

Jérôme Fortin – 536 996 920

Nicolas Campeau – 536 781 691

Juliette Pourrain – 536 983 589

Daniel Ilboudo – 536 773 884

Nesrine Imloul- 111 290 248

## **Devoir #3 – Métataxonomie**

Rapport présenté à Nicolas Derome

Dans le cadre du cours

Statistiques génétiques : concepts et analyse

BIF-4002

25-04-2025



## Introduction

L'étude de la diversité microbienne dans les environnements aquatiques, notamment en aquaculture, est devenue un enjeu majeur tant pour la recherche que pour les pratiques d'élevage durable. Parmi les espèces les plus étudiées figure le saumon atlantique (*Salmo salar*), pilier de l'industrie aquacole dans de nombreux pays, dont le Canada. Le microbiote intestinal de ce poisson joue un rôle essentiel dans sa croissance, sa santé immunitaire et son adaptation à différents environnements (Gahan et al., 2024). Afin de mieux comprendre les communautés microbiennes associées au saumon, des approches modernes comme la métagénomique et plus spécifiquement la métataxonomie sont utilisées.

La métataxonomie repose sur le séquençage d'un gène marqueur — souvent le gène 16S rRNA chez les bactéries — afin de caractériser la composition taxonomique des communautés microbiennes présentes dans un échantillon. Cette approche permet d'identifier et de comparer les profils microbiens sans passer par des étapes de culture, ce qui est crucial dans le cas de micro-organismes difficiles à cultiver ou inconnus (Quast et al., 2013). Grâce à l'utilisation de bases de données comme SILVA, les séquences obtenues peuvent être associées à des taxons connus jusqu'au niveau du genre ou de l'espèce, selon le degré de similarité (Yarza et al., 2014).

Dans le domaine de l'aquaculture, l'analyse du microbiote intestinal des poissons comme le saumon revêt un intérêt particulier. Le microbiome joue un rôle crucial dans la digestion des nutriments, la régulation du système immunitaire, la résistance aux pathogènes, ainsi que dans la croissance et le bien-être général des animaux (Garhan et al., 2024 ; Schmidt et al., 2024). Une perturbation de cet équilibre microbien – appelée dysbiose – peut entraîner une sensibilité accrue aux maladies, des problèmes métaboliques, voire une baisse de performance dans les fermes piscicoles. C'est pourquoi de plus en plus d'études cherchent à comprendre les facteurs qui influencent la composition et la stabilité du microbiote intestinal chez les espèces aquacoles.

Les recherches récentes ont mis en évidence que des facteurs comme l'alimentation, le stade de développement, l'environnement (eau douce vs eau salée), les traitements prophylactiques ou encore la densité d'élevage peuvent influencer significativement les communautés microbiennes présentes dans l'intestin du saumon (Schmidt et al., 2024). Par exemple, des changements dans le régime alimentaire peuvent favoriser certaines familles bactériennes aux dépens d'autres, ce qui peut avoir des effets en cascade sur l'immunité ou la

digestion. Comprendre ces interactions est donc fondamental pour améliorer les pratiques aquacoles.

Le développement des technologies de séquençage haut débit a permis d'obtenir un aperçu de plus en plus précis de ces communautés microbiennes. En particulier, les plateformes Illumina offrent la possibilité de générer des millions de lectures courtes à partir d'échantillons biologiques, ce qui facilite la détection de micro-organismes minoritaires, la comparaison entre conditions expérimentales, et l'analyse de la diversité alpha et bêta des communautés bactériennes (Caporasso et al., 2010). Couplée à des logiciels spécialisés comme QIIME2, cette technologie permet une analyse rigoureuse et reproductible des profils microbiens, de l'extraction des séquences à leur interprétation statistique.

Dans ce contexte, le présent devoir a pour objectif de caractériser et de comparer les communautés microbiennes intestinales associées à différents échantillons de saumons atlantiques à l'aide d'une approche métataxonomique. Les données utilisées sont issues d'un séquençage 16S rRNA ciblant une région hypervariable du gène bactérien, permettant une assignation taxonomique précise à l'aide d'outils bioinformatiques reconnus. Les échantillons analysés représentent diverses conditions biologiques ou environnementales, permettant d'explorer la variabilité des profils microbiens selon des facteurs spécifiques définis dans le fichier de métadonnées.

L'objectif de ce travail est de mettre en pratique une démarche d'analyse métataxonomique complète et rigoureuse à partir de données réelles, en vue de décrire la composition microbienne intestinale du saumon et d'explorer sa variabilité selon différents groupes définis expérimentalement.

## Matériel et méthodes

L'étude a été réalisée à partir de quatre populations de saumons atlantiques : Malbaie Captive (MC), Malbaie Sauvage (MW), Rimouski Captive (RC) et Rimouski Sauvage (RW). Les individus captifs (MC et RC) ont été élevés dans un environnement commun, la pisciculture de Tadoussac, tandis que les individus sauvages (MW et RW) ont été prélevés dans deux environnements naturels distincts : la rivière Malbaie et la rivière Rimouski, respectivement.

L'ADN du microbiote intestinal a été extrait à partir des échantillons intestinaux de chaque individu. L'analyse a ciblé la région V3-V4 (~456 pb) du gène 16S de l'ARN ribosomique bactérien, couramment utilisée pour l'identification taxonomique des bactéries. Le séquençage a été effectué sur la plateforme *MiSeq d'Illumina*.

Le traitement des données a été réalisé dans R en utilisant les bibliothèques *DADA2* (Callahan et al., 2025) et *phyloseq* (J. McMurdie et al., 2025). Les lectures brutes ont été filtrées en fonction de leur qualité (score Phred  $\geq 20-25$ ), corrigées pour les erreurs de séquençage, et les séquences chimériques ont été éliminées. Ce pipeline a permis de générer une table d'abondance des ASVs (*Amplicon Sequence Variants*), correspondant aux séquences uniques détectées.

Les ASVs ont été annotées taxonomiquement à l'aide de la base *SILVA* (SILVA, 2024), et les abondances ont été normalisées avec phyloseq. Les 30 ASVs les plus abondants ont été visualisées aux niveaux taxonomiques de l'ordre, de la famille et du genre. Enfin, la répartition des 20 ASVs les plus fréquents a été analysée par population.

Ensuite, la diversité alpha des échantillons a été évaluée selon quatre mesures : la richesse spécifique (Chao1), l'entropie de Shannon, l'inverse de l'entropie de Simpson et l'indice de Fisher. La richesse spécifique (S) est représentée par l'équation suivante :

$$S_{Chao1} = S_{obs} + \frac{n_1(n_1-1)}{2(n_2+1)}$$

$S_{obs}$  est le nombre d'espèces différentes observées,  $n_1$  est le nombre d'espèces observées une fois (singletons) et  $n_2$  est le nombre d'espèces observées deux fois (doublons). L'entropie de Shannon (H) est définie par :

$$H = -\sum_{s=1}^S p_s \ln p_s$$

$p_s$  représente la probabilité qu'un individu tiré au hasard appartienne à l'espèce s. L'inverse de l'entropie de Simpson, une dérivation de l'indice de Simpson, est définie par (Lynch, n.d.):

$$InvE = \frac{1}{\sum_{s=1}^S p_s^2}$$

Finalement, l'indice de Fisher est déterminé par (Styan et al., 2005) :

$$S = \alpha \log_e \left( 1 + \frac{n}{\alpha} \right)$$

$\alpha$  est l'indice de diversité Fisher, n le nombre total d'individu et S est le nombre d'espèces différentes.

Par la suite, afin de déterminer si le test ANOVA peut être utilisé, deux postulats doivent être respectés : l'homogénéité des variances et la normalité de la distribution des données (DATatab Team, 2025). Pour les vérifier, deux tests ont été réalisés : le test de Shapiro-Wilk pour la normalité, et le test de Levene pour l'homogénéité des variances.

Lorsque ces deux postulats sont respectés, l'analyse de variance peut être effectuée. Dans cette étude, une ANOVA à un facteur a été utilisée, suivie d'un test post-hoc de Tukey pour comparer les groupes. Le test ANOVA à un facteur repose sur le modèle suivant (National Institute of Standards and Technology [NIST], n.d.) :

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

où  $Y_{ij}$  représente la réponse observée pour l'observation j dans le niveau i,  $\mu$  la moyenne générale,  $\tau_i$  l'effet du niveau i du facteur et  $\epsilon_{ij}$  est l'erreur aléatoire associée à l'observation.

Cependant, si au moins un des deux postulats n'est pas respecté, deux tests non paramétriques peuvent être réalisés : le test de Kruskal-Wallis et le test de Wilcoxon. Lorsque l'hypothèse de normalité n'est pas vérifiée pour un jeu de données, le test de Kruskal-Wallis (ou test H) est utilisé. Celui-ci est défini par l'équation suivante (DATatab Team, 2025b) :

$$H = \frac{n-1}{n} \cdot \sum_{s=1}^S \frac{n_s \cdot (\bar{R} - E_R)^2}{\sigma^2}$$

$$\text{où } \sigma^2 = \frac{n^2-1}{12} \text{ et } E_R = \frac{n+1}{2}$$

La valeur H correspond à la statistique de test, qui suit une loi du  $\chi^2$ .  $n$  représente le nombre le nombre de groupes,  $\bar{R}$  est la somme des rangs pour le groupe i,  $\sigma^2$  est la variance des rangs et  $E_R$  est la valeur attendue du rang moyen.

Par la suite, lorsque l'ANOVA ne peut pas être utilisée, le test de Wilcoxon remplace le test post-hoc de Tukey. Ce test permet de comparer deux groupes à la fois en se basant sur les rangs des observations.

La diversité bêta a été analysée à l'aide de la méthode NMDS (Positionnement Multidimensionnel Non-métrique) et PCoA (Analyse en Coordonnées Principales), appliquées sur des matrices de dissimilarité Bray-Curtis, UniFrac et UniFrac pondéré.

Le NMDS repose sur la minimisation d'un indice de stress (S), défini par l'équation suivante :

$$S = \sqrt{\frac{\sum_{i \neq j} [\theta(d_{ij}) - \tilde{d}_{ij}]^2}{\sum_{i \neq j} \tilde{d}_{ij}^2}}$$

$d_{ij}$  représente la distance observée entre les échantillons i et j,  $\tilde{d}_{ij}$  est la distance entre i et j dans l'espace réduit et  $\theta(d_{ij})$  est une transformation monotone non linéaire des différences observées.

Finalement, des tests PERMANOVA (via la fonction ADONIS) ont été réalisés sur les matrices de diversité bêta afin de vérifier statistiquement la présence de différences significatives entre les groupes. Le test PERMANOVA repose sur un modèle linéaire multivarié de la forme suivante :

$$Y = \mu + \sum_{i=1}^p \alpha_i X_i + \varepsilon$$

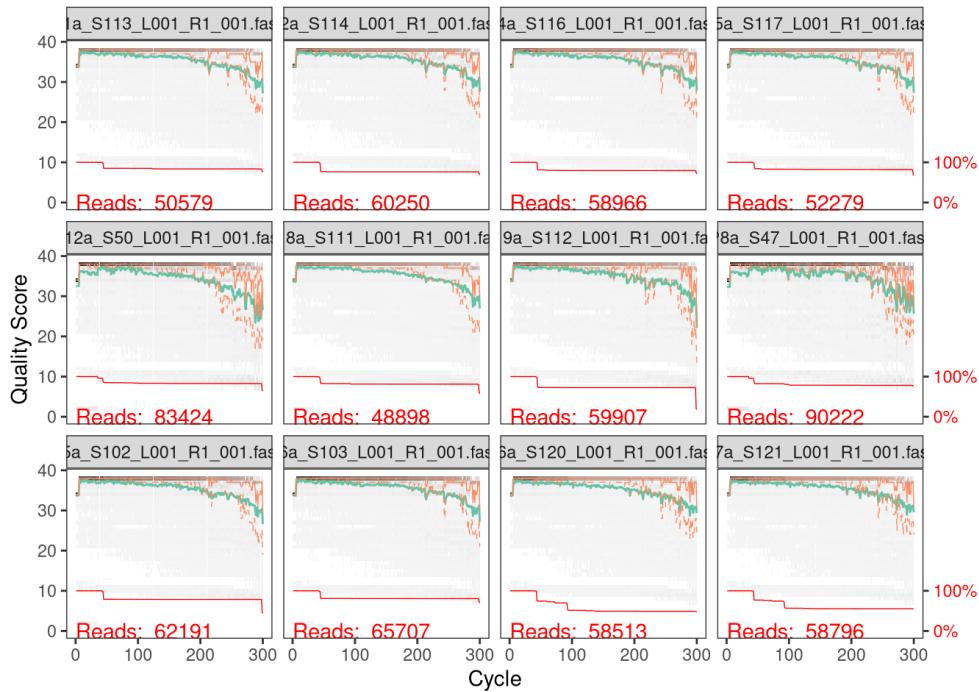
où  $Y$  représente les distances entre les échantillons,  $X_i$  est le prédicteur (Marchand, 2021),  $\alpha_i$  les effets de ces facteurs,  $\varepsilon$  la variabilité aléatoire et p le nombre total de variables explicatives.

Ensuite, le test PERMANOVA a été refait, mais cette fois pour un sous-échantillon incluant uniquement les 100 ASVs les plus abondants.

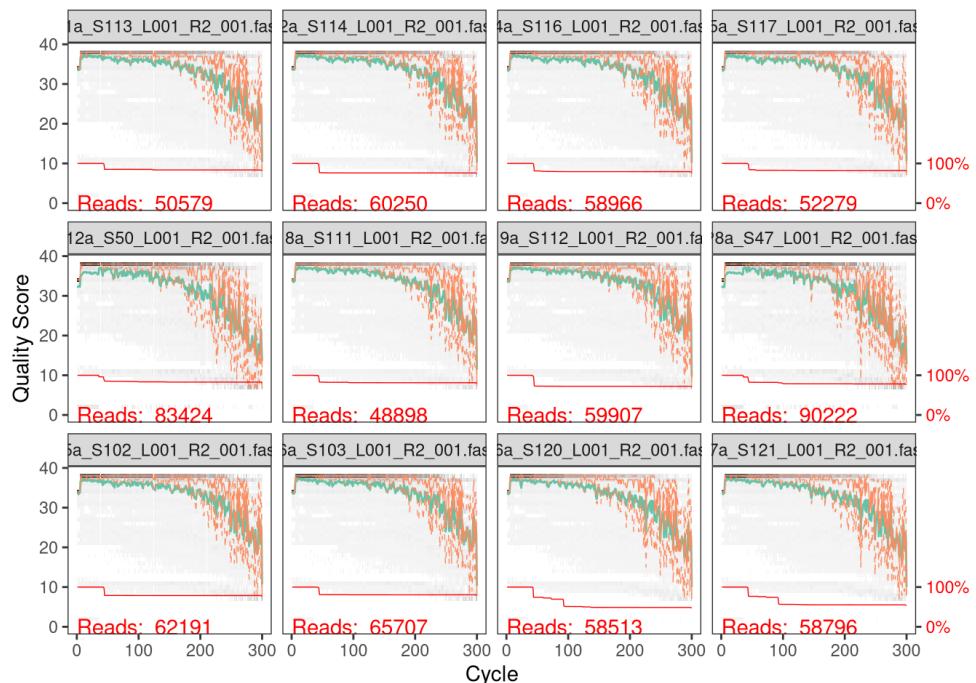


## Résultats

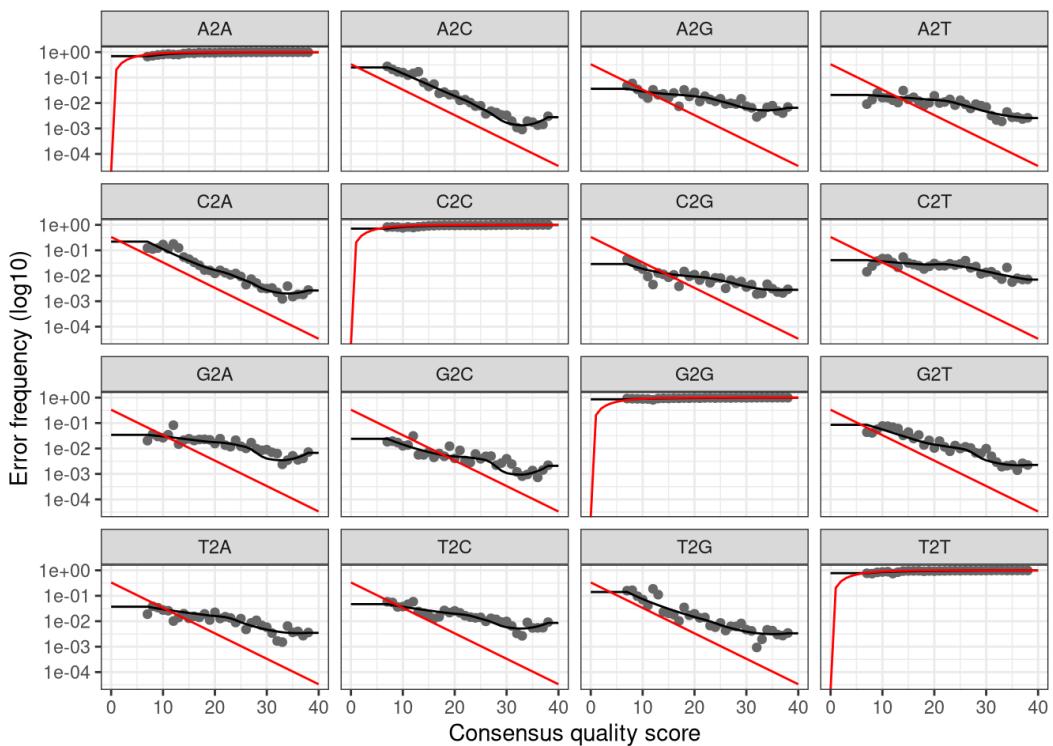
**Figure 1 :** Visualisation de la qualité des séquences dans le sens de lecture direct (*forward*) pour la population des quatres saumons.



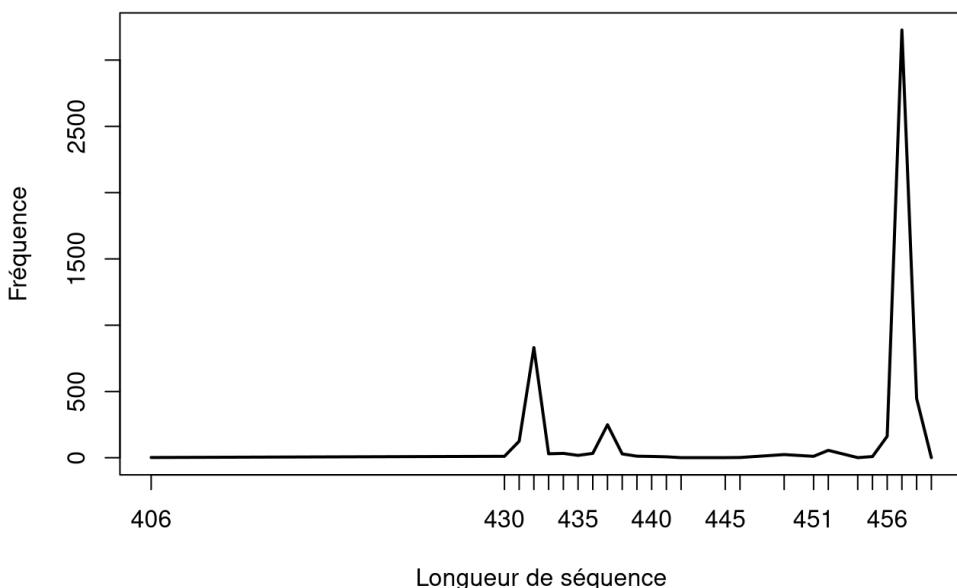
**Figure 2 :** Visualisation de la qualité des séquences dans le sens de lecture complémentaire (*reverse*) pour la population des quatres saumons.



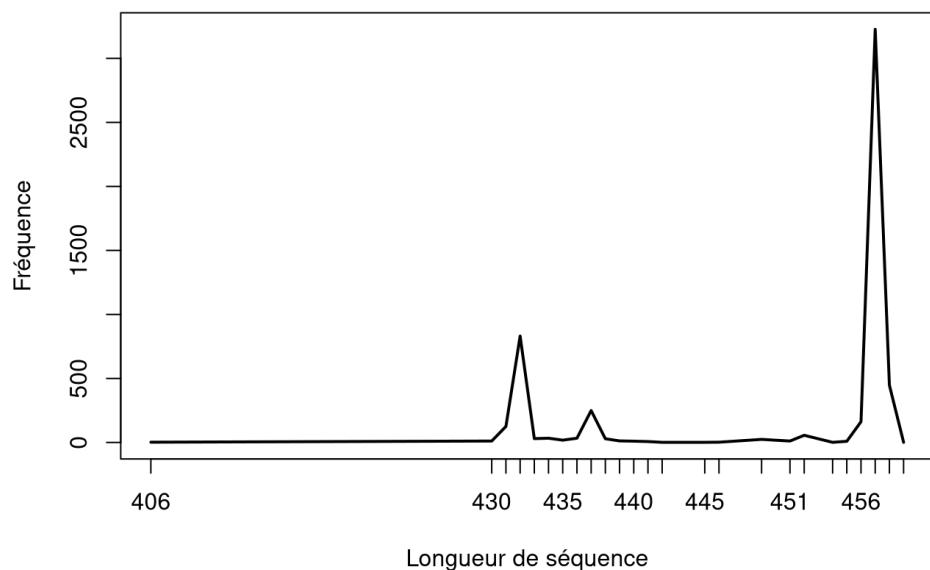
**Figure 3 :** Visualisation des modèles d'erreurs pour les séquences dans le sens de lecture direct (*forward*) des quatres populations de saumons.



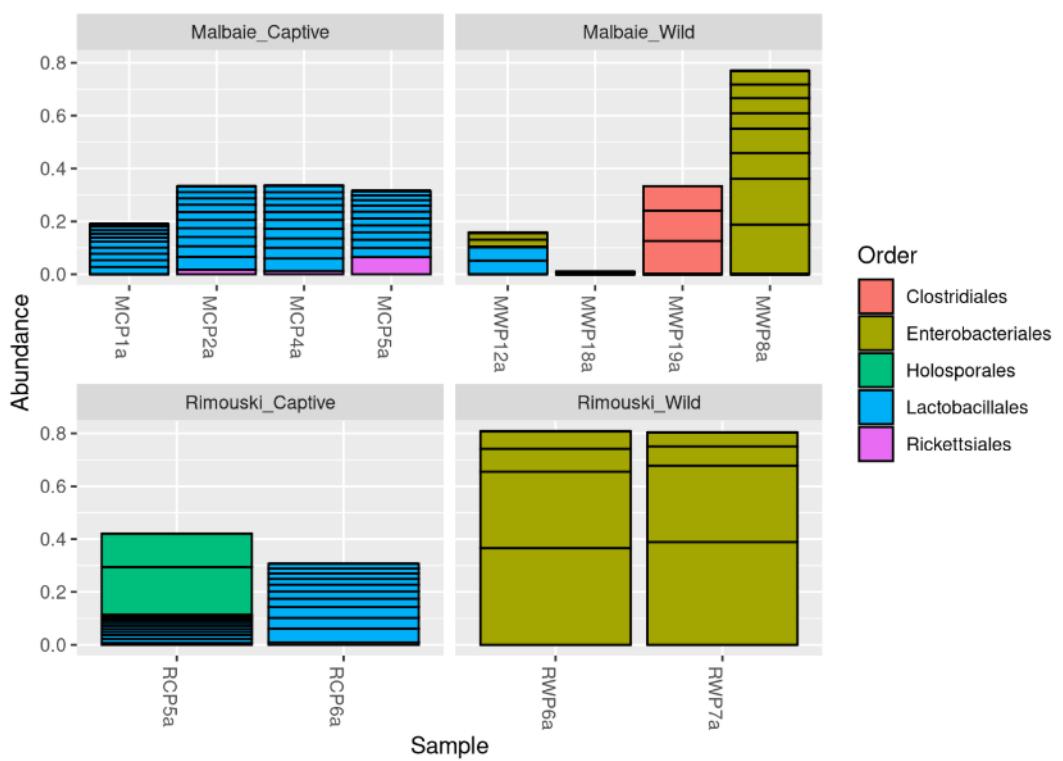
**Figure 4 :** Visualisation du tableau d'abondance avant suppression des chimères.



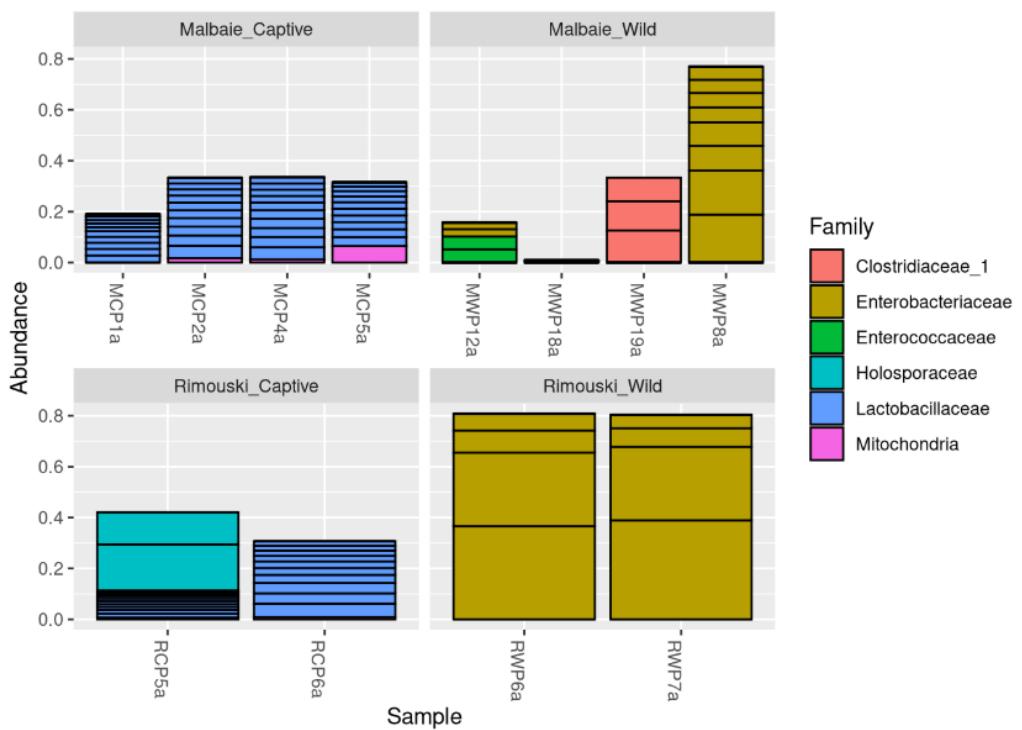
**Figure 5 :** Visualisation du tableau d'abondance après la suppression des chimères.



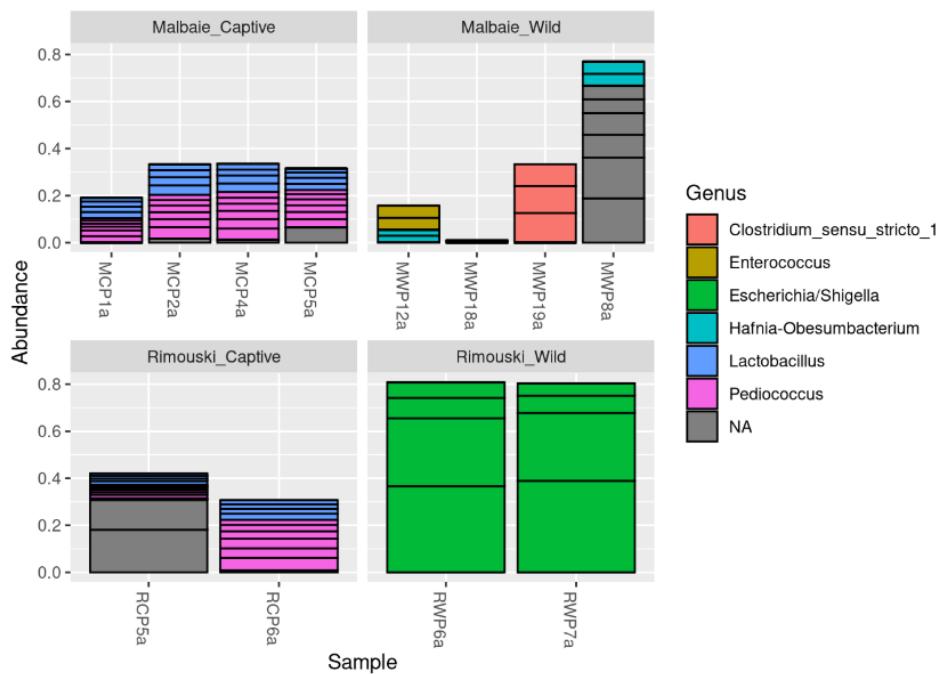
**Figure 6 :** Abondance relative des 30 ASVs/OTUs les plus abondants par ordre taxonomique pour les quatre populations de saumons.



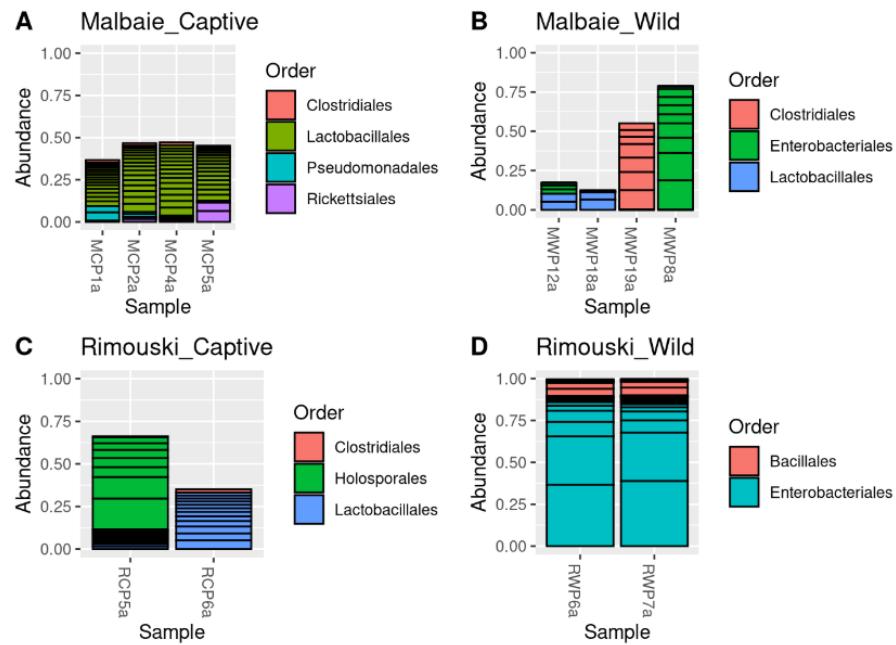
**Figure 7 :** Abondance relative des 30 ASVs/OTUs les plus abondants par famille taxonomique pour les quatre populations de saumons.



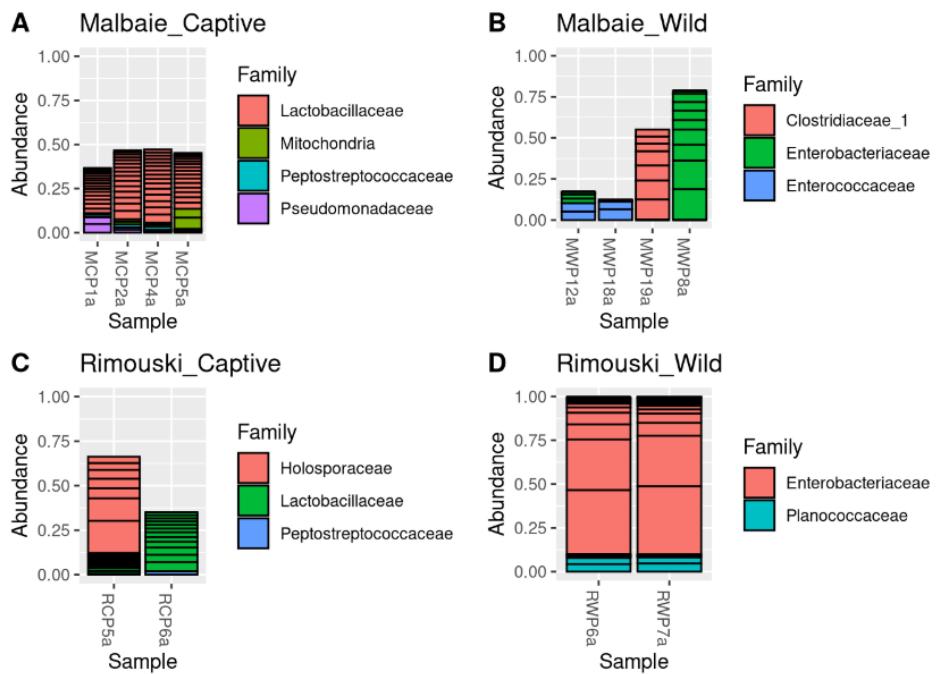
**Figure 8 :** Abondance relative des 30 ASVs/OTUs les plus abondants par genre taxonomique pour les quatre populations de saumons.



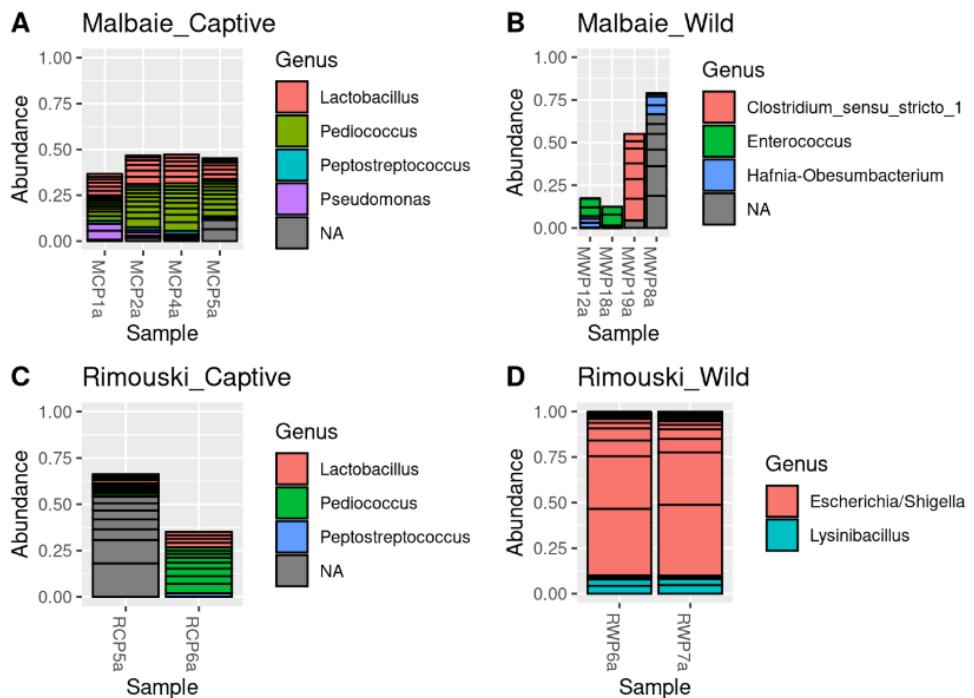
**Figure 9 :** Répartition des 20 ASVs/OTUs les plus abondants par ordre taxonomique pour chaque population de saumons.



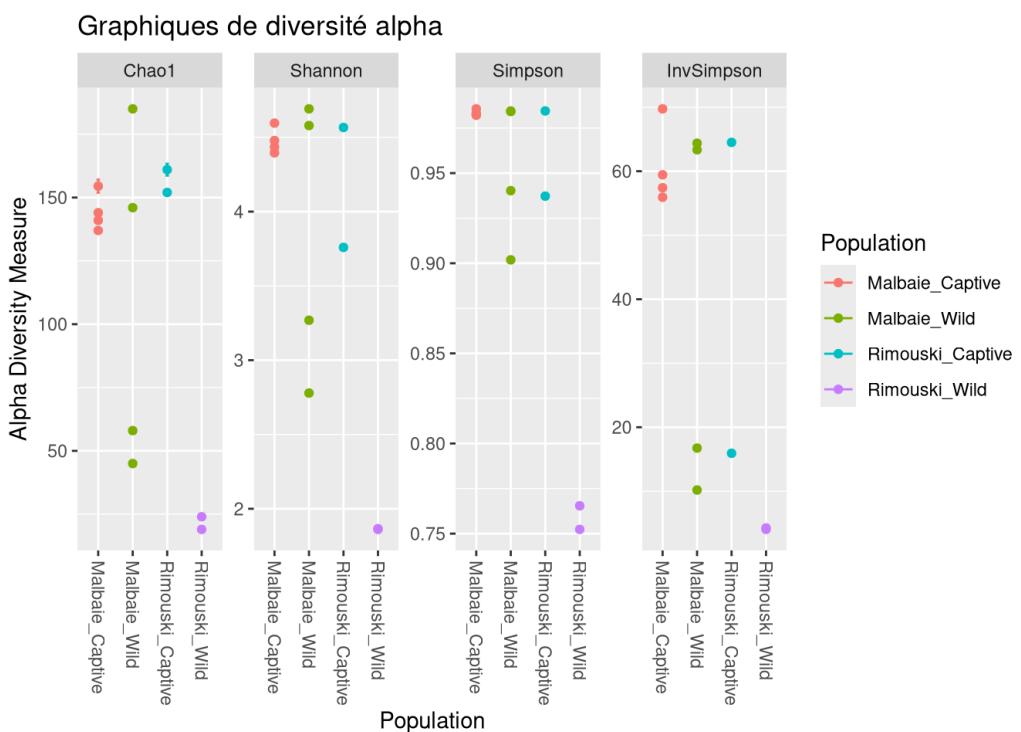
**Figure 10 :** Répartition des 20 ASVs/OTUs les plus abondants par famille taxonomique pour chaque population de saumons.



**Figure 11 :** Répartition des 20 ASVs/OTUs les plus abondants par genre taxonomique pour chaque population de saumons.



**Figure 12 :** Visualisation de la diversité alpha par population de saumon.



**Tableau 1** : Test statistique de Shapiro-Wilk (normalité) pour 4 indices de diversité alpha.

Indices de diversité alpha	W	P-value
Chaos1	0,82192	0,01681
Shannon	0,78564	0,006462
InvSimpson	0,79131	0,007475
Fisher	0,76709	0,004055

**Tableau 2** : Test statistique de Levene (homogénéité des variances) pour 4 indices de diversité alpha.

Indices de diversité alpha	Statistique F	P-value
Chaos1	18,172	6,252e <sup>-04</sup>
Shannon	21,974	3,226e <sup>-04</sup>
InvSimpson	38,784	4,106e <sup>-05</sup>
Fisher	37,075	4,853e <sup>-05</sup>

**Tableau 3** : Test statistique de Kruskal-Wallis pour 4 indices de diversité alpha.

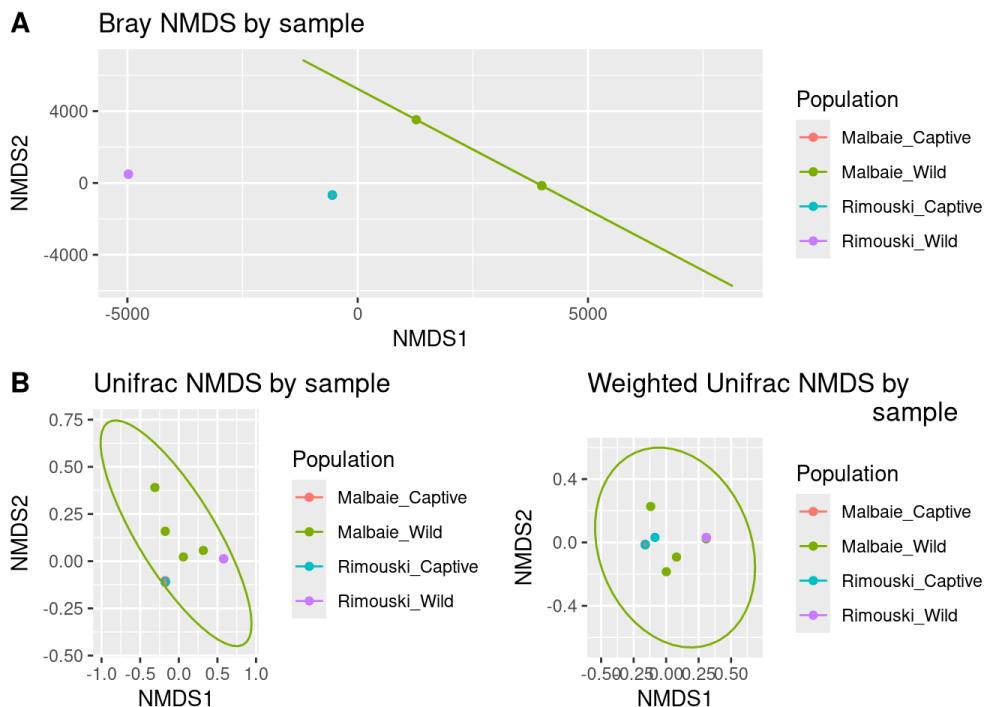
Indices de diversité alpha	$\chi^2$	df	P-value
Chao1	5,8269	3	0,1203
Shannon	4,7500	3	0,1910
InvShannon	4,9615	3	0,1746
Fisher	5,4231	3	0,1433

**Tableau 4** : Test statistique de Wilcoxon pour 4 indices de diversité alpha.

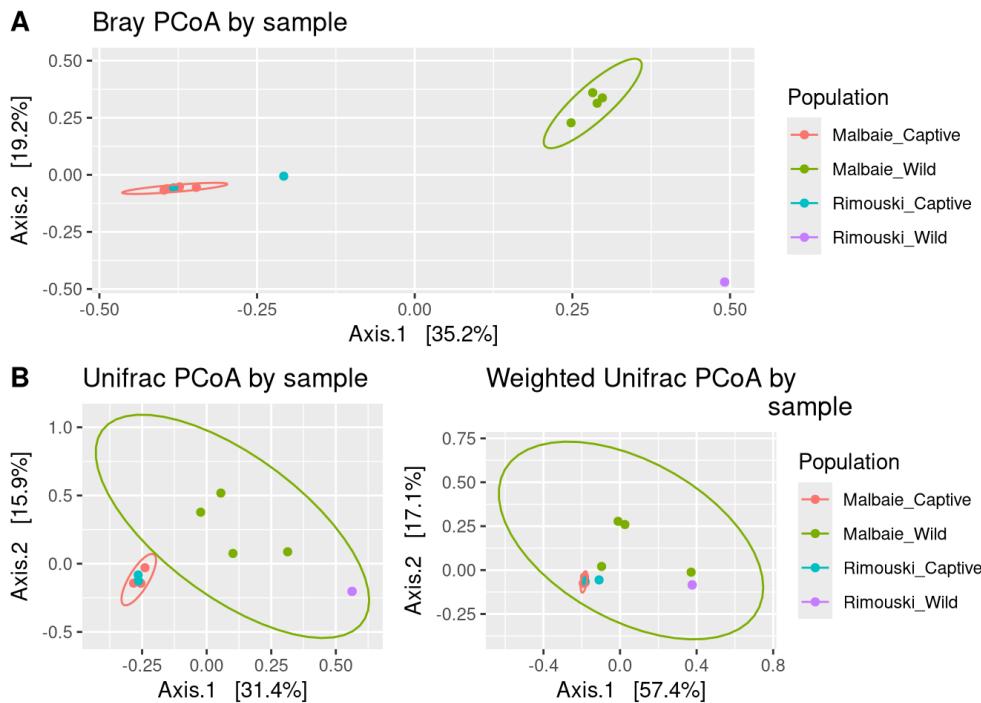
Chao1			
	Malbaie_Captive	Malbaie_Wild	Rimouski_Captive
Malbaie_Wild	0,89	-	-
Rimouski_Captive	0,50	0,64	-
Rimouski_Wild	0,40	0,40	0,50
Shannon			
	Malbaie_Captive	Malbaie_Wild	Rimouski_Captive
Malbaie_Wild	1,00	-	-
Rimouski_Captive	1,00	1,00	-

Rimouski_Wild	0,40	0,40	0,67
InvSimpson			
	Malbaie_Captive	Malbaie_Wild	Rimouski_Captive
Malbaie_Wild	0,80	-	-
Rimouski_Captive	0,80	0,80	-
Rimouski_Wild	0,40	0,40	0,67
Fisher			
	Malbaie_Captive	Malbaie_Wild	Rimouski_Captive
Malbaie_Wild	0,89	-	-
Rimouski_Captive	0,64	0,64	-
Rimouski_Wild	0,40	0,40	0,64

**Figure 13 :** Représentation NMDS des échantillons basée sur la distance de Bray-Curtis, Unifrac et Unifrac pondérée pour la population des 4 saumons.



**Figure 14 :** Représentation PCoA des échantillons basée sur la distance de Bray-Curtis, Unifrac et Unifrac pondérée pour la population des 4 saumons.



**Tableau 5 :** Test statistique Permanova sur les matrices de diversité bêta (ADONIS) basées sur la distance de Unifrac.

Source	Df	Somme des carrés	Carrés moyens	Statistique F	R <sup>2</sup>	P-value
Population	3	1,8766	0,62554	2,6817	0,5014	0,001998
Résidus	8	1,8661	0,23327	-	0,4986	-
Total	11	3,7427	-	-	1,0000	-

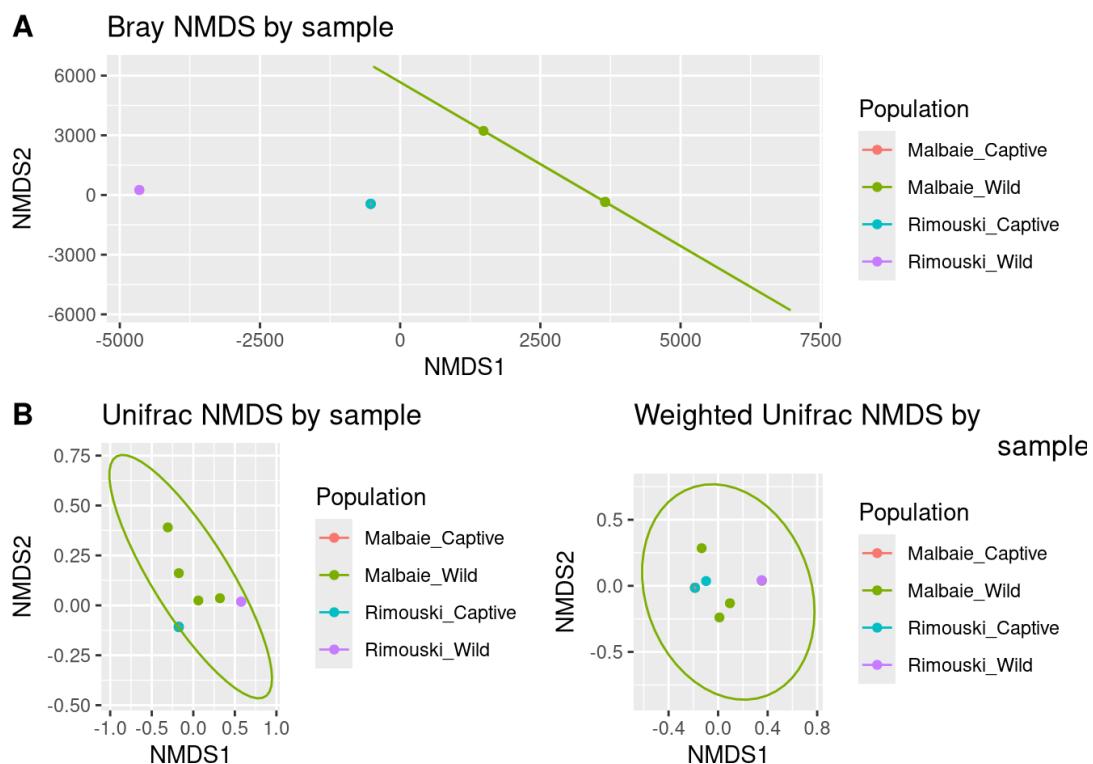
**Tableau 6 :** Test statistique Permanova sur les matrices de diversité bêta (ADONIS) basées sur la distance de Unifrac pondérée.

Source	Df	Somme des carrés	Carrés moyens	Statistique F	R <sup>2</sup>	P-value
Population	3	0,38256	0,12751	4,1798	0,6105	0,001998
Résidus	8	0,24407	0,03051	-	0,3895	-
Total	11	0,62662	-	-	1,0000	-

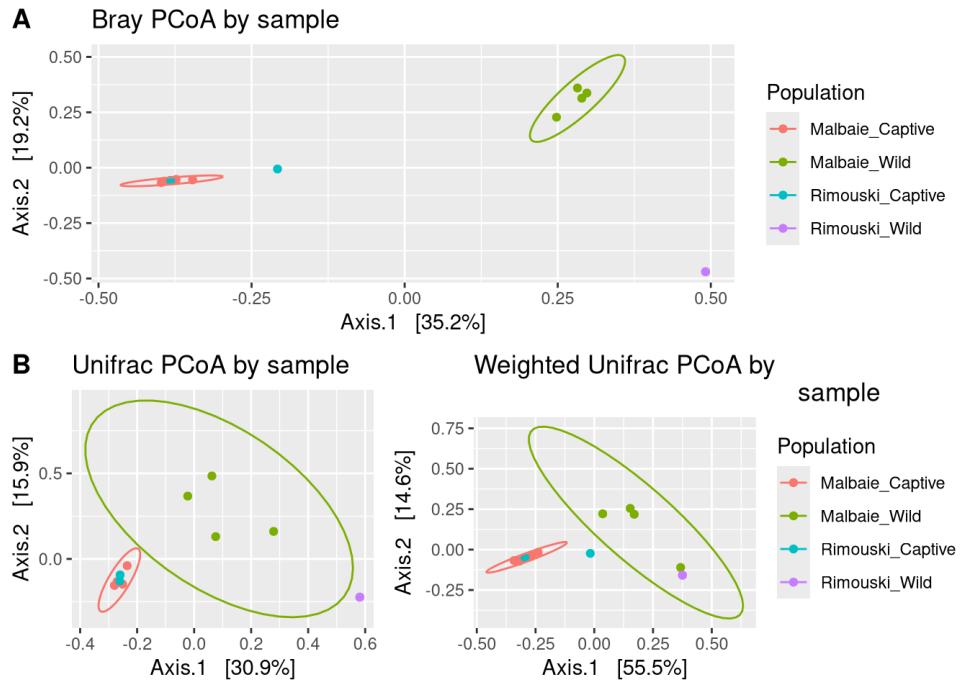
**Tableau 7** : Test statistique Permanova sur les matrices de diversité bêta (ADONIS) basées sur la distance de Bray-Curtis.

Source	Df	Somme des carrés	Carrés moyens	Statistique F	R <sup>2</sup>	P-value
Population	3	2,5707	0,85689	3,6923	0,58064	0,000999
Résidus	8	1,8566	0,23208	-	0,41936	-
Total	11	4,4273	-	-	1,00000	-

**Figure 15** : Représentation NMDS des échantillons basée sur la distance de Bray-Curtis, Unifrac et Unifrac pondérée pour la population des 4 saumons des 100 ASVs les plus abondants.



**Figure 16 :** Représentation PCoA des échantillons basée sur la distance de Bray-Curtis, Unifrac et Unifrac pondérée pour la population des 4 saumons des 100 ASVs les plus abondants.



**Tableau 8 :** Test statistique Permanova sur les matrices de diversité bêta (ADONIS) basées sur la distance de Unifrac des 100 ASVs les plus abondants.

Source	Df	Somme des carrés	Carrés moyens	Statistique F	R <sup>2</sup>	P-value
Population	3	2,05227	0,68409	7,3684	0,73426	0,001998
Résidus	8	0,74273	0,09284	-	0,26574	-
Total	11	2,79500	-	-	1,00000	-

**Tableau 9 :** Test statistique Permanova sur les matrices de diversité bêta (ADONIS) basées sur la distance de Unifrac pondérée des 100 ASVs les plus abondants.

Source	Df	Somme des carrés	Carrés moyens	Statistique F	R <sup>2</sup>	P-value
Population	3	1,39448	0,46483	4,8978	0,64747	0,004995
Résidus	8	0,75924	0,09491	-	0,35253	-
Total	11	2,15373	-	-	1,00000	-

**Tableau 10 :** Test statistique Permanova sur les matrices de diversité bêta (ADONIS) basées sur la distance de Bray-Curtis des 100 ASVs les plus abondants.

Source	Df	Somme des carrés	Carrés moyens	Statistique F	R <sup>2</sup>	P-value

Population	3	2,6841	0,89471	4,5011	0,62796	0,000999
Résidus	8	1,5902	0,19878	-	0,37204	-
Total	11	4,2744	-	-	1,00000	-

## Discussion

### Partie 1 :

L'objectif à l'étape du contrôle qualité fut d'établir un seuil de qualité permettant d'obtenir des séquences de longueur relativement similaire à celle des lectures originales. Un score phred de 20 à 25 a ainsi été choisi comme seuil, ce qui correspond à une longueur de troncation de 290 nucléotides pour les lectures avant et de 270 nucléotides pour les lectures arrière (**Figures 1-2**). La majeure partie des lectures originale a été conservée étant donné que celles-ci font toutes 300 nucléotides de long. Les lectures comportant plus de 2 erreurs attendues après l'étape de filtration ont également été retirés du jeu de données, et ce tant pour les lectures avant que les lectures arrière. L'inférence des variants d'amplicon a ensuite été réalisée à partir de taux d'erreurs estimés par la fonction learnErrors (**Figure 3**) et l'annotation taxonomique a été réalisée à partir de la base de données Silva. Le reste de cette étape de l'analyse s'est fait avec les paramètres par défaut offerts par dada2.

### Partie 2 :

La visualisation des 30 ASVs les plus abondants a été réalisée à trois niveaux taxonomiques, soit l'ordre, la famille et le genre, à partir du jeu de données complet ainsi qu'à partir de sous-échantillons correspondant aux quatre populations de saumons analysées (Malbaie\_Captive, Malbaie\_Wild, Rimouski\_Captive, Rimouski\_Wild).

L'analyse globale (**Figures 6-7-8**) met en évidence certains taxons dominants, mais masque les différences spécifiques entre populations. À l'échelle de l'ordre, *Lactobacillales* domine à Malbaie\_Captive, alors que *Clostridiales* et *Enterobacteriales* sont majoritaires à Malbaie\_Wild, et que Rimouski\_Wild est presque exclusivement dominé par *Enterobacteriales*. Rimouski\_Captive, quant à lui, présente un mélange de *Holosporales* et de *Lactobacillales*.

Au niveau de la famille, on observe la prédominance de *Lactobacillaceae* à Malbaie\_Captive, une co-dominance de *Clostridiaceae\_1* et *Enterobacteriaceae* à Malbaie\_Wild, une combinaison de *Holosporaceae* et *Lactobacillaceae* à Rimouski\_Captive, et une domination quasi exclusive de *Enterobacteriaceae* à Rimouski\_Wild.

Enfin, au niveau du genre, *Lactobacillus* et *Pediococcus* sont les plus fréquents à Malbaie\_Captive, tandis que *Clostridium sensu stricto 1*, *Enterococcus* et des genres non assignés (*NA*) sont abondants à Malbaie\_Wild. Rimouski\_Captive montre un mélange de *Lactobacillus*, *Pediococcus* et des *NA*, alors que Rimouski\_Wild est dominé par *Escherichia/Shigella* (~90 %).

Lorsque les mêmes données sont visualisées par population (**Figures 9-10-11**), les différences apparaissent plus clairement. Malbaie\_Captive se distingue par une plus grande diversité de genres potentiellement bénéfiques, tandis que Malbaie\_Wild est dominée par des taxons fermentaires. Rimouski\_Captive présente une composition mixte, et Rimouski\_Wild est caractérisé par une très faible diversité, avec une prédominance d'un seul taxon.

Ces résultats démontrent qu'il existe une différence significative entre les analyses effectuées à partir du jeu complet et celles faites à partir des sous-échantillons : l'approche globale tend à lisser les variations entre populations, ce qui peut masquer des profils microbiens spécifiques et importants sur le plan biologique.

Par ailleurs, un effet environnemental clair est observé entre les sites de Malbaie et de Rimouski. Les saumons provenant de Malbaie, surtout en captivité, présentent une diversité microbienne plus riche et équilibrée, incluant des genres associés à une flore intestinale saine. En revanche, ceux de Rimouski, en particulier les individus sauvages, présentent un microbiote dominé par *Escherichia/Shigella*, un genre souvent associé à des déséquilibres intestinaux ou à une qualité environnementale inférieure.

Ces différences suggèrent que l'environnement local a une influence marquée sur la structure des communautés microbiennes intestinales, ce qui peut avoir des répercussions importantes sur la santé des poissons et sur les pratiques aquacoles.

### Partie 3 :

Le graphique de diversité alpha (**Figure 12**) permet d'observer différents tests, qui nous indique des informations sur les différentes populations comme leurs richesses et leurs équitabilités, on observe par exemple avec le test de Chao1 une estimation du nombre d'espèce non observé, ce qui communique de l'information sur le nombre d'espèce rare et donc de la richesse de la population. Avec cette information on est capable de voir que les populations de Malbaie\_Captive et de Rimouski\_Captive ont une très grande richesse pour tous les échantillons, alors que Malbaie\_Wild a une très grande disparité entre ses échantillons ce qui

nous empêche d'avoir une bonne conclusion. Finalement la population de Rimouski\_Wild a un score très bas, ce qui indique une richesse minimale avec peu de taxon rare.

On peut ensuite observer les tests de Shannon et de Simpson qui nous communiquent des informations très similaires, soit l'équitableté, cela nous permet donc de savoir si chaque espèce de la population contient un nombre similaire d'individu. On observe donc ici que Malbaie\_Captive à la plus haute équitableté avec un score de Simpson très proche de 1 et un score de Shannon proche de 5. On a ensuite Rimouski\_Captive qui a aussi un très haut score dans les deux tests, mais avec une plus grande distance entre les échantillons, ce qui nous empêche de bien conclure. Cet effet est encore plus présent dans la population de Malbaie\_Wild il est très difficile de voir l'équitableté à partir de ce graphique. On a finalement Rimouski\_Wild qui a des scores très bas ce qui indique que la population contient probablement une espèce en très grande majorité.

En conclusion on observe que les populations en captivité ont presque toujours une plus grande richesse et une meilleure équitableté, que les populations sauvages. On observe aussi que les populations de la Malbaie ont une meilleure diversité que celle de la Rimouski.

Finalement pour le choix de nos mesures de diversité et de tests statistiques, le test de Shapiro-Wilk, qui nous donne la normalité (**Tableau 1**), les valeurs de W ne sont pas assez proches de 1 pour conclure une normalité, mais les P-value sont en-dessous de 0,05, donc les résultats sont valides. Le test de Levene, qui nous donne l'homogénéité des variances (**Tableau 2**) nous donne des F trop élevés, ce qui indique de trop grande variabilité et avec les P-value en-dessous de 0,05 ça confirme que l'homogénéité des variances n'est pas atteinte. Ensuite le test statistique de Kruskal-Wallis (**Figure 3**) où les P-value sont au-dessus de 0,05, donc l'hypothèse nulle n'est pas rejetée, mais les scores de khi-deux sont en-dessous de la valeur critique dans le tableau des valeurs critiques des khi-deux, donc ils sont acceptables. On a finalement le test de Wilcoxon (**Figure 4**) qui permet de voir la différence de tendance centrale entre les différentes populations pour chaque test de diversité alpha. Nos choix de tests finals sont donc qu'il serait préférable de soit prendre le test de Kruskal-Wallis ou celui de Wilcoxon.

#### **Partie 4 :**

On a ensuite les tests de diversité Beta (**Figures 13-14**), qui permettent de voir la similarité entre les différentes populations. La figure 13, qui compare Unifrac pondéré et non-pondéré, permet d'observer le nombre d'OTUs/ASVs partagés entre les communautés et leurs degrés d'apparentement. S'ils sont proches dans le graphique, cela indique un grand nombre

d'OTUs/ASVs en commun et un grand degré d'appartenance. Sans pondération, les populations de Malbaie\_Captive et de Rimouski\_Captive sont identiques. Malbaie\_Wild est très éparpillée mais est proche de toutes les populations, tandis que Rimouski\_Wild est loin des deux populations captives tout en étant assez proche de Malbaie\_Wild. Lorsque Unifrac est pondéré, les résultats restent similaires sauf le fait que Malbaie\_Wild est un peu plus éloigné de Rimouski\_Wild et plus proche des deux autres populations. Avec la représentation de PCoA les résultats restent similaires avec une plus expansion dans chaque population, mais particulièrement celle de Malbaie\_Wild.

La conclusion sur les populations est donc que les populations captives sont beaucoup plus similaires entre elle-même que les populations sauvages et les populations de Rimouski sont très différente les unes des autres, alors que celle de Malbaie ne sont pas similaire, mais pas trop loin.

Ensuite, les tests de Permanova nous permettent de choisir quelle distance le plus approprier à notre étude (**Tableaux 5-6-7**). Les trois distances ont une P-value plus basse de 0,05, donc nous rejetons l'hypothèse nul pour les trois distances. Il faut ensuite observer la statistique F, qui nous informe de la signification des résultats. À partir de cela il est possible de conclure que la distance de Unifrac pondérée est la plus significative.

## Partie 5

Afin d'évaluer l'influence des ASVs dominants sur la structuration du microbiote, une nouvelle analyse a été effectuée à partir d'un jeu de données restreint aux 100 ASVs les plus abondants (Figure 15 et 16, Tableau 8 à 10). Cette approche visait à déterminer si la structure observée dans les analyses précédentes était principalement portée par les taxons les plus fréquents, ou si les taxons rares contribuaient également à la différenciation entre les populations.

Les distances tests PERMANOVA réalisés sur les distances de Unifrac, Unifrac pondérée et Bray-Curtis ont tous révélé des différences significatives entre les populations ( $p < 0,005$ ) avec des valeurs de  $R^2$  élevées (jusqu'à 73,4% pour Unifrac). Ces résultats suggèrent que les 100 ASVs les plus abondant suffisent à expliquer une part majeure de la variabilité observée entre les groupes, et que les conclusions précédentes demeurent robustes, même en l'absence des ASVs rares.

Néanmoins, l'approche ciblée sur les ASVs les plus fréquents a pour effet de réduire la complexité écologique du microbiote, en excluant des séquences potentiellement rares mais biologiquement pertinentes. Par exemple, certains genres peu représentés globalement, mais spécifiques à un échantillon ou à une condition (p. Ex. *Holosporaceae* à Rimouski\_Captive), peuvent être omis. Ainsi, bien que les tendances générales soient conservées, l'analyse complète reste préférable pour détecter les signatures microbiennes fines et les effets subtils de l'environnement ou de la condition de vie.

## Conclusion

L'ensemble des analyses menées dans cette étude permet de conclure que la composition du microbiote intestinal chez le saumon atlantique est fortement influencée par trois facteurs principaux : la population d'origine, l'environnement géographique (Malbaie vs Rimouski) et la situation de vie (Captif vs Sauvage).

Les analyses de diversité alpha ont montré une tendance à une diversité plus élevée chez les poissons captifs, en particulier ceux de Malbaie, bien que cette différence ne soit pas significative statistiquement. En revanche, les résultats de diversité beta (analyses NMDS, PCoA et PERMANOVA) ont mis en évidence des différences marquées entre les groupes, confirmant l'importance de la structure taxonomique pour discriminer les populations.

Les analyses taxonomiques ont montré que:

- Les saumons captifs de Malbaie possèdent une communauté microbienne diverse et équilibrée, dominée par des genres bénéfiques tels que *Lactobacillus* et *Pediococcus*.
- Les saumons sauvages de Rimouski présentent une très faible diversité, avec une domination quasi exclusive d'*Escherichia/shigella*, un profil potentiellement associé à une altération de la santé intestinale ou à des conditions environnementales moins favorables.

L'effet de l'environnement géographique s'avère donc majeur. Les différences observées entre Malbaie et Rimouski, même pour les poissons dans les conditions de captivité similaires, indiquent que le milieu local influence fortement la structuration du microbiote, indépendamment du statut sauvage ou captif.

Enfin, la comparaison avec les 100 ASVs les plus abondant démontre que ces derniers suffisent à capturer la structure globale des communautés microbiennes. Toutefois, l'analyse

complète, incluant les ASVs rares, demeure essentielle pour détecter des signatures microbiennes spécifiques et pour mieux comprendre la complexité du microbiote intestinal.

Ces résultats ont des implications importantes en aquaculture, notamment pour le suivi de la santé intestinale des saumons et pour le développement de stratégies de gestion personnalisées selon le site de production. Ces observations soulignent l'importance de considérer le microbiote comme un biomarqueur potentiel de la qualité de l'environnement aquatique. Elles encouragent également l'intégration d'analyses microbiologiques dans les pratiques de gestion durable des élevages de saumons, afin d'optimiser la santé des poissons et la performance des sites d'élevage.

## Bibliographie

Callahan, B., McMurdie, P., & Holmes, S. (2025). *DADA2: Fast and accurate sample inference from amplicon data with single-nucleotide resolution* (1.34.0) [Software]. <https://benjneb.github.io/dada2/index.html>

Caporaso, J. G. et al. (2010). *QIIME allows analysis of high-throughput community sequencing data*. Nature Methods.

DATatab Team. (2025a). *ANOVA à deux facteurs (sans répétition)*. DATatab: Online Statistics Calculator. <https://datatab.fr/tutorial/two-factorial-anova-without-repeated-measures>

DATatab Team. (2025b). *Test de Kruskal-Wallis*. DATatab: Online Statistics Calculator. <https://datatab.fr/tutorial/kruskal-wallis-test>

Gahan, J. M. et al. (2024). *Genomic and functional characterization of the Atlantic salmon gut microbiome in relation to nutrition and health*. Nature Microbiology. <https://www.nature.com/articles/s41564-024-01830-7>

J. McMurdie, P., Holmes, S., Jordan, G., & Chamberlain, S. (2025). *phyloseq* [Software]. In *Bioconductor* (1.50.0). <https://www.bioconductor.org/packages/release/bioc/html/phyloseq.html>

Lynch, D. (n.d.). *Alpha Diversity*. One Codex Docs. <https://docs.onecodex.com/en/articles/4136553-alpha-diversity>

Marchand, P. (2021, October 12). *Régression linéaire multiple*. [https://pmarchand1.github.io/ECL7102/notes\\_cours/7-Regression\\_multiple.html](https://pmarchand1.github.io/ECL7102/notes_cours/7-Regression_multiple.html)

National Institute of Standards and Technology [NIST]. (n.d.). *7.4.3.2. The one-way ANOVA model and assumptions*. NIST - Engineering Statistics Handbook.

<https://www.itl.nist.gov/div898/handbook/prc/section4/prc432.htm#:~:text=The%20mathematical%20model%20that%20describes,%2C%20%E2%80%A6%20%2C%20k%20levels>

Quast, C. et al. (2013). *The SILVA ribosomal RNA gene database project: improved data processing and web-based tools*. Nucleic Acids Research.

<https://academic.oup.com/nar/article/41/D1/D590/1067993>

Schmidt, V. T. et al. (2024). *A meta-analysis revealing the technical, environmental, and host-associated factors influencing the gut microbiota of salmon and trout*. Reviews in Aquaculture. <https://onlinelibrary.wiley.com/doi/full/10.1111/raq.12913>

SILVA. (2024). Silva [Software]. In *SILVA* (138.2). <https://www.arb-silva.de/>

Styan, G. P. H., †1, McGill University, Puntanen, S., & University of Tampere. (2005). *FISHER'S ALPHA INDEX OF BIODIVERSITY: 1943–2005* [Journal-article]. <https://jhanley.biostat.mcgill.ca/IBC2006/Abstracts/pdf/453.pdf>

Yarza, P. et al. (2014). *Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences*. Nature Reviews Microbiology.