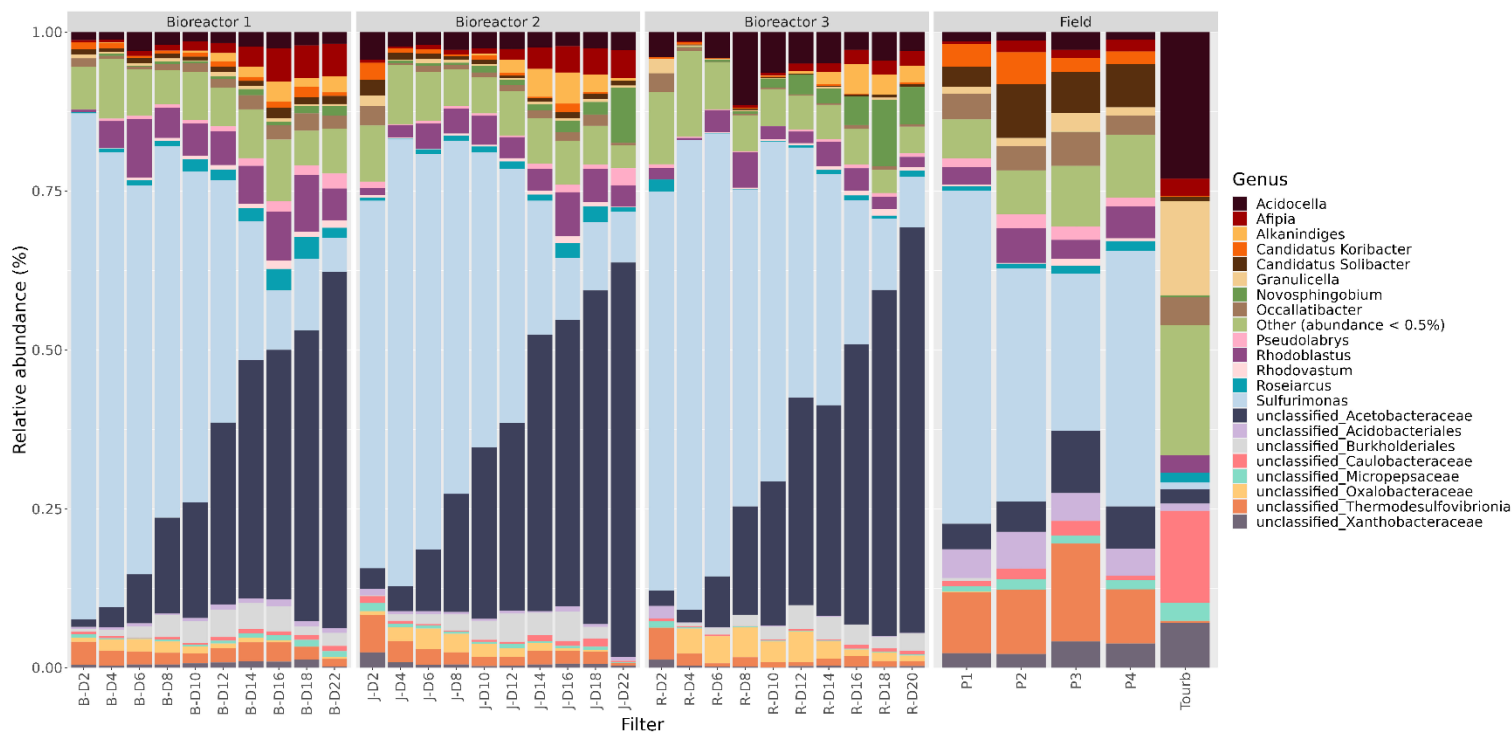


Figure P. Relative abundance (%) of bacterial genera in our sessile community. Genera with an average relative abundance of less than 0.5% were categorized as “Other”. Going from left to right, the histogram is divided into 3 groups: the first bioreactor, the second one and the third one. There are no field data here.

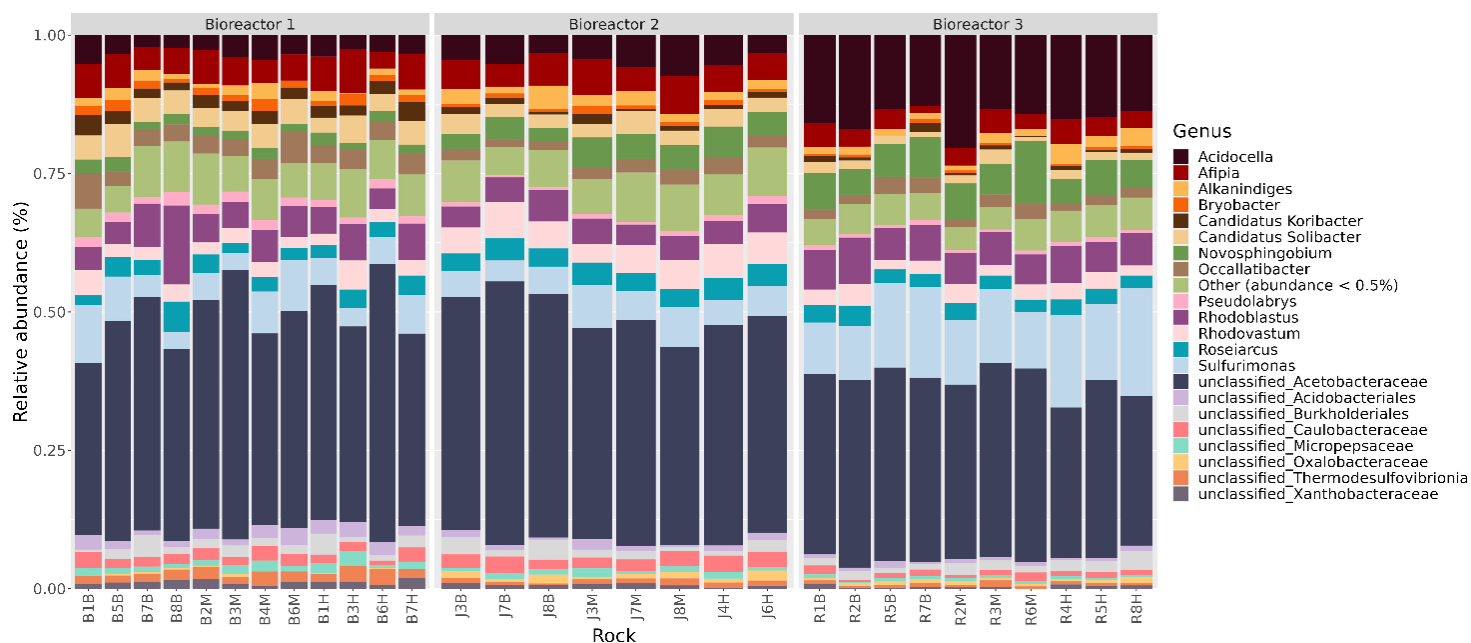


Figure Q. Temporal changes in the relative abundance (%) of the eukaryotic genera in our planktonic community. Genera with an average relative abundance of less than 0.1% were categorized as “Other”. Going from left to right, the histogram is divided into 4 groups: the first bioreactor, the second one, the third one and the field data.

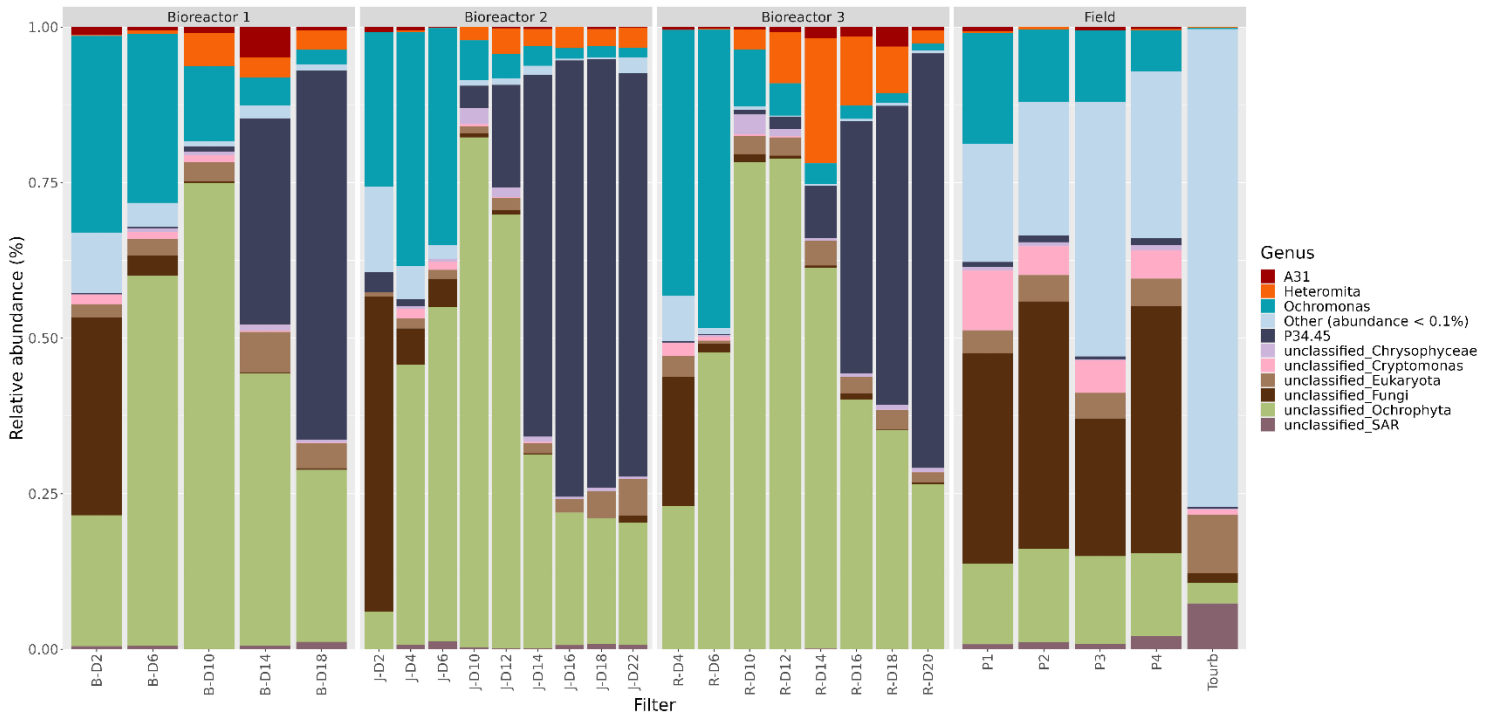


Figure R. Relative abundance (%) of eukaryotic genera in our sessile community. Genera with an average relative abundance of less than 0.1% were categorized as “Other”. Going from left to right, the histogram is divided into 3 groups: the first bioreactor, the second one and the third one. There are no field data here.

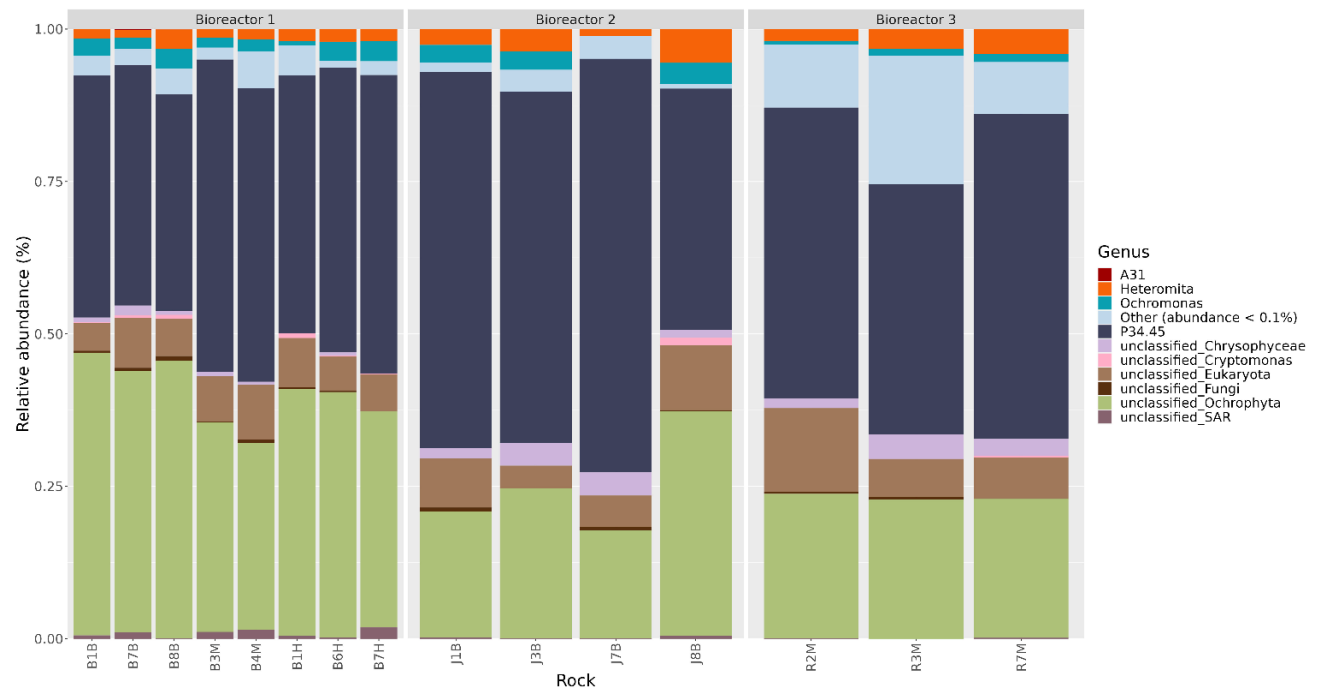


Figure S. Relative abundance (%) of archaeal genera in our sessile community. Genera with an average relative abundance of less than 0.1% were categorized as “Other”. Going from left to right, the histogram is divided into 4 groups: the first bioreactor, the second one, the third one and the field data.

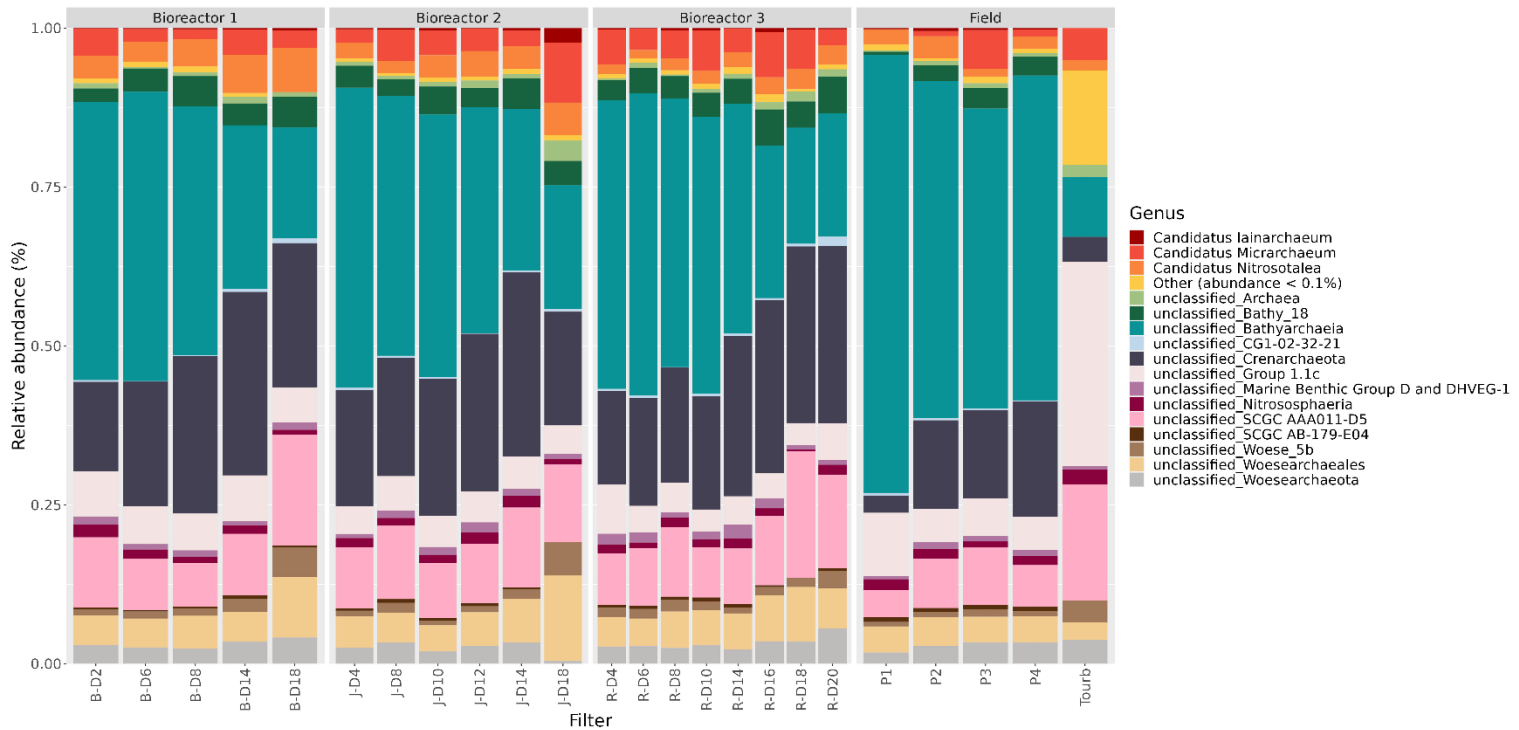


Figure T. Percentage of genera in the sessile and planktonic community of the Bacteria domain shown in a Venn diagram.

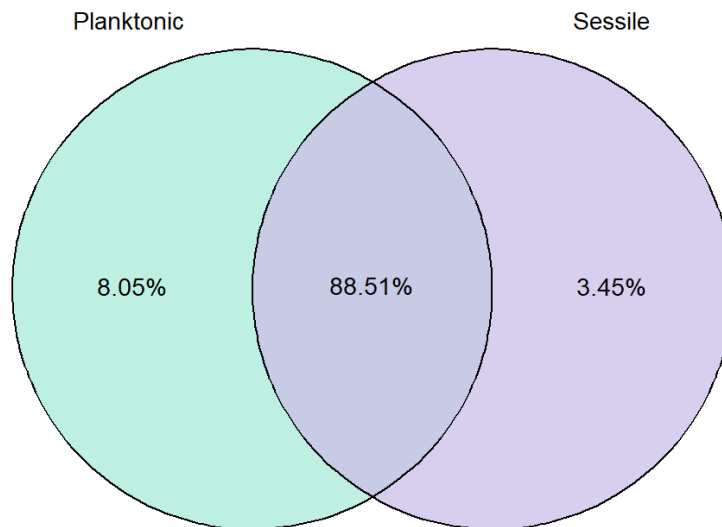


Figure U. Percentage of genera in the sessile and planktonic community of the Eukaryote domain shown in a Venn diagram.

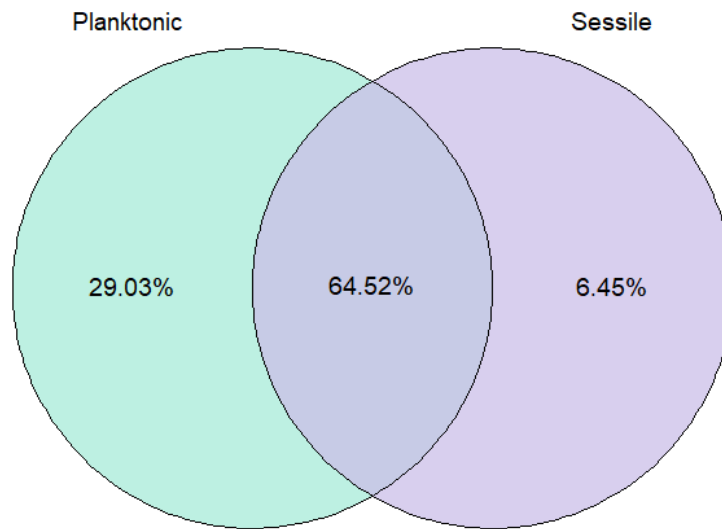
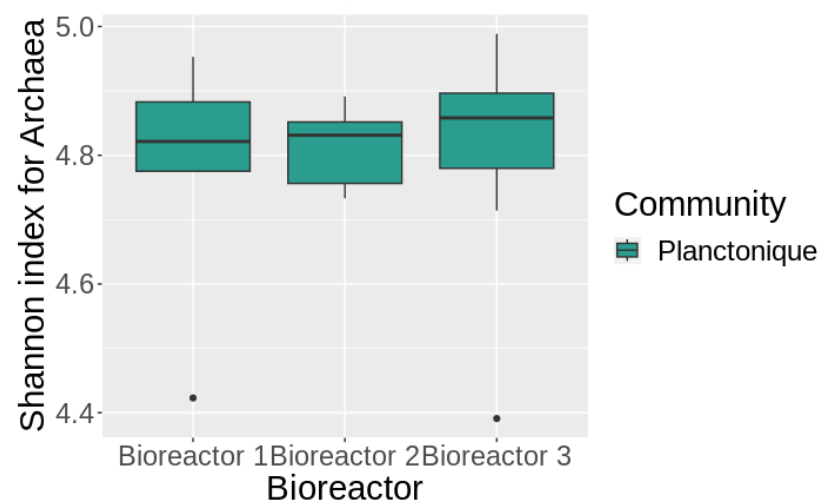


Figure V. Alpha-diversity (Shannon index) for the planktonic and sessile community of the Achaea kingdom. The three bioreactors are divided by box.



PCoA 2 (24.44% of variation)

PCoA 1 (28.74% of variation)

Bioreactor

- Bioreactor 1
- ▲ Bioreactor 2
- Bioreactor 3
- ◆ Field

Sample Type

- Field
- Filter

Bioreactor

- Bio1
- Bio2
- Bio3

Annexe B: Supplementary tables

Table 1. Component present in the 2 x AE buffer.

Components	Concentration (in M)	Volume (in ml)
EDTA	0.5	20
EGTA	1.5	50
NaCl	1	40
Tris (pH 8.0)	1	40
mili-Q Water	---	50
Total		200

Table 2. Results of multiple Kruskal-Wallis test between geochemical properties as the rock properties between bioreactors.

Parameter	chi-squared	df	p-value
DIC	0.54799	2	0.7603
DOC	3.4245	2	0.1805
NO ₂ ⁻	0.43457	2	0.8047
NO ₃ ⁻	2.1625	2	0.3392
NH ₄	0.7513	2	0.6868
% dissolved O ₂	0.41282	2	0.8135
pH	5.7204	2	0.05726
Iron Oxide	3.827	2	0.1476
Quartz	1.1695	2	0.5572
Red Feldspar	1.1945	2	0.5503
White Feldspar	3.5174	2	0.1723
Mica	0.77422	2	0.679
Roughness	2.1901	2	0.3345

Table 3. Results of the linear regression done on the bacterial abundance and the roughness of the rocks.

Parameters	Estimate Std	Error	T-value	Pr(> t)
Intercept	971388	159807	6.079	1.48×10^{-6}
Roughness 1	192715	161326	1.195	0.242
Intercept	958308	186902	5.127	1.96×10^{-5}
Roughness 2	188995	231073	0.818	0.42

Table 4. Results of the Kruskal-Wallis test comparing the rock properties between bioreactors.

Parameter	chi-squared	df	p-value
Iron Oxide	3.827	2	0.1476
Quartz	1.1695	2	0.5572
Red Feldspar	1.1945	2	0.5503
White Feldspar	3.5174	2	0.1723
Mica	0.77422	2	0.679
Roughness	2.1901	2	0.3345

Table 5. Result of multiple Pearson correlation test between the mineralogy and the bacterial abundance.

Mineral	Correlation	T	df	p-value
Iron Oxide	0.1111423	0.85171	58	0.3979
Quartz	-0.08373164	-0.63993	58	0.5247
Red Feldspar	-0.01513301	-0.11526	58	0.9086
White Feldspar	-0.01751817	-0.13343	58	0.8943
Mica	0.2213282	1.7285	58	0.08923

Table 6. Results of the generalized linear model with the α -diversity. The p-values < 0.05 are in bold.

Domain	Coefficients	Estimate	Std. Error	T value	Pr(> t)
Bacteria	Intercept	3.420101	0.205041	16.680	3.35×10^{-13}
	DIC	-0.218193	0.288054	-0.757	0.458
	DOC	0.003226	0.003560	0.906	0.376
	NH ₄	-4.117057	1.992979	-2.066	0.052
Eukaryote	Intercept	1.680021	0.166729	10.076	8.5×10^{-8}
	DIC	1.188771	0.273971	4.339	0.00068
	DOC	-0.002053	0.003018	-0.680	0.50734
	NH ₄	-2.905962	1.442604	-2.014	0.06360
Archaea	Intercept	4.8960651	0.0326732	149.850	$< 2 \times 10^{-16}$
	DIC	-0.5902471	0.0616987	-9.567	5.77×10^{-7}
	DOC	0.0039372	0.0005721	6.882	1.69×10^{-5}
	NH ₄	0.9316038	0.3635072	2.563	0.0249

Table 7. Result of multiple Pearson correlation test between the α -diversity and the rock properties.

Domain	Parameter	Correlation	T	df	p-value
Bacteria	Roughness1	-0.03575195	-0.1893	28	0.8512
	Roughness2	0.1337367	0.71408	28	0.4811
	Iron Oxide	-0.0244997	-0.18664	58	0.8526
	Quartz	0.2051354	1.5962	58	0.1159
	Red Feldspar	0.09687343	0.74125	58	0.4615
	White Feldspar	-0.1811629	-1.4029	58	0.166
	Mica	-0.2731739	-2.1627	58	0.0347
Eukaryote	Roughness1	0.1087482	0.39444	13	0.6997
	Roughness2	0.2548637	0.95031	13	0.3593

	Iron Oxide	0.1835409	0.98799	28	0.3316
	Quartz	0.04268992	0.2261	28	0.8228
	Red Feldspar	0.09011004	0.47877	28	0.6358
	White Feldspar	-0.05250312	-0.2782	28	0.7829
	Mica	-0.1390769	-0.74315	28	0.4636

Table 8. Results of the PERMANOVA analysis from the three domains. The p-values < 0.05 are in bold.

Domain	Clusters	Parameters	df	SumOfSqs	R ²	F	P-value
Bacteria	D2-D8	Bioreactor	2	0.09800	0.18141	1.4933	0.116
		Day	1	0.17972	0.33267	5.4769	0.001
		Residual	8	0.26251	0.48592		
		Total	11	0.54024	1.00000		
	D10-D12	Bioreactor	2	0.109700	0.52522	3.8434	0.01667
		Day	1	0.070620	0.33812	4.9484	0.03333
		Residual	2	0.028543	0.13666		
		Total	5	0.208863	1.00000		
	D14-D24	Bioreactor	2	0.28883	0.31463	3.1194	0.001
		Day	1	0.25535	0.27816	5.5156	0.001
		Residual	8	0.37037	0.40345		
		Total	11	0.91799	1.00000		
	Rock	Bioreactor	2	1.0687	0.45537	11.287	0.001
		Residual	27	1.2782	0.54463		
		Total	29	2.3469	1.00000		
	All	SampleType	1	3.8614	0.32328	30.096	0.001
		Residuals	63	8.0830	0.67672		
		Total	64	11.9444	1.00000		
Eukaryote	D2-D6	Bioreactor	2	0.15636	0.25108	1.8664	0.220
		Day	1	0.35070	0.56314	8.3724	0.001
		Residual	3	0.12566	0.20178		
		Total	6	0.62276	1.00000		
	D10-D12	Bioreactor	2	0.026509	0.47017	1.1583	0.4667

		Day	1	0.014892	0.26413	1.3014	0.3000
		Residual	1	0.011443	0.20296		
		Total	4	0.056382	1.00000		
	D14-D24	Bioreactor	2	0.19739	0.29673	2.8440	0.065
		Day	1	0.23594	0.35469	6.7988	0.007
		Residual	6	0.20822	0.31301		
		Total	9	0.66521	1.00000		
	Rock	Bioreactor	2	0.21053	0.46479	5.2105	0.001
		Residual	12	0.24243	0.53521		
		Total	14	0.45296	1.00000		
	All	SampleType	1	1.5632	0.18845	9.2882	0.001
		Residuals	40	6.7319	0.81155		
		Total	41	8.2950	1.00000		
Archaea	D2-D12	Bioreactor	2	0.069403	0.32325	1.3440	0.023
		Day	1	0.031776	0.14800	1.2307	0.145
		Residuals	4	0.103275	0.48101		
		Total	7	0.214704	1.00000		
	D14-D20	Bioreactor	2	0.23328	0.36083	1.4431	0.094
		Day	1	0.11423	0.17668	1.4132	0.115
		Residuals	4	0.32331	0.50009		
		Total	7	0.64651	1.00000		

Table 9. Results of the dbRDA analysis based on geochemical parameters of the water in the bioreactors and based on the mineralogy of the rock surfaces for the Bacteria and Eukaryote domain.

Domain	dbRDA	Parameters	Df	SumofSqs	F	P-value
Bacteria	Geochemical parameters	DIC	1	0.41656	3.8056	0.003
		NH ₄	1	0.20453	1.8685	0.056
		Residual	21	2.29864		
	Mineralogy and roughness	Quartz	1	0.1277	1.3816	0.100
		Red Feldspar	1	0.0772	0.8347	0.684
		Mica	1	0.1175	1.2710	0.155
		Roughness	1	0.1296	1.4024	0.110
		Residual	55	5.0840		
Eukaryote	Geochemical parameters	DIC	1	0.47291	5.0829	0.006
		NH ₄	1	0.22868	2.4579	0.069
		Residual	15	1.39559		
	Mineralogy and roughness	Iron Oxide	1	0.04724	0.6815	0.744
		Red Feldspar	1	0.02318	0.3345	0.986
		White Feldspar	1	0.10471	1.5107	0.142
		Mica	1	0.06520	0.9406	0.489
		Roughness	1	0.08419	1.2146	0.257
		Residual	24	1.66350		
Archaea	Geochemical parameters	DIC	1	0.21665	4.3134	0.001
		DOC	1	0.22572	4.4940	0.002
		Residual	13	0.65296		

Table 10. Results of the dbRDA analysis based on the mineralogy of the rock surfaces for the three domains. The p-values < 0.05 are in bold.

Domain	Parameter	Df	SumOfSqs	F	Pr(>F)
Bacteria	Quartz	1	0.1277	1.3816	0.100
	Red Feldspar	1	0.0772	0.8347	0.684
	Mica	1	0.1175	1.2710	0.155
	Roughness	1	0.1296	1.4024	0.110
	Residual	55	5.0840		
Eukaryote	Iron Oxide	1	0.04724	0.6815	0.744
	Red Feldspar	1	0.02318	0.3345	0.986
	White Feldspar	1	0.10471	1.5107	0.142
	Mica	1	0.06520	0.9406	0.489
	Roughness	1	0.08419	1.2146	0.257
	Residual	24	1.66350		
Archaea	DIC	1	0.21665	4.3134	0.001
	DOC	1	0.22572	4.4940	0.002
	Residual	13	0.65296		

Table 11. Bacterial abundance and roughness of each rock

Sample	Abundance	Roughness 1	Roughness 2
B1B	1332608	0.50956	0.52814
B5B	1649984	2.18455	1.20897
B7B	818032	1.47514	1.05115
B8B	729321	0.41211	0.49433
B1H	928812	0.46088	0.26897
B3H	67110.35	0.35313	0.46088
B6H	1428149	0.49037	1.0858
B7H	453429.5	0.42418	0.45076
B2M	1132813	0.20488	0.56445
B3M	476948.5	0.20872	0.43604
B4M	809970	1.31376	0.48557
B6M	2392239	0.77908	0.48183
J3B	365066.5	0.41886	0.30142
J7B	43592.8	0.33612	0.71661
J8B	545134.8	0.19202	0.26126
J4H	194.3	0.73391	0.56564
J6H	453509.3	0.18398	0.17993
J3M	141589.6	0.6446	1.19519
J7M	652659.5	0.89637	0.29622
J8M	553465	0.60504	0.67945
R1B	1376036	1.06712	0.34994
R2B	1506898	0.27469	0.92691
R5B	1572235	0.14353	0.2723
R7B	2386062	0.98089	0.65969

R4H	748548	0.74837	0.60108
R5H	446534.8	1.59695	0.94036
R8H	1046161	0.34037	0.99329
R2M	498009.8	0.44693	0.62852
R3M	463507	0.76171	0.64184
R6M	142764.1	0.14662	0.26944

Table 12. ASVs removed during the decontamination step 1.16 for each domain.

Domain	Number of ASVs removed	Number of ASVs kept
Bacteria	2	5653
Eukaryote	1	576
Archaea	1	920

Table 13. ASVs found in the control samples.

Domain	Sample name	Number of unique ASVs
Bacteria	BlancKit	59
	CRTL-B4B	0
	CRTL-B5H	0
	CRTL-J1B	1
	CRTL-J2M	2
	CRTL-R3H	2
	CRTL-R6M	0
	CTRL-PCR	0
Eukaryote	BlancKit	1
	CRTL-B4B	3
	CRTL-B5H	0
	CRTL-J1B	3
	CRTL-J2M	2
	CRTL-R3H	2
	CRTL-R6M	4
Archaea	BlancKit	2
	ctrl-PCR	3

Annexe C: PCR and dPCR primers

Table 14. PCR primer pairs used for 16S/18S rRNA gene amplification and sequencing.

Domain	Primer	Sequence (5'-3')
Bacteria	B341F	CCTACGGGAGGCAGCAG
	B785R	GACTACHVGGGTATCTAATCC
Eukaryote	A340F	CCCTACGGGGYGCASCAG
	A915R	GTGCTCCCCGCCAATTCCT
Archaea	E960F	GGCTTAATTTGACTCAACRCG
	E1438R	GGGCATCACAGACCTGTTAT

Table 15. Primer pairs and probe used for the dPCR assays.

Primer	Type	Sequence (5'-3')
B341F	Forward primer	CCTACGGGAGGCAGCAG
B515	Probe	GTGYCAGCMGCCGCGGTAA
B785R	Reverse primer	GACTACHVGGGTATCTAATCC

Annexe D: PCR and dPCR conditions

Table 16. PCR conditions used for 16S/18S rRNA gene amplification and sequencing.

Domain		Denaturation	Annealing	Extension	Final extension
Bacteria	Nb of cycles	35			1
	Time	30 sec	1 min	1 min	10 min
	Temperature	98°C	57°C	72°C	72°C
Eukaryote	Nb of cycles	35			1
	Time	30 sec	30 sec	1 min	10 min
	Temperature	98°C	61°C	72°C	72°C
Archaea	Nb of cycles	35			1
	Time	30 sec	30 sec	1 min	10 min
	Temperature	98°C	67°C	72°C	72°C

Table 17. Conditions used for the dPCR assays.

	Activation	Hybridization/Elongation	Denaturation	Final elongation
Nb of cycles	1	39		1
Time	10 min	3 min	30 sec	2 min
Temperature	96°C	56°C	98°C	56°C

Annexe E: Scripts

1. SEQUENCE PROCESSING (DADA PIPELINE)

Here, I used a modified version of the DADA pipeline which can be found in the following link: <https://benjjneb.github.io/dada2/tutorial.html>. We start off with the Illumina-sequenced paired-end fastq files.

1.1 Bacterial sequences processing

1.1.1 Opening DADA2 and setting our working directory for our Bacteria files.

```
library(dada2); packageVersion("dada2")  
#Setting a path to the FASTQ files  
#Here I'm focusing on my Bacterias  
path <- "~/sequences/Bac/bac_fastq"  
#Just going to make sure the path is leading to the. fastq files  
list.files(path)
```

1.2 Separate the Forward and Reverse fastq files.

```
# Identify forward and reverse reads  
# Filenames containing R1 are forward reads, R2 are reverse reads.  
fnFs <- sort(list.files(path, pattern= "R1_001.fastq", full.names = TRUE))  
fnRs <- sort(list.files(path, pattern= "R2_001.fastq", full.names = TRUE))
```

1.3 Assign the filenames for the filtered fastq.gz files and place in filtered subdirectory.

```
sample.names <- sapply(strsplit(basename(fnFs), "_"), `[`, 1)  
# We'll still have the "-bac" at the end of the file names.  
#To remove that, we'll use gsub()  
sample.names <- gsub("-bac", "", sample.names)  
#Let's put the files in a subdirectory  
filtFs <- file.path("~/sequences/Bac/filtered", paste0(sample.names, "_F_filt.fastq.gz"))  
filtRs <- file.path("~/sequences/Bac/filtered", paste0(sample.names, "_R_filt.fastq.gz"))  
  
names(filtFs) <- sample.names  
names(filtRs) <- sample.names
```

1.4 Let's check the quality of our sequences.

#First, we can look at some of the Forwards sequences for Filter samples

```
plotQualityProfile(fnFs[9:18])
```

#Then, some of the Forwards sequences for Rock samples

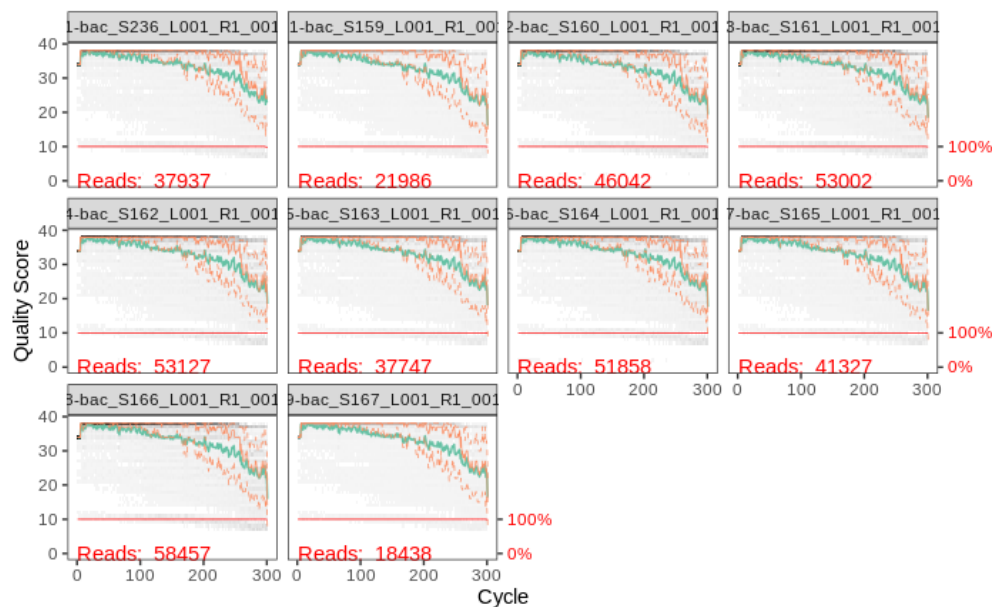
```
plotQualityProfile(fnFs[40:50])
```

#Now we can look at some of the Reverse sequences for Filter samples

```
plotQualityProfile(fnRs[9:18])
```

#And finally, some of the Reverse sequences for Rock samples

```
plotQualityProfile(fnRs[40:50])
```



1.5 Trim the reads thanks to the quality graphs (see previous section of code)

From the previous section, it looks like the quality diminishes at the 250 mark. For the forward sequences, we'll remove the first 17 nucleotides (size of the primer). For the reverse, we'll remove the first 21 nucleotides (size of the primer). We'll trim the last nucleotides at position 250 (trimming the last 50 nucleotides) to avoid errors.

For the Eucaryotes, it looks like the quality diminishes at the 270 mark. For the forward sequences, we'll remove the first 17 nucleotides (size of the primer). For the reverse, we'll remove

the first 21 nucleotides (size of the primer). We'll trim the last nucleotides at position 270 (trimming the last 70 nucleotides) to avoid errors.

For the Archaea, it looks like the quality diminishes at the 230 mark. For the forward sequences, we'll remove the first 17 nucleotides (size of the primer). For the reverse, we'll remove the first 21 nucleotides (size of the primer). We'll trim the last nucleotides at position 230 (trimming the last 70 nucleotides) to avoid errors.

```
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, truncLen=c(250,200),
                    maxN=0, maxEE=c(5,5), truncQ=2, trimLeft=c(17,21), compress=TRUE, multithread=TRUE)
out
```

1.6 Looking at the error rate.

```
errF <- learnErrors(filtFs, multithread=TRUE, randomize=TRUE)
#101502023 total bases in 435631 reads from 18 samples will be used for learning the error rates.
errR <- learnErrors(filtRs, multithread=TRUE, randomize=TRUE)
#101447176 total bases in 566744 reads from 24 samples will be used for learning the error rates.
```

1.7 Time to dereplicate: removing replicate of identical sequences from each sample.

```
derepFs <- derepFastq(filtFs, verbose= TRUE)
derepRs <- derepFastq(filtRs, verbose= TRUE)
# Name the derep-class objects by the sample names
names(derepFs) <- sample.names
names(derepRs) <- sample.names
# Apply the core sample inference algorithm to the filtered and trimmed sequence data
dadaFs <- dada(filtFs, err=errF, pool= "pseudo", multithread= TRUE)
dadaRs <- dada(filtRs, err=errR, pool="pseudo", multithread= TRUE)
```

1.8 Create the ASVs

```
# First, we merge the paired reads
mergers <- mergePairs(dadaFs, filtFs, dadaRs, filtRs)
# Construct an amplicon sequence variant table (ASV) table and check the dimension of it
```

```

seqtab <- makeSequenceTable(mergers)
# Inspect distribution of sequence lengths
table(nchar(getSequences(seqtab)))
#The majority of the sequences are 400 nucleotides
# Remove chimeras
seqtab.nochim <- removeBimeraDenovo(seqtab, method="consensus", multithread= TRUE)
#View dimension of your matrices
dim(seqtab.nochim)
sum(seqtab.nochim)/sum(seqtab)

[1] 79 5712
[1] 0.8088081

```

1.9 Track reads through the pipeline.

```

# Obtain the count of how many sequences were deleted at each steps.
getN <- function(x) sum(getUniques(x))
track <- cbind(out, sapply(dadaFs, getN), sapply(dadaRs, getN), sapply(mergers, getN), rowSums(seqtab.nochim))
colnames(track) <- c("input", "filtered", "denoisedF", "denoisedR", "merged", "nonchim")
rownames(track) <- sample.names
track

```

1.10 Assign the taxonomy to our ASVs.

```

# Identifying the taxonomy
taxa <- assignTaxonomy(seqtab.nochim, "~/silva_nr99_v138.1_train_set.fa.gz?download=1", multithread=TRUE, tryRC= TRUE)
# Adding the species for each sequence
taxa.sp <- addSpecies(taxa, "~/silva_species_assignment_v138.1.fa.gz?download=1", allowMultiple = TRUE, tryRC = TRUE)
# Removing sequence rownames for display only
taxa.print <- taxa.sp
rownames(taxa.print) <- NULL
head(taxa.print, n=20)

```

1.11 Save as .CSV files and save the sequence table.

```
# For the taxonomy table
write.csv(as.data.frame(taxa.print), file = file.path("~/sequences/Bac/bac_csv/BAC_S1_taxonomy_table.csv"))

# For the ASV table
write.csv(as.data.frame(seqtab.nochim),
          file = file.path("~/sequences/Bac/bac_csv/ASV_bac1_table.csv"))
write.csv(as.data.frame(t(seqtab.nochim)),
          file = file.path("~/sequences/Bac/bac_csv/ASV_S1_t_table.csv"))

# Save sequence table and taxonomic annotations to individual files
saveRDS(seqtab.nochim, file = file.path("~/sequences/Bac/bac1_seqtab_nochim.rds"))
saveRDS(taxa.sp, file = file.path("~/sequences/Bac/bac1_taxa_sp.rds"))
```

1.12 Loading libraries and the DADA2 results for the cleaning and summarizing.

```
library(picante)
library(vegan)
library(ggplot2)
library(reshape2)
library(ggpubr)
library(DESeq2)
library(pheatmap)
library(phyloseq)
library(Biostrings); packageVersion("Biostrings")
library(decontam)
library(phyloseq)

# load sequence table (nonchimeric ASV abundances in samples)
seqtab.nochim <- readRDS("~/sequences//Bac/bac1_seqtab_nochim.rds")

# load taxonomic annotations (taxonomic ID of each ASV)
```

```
taxa.sp <- readRDS("~/sequences/Bac/bac1_taxa_sp.rds")
```

1.13 Creating a community and taxonomy object.

```
# community object from nonchimeric ASV sequence table
comm <- seqtab.nochim

# taxonomy object - make sure order of ASVs match between community and taxonomy
taxo <- taxa.sp[colnames(comm),]
```

1.14 Doing some cleaning

```
# First, let's make sure we only have bacteria phylums in our taxonomy
table(taxo[, "Kingdom"])

# Well, we have 4 Eukaryota. So those will need to be removed. Let's also put our taxo as a data frame to
# be able to use phyloseq. We can also use this opportunity to put "NAs".
taxid <- as.data.frame(t(taxo))
taxid[] <- lapply(taxid, as.character)
taxid2 <- tidyr::fill(taxid, colnames(taxid), .direction = "down")
taxid2 <- sapply(taxid2, function(x){paste0("unclassified_", x)})
taxid[is.na(taxid)] <- taxid2[is.na(taxid)]
taxid <- t(taxid)
taxid[taxid == "unclassified_NA"] <- NA

# Remove unclassified ASV from the taxonomy matrix then from the ASV matrix
taxid <- subset(as.data.frame(taxid), Kingdom == "Bacteria")
comm <- seqtab.nochim[, colnames(comm) %in% rownames(taxid)]

# Let's make sure we only have Bacterias
table(taxid[, "Kingdom"])
```

1.15 Create phyloseq containing the phylogenetic tree and the metadata.

```
# load metadata
metadata <- read.csv("/home/dpatel/metadata_Bioreacteur1.csv", header=TRUE, sep=";", row.names =
1)

#Phyloseq
```

```

ps <- phyloseq(otu_table(t(comm), taxa_are_rows=TRUE), tax_table(as.matrix(taxid)), sample_data(met
adata))

# Add sequence table

dna <- Biostrings::DNAStringSet(taxa_names(ps))
names(dna) <- taxa_names(ps)
ps <- merge_phyloseq(ps, dna)

# Make ASV ID shorter

taxa_names(ps) <- paste0("ASV", seq(ntaxa(ps)))

```

1.16 Remove the negative controls (decontam)

```

sample_data(ps)$is.neg <- sample_data(ps)$SampleType == "Controle"
contamdf.prev <- isContaminant(ps, method="prevalence", neg="is.neg", threshold=0.5)
table(contamdf.prev$contaminant)
ps.noncontam <- prune_taxa(!contamdf.prev$contaminant, ps)
final.ps=subset_samples(ps.noncontam, !SampleType=="Controle")

#Save the ASVs, taxa and metadata

write.csv(as.data.frame(as(tax_table(final.ps), "matrix")), file = "~/sequences/Bac/bac_csv/noncontam_t
axa.csv")

write.csv(as.data.frame(as(otu_table(final.ps), "matrix")),file = "~/sequences/Bac/bac_csv/noncontam_a
sv.csv")

write.csv(as.data.frame(as(sample_data(final.ps), "matrix")), file= "~/sequences/Bac/bac_csv/nonconta
m_meta.csv")

```

1.17 Making rarefaction curves

```

# What is the smallest number of sequences/sample for subset of samples?
min(rowSums(noncontam_comm))
rowSums(noncontam_comm)

# Rarefaction curve for subset of samples
rarecurve(noncontam_comm, step=200, label=TRUE)
rarecurve(noncontam_comm, step=200, label=TRUE, xlim=c(0,3000))

# Save the graphs

```

```

rarecurve(noncontam_comm, step=200, label=TRUE)
dev.print(width = 24, height = 10, pdf, "/home/dpatel/sequences/Bac/bac_Figures/rarefaction_curve_all.pdf")
rarecurve(noncontam_comm, step=200, label=TRUE, xlim=c(0,2000))
dev.print(width = 24, height = 10, pdf, "/home/dpatel/sequences/Bac/bac_Figures/rarefaction_curve_short.pdf")
minTotRelAbun = 5e-5
x = taxa_sums(final.ps)
keepTaxa = (x / sum(x)) > minTotRelAbun
prunedSet = prune_taxa(keepTaxa, final.ps)
sample_sums(prunedSet)
sample_sums(final.ps)
rarecurve(t(otu_table(final.ps)), step=50, cex=0.5)
ps_raref <- rarefy_even_depth(prunedSet, 2300, rngseed=112)

```

For the eucaryotes, we rarefied at 3000 (prunedSet, 3000, rngseed=112) while we rarefied at 4000 for the archaea (prunedSet, 3000, rngseed=112).

1.18 Save asv, taxa, metadata rarefied.

```

write.csv(as.data.frame(as(tax_table(ps_raref), "matrix")), file = "~/sequences/Bac/bac_csv/rarefied_taxa.csv")
write.csv(as.data.frame(as(otu_table(ps_raref), "matrix")), file = "~/sequences/Bac/bac_csv/rarefied_asv.csv")
write.csv(as.data.frame(as(sample_data(ps_raref), "matrix")), file = "~/sequences/Bac/bac_csv/rarefied_meta.csv")

```

The rest of the scripts were done for all three domains. Thus, only the script for the Bacteria samples will be shown here. The only exceptions are the parts about the sessile community: since we did not have the data for the archaeal sessile community, those tests and plots we excluded.

2. CALCULATION OF THE ALPHA-DIVERSITY AND DOING ANALYSIS

2.1 Loading necessary libraries

```

library(vegan)
library(tidyr)

```

```
library(dplyr)
library(FSA)
library(ggplot2)
```

2.2 Loading all files needed here and separating certain dataframes.

```
# Load ASV table and metadata
ASV_table <- t(read.table("~/sequences/Bac/bac_csv/rarefied_asv.csv", header = TRUE, row.names = 1,
sep = ","))
metadata <- read.table("~/sequences/Bac/bac_csv/rarefied_meta.csv", header=TRUE, sep=";", row.names = NULL)
# Since our sample names differ in our ASV and Metadata table, we'll have to replace the "-" to "."
# We'll also change the name of the first column to "Nom"
metadata$X <- gsub("-", ".", metadata$X)
colnames(metadata)[1] <- "Nom"
```

2.3 Calculate the alpha diversity.

```
# Calculate the alpha diversity and put it in the metadata dataframe
metadata$alpha_diversity <- diversity(ASV_table, index = "shannon")
metadata$richness <- specnumber(ASV_table)
```

2.4 Separate the Filter samples from the Rock samples.

```
# Starting with the Filters
Filters_all <- metadata[(metadata$SampleType == "Filtre"), ]
Filters_all$Nom <- gsub("F-", "", Filters_all$Nom)
Filters_noT <- Filters_all[(Filters_all$Bioreacteur == "Terrain"), ]
# Doing with the Rocks
Rocks_noT <- metadata[(metadata$SampleType == "Roche"), ]
Rocks_noT$Nom <- gsub("R-", "", Rocks_noT$Nom)
# And let's make a dataframe with both Filters and Rocks
ROnly <- metadata[!(metadata$Bioreacteur == "Terrain"), ]
# To merge our geochemical data to the reste of our metadata, I'll have to make another dataframe
```

```
Analyse <- read.table("/home/dpatel/Analyse_geo_merged.csv", header=TRUE, sep=";", row.names = N
ULL)
Filtres_noT_geo <- merge(Filtres_noT, Analyse, by = c("Bioreacteur", "Jour"), sort = TRUE)
```

2.5 Visualizing our alpha diversity with BOTH Filters and Rocks

```
# Change the names in the Bioreacteur column to Bioreactor 1, Bioreactor 2, Bioreactor 3 and Field
ROnly$Bioreacteur <- ifelse(grepl("^B", ROnly$Bioreacteur), "Bioreactor 1",
  ifelse(grepl("^J", ROnly$Bioreacteur), "Bioreactor 2",
    ifelse(grepl("^R", ROnly$Bioreacteur), "Bioreactor 3", "Field"))))
# Doing the boxplot for our Filters and Rocks WITHOUT our Terrain data
RF_box <- ggplot(ROnly, aes(x = Bioreacteur, y = alpha_diversity, fill = SampleType)) +
  geom_boxplot() +
  labs(x = "Bioreactors", y = "Shannon index for Bacteria") +
  scale_fill_manual(labels = c("Planktonic", "Sessile"), values = c("#2A9D8F", "#F4A261"), name = "Comm
unity") +
  theme(text = element_text(size = 22), legend.text = element_text(size = 18))
RF_box
```

2.6 Comparing alpha-diversity between lifestyles and bioreactor.

```
# Looking at the Normality
shapiro.test(ROnly$alpha_diversity)
# Checking the Homogeneity of Variance
fligner.test(alpha_diversity ~ SampleType, data = ROnly)
kruskal.test(alpha_diversity ~ SampleType, data = ROnly)
# Kruskal-Wallis chi-squared = 31.729, df = 1, p-value = 1.772e-08
# There is a significant difference between the SampleType

# To compare our Bioreactors for our Rock and Filter, we'll do a Kruskal Wallis test
kruskal.test(alpha_diversity ~ Bioreacteur, data = Filtres_noT)
kruskal.test(alpha_diversity ~ Bioreacteur, data = Rocks_noT)
```

2.7 Regression between our environmental factors and our alpha diversity.

For this part, we're just using our Filters data without our Terrain in it. I'll have to make another dataframe since I'm missing some data.

```
Analyse <- read.table("/home/dpatel/Analyse_geo_merged.csv", header=TRUE, sep=";", row.names = NULL)
```

```
Filtres_noT_geo <- merge(Filters_noT, Analyse, by = c("Bioreacteur", "Jour"), sort = TRUE)
```

We'll do multiple models with the variables

```
geo_glm1 <- glm(alpha_diversity ~ Moy_DIC + Moy_DOC + Moy_NH4 + Moy_NO3 + Moy_NO2 + Moy_NOx + pH, data = Filtres_noT_geo)
```

```
geo_glm2 <- glm(alpha_diversity ~ Moy_DIC + Moy_DOC + Moy_NH4 + Moy_NO3 + Moy_NO2 + Moy_NOx, data = Filtres_noT_geo)
```

```
geo_glm3 <- glm(alpha_diversity ~ Moy_DIC + Moy_DOC + Moy_NH4 + Moy_NO3 + Moy_NO2, data = Filtres_noT_geo)
```

```
geo_glm4 <- glm(alpha_diversity ~ Moy_DIC + Moy_DOC + Moy_NH4 + Moy_NO3, data = Filtres_noT_geo)
```

```
geo_glm5 <- glm(alpha_diversity ~ Moy_DIC + Moy_DOC + Moy_NH4, data = Filtres_noT_geo)
```

```
geo_glm6 <- glm(alpha_diversity ~ Moy_DIC + Moy_DOC, data = Filtres_noT_geo)
```

```
geo_glm7 <- glm(alpha_diversity ~ Moy_DIC, data = Filtres_noT_geo)
```

Checking which models is the best

```
AIC(geo_glm1, geo_glm2, geo_glm3, geo_glm4, geo_glm5, geo_glm6, geo_glm7)
```

```
##           df    AIC
## geo_glm1  9 36.45443
## geo_glm2  8 34.45444
## geo_glm3  7 33.78771
## geo_glm4  6 33.10867
## geo_glm5  5 31.10966
## geo_glm6  4 33.75135
## geo_glm7  3 33.12171
```

The best model seems to be our model 5 (alpha_diversity ~ Moy_DIC + Moy_DOC + Moy_NH4).

Let's compare the model to the other ones.

```
anova(geo_glm1, geo_glm5, test="Chisq")
```

```
anova(geo_glm2, geo_glm5, test="Chisq")
```

```
anova(geo_glm3, geo_glm5, test="Chisq")
```

```
anova(geo_glm6, geo_glm5, test="Chisq")
anova(geo_glm7, geo_glm5, test="Chisq")

# The geo_glm4 is not statistically different from the model 1, 2, 3 and 7, but is statistically different from
the model 6 (not by much). So, it seems like the NH4 is an important parameter and that we must keep it

# Now, let's look at the results of our glm and the Residuals

summary(geo_glm5)
plot(geo_glm5)
```

3. CALCULATION OF THE BETA-DIVERSITY AND DOING ANALYSIS

3.1 Load libraries

```
library(vegan)
library(tidyr)
library(dplyr)
library(ggplot2)
```

3.2 Doing the dissimilarity matrix

```
# Load ASV table and metadata
ASV_table <- t(read.table("/home/dpatel/sequences/Bac/bac_csv/rarefied_asv.csv", header = TRUE, ro
w.names = 1, sep = ","))

# Let's start by doing a dissimilarity matrix (Bray-Kurtis)
beta_dist <- vegdist(ASV_table, method = "bray")
write.csv(as.matrix(beta_dist), "/home/dpatel/sequences/Bac/bac_csv/dist_matrix.csv")
beta_dist <- read.table("/home/dpatel/sequences/Bac/bac_csv/dist_matrix.csv",
```

3.3 PERMANOVA for Filters

```
# After taking a look at the PCoA, we'll separate the data based on the Jour. So, the days will be J2-J8, J1
0-J12 and J14-J16.
rm(list=ls())

# Load ASV table and metadatas
ASV_table <- t(read.table("~/sequences/Bac/bac_csv/rarefied_asv.csv", header = TRUE, row.names = 1,
sep = ","))
metadata <- read.table("~/sequences/Bac/bac_csv/rarefied_meta.csv", header=TRUE, sep=",", row.nam
es = NULL)
```

```

beta_dist <- read.table("/home/dpatel/sequences/Bac/bac_csv/dist_matrix.csv", header=TRUE, sep=";",
row.names = NULL)

colnames(beta_dist) <- c("Nom", beta_dist$X)

# Remove the unnecessary column in metadata

metadata <- metadata %>% select(-is.neg, -Rugosite1, -Rugosite2, -concADN)

# Since our sample names differ in our ASV and Metadata table, we'll have to replace the "-" to "."
# We'll also change the name of the first column to "Nom"

metadata$X <- gsub("-", ".", metadata$X)

colnames(metadata)[1] <- "Nom"

# Doing the PERMANOVA

# Separate the Filtres based on the Jours

meta14 <- metadata[metadata$Jour %in% c("2", "4", "6", "8"), ]
meta56 <- metadata[metadata$Jour %in% c("10", "12"), ]
meta711 <- metadata[metadata$Jour %in% c("14", "16", "18", "20", "22"), ]

# Now we have to join the metadata dataframes with the matrix and keep the lower half of the dist

meta14_dist <- inner_join(meta14, beta_dist, by = "Nom")
dist14 <- meta14_dist %>% select(all_of(meta14_dist$Nom)) %>% as.dist()

meta56_dist <- inner_join(meta56, beta_dist, by = "Nom")
dist56 <- meta56_dist %>% select(all_of(meta56_dist$Nom)) %>% as.dist()

meta711_dist <- inner_join(meta711, beta_dist, by = "Nom")
dist711 <- meta711_dist %>% select(all_of(meta711_dist$Nom)) %>% as.dist()

# Run the adonis2 function and look at the results

Corr14 <- adonis2(dist14~Bioreacteur + Jour, data=meta14_dist, by="margin")
Corr56 <- adonis2(dist56~Bioreacteur + Jour, data=meta56_dist, by="margin")
Corr711 <- adonis2(dist711~Bioreacteur + Jour, data=meta711_dist, by="margin")

```

3.4 PERMANOVA for Rocks

```
# Load ASV table and metadata
rm(list=ls())

ASV_table <- t(read.table("~/sequences/Bac/bac_csv/rarefied_asv.csv", header = TRUE, row.names = 1,
sep = ","))

metadata <- read.table("~/sequences/Bac/bac_csv/rarefied_meta.csv", header=TRUE, sep=",", row.names = NULL)

beta_dist <- read.table("/home/dpatel/sequences/Bac/bac_csv/dist_matrix.csv", header=TRUE, sep=",",
row.names = NULL)

colnames(beta_dist) <- c("Nom", beta_dist$X)

# Remove the unnecessary column in metadata
metadata <- metadata %>% select(-is.neg, -concADN, -Jour)

# Since our sample names differ in our ASV and Metadata table, we'll have to replace the "-" to "."
# We'll also change the name of the first column to "Nom"
metadata$X <- gsub("-", ".", metadata$X)
colnames(metadata)[1] <- "Nom"

# Remove the 'Terrain' data
metadata <- metadata[metadata$SampleType == "Roche", ]

# Now we have to join the metadata dataframe with the matrix and keep the lower half of the dist
meta_dist <- inner_join(metadata, beta_dist, by = "Nom")
all_dist <- meta_dist %>% select(all_of(meta_dist$Nom)) %>% as.dist()

# Run the adonis2 function and look at the results
Corr <- adonis2(all_dist ~ Bioreacteur + Rugosite1 + Rugosite2, data = meta_dist, by="margin")
print(Corr)
```

3.5 PERMANOVA for lifestyle

```
meta_dist <- inner_join(metadata, beta_dist, by = "Nom")
dist <- meta_dist %>% select(all_of(meta_dist$Nom)) %>% as.dist()
Corr <- adonis2(dist ~ SampleType, data = meta_dist, by="margin")
```

3.6 Doing the dbRDA: Filters only

```
# For the metadata dataframe, it should have samples as rows and metadata variables as columns.
rm(list=ls())

# Load the ASV table and the metadata
ASV_table <- t(read.table("~/sequences/Bac/bac_csv/rarefied_asv.csv", header = TRUE, row.names = 1,
sep = ","))

metadata <- read.table("~/sequences/Bac/bac_csv/rarefied_meta.csv", header=TRUE, sep="," , row.names = NULL)

# Remove the unnecessary column in metadata
metadata <- metadata %>% select(-is.neg,-Rugosite1,-Rugosite2)

# Since our sample names differ in our ASV and Metadata table, we'll have to replace the "-" to "."
# We'll also change the name of the first column to "Nom"
metadata$X <- gsub("-", ".", metadata$X)
colnames(metadata)[1] <- "Nom"

# Taking all my Filter data from metadata
Meta_filters_all <- metadata[(metadata$SampleType == "Filtre"), ]

# Merging the geochemical data to the reste of our metadata.
Analyse <- read.table("/home/dpatel/Analyse_geo_merged.csv", header=TRUE, sep="," , row.names = NULL)

Meta_filters_all <- merge(Meta_filters_all, Analyse, by = c("Bioreacteur", "Jour"), sort = TRUE)

# Change the names in the Bioreacteur column to Bioreactor 1, Bioreactor 2, Bioreactor 3 and Field
Meta_filters_all$Bioreacteur <- ifelse(grepl("^B", Meta_filters_all$Bioreacteur), "Bio1",
  ifelse(grepl("^J", Meta_filters_all$Bioreacteur), "Bio2",
    ifelse(grepl("^R", Meta_filters_all$Bioreacteur), "Bio3", "Field"))))

# Taking the ASV table and keeping the same samples as in the Filtres_noT_geo metadata. We'll also transform the raw ASV data into relative abundance.
ASV_table <- ASV_table[Meta_filters_all$Nom, ]
ASV_table <- decostand(ASV_table, method = "hellinger")

# Removing "F." from the sample names for the ASV table and the Meta_filters_all dataframe
rownames(ASV_table) <- gsub("F\\.", "", rownames(ASV_table))
Meta_filters_all$Nom <- gsub("F\\.", "", Meta_filters_all$Nom)

# Change the J of Jour to D (day)
```

```
Meta_filters_all$Nom <- gsub(".J", "-D", Meta_filters_all$Nom)
```

3.6.1 Doing the Distance matrix and dbRDA

```
beta_dist <- vegdist(ASV_table, method = "bray")  
  
dbRDA_result <- capscale(beta_dist ~ Moy_DIC + Moy_DOC + Moy_NO2 + Moy_NOx + Moy_NH4 + Moy_NO3 + pH, data = Meta_filters_all)  
  
# Let's see if we can simplify our dbRDA  
  
# So step() is going to find the simplest models we can have: taking away variables that aren't too important  
  
mod1 <- step(dbRDA_result, scope = formula(dbRDA_result), criteria = "AIC", test = "perm", direction = "both")  
  
summary(mod1)  
  
mod2 <- capscale(beta_dist ~ Moy_DIC + Moy_NH4, data = Meta_filters_all)  
  
# Doing statistical analysis on the simpler model
```

3.6.2 Doing some statistical test

```
# Set a random seed for reproducibility  
set.seed(123)  
  
# Looking for some collinearity  
vif.cca(mod2)  
  
## Moy_DIC Moy_NH4  
## 1.040228 1.040228  
  
# Ok, so we don't see any "inflation" in the numbers so there isn't any collinearity  
  
# test for significant variables from the metadata  
anova(mod2, by = "terms", permu = 1000)
```

3.7 Doing the dbRDA: Rock only

```
# Loading all files needed here  
rm(list=ls())  
  
# Load ASV table and metadatas  
  
ASV_table <- t(read.table("~/sequences/Bac/bac_csv/rarefied_asv.csv", header = TRUE, row.names = 1, sep = ","))
```

```

metadata <- read.table("~/sequences/Bac/bac_csv/rarefied_meta.csv", header=TRUE, sep=";", row.names = NULL)

mineral <- read.table("~/Mineralogie_geo_merged.csv", header=TRUE, sep=";", row.names = NULL)

# Since our sample names differ in our ASV and Metadata table, we'll have to replace the "-" to "."
# We'll also change the name of the first column to "Nom"
metadata$X <- gsub("-", ".", metadata$X)
colnames(metadata)[1] <- "Nom"

# Since our mineralogy is in percentage, we'll divide the values by 100.
mineral <- mineral %>% mutate_at(vars(X.IronOxide, X.Quartz, X.BlackQuartz, X.RedFeldspath, X.WhiteFeldspath, X.Mica, X.Undetermined), ~ . / 100)

# Also, since BlackQuartz and Quartz have the same molecular composition, we're going to add them into the Quartz column.
mineral <- mineral %>% mutate(X.Quartz = X.Quartz + X.BlackQuartz)

# Merging metadata with mineral dataframe so everything is together
# We'll then just keep the Roche samples
meta_mineral <- merge(metadata, mineral, by = "Nom", all.x = TRUE)
meta_mineral <- meta_mineral[(meta_mineral$SampleType == "Roche"), ]

# Remove the unnecessary column
meta_mineral <- meta_mineral %>% select(-is.neg, -Rugosite1, -Rugosite2, -X.BlackQuartz, -Total, -Jour, -SampleType, -concADN)

# Change the names in the Bioreacteur column to Bioreactor 1, Bioreactor 2, Bioreactor 3 and Field
meta_mineral$Bioreacteur <- ifelse(grepl("^B", meta_mineral$Bioreacteur), "Bio1",
  ifelse(grepl("^J", meta_mineral$Bioreacteur), "Bio2",
    ifelse(grepl("^R", meta_mineral$Bioreacteur), "Bio3", "")))

# Taking the ASV table and keeping the same samples as in the meta_mineral
ASV_table <- ASV_table[meta_mineral$Nom, ]

# Removing "R." from the sample names for the meta_mineral dataframe
meta_mineral$Nom <- gsub("R\\.", "", meta_mineral$Nom)

# We'll also transform the raw ASV data into relative abundance
ASV_table <- decostand(ASV_table, method = "hellinger")

```

3.7.1 Doing the Distance matrix and dbRDA

```
beta_dist <- vegdist(ASV_table, method = "bray")

dbRDA_result <- capscale(beta_dist ~ X.IronOxide + X.Quartz + X.RedFeldpath + X.WhiteFeldpath + X.Mica + X.Undetermined + Roughness, data = meta_mineral)

# Let's see if we can simplify our dbRDA
set.seed(123)
mod1 <- step(dbRDA_result, scope = formula(dbRDA_result), test = "perm")

## X.IronOxide X.Quartz X.RedFeldpath X.WhiteFeldpath X.Mica
## 2.697464 624.322794 45.167683 167.700033 142.402159

###Ok, so we see an "inflation" in the numbers so there is collinearity

mod2 <- capscale(beta_dist ~ X.Quartz + X.RedFeldpath + X.Mica, data = meta_mineral)
vif.cca(mod2)

## X.Quartz X.RedFeldpath X.Mica
## 5.546982 1.889650 4.133549

###This is better. We'll use this model.
anova(mod2, by="terms", permu=1000)
```

4. COMPARING GEOCHEMICAL PARAMETERS BETWEEN BIOREACTORS

4.1 Load libraries and document

```
library(MASS)
library(tidyr)
library(dplyr)
library(ggplot2)
library(stats)

Analyse <- read.csv("Analyse_geo_Test1.csv", header = TRUE, sep = ",", dec = ".")
```

The following steps will be done for every parameters.

4.2 Doing a dataframe with the averages and standard deviation of DIC based on the Bioreactor and Day

```
df <- data.frame()

for(i in unique(Analyse$Jour)) {
  subset(Analyse, Jour == i) -> data_Jour
  aggregate(DIC.mg.L ~ Bioreacteur, data = data_Jour, mean, na.rm = TRUE) -> DIC_moy
  aggregate(DIC.mg.L ~ Bioreacteur, data = data_Jour, sd, na.rm = TRUE) -> DIC_sd
  df <- rbind(df, data.frame(DIC_moy, DIC_sd))
}

# Rearrange the rows to have all "Bleu", "Jaune", "Rouge" in that order + Remove the Tourb and Puit row
df <- df %>% arrange(factor(Bioreacteur, levels = c("Bleu", "Jaune", "Rouge")))
df <- df[-c(31,32),]

## Let's rename the Bioreacteur.1 column as "Jour" since that column is pretty useless for us. While we're
at it, we can also change the name of our other column so that they represent the values properly.

colnames(df)[colnames(df) == "Bioreacteur.1"] <- "Jour"
colnames(df)[colnames(df) == "DIC.mg.L."] <- "Moy"
colnames(df)[colnames(df) == "DIC.mg.L..1"] <- "Sd"

# Now, we have to put the number of the days.

df$Jour <- as.numeric(c("6", "8", "10", "12", "14", "16", "18", "20", "22", "24", "6", "8", "10", "12", "14", "16", "18",
"20", "22", "24", "6", "8", "10", "12", "14", "16", "18", "20", "22", "24"))

# We'll change the names in the Bioreacteur column to Bioreactor1, Bioreactor2, Bioreactor3 and Field
df$Bioreacteur <- ifelse(grepl("^B", df$Bioreacteur), "Bioreactor 1",
  ifelse(grepl("^J", df$Bioreacteur), "Bioreactor 2",
    ifelse(grepl("^R", df$Bioreacteur), "Bioreactor 3", " ")))
```

4.3 Comparing the data between bioreactors.

```
shapiro.test(df$Moy)
fligner.test(Moy ~ Bioreacteur, data = df)
kruskal.test(Moy ~ Bioreacteur, data = df)
```

5. ROCK PROPERTIES ANALYSIS

5.1 Load libraries and document

```
library(tidyr)
library(dplyr)
library(ggplot2)
library(stats)
library(vegan)
library(FSA)
library(lmtest)
library(correlation)

# Load metadata
mineral <- read.table("/home/dpatel/sequences/Divya_Mineralogie.csv", header=TRUE, sep=";", row.names = NULL)

# Since our mineralogy is in percentage, we'll divide the values by 100.
mineral <- mineral %>% mutate_at(vars(X.IronOxide, X.Quartz, X.BlackQuartz, X.RedFeldpath, X.WhiteFeldpath, X.Mica, X.Undetermined), ~ . / 100)

# Also, since BlackQuartz and Quartz have the same molecular composition, we're going to add them into the Quartz column.
mineral <- mineral %>% mutate(X.Quartz = X.Quartz + X.BlackQuartz)
```

5.2 Test correlation between bacterial abundance and the minerals

```
cor.test(mineral$Copies, mineral$X.IronOxide, method = "pearson")
cor.test(mineral$Copies, mineral$X.Quartz, method = "pearson")
cor.test(mineral$Copies, mineral$X.RedFeldpath, method = "pearson")
cor.test(mineral$Copies, mineral$X.WhiteFeldpath, method = "pearson")
cor.test(mineral$Copies, mineral$X.Mica, method = "pearson")
```

5.3 Test correlation between bacterial α -diversity and the minerals

```
# Load ASV table and metadata
ASV_table <- t(read.table("~/sequences/Bac/bac_csv/rarefied_asv.csv", header = TRUE, row.names = 1, sep = ";"))
```

```

metadata <- read.table("~/sequences/Bac/bac_csv/rarefied_meta.csv", header=TRUE, sep=";", row.names = NULL)

metadata$X <- gsub("-", ".", metadata$X)
colnames(metadata)[1] <- "Nom"

# Calculate the alpha diversity and put it in the metadata data frame
metadata$alpha_diversity <- diversity(ASV_table, index = "shannon")
metadata$richness <- specnumber(ASV_table)

# Keep rock samples only
Rock <- metadata[(metadata$SampleType == "Roche"), ]
Rock$Nom <- gsub("R\\.", "", Rock$Nom)
meta_dist <- inner_join(mineral, Rock, by = "Nom")

# Doing the correlations
cor.test(meta_dist$alpha_diversity, meta_dist$X.IronOxide, method = "pearson")
cor.test(meta_dist$alpha_diversity, meta_dist$X.Quartz, method = "pearson")
cor.test(meta_dist$alpha_diversity, meta_dist$X.RedFeldpath, method = "pearson")
cor.test(meta_dist$alpha_diversity, meta_dist$X.WhiteFeldpath, method = "pearson")
cor.test(meta_dist$alpha_diversity, meta_dist$X.Mica, method = "pearson")

```

5.4 Comparing mineralogy and roughness between bioreactors

```

rm(list=ls())

# Load metadata
mineral <- read.table("/home/dpatel/sequences/Divya_Mineralogie.csv", header=TRUE, sep=";", row.names = NULL)

# Since our mineralogy is in percentage, we'll divide the values by 100.
mineral <- mineral %>% mutate_at(vars(X.IronOxide, X.Quartz, X.BlackQuartz, X.RedFeldpath, X.WhiteFeldpath, X.Mica, X.Undetermined), ~ . / 100)

# Also, since BlackQuartz and Quartz have the same molecular composition, we're going to add them into the Quartz column.
mineral <- mineral %>% mutate(X.Quartz = X.Quartz + X.BlackQuartz)

#normality test

```

```

shapiro.test(mineral$X.IronOxide)
shapiro.test(mineral$X.Quartz)
shapiro.test(mineral$X.RedFeldpath)
shapiro.test(mineral$X.WhiteFeldpath)
shapiro.test(mineral$X.Mica)
shapiro.test(mineral$Roughness)

### all of the data follows a normal distribution

# Add bioreactor column
mineral$Bioreacteur <- ifelse(grepl("^B", mineral$Nom), "Bioreactor 1",
                             ifelse(grepl("^J", mineral$Nom), "Bioreactor 2",
                                     ifelse(grepl("^R", mineral$Nom), "Bioreactor 3", "")))

# Variance test
leveneTest(X.IronOxide ~ Bioreacteur, data = mineral)
leveneTest(X.Quartz ~ Bioreacteur, data = mineral)
leveneTest(X.RedFeldpath ~ Bioreacteur, data = mineral)
leveneTest(X.WhiteFeldpath ~ Bioreacteur, data = mineral)
leveneTest(X.Mica ~ Bioreacteur, data = mineral)
leveneTest(Roughness ~ Bioreacteur, data = mineral)

##all variances are homogenous

# Kruskal Wallis test
kruskal.test(X.IronOxide ~ Bioreacteur, data = mineral)
kruskal.test(X.Quartz ~ Bioreacteur, data = mineral)
kruskal.test(X.RedFeldpath ~ Bioreacteur, data = mineral)
kruskal.test(X.WhiteFeldpath ~ Bioreacteur, data = mineral)
kruskal.test(X.Mica ~ Bioreacteur, data = mineral)
kruskal.test(Roughness ~ Bioreacteur, data = mineral)

```

6. BACTERIAL ABUNDANCE AND ROUGHNESS

6.1 Load libraries and documents

```

library(vegan)
library(tidyr)

```

```
library(dplyr)
library(ggplot2)
library(stats)
library(gridExtra)
library(lmtest)

dpcr <- read.table("/home/dpatel/dPCR.csv", header=TRUE, sep="," , row.names = NULL)
```

6.2 Doing a linear regression

```
shapiro.test(dpcr$Roughness.1)
shapiro.test(dpcr$Roughness.2)
# Ok there's no normal distribution
# Normalize the data with a logarithmic transformation on Roughness.1 and Roughness.2
dpcr_transformed <- dpcr %>%
  mutate(
    Roughness.1_log = log(Roughness.1),
    Roughness.2_log = log(Roughness.2)
  )
shapiro.test(dpcr_transformed$Roughness.1_log)
shapiro.test(dpcr_transformed$Roughness.2_log)
lm_model1 <- lm(Copies~Roughness.1_log , data = dpcr_transformed)
lm_model2 <- lm(Copies~ Roughness.2_log , data = dpcr_transformed)
# Display the summary of the linear regression
summary(lm_model1)
summary(lm_model2)
# Calculate correlation between Copies and Roughness.1_log and Roughness.2_log
cor(dpcr_transformed$Copies, dpcr_transformed$Roughness.1_log)
cor(dpcr_transformed$Copies, dpcr_transformed$Roughness.2_log)
```

7. ROCK ROUGHNESS (MATLAB)

```
% script to load in 3-D data corresponding to one chip and calculate the
% RMS roughness of the chip. Navigate to the folder containing the ascii files of the data in the Current Folder to left
% the script works by creating a list of all the filenames in the directory then running the analysis on each file in turn. Filenames need to have the same number of characters and contain no spaces
% OUTPUTS are vectors of Sa, RMS, and relief, each containing one value per sample in a list must have the Statistics and Machine Learning Toolbox Add-on as well as the Antenna Toolbox
% initialize the list of files - the files need to all be in the same directory
inFiles = dir('C:/Users/patel/OneDrive/Documents/Maitrise/1.Test_1/3. Cloud Compare/Recto'); % creates a structure a with the file details
% initialize the output vectors - the values of the roughness parameters
% will be entered in here in the same alphabetical order as the file list
Sa = ones(length(inFiles)-2,1);
RMS = ones(length(inFiles)-2,1);
relief = ones(length(inFiles)-2,1);
% run the loop
for i = 1:length(inFiles)-2
    f = (inFiles(i).name); % define names of the file at this position in the loop
% load the data
a = load(f);
% isolate the position coordinates
x = a(:,1);
y = a(:,2);
z = a(:,3);
% create the array of points corresponding to the position coordinates
dx = 0.055;
xvec = min(x):dx:max(x);
yvec = min(y):dx:max(y);
[X,Y] = meshgrid(xvec,yvec);
Z = griddata(x,y,z,X,Y); % Z is the array that contains the topography
```

```

% re-center the topography to vary around 0
Z2 = Z-mean(Z(:),"omitnan");

%check if the data look ok

figure
imagesc(xvec-min(xvec),yvec-min(yvec),Z2)
colorbar
axis equal tight
title 'Map of the surface topography of the chip - units are mm'
Sa(i) = mean(abs(Z(:),"omitnan");
RMS(i) = std(Z(:),"omitnan"); %
relief(i) = range(Z,'all'); % this is the difference between the maximum and minimum height on the sample, might also be useful
end
figure
plot(RMS,Sa,','MarkerSize',12)
xlabel 'RMS'
ylabel 'Sa'
xlswrite('NAME.xls',[RMS(:),Sa(:)])

```

Annexe F: 3D Scanning methodology and chip isolation

INSTRUCTIONS FOR RUNNING THE C210 SCANNER VIA FLEXSCAN3D (FDA033A)

1. Connect the scanner to power. A blue light should appear at the side of the scanner.
2. Launch Flexscan3D 3.3. Wait for the connecting to scanner progress bar to complete. Check the scanner is connected by clicking the “Scanners Pannel” button on the main tool bar. The Status should read “connected” and the live windows to the right should show an image.
3. The sample needs to be around 16 cm below the scanner to be captured by the scanner. Ideally, the surface of interest should be horizontal.
4. The scanner performs best with low ambient light. Consider turning off the ceiling lights in the lab.
5. Go to the project tab in Flexscan3D. Create a new project with the “New” button in the Project section of the toolbar.
6. Make sure the first button on the left of the toolbar (in the Mode section) displays “Scan”. This drop-down menu opens up other tools for editing the data after the scan is complete. On the left side of the Flexscan3D window, choose the following settings from the drop-down menus:
 - Generate: Mesh
 - Alignment: None
 - Clean up: None
 - Smoothing: 1
 - Mesh density: scroll bar to maximum value on the right
 - HDR: unchecked
7. On the right-hand side of the Flexscan3D window, increase the exposure time of the cameras until the live display below the adjustment bar start to show red dots. On the live displays, red dots are over-exposed, greys are ok for data collection, blue means no data collection.

8. Hit “Scan”. Visually inspect scan to check for completeness. Note the number of Vertices in the status bar above the 3D scan data representation (higher number is better). If there are holes in the data, try adjusting the exposure time and/or the height of the sample.
9. Move the sample a small amount (e.g. 1-3 mm) and re-scan. Second scan should appear in the 3D data representation overlapping the earlier scan(s).
 - 1-2 scans are usually sufficient to represent surfaces (more may be needed for complex 3D objects). 3 is ok, but don’t go above 4 scans.
10. After collecting scans, the next step is to align the scans. Select scans you want to align by holding down Ctrl and clicking on each scan in the data tree. When selected, the scan preview in the data tree shows red. Under the Align drop down menu in the Alignment section of the Flexscan3D toolbar, select “Fine Alignment” (NOTE this only works if the scans are already quite closely matched so only move the sample by a small amount between scans). Click the Align button. The alignment process automatically locks the selected scans when it is finished.
11. Combine the aligned scans by clicking “Combine” in the Combine section of the main toolbar.
12. Finalize the merge by clicking “Finalize” in the Combine section of the main toolbar. Check Precise Merge. Uncheck all other options. Click OK.
13. Remove unwanted parts of the dataset:
 - Hold down Ctrl and click and drag to select area to remove in the 3D data representation window. The area turns yellow. Press delete and repeat for all the unwanted data.
14. Save the project by clicking Save All in the Scan section of the main toolbar.
15. To Export the data, click the combined dataset in the data tree so that it turns red. Select the drop-down menu from the Export button in the Scan section of the main toolbar - choose desired file format: ASCII file.

STEPS TO PREPARE DATA IN CLOUDCOMPARE

1. Launch CloudCompare.
2. File → Open → navigate to your ascii file that you exported from FlexScan3D.
3. Hit Open then Apply All. The data will load into the 3D data viewer. To view color for any dataset, select the data in the data tree at left then Edit → Colors → Height Ramp → OK.
4. Check if the data are upside down: Click the  button on the left vertical toolbar to view the top of the data. Rotate the data with the left mouse button to see if it is the right way up.
5. Crop the data: click  in the top horizontal toolbar. A new pop-up toolbar appears. Select the data by using  (either in rectangular selection or free form from the drop-down menu). Isolate one chip by clicking , the segment in button. Click the green arrow on the pop-up toolbar.
6. In the data tree on the left, de-select the original dataset so that only the chip you isolated is showing. Tools → Fit → Plane. A 3D rotation matrix appears in the console. Select the rotation matrix and copy it.
7. Make sure the data corresponding to the chip is selected in the data tree. Edit → Apply Transformation. In the pop-up window, go to the Matrix 4x4 tab and delete whatever values are in there. Paste the values that you copied from the console and hit OK.
8. Expand the data corresponding to the chip of interest and deselect the plane that was fit to the data (Dip: 004 deg. Etc)
9. Select the data corresponding to the chip of interest in the data tree and nothing else. Then File → Save. Navigate to your folder of choice and save the data as ASCII cloud (*.txt) format.
10. In the data tree, deselect the chip of interest. Re-select the original dataset (it should show all the chips minus the one you just worked on). Isolate another chip and save by repeating steps 5 to 9. Repeat for all chips until you have one ASCII file for each chip.

Annexe G: Kit methodology

For the DNA and RNA extraction, we used two kits that works in conjunction in order to get DNA and RNA from the same sample: the RNeasy PowerSoil Total RNA kit and the RNeasy PowerSoil DNA elution kit. The following protocol is how we used both kits.

1. Pour avoir l'ARN, mettre le filtre en dans un tube PowerBead de 15mL.
2. Add 2.5 ml of PowerBead Solution + 0.25 ml of Solution SR1 + 0.8 ml of Solution IRS.
3. Add 3.5 ml of phenol/chloroform/isoamyl alcohol (pH 6.5–8.0). Vortex the Tube until the biphasic layer disappears.
4. Vortex at maximum speed for 15 min.
5. Centrifuge at 2500 x g for 10 min.
6. Transfer the upper aqueous phase (avoid the interphase and lower phenol layer) to a clean 15 ml Collection Tube.
7. Add 1.5 ml of Solution SR3 to the aqueous phase.
8. Vortex to mix.
9. Incubate at 2-8°C for 10 min.
10. Centrifuge at 2,500 x g for 10 min at room temperature. Transfer the supernatant to a new 15 ml Collection Tube.
11. Add 5 ml of Solution SR4 to the supernatant in the Tube and vortex to mix.
12. Incubate at room temperature for 30 min.
13. Centrifuge at 2500 x g for 30 min. Decant the supernatant.
14. Invert the Tube on a paper towel for 5 min to remove the rest of the supernatant.
15. Shake the Solution SR5 to mix and add 1 ml to the Collection Tube. Vortex to re-suspend the pellet.
16. Prepare one JetStar Mini Column.
 - a) Place the JetStar Mini Column inside a new 15 ml Collection Tube to extract the DNA.

- b) Add 2 ml of Solution SR5 to the JetStar Mini Column. Wait for it to completely flow through the column.
 - c) Add the solution from the step 15 through the column and wait for it to pass. This step might have to be done two times. You must discard the liquid each time.
17. Add 1 ml of Solution SR5 to the JetStar Mini Column and allow it to completely flow through. When there's no solution left in the column, transfer the column into a new 15 ml Collection Tube.
 18. Add 1 ml of Solution SR8 to the Column to elute the bound DNA. Allow Solution SR8 to flow into the Collection Tube.
 19. Transfer the eluted DNA to a 2.2 ml Collection Tube and add 1 ml of Solution SR4.
 20. Transfer the column into a new 15 ml Collection Tube to extract the RNA.
 21. Shake Solution SR6 to mix and then add 1 ml to the JetStar Mini Column to elute the RNA. Allow Solution SR8 to flow into the Collection Tube.
 22. Transfer the eluted DNA to a 2.2 ml Collection Tube and add 1 ml of Solution SR4.
 23. Invert both 2.2 ml Collection Tube (the DNA and RNA) multiple times to mix.
 24. Incubate both tubes for at least 10 min at -15 to -30°C.
 25. Centrifuge the tubes at 13,000 x g for 15 min to pellet the DNA and RNA.
 26. Decant the supernatant and invert the tubes onto a paper towel for 10 min to air dry the pellet.
 27. Add 100µL of Solution SR7 in each tube to re-suspend the pellets.

Références

- Abed, R. M., Al Fahdi, D., & Muthukrishnan, T. (2019). Short-term succession of marine microbial fouling communities and the identification of primary and secondary colonizers. *Biofouling*, 35(5), 526-540.
- Alfreider, A., Krössbacher, M., & Psenner, R. (1997). Groundwater samples do not reflect bacterial densities and activity in subsurface systems. *Water Research*, 31(4), 832-840.
- Amalfitano, S., Del Bon, A., Zoppini, A., Ghergo, S., Fazi, S., Parrone, D., ... & Preziosi, E. (2014). Groundwater geochemistry and microbial community structure in the aquifer transition from volcanic to alluvial areas. *Water research*, 65, 384-394.
- Amy, P. S., Haldeman, D. L., Ringelberg, D., Hall, D. H., & Russell, C. (1992). Comparison of identification systems for classification of bacteria isolated from water and endolithic habitats within the deep subsurface. *Applied and Environmental Microbiology*, 58(10), 3367-3373.
- Anandham, R., Kim, S. J., Moon, J. Y., Weon, H. Y., & Kwon, S. W. (2013). Paraherbaspirillum soli gen. nov., sp. nov. isolated from soil. *Journal of Microbiology*, 51, 262-267.
- Anantharaman, K., Brown, C. T., Hug, L. A., Sharon, I., Castelle, C. J., Probst, A. J., ... & Banfield, J. F. (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nature communications*, 7(1), 1-11.
- Benedek, T., Táncsics, A., Szabó, I., Farkas, M., Szoboszlai, S., Fábrián, K., ... & Kriszt, B. (2016). Polyphasic analysis of an Azoarcus-Leptothrix-dominated bacterial biofilm developed on stainless steel surface in a gasoline-contaminated hypoxic groundwater. *Environmental Science and Pollution Research*, 23, 9019-9035.
- Bennett, P. C., Rogers, J. R., Choi, W. J., & Hiebert, F. K. (2001). Silicates, silicate weathering, and microbial ecology. *Geomicrobiology Journal*, 18(1), 3-19.
- Beveridge, T. J., Makin, S. A., Kadurugamuwa, J. L., & Li, Z. (1997). Interactions between biofilms and the environment. *FEMS Microbiology reviews*, 20(3-4), 291-303.
- Böhme, A., Risse-Buhl, U., & Küsel, K. (2009). Protists with different feeding modes change biofilm morphology. *FEMS microbiology ecology*, 69(2), 158-169.
- Bos, R., Van der Mei, H. C., & Busscher, H. J. (1999). Physico-chemistry of initial microbial adhesive interactions—its mechanisms and methods for study. *FEMS microbiology reviews*, 23(2), 179-230.
- Breugelmans, P., Barken, K. B., Tolker-Nielsen, T., Hofkens, J., Dejonghe, W., & Springael, D. (2008). Architecture and spatial organization in a triple-species bacterial biofilm synergistically degrading the phenylurea herbicide linuron. *FEMS microbiology ecology*, 64(2), 271-282.

- Bryce, C. C., Le Bihan, T., Martin, S. F., Harrison, J. P., Bush, T., Spears, B., ... & Cockell, C. S. (2016). Rock geochemistry induces stress and starvation responses in the bacterial proteome. *Environmental microbiology*, 18(4), 1110-1121.
- Chandler, L., Harford, A. J., Hose, G. C., Humphrey, C. L., Chariton, A., Greenfield, P., & Davis, J. (2021). Saline mine-water alters the structure and function of prokaryote communities in shallow groundwater below a tropical stream. *Environmental Pollution*, 284, 117318.
- Ciston, S., Lueptow, R. M., & Gray, K. A. (2008). Bacterial attachment on reactive ceramic ultrafiltration membranes. *Journal of Membrane Science*, 320(1-2), 101-107.
- Corliss, J. O., & Esser, S. C. (1974). Comments on the role of the cyst in the life cycle and survival of free-living protozoa. *Transactions of the American Microscopical Society*, 578-593.
- Dar, R., Bandh, S. A., Shafi, S., & Shameem, N. (2019). Bacterial diversity of the rock-water interface in freshwater ecosystem. In *Freshwater Microbiology* (pp. 73-104). Academic Press.
- Dedysh, S. N., Kulichevskaya, I. S., Huber, K. J., & Overmann, J. (2017). Defining the taxonomic status of described subdivision 3 Acidobacteria: proposal of Bryobacteraceae fam. nov. *International journal of systematic and evolutionary microbiology*, 67(2), 498-501.
- Dejsirilert, S., Kondo, E., Chiewsilp, D., & Kanai, K. (1991). Growth and survival of *Pseudomonas pseudomallei* in acidic environments. *Japanese Journal of Medical Science and Biology*, 44(2), 63-74.
- Dubois, I. E., Holgersson, S., Allard, S., & Malmström, M. (2010). Correlation between particle size and surface area for chlorite and K-feldspar. In *13th International Conference on Water-Rock Interaction, WRI-13; Guanajuato; Mexico; 16 August 2010 through 20 August 2010* (pp. 717-720).
- Dobrovolskaya, T. G., Golovchenko, A. V., Lysak, L. V., & Yurchenko, E. N. (2020). Taxonomic structure of bacterial communities of Rhizospheric soil under Bogs' plants. *Moscow University Soil Science Bulletin*, 75, 93-100.
- Fierer, N., Nemergut, D., Knight, R., & Craine, J. M. (2010). Changes through time: integrating microorganisms into the study of succession. *Research in microbiology*, 161(8), 635-642.
- Flynn, T. M., Sanford, R. A., Ryu, H., Bethke, C. M., Levine, A. D., Ashbolt, N. J., & Santo Domingo, J. W. (2013). Functional microbial diversity explains groundwater chemistry in a pristine aquifer. *BMC microbiology*, 13, 1-15.
- Foster, S. S. D., & Chilton, P. J. (2003). Groundwater: the processes and global significance of aquifer degradation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1440), 1957-1972.
- Griebler, C., & Lueders, T. (2009). Microbial biodiversity in groundwater ecosystems. *Freshwater Biology*, 54(4), 649-677.

- Griebler, C., Mindl, B., Slezak, D., & Geiger-Kaiser, M. (2002). Distribution patterns of attached and suspended bacteria in pristine and contaminated shallow aquifers studied with an in situ sediment exposure microcosm. *Aquatic microbial ecology*, 28(2), 117-129.
- Grösbacher, M., Spicher, C., Bayer, A., Obst, M., Karwautz, C., Pilloni, G., ... & Griebler, C. (2016). Organic contamination versus mineral properties: competing selective forces shaping bacterial community assembly in aquifer sediments. *Aquatic Microbial Ecology*, 76(3), 243-255.
- Groult, B., St-Jean, V., & Lazar, C. S. (2023). Linking Groundwater to Surface Discharge Ecosystems: Archaeal, Bacterial, and Eukaryotic Community Diversity and Structure in Quebec (Canada). *Microorganisms*, 11(7), 1674.
- Han, Y., & Perner, M. (2015). The globally widespread genus *Sulfurimonas*: versatile energy metabolisms and adaptations to redox clines. *Frontiers in microbiology*, 6, 989.
- Harvey, R. W., Harms, H., & Landkamer, L. (2007). Transport of microorganisms in the terrestrial subsurface: In situ and laboratory methods. In *Manual of Environmental Microbiology, Third Edition* (pp. 872-897). American Society of Microbiology.
- Herrmann, M., Geesink, P., Yan, L., Lehmann, R., Totsche, K. U., & Küsel, K. (2020). Complex food webs coincide with high genetic potential for chemolithoautotrophy in fractured bedrock groundwater. *Water research*, 170, 115306.
- Hess, S., & Melkonian, M. (2013). The mystery of clade X: *Orciraptor* gen. nov. and *Viridiraptor* gen. nov. are highly specialised, algivorous amoeboflagellates (Glissomonadida, Cercozoa). *Protist*, 164(5), 706-747.
- Hiraishi, A. (2015). Acidocella. *Bergey's Manual of Systematics of Archaea and Bacteria*, 1-6.
- Inagaki, F., Takai, K., Kobayashi, H., Nealson, K. H., and Horikoshi, K. (2003). *Sulfurimonas autotrophica* gen. nov., sp. Nov., a novel sulfur-oxidizing ϵ -proteobacterium isolated from hydrothermal sediments in the Mid-Okinawa Trough. *Int. J. Syst. Evol. Microbiol.* 53, 1801–1805.
- Johnson, W. P., & Logan, B. E. (1996). Enhanced transport of bacteria in porous media by sediment-phase and aqueous-phase natural organic matter. *Water Research*, 30(4), 923-931.
- Jones, A. A., & Bennett, P. C. (2014). Mineral microniches control the diversity of subsurface microbial populations. *Geomicrobiology Journal*, 31(3), 246-261.
- Jones, A. A., & Bennett, P. C. (2017). Mineral ecology: surface specific colonization and geochemical drivers of biofilm accumulation, composition, and phylogeny. *Frontiers in microbiology*, 8, 491.
- Katechakis, A., & Stibor, H. (2006). The mixotroph *Ochromonas tuberculata* may invade and suppress specialist phago- and phototroph plankton communities depending on nutrient conditions. *Oecologia*, 148, 692-701.

- Kaur, G., Kaur, G., Krol, M., & Brar, S. K. (2022). Unraveling the mystery of subsurface microorganisms in bioremediation. *Current Research in Biotechnology*.
- Kim, J. I., Jeong, M., Archibald, J. M., & Shin, W. (2020). Comparative plastid genomics of non-photosynthetic chrysophytes: genome reduction and compaction. *Frontiers in Plant Science*, *11*, 572703.
- Kjelleberg, S., Humphrey, B. A., & Marshall, K. C. (1982). Effect of interfaces on small, starved marine bacteria. *Applied and Environmental Microbiology*, *43*(5), 1166-1172.
- Knotek-Smith, H. M., Jung, C. M., Harrelson, D. W., Poda, A. R., & Bednar, A. J. (2020). Biopolymer production in the aquifer of a groundwater pump-and-treat system.
- Lawson, P. (2016). The taxonomy of the genus *Clostridium*: current status and future perspectives. *微生物学通报*, *43*(5), 1070-1074.
- Lazar, C. S., Lehmann, R., Stoll, W., Rosenberger, J., Totsche, K. U., & Küsel, K. (2019). The endolithic bacterial diversity of shallow bedrock ecosystems. *Science of The Total Environment*, *679*, 35-44.
- Lee, J. W., Nam, J. H., Kim, Y. H., Lee, K. H., & Lee, D. H. (2008). Bacterial communities in the initial stage of marine biofilm formation on artificial surfaces. *The journal of microbiology*, *46*, 174-182.
- Lefranc, M., Thénot, A., Lepere, C., & Debroas, D. (2005). Genetic diversity of small eukaryotes in lakes differing by their trophic status. *Applied and environmental microbiology*, *71*(10), 5935-5942.
- Li, B., & Logan, B. E. (2004). Bacterial adhesion to glass and metal-oxide surfaces. *Colloids and surfaces B: Biointerfaces*, *36*(2), 81-90.
- Lehosmaa, K., Muotka, T., Pirttilä, A. M., Jaakola, I., Rossi, P. M., & Jyväsjärvi, J. (2021). Bacterial communities at a groundwater-surface water ecotone: gradual change or abrupt transition points along a contamination gradient?. *Environmental Microbiology*, *23*(11), 6694-6706.
- Lessmann, D., Fyson, A., & Nixdorf, B. (2000). Phytoplankton of the extremely acidic mining lakes of Lusatia (Germany) with pH ≤ 3. *Hydrobiologia*, *433*(1-3), 123-128.
- Lopez de Victoria, G. (1989). *Chemotactic behavior of deep subsurface bacteria toward carbohydrates, amino acids and a chlorinated alkene* (No. DP-1779). Savannah River Lab., Aiken, SC (USA).
- Loquay, N., Wylezich, C., & Arndt, H. (2009). Composition of groundwater nanoprotist communities in different aquifers based on aliquot cultivation and genotype assessment of cercozoans. *Fundamental and applied limnology*, *174*(3), 261.

- Ma, Y., Zhao, H., Shan, Q., Xu, Y., Yu, M., Cui, J., ... & He, X. (2021). K-strategy species plays a pivotal role in the natural attenuation of petroleum hydrocarbon pollution in aquifers. *Journal of Hazardous Materials*, 420, 126559.
- McMahon, S., & Parnell, J. (2014). Weighing the deep continental biosphere. *FEMS Microbiology Ecology*, 87(1), 113-120.
- Nuppenen-Puputti, M., Kietäväinen, R., Kukkonen, I., & Bomberg, M. (2023). Implications of a short carbon pulse on biofilm formation on mica schist in microcosms with deep crystalline bedrock groundwater. *Frontiers in Microbiology*, 14, 1054084.
- Ormeño-Orrillo, E., & Martínez-Romero, E. (2019). A genomotaxonomy view of the *Bradyrhizobium* genus. *Frontiers in microbiology*, 10, 1334.
- Pagnier, I., Raoult, D., & La Scola, B. (2011). Isolation and characterization of *Reyranella massiliensis* gen. nov., sp. nov. from freshwater samples by using an amoeba co-culture procedure. *International Journal of Systematic and Evolutionary Microbiology*, 61(9), 2151-2154.
- Park, J., Sanford, R. A., & Bethke, C. M. (2009). Microbial activity and chemical weathering in the Middendorf aquifer, South Carolina. *Chemical geology*, 258(3-4), 232-241.
- Probst, A. J., Holman, H. Y. N., DeSantis, T. Z., Andersen, G. L., Birarda, G., Bechtel, H. A., ... & Moissl-Eichinger, C. (2013). Tackling the minority: sulfate-reducing bacteria in an archaea-dominated subsurface biofilm. *The ISME journal*, 7(3), 635-651.
- Purkamo, L., Bomberg, M., Nyssönen, M., Ahonen, L., Kukkonen, I., & Itävaara, M. (2017). Response of deep subsurface microbial community to different carbon sources and electron acceptors during ~ 2 months incubation in microcosms. *Frontiers in Microbiology*, 8, 232.
- Rempfert, K. R., Miller, H. M., Bompard, N., Nothaft, D., Matter, J. M., Kelemen, P., ... & Templeton, A. S. (2017). Geological and geochemical controls on subsurface microbial life in the Samail Ophiolite, Oman. *Frontiers in microbiology*, 8, 56.
- Roberts, J. A. (2004). Inhibition and enhancement of microbial surface colonization: the role of silicate composition. *Chemical Geology*, 212(3-4), 313-327.
- Rogers, J. R., & Bennett, P. C. (2004). Mineral stimulation of subsurface microorganisms: release of limiting nutrients from silicates. *Chemical Geology*, 203(1-2), 91-108.
- Rogers, J. R., Bennett, P. C., & Choi, W. J. (1998). Feldspars as a source of nutrients for microorganisms. *American Mineralogist*, 83(11-12_Part_2), 1532-1540.
- Rolf K. Hommel, Peter Ahnert, ACETOBACTER, Editor(s): Richard K. Robinson, Encyclopedia of Food Microbiology, Elsevier, 1999, Pages 1-7, <https://doi.org/10.1006/rwfm.1999.0005>.

- Rothhaupt, K. O. (1996). Laboratory experiments with a mixotrophic chrysophyte and obligately phagotrophic and phototrophic competitors. *Ecology*, 77(3), 716-724.
- Ross, N., Villemur, R., Marcandella, E., & Deschenes, L. (2001). Assessment of changes in biodiversity when a community of ultramicrobacteria isolated from groundwater is stimulated to form a biofilm. *Microbial ecology*, 42, 56-68.
- Sang, S., Zhang, X., Dai, H., Hu, B. X., Ou, H., & Sun, L. (2018). Diversity and predictive metabolic pathways of the prokaryotic microbial community along a groundwater salinity gradient of the Pearl River Delta, China. *Scientific reports*, 8(1), 17317.
- Scholl, M. A., Mills, A. L., Herman, J. S., & Hornberger, G. M. (1990). The influence of mineralogy and solution chemistry on the attachment of bacteria to representative aquifer materials. *Journal of Contaminant Hydrology*, 6(4), 321-336.
- Sievert, S. M., Scott, K. M., Klotz, M. G., Chain, P. S., Hauser, L. J., Hemp, J., ... & USF Genomics Class. (2008). Genome of the epsilonproteobacterial chemolithoautotroph *Sulfurimonas denitrificans*. *Applied and environmental microbiology*, 74(4), 1145-1156.
- Smith, A. R., Kieft, B., Mueller, R., Fisk, M. R., Mason, O. U., Popa, R., & Colwell, F. S. (2019). Carbon fixation and energy metabolisms of a subseafloor olivine biofilm. *The ISME Journal*, 13(7), 1737-1749.
- Smith, H. J., Zelaya, A. J., De León, K. B., Chakraborty, R., Elias, D. A., Hazen, T. C., ... & Fields, M. W. (2018). Impact of hydrologic boundaries on microbial planktonic and biofilm communities in shallow terrestrial subsurface environments. *FEMS microbiology ecology*, 94(12), 191.
- Smith, M. C., Bowman, J. P., Scott, F. J., & Line, M. A. (2000). Sublithic bacteria associated with Antarctic quartz stones. *Antarctic Science*, 12(2), 177-184.
- Stevens, T. O., & McKinley, J. P. (1995). Lithoautotrophic microbial ecosystems in deep basalt aquifers. *Science*, 270(5235), 450-455.
- Taylor, R. L., Verran, J., Lees, G. C., & Ward, A. P. (1998). The influence of substratum topography on bacterial adhesion to polymethyl methacrylate. *Journal of Materials Science: Materials in Medicine*, 9(1), 17-22.
- Terrado, R., Pasulka, A. L., Lie, A. A., Orphan, V. J., Heidelberg, K. B., & Caron, D. A. (2017). Autotrophic and heterotrophic acquisition of carbon and nitrogen by a mixotrophic chrysophyte established through stable isotope analysis. *The ISME journal*, 11(9), 2022-2034.
- Uroz, S., Kelly, L. C., Turpault, M. P., Lepleux, C., & Frey-Klett, P. (2015). The mineralosphere concept: mineralogical control of the distribution and function of mineral-associated bacterial communities. *Trends in microbiology*, 23(12), 751-762.

- Verran, J., & Whitehead, K. (2005). Factors affecting microbial adhesion to stainless steel and other materials used in medical devices. *The International journal of artificial organs*, 28(11), 1138-1145.
- Villeneuve K, Violette M, Lazar CS. From Recharge, to Groundwater, to Discharge Areas in Aquifer Systems in Quebec (Canada): Shaping of Microbial Diversity and Community Structure by Environmental Factors. *Genes* (Basel). 2022 Dec 20;14(1):1. doi: 10.3390/genes14010001. PMID: 36672742; PMCID: PMC9858702.
- Viles, H. (1995). Ecological perspectives on rock surface weathering: towards a conceptual model. *Geomorphology*, 13(1-4), 21-35.
- Vishnyakova, A., Popova, N., Artemiev, G., Botchkova, E., Litt, Y., & Safonov, A. (2022). Effect of Mineral Carriers on Biofilm Formation and Nitrogen Removal Activity by an Indigenous Anammox Community from Cold Groundwater Ecosystem Alone and Bioaugmented with Biomass from a “Warm” Anammox Reactor. *Biology*, 11(10), 1421.
- Wackett, L. P., Dodge, A. G., & Ellis, L. B. (2004). Microbial genomics and the periodic table. *Applied and environmental microbiology*, 70(2), 647-655.
- Walker, L. R., & del Moral, R. (2011). Primary succession. *eLS*.
- Whitehead, K. A., & Verran, J. (2006). The effect of surface topography on the retention of microorganisms. *Food and bioprocesses processing*, 84(4), 253-259.
- Whitman, W. B., Coleman, D. C., & Wiebe, W. J. (1998). Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences*, 95(12), 6578-6583.
- Wollmann, K., Deneke, R., Nixdorf, B., & Packroff, G. (2000). Dynamics of planktonic food webs in three mining lakes across a pH gradient (pH 2–4). *Hydrobiologia*, 433(1-3), 3-14.
- Yin, W., Wang, Y., Liu, L., & He, J. (2019). Biofilms: the microbial “protective clothing” in extreme environments. *International journal of molecular sciences*, 20(14), 3423.
- Zhang, H., Gao, Z., Shi, M., & Fang, S. (2020). Soil bacterial diversity and its relationship with soil CO₂ and mineral composition: a case study of the Laiwu experimental site. *International journal of environmental research and public health*, 17(16), 5699.