

Université de Montréal

**Utilisations du microbiome en sciences judiciaires**

*Par*

Audrey-Anne Leblanc

Département de sciences biologiques

Faculté des arts et des sciences

Mémoire présenté en vue de l'obtention du grade de maîtrise ès sciences (M.Sc.)

en sciences biologiques

Août 2023

© Audrey-Anne Leblanc, 2023

Université de Montréal

Département de sciences biologiques, Faculté des arts et des sciences

---

*Ce mémoire intitulé*

**Utilisations du microbiome en sciences judiciaires**

*Présenté par*

**Audrey-Anne Leblanc**

*A été évalué(e) par un jury composé des personnes suivantes*

**Matthew Regan**

Président-rapporteur

**François-Joseph Lapointe**

Directeur de recherche

**Shari Forbes**

Co-directrice

**Joseph Orkin**

Membre du jury

## Résumé

La microbiologie judiciaire consiste à utiliser des microorganismes comme éléments de preuves dans une enquête judiciaire. Elle est appliquée dans plusieurs buts, dont l'identification de fluides biologiques et la compréhension des mécanismes de décomposition des cadavres entre autres. L'objectif de ce mémoire est de tester l'application de nouvelles méthodes de prédictions utilisant la microbiologie judiciaire et d'évaluer leur potentiel pour le domaine.

Dans un premier temps, au chapitre 2, nous avons étudié une nouvelle méthode pour identifier les fluides vaginaux par le microbiome. Alors que plusieurs méthodes existent pour identifier des sources pures de fluides vaginaux, nous avons testé un modèle sur des mélanges de fluides pour voir s'il était possible de reconnaître la présence de sources vaginales dans ceux-ci. Nos résultats montrent que le modèle utilisé avec le seuil choisi a une spécificité de 100% en ce qui a trait à l'identification de fluides vaginaux dans des mélanges et une sensibilité de 70%. Ainsi, nos résultats démontrent qu'il serait possible d'utiliser éventuellement cette méthode en sciences judiciaires.

Dans un deuxième temps, au chapitre 3, nous avons exploré le changement du microbiome épiniécrotique sur une période de sept mois comprenant un hiver complet. Nous avons pu identifier les taxons spécifiques qui ont une différence d'abondance significative entre les saisons et ainsi développer un modèle de prédiction qui peut déterminer si un corps a passé l'hiver ou non. Notre modèle n'a jamais prédit qu'un corps n'avait pas passé l'hiver quand en réalité c'était le cas, mais a parfois surestimé les résultats.

**Mots-clés :** Microbiologie judiciaire, Sciences judiciaires, Fluide vaginal, Microbiome, Microbiome épiniécrotique, Décomposition humaine, Nérobiome

## **Abstract**

Microbial forensics uses microorganisms as physical evidence in different investigations. It is used in multiple ways, such as the identification of body fluids or the understanding of the different mechanisms of body decomposition. The objectives of this research are to apply novel prediction methods using microbial forensics and assess their potential for forensic science.

First, in chapter 2, we studied a novel method to identify vaginal fluids using the microbiome. Similar methods already exist to identify pure sources of body fluids, but we tested a model that could find vaginal fluids in mixture samples. Our results show that the model used with the chosen threshold has a specificity of 100% and a sensitivity of 70% for the identification of vaginal fluids in mixtures. Therefore, our results show that it would be possible to use this method in the future in forensic science.

Second, in chapter 3, we explored the change of the epinecrotic microbiome over a 7-months period including a full winter. We could identify specific taxa that were differentially abundant before and after winter and developed a prediction model that would calculate if a body went through winter or not. The model never predicted that a body did not go through winter when it did, but it sometimes overestimated the results predicting it went through winter when it did not.

**Keywords:** Microbial forensic, Forensic science, Vaginal fluids, Microbiome, Epinecrotic microbiome, Human decomposition, Necrobiome

## Table des matières

Résumé.....	1
Abstract .....	2
Table des matières .....	3
Liste des tableaux.....	5
Liste des figures.....	6
Liste des sigles et abréviations.....	7
Remerciements .....	8
Chapitre 1 – Introduction .....	9
1.1 La microbiologie judiciaire .....	9
1.2 Identification des sécrétions biologiques par le microbiome .....	10
1.2.1 Le microbiome humain .....	10
1.2.2 Méthodes existantes pour l’identification des sécrétions biologiques .....	11
1.2.3 Identification des sécrétions biologiques par le microbiome.....	12
1.3 La science de la décomposition.....	13
1.3.1 Le microbiome et la décomposition.....	15
1.3.2 Les taxons pertinents du microbiome épinérotique.....	16
1.3.3 Conditions d’études de la décomposition .....	17
1.3.4 Le microbiome environnemental en hiver.....	17
1.4 Analyses bio-informatiques dans l’étude du microbiome .....	18
1.4.1 Bio-informatique et statistiques .....	18
1.5 Problématique.....	20
1.6 Objectifs .....	21

1.7 Organisation du mémoire .....	21
Chapitre 2 – Identification of vaginal fluids using microbiome for forensic purposes: A proof of concept.....	23
2.1 Résumé.....	24
2.2 Abstract .....	25
2.3 Introduction.....	26
2.4 Materials and Methods.....	29
2.5 Results .....	31
2.6 Discussion .....	35
2.7 Supplementary files.....	37
Chapitre 3 – Impact of winter on the epinecrotic microbiome in a humid, continental (Dfb) climate in Quebec, Canada .....	40
3.1 Résumé.....	41
3.2 Abstract .....	42
3.3 Introduction.....	43
3.4 Materials and Methods.....	45
3.5 Results .....	47
3.6 Discussion .....	53
3.7 Conclusion .....	55
3.8 Supplementary files.....	56
Chapitre 4: Conclusion .....	57
Références bibliographiques.....	59

## Liste des tableaux

<b>Table S2.1.</b> Information asked to participants with a potential to have an impact on the microbiome composition .....	37
<b>Table S2.2.</b> Participant's contribution of fluids for each sample .....	38
<b>Table 3.1</b> - PERMANOVA analysis calculating distances between samples from the three different seasons. ....	49
<b>Table 3.2.</b> - Families of bacteria detected by ANCOM to be differently abundant for each body site among seasons. ....	50

## Liste des figures

<b>Figure 2.1</b> - Principal coordinate analysis (PCoA) based on unweighted UniFrac distances generated from all 50 collected samples. The samples representing pure sources are identified by different symbols, whereas mixture samples containing two different body fluids are identified by stars (*). The 5 samples containing a mixture of vaginal fluids from two different women are considered pure sources and have been identified in the vaginal group.....	32
<b>Figure 2.2</b> - SourceTracker results of the source contribution including the 6 mixture samples that the model could not predict correctly the vaginal fluid contribution of it. These samples are identified by a red arrow.....	33
<b>Figure 2.3</b> - Relative Frequency of bacterial orders of the samples containing vaginal fluids that were never found by the predictive model from SourceTracker2.6 Discussion .....	34
<b>Figure 3.1</b> - Bacterial composition differences between seasons and between body sites. Alpha diversity differences between the different body sites sampled and between the different seasons represented by Shannon entropy. Grey dots represent outliers. The different seasons were analyzed independently of the body sites and body sites were analyzed independently of the seasons.....	48
<b>Figure 3.2</b> - PCoA visualization of unweighted UniFrac distances based on (A) sampling sites and (B) seasons.....	48
<b>Figure 3.3</b> - Relative abundance through time of the 3 of the 5 most important bacterial taxa identified by ANCOM. The 3 taxa represented here show the strongest difference between “before winter” and “after winter” .....	51
<b>Figure 3.4</b> - Random forest regression results of the season prediction model. The confusion matrix shows the proportion of samples that were rightly and wrongly predicted to have been sampled either before winter, during winter or after winter by the regression model versus the true number of samples collected for each season.....	52
<b>Figure S3.1.</b> Volatility results of <i>Bacillaceae</i> and <i>Pseudomonadaceae</i> showing the relative abundance of the 2 families of bacteria through time. ....	56



## Liste des sigles et abréviations

ANCOM : Analysis of compositions of microbes

ARN : Acide ribonucléique

ARNr : Acide ribonucléique ribosomique

ARNt : Acide ribonucléique de transfert

ASV : Amplicon sequence variant

IPM: Intervalle post-mortem

KADD: Kelvin accumulation degree day

mARN: ARN messenger

miARN: Micro ARN

OTU : Operational taxonomic unit

PCR : Polymerase chain reaction

PSA: Prostate specific antigen

REST(ES): Recherche en Sciences Thanatologiques Expérimentales et Sociales

RFC : Random forest classifier

STR : Short tandem repeat

## Remerciements

J'aimerais remercier mon directeur de recherche François-Joseph Lapointe, pour m'avoir permis de faire ce projet de recherche hors de l'ordinaire. Ce projet rêvé pour moi fût possible grâce à quelques pirouettes ici et là, mais le support a toujours été présent pour y arriver. Les nombreux conseils et le soutien furent d'une aide immense.

Merci aussi à ma co-directrice, Shari Forbes, pour sa sagesse et sa bonté tout au long du projet. Peu de gens étudient dans ce domaine et les conseils et le savoir donnés furent très éclairants, rassurants et valorisant pour moi.

J'aimerais aussi remercier tous les membres de Rest(es), votre soutien moral était précieux. Votre aide tout au long du projet fut à chaque fois présent et tellement apprécié.

Merci au personnel du laboratoire de sciences judiciaires et de médecine légale de m'avoir permis de faire ce projet avec eux et de m'avoir éclairé sur les rudiments de la réalité des sciences judiciaires appliquées.

Un gros merci à mon amie Caroline Seyer, qui malgré que pour elle, Excel soit un outil simple et rapide, pour moi ce n'est pas tant le cas. Son aide pour classer certaines de mes données fut peut-être courte, mais sans elle, le même résultat aurait pris des heures de plus pour la non-initiée que je suis.

Un immense merci à mon frère, expert en informatique qui m'a dépanné de nombreuses fois dans mes problèmes de codage bio-informatiques. Une aide précieuse qui m'a fait sauver beaucoup de temps et de quiétude mentale.

Finalement, un immense merci aux donateurs de Rest(es) et à leur famille. Ces gens ont fait don de leur corps pour aider la science et ce geste est d'une générosité immense. Grâce à eux, des scientifiques de plusieurs disciplines peuvent étudier la science de la décomposition pour pouvoir aider dans le futur la résolution d'enquêtes et aider ainsi à apporter justice aux victimes et à leurs proches.

# Chapitre 1 – Introduction

## 1.1 La microbiologie judiciaire

La microbiologie judiciaire est une science qui se développe très rapidement depuis les dernières années. Elle consiste à étudier et analyser les microorganismes pour potentiellement les utiliser comme éléments de preuve pour aider à un large éventail d'enquêtes (Budowle et al., 2014). L'importance de la microbiologie judiciaire fût mise de l'avant initialement lors de l'enquête sur les attaques à l'anthrax, un agent de bioterrorisme, en 2001 aux États-Unis (Carter et al., 2017). Depuis, cette science s'est élargie et elle n'est plus seulement employée lors d'enquêtes de bioterrorisme, mais aussi lors de biocrimes, certains cas de négligence ou encore d'agressions sexuelles (Budowle et al., 2019). Les avancées sur les connaissances du microbiome humain ont aussi permis de faire avancer la microbiologie judiciaire avec des projets de recherche sur l'identification des individus par le microbiome, la détermination de certaines causes de la mort, ou encore l'évaluation de l'intervalle post-mortem (Oliviera et Amorim, 2018). L'intervalle post-mortem (IPM) se définit comme l'intervalle de temps entre la mort d'une personne et le moment où l'on retrouve le corps (Metcalf et al., 2013). La détermination de l'IPM reste encore à ce jour difficile à déterminer de façon précise et l'ajout de nouveaux outils pour appuyer ceux déjà existants demeure essentiel (Byard, 2016). Aussi, l'identification spécifique de l'origine anatomique de fluides biologiques n'est pas toujours possible et des méthodes complémentaires pourraient faciliter le travail des biologistes judiciaires et potentiellement améliorer leur rendement. C'est ainsi que l'étude des microorganismes présents sur les pièces à conviction ainsi que le corps et les alentours de celui-ci tout au long de la décomposition pourrait aider à consolider la précision et l'exactitude des estimations d'intervalle post-mortem et de l'identification des fluides biologiques. La microbiologie judiciaire demeure encore une science à développer, mais son potentiel pour l'amélioration du rendement et de la précision du travail des experts médico-légal est très important et l'exploration de cette science reste pertinente.

## 1.2 Identification des sécrétions biologiques par le microbiome

### 1.2.1 Le microbiome humain

Le microbiome est défini comme l'ensemble génétique de la communauté microbienne qui vit à l'intérieur et sur les organismes vivants à la différence du microbiote qui se définit comme l'ensemble des individus qui forme cette communauté microbienne (Turnbaugh et al., 2007). La composition du microbiome change entre les différents individus, mais aussi selon les différentes parties du corps (Costello et al., 2009; Human microbiome project consortium, 2012). Plusieurs facteurs influencent sa composition, dont la génétique, les différentes maladies, la diète d'une personne, ou encore la prise de drogue ou de médicaments (Suzuki et al., 2019).

Les différentes parties du corps offrent aux microorganismes qui les colonisent des conditions bien distinctes ce qui permet d'observer des communautés de microorganismes avec des compositions différentes entre les sites anatomiques. Par exemple, le microbiome vaginal est composé d'au moins 70% de bactéries du genre *Lactobacillus* dont le produit métabolique est l'acide lactique (Miller et al., 2016). La grande production d'acide lactique par ces bactéries résulte en un environnement vaginal acide dont le pH est de 4.5 ou moins. Cette acidité permettrait, entre autres, de diminuer le nombre de microorganismes capables de coloniser l'environnement vaginal et aurait ainsi, un rôle de protection contre les infections (Miller et al., 2016). D'un autre côté, le système digestif offre aux bactéries un environnement totalement différent et qui change tout au long de celui-ci. Par exemple, l'estomac est un environnement très acide et dont peu de bactéries ont la capacité adaptative de le coloniser. Par contre, le colon (dernière partie du système digestif) est la partie du système dont la colonisation est la plus grande (Cresci et Izzo, 2019). L'oxygène se faisant rare dans le système digestif, celui-ci est surtout colonisé par des bactéries anaérobiques.

Considérant les changements dans la composition du microbiome selon le site anatomique, il est alors possible de penser que le prélèvement d'échantillons de microbiome pourrait nous donner des indices sur les sources anatomiques desquelles ceux-ci proviennent. L'identification du site anatomique peut donner des informations précieuses pour reconstituer la suite des événements lors d'enquêtes.

### **1.2.2 Méthodes existantes pour l'identification des sécrétions biologiques**

Présentement il existe plusieurs tests présomptifs pour l'identification du matériel biologique. Malgré la rapidité pratique de ces tests, ceux-ci ne sont pas spécifiques aux fluides et donc, ne répondent pas strictement au matériel biologique testé. Par exemple, un des tests présomptifs existant pour le sang est un test catalytique à base de phénolphtaléine qui se nomme le test Kastle-Meyer (KM). Dans le test KM, l'activité de type peroxydase de l'hémoglobine réduit le peroxyde d'hydrogène déposé sur l'échantillon, ce qui relâche des molécules d'oxygène et ainsi oxyde la phénolphtaléine en phénolphtaléine apportant un changement de couleur du transparent au rosé. Cependant, l'hémoglobine n'est pas spécifique à l'humain. Plusieurs autres espèces ont de l'hémoglobine. De plus, certains produits nettoyants ainsi que des végétaux provoquent la réaction catalysante lors du test KM (Peterson et Kovacs, 2014; Vennemann et al., 2014). En conséquence, ce test n'est pas spécifique et d'autres produits peuvent produire un résultat faussement positif. Les mêmes problèmes de spécificité sont présents pour d'autres fluides biologiques tels que le sperme ou encore la salive (Vennemann et al., 2014). Il est donc important d'avoir des méthodes qui confirment les fluides et matériaux biologiques avec une spécificité de 100%. Autrement, des échantillons seront pris inutilement sur des pièces à conviction dont le test présomptif est positif pour un fluide biologique afin de tester la présence d'ADN pour obtenir un profil ADN de cette pièce à conviction. Ceci entraîne des coûts inutiles pour le laboratoire et une perte de temps pour les employés.

Les tests qui ont une spécificité de 100% sont des tests confirmatifs, habituellement plus longs et coûteux. Par exemple, pour le sperme, il s'agit d'observer au microscope des échantillons pour confirmer la présence de spermatozoïdes. Cette méthode peut s'avérer ardue lorsqu'il y a plusieurs échantillons à évaluer. Ainsi, la recherche de méthodes spécifiques d'identification des matériaux biologiques rapides et fiables est importante et pertinente pour le milieu des sciences judiciaires. Présentement les méthodes supplémentaires explorées incluent, entre autres, la caractérisation des tissus biologiques par l'ARN (Blackman et al., 2018; Flemming et Harbison, 2010; Jusuola et Ballantyne, 2007), la méthylation de l'ADN (Frumkin et al., 2011) et l'évaluation du microbiome (Doi et al., 2017; Hanssen et al., 2017).

### 1.2.3 Identification des sécrétions biologiques par le microbiome

Différentes parties du corps abritent différents types de bactéries et cette caractéristique du microbiome humain pourrait aider la science judiciaire. Ainsi, une trace de salive sur une pièce à conviction n'aurait pas la même composition microbiologique qu'une trace contenant des fluides vaginaux. Les différentes études qui se penchent sur l'identification de fluides biologiques par le microbiome reposent sur des méthodes différentes. L'une des méthodes utilisées pour identifier principalement les fluides vaginaux est de déterminer les taxons indicateurs spécifiques à ce type de fluide. La forte composition en Lactobacilles peut être suffisante pour l'identification de fluides vaginaux provenant de sources pures (Doi et al., 2014 ; Giampaoli et al., 2016). Certaines espèces de Lactobacilles comme *L. crispatus* et *L. jensenii* peuvent être indicatives de fluides vaginaux si elles sont présentes, cependant l'espèce *L. iners* n'est pas spécifique de la région vaginale et ne peut donc pas être utilisée à des fins d'identification de fluides biologiques (Akutsu et al., 2012). L'espèce *L. gasseri* est encore controversée dans la littérature à savoir si elle est spécifique de fluide vaginal et ne devrait donc pas être utilisée pour l'identification de fluide vaginal pour l'instant (Choi et al. 2014 ; Doi et al., 2014).

D'autres méthodes utilisent l'abondance relative des échantillons pour identifier les fluides biologiques ou encore de l'apprentissage par machine (*machine learning*) avec de hauts pourcentages de spécificité. Des méthodes statistiques comme le Random Forest Classifier (RFC) est une façon légitime de déterminer la source de fluides biologiques. Cependant, la base de données requise pour entraîner le modèle doit être vaste et diversifiée pour produire des résultats représentatifs de la population (Tackmann et al., 2018). Singh et Publow (2021) ont utilisé un *support vector machine* (SVM) pour prédire les différents fluides biologiques testés. À l'exception du sang menstruel indistinguishable des fluides vaginaux et de l'urine, le pourcentage d'exactitude prédictive obtenu était de près de 90%. Cependant, cette étude testait des échantillons de sources pures, ce qui est peu représentatif des échantillons qui seraient prélevés sur une scène de crime où des mélanges de microbiomes provenant de différents sites anatomiques sont habituellement prélevés. De plus, l'abondance relative peut être utilisée pour évaluer l'effet du temps sur les échantillons de microbiome. Comme des prélèvements ne seraient pas nécessairement pris quelques minutes après le contact de l'objet et du site

anatomique, mais potentiellement plusieurs jours après le contact, il est important de connaître l'impact du temps sur la composition du microbiome. Celui-ci ne semblerait pas assez modifié, même après 1 mois dans des conditions intérieures, pour empêcher l'identification du fluide biologique, du moins, en ce qui concerne le microbiome de sperme (Dobay et al., 2019; Yao et al., 2020). Des méthodes par apprentissage profond sont aussi utilisées et certaines ont été rendues publiques, ce qui pourrait standardiser la méthode d'identification des fluides biologiques par le microbiome et ainsi aider à l'acceptabilité de la méthode dans les cours de justice (Lopez et al., 2019).

Pour l'expérimentation présentée, la méthode de prédiction des fluides biologiques est celle de l'utilisation du *plugin* SourceTracker2 du logiciel Qiime2. SourceTracker2 a été développé au départ pour retracer les sources de contamination dans les puits de plaques à PCR. Le logiciel sert à calculer la proportion de contamination dans les puits (Knights et al., 2011). Cette méthode semble appropriée pour identifier la contribution de différentes sources dans un échantillon, car elle permet de tester des échantillons composés d'un mélange de sources (ex : sources vaginale et cutanée) ce qui a été très peu testé dans la littérature jusqu'à présent.

### **1.3 La science de la décomposition**

La décomposition humaine est un phénomène naturel et complexe qui implique plusieurs processus distincts, tous influencés par différents facteurs, tant biotiques qu'abiotiques (Hayman et Oxenham, 2016). On sépare habituellement la décomposition des corps en plusieurs stades, quoique le consensus scientifique sur le nombre de stades ne soit pas encore établi. La moyenne vacille entre quatre et six stades au total (Carter et al., 2006; Finley et al., 2015; Vaas, 2011). Les principaux stades de décomposition généralement acceptés sont : (1) frais (2) gonflé (3) décomposition active (4) décomposition avancée, et (5) sec. L'état frais est le premier stade après la mort. Il se caractérise par une diminution de l'oxygène disponible pour les cellules qui mène à la destruction de celles-ci par autolyse. C'est aussi à ce stade que les premiers insectes apparaissent pour venir pondre leurs œufs et commencer la colonisation (Carter et al., 2006). Par la suite, l'activité métabolique des bactéries du système intestinal, entre autres, produit des gaz

qui amènent le corps à se gonfler (Finley et al., 2015). Il s'agit ici du 2<sup>e</sup> stade de décomposition. Le stade de décomposition active se caractérise par une perte de biomasse rapide due à une grande activité entomologique ainsi que la perte de fluides corporels dans l'environnement immédiat (Carter et al., 2006; Finley et al., 2015). La décomposition avancée est un stade où la perte de biomasse ralentie jusqu'à l'obtention de restes secs constitués d'os et possiblement de peau sèche qui caractérise le dernier stade de décomposition (Finley et al., 2015).

Les principaux facteurs qui influencent la vitesse de décomposition sont la température, l'humidité, les précipitations et le pH de l'environnement de surface où les corps sont enterrés ou déposés (Vaas, 2011). L'intérêt des recherches sur la décomposition humaine est multiple, mais l'un des principaux objectifs est d'aider les forces de l'ordre dans leurs enquêtes, notamment pour la détermination de l'intervalle post-mortem (IPM). Il existe plusieurs méthodes qui permettent d'évaluer l'IPM. Premièrement, lors du stade frais de la décomposition, les processus de *rigor mortis* (raideur des muscles), *livor mortis* (décoloration de la peau causée par le mouvement du sang dû à la gravité) et d'*algor mortis* (changement de température du corps) peuvent fournir des indices sur l'IPM (Byard, 2017; Goff, 2009). L'analyse post-mortem de l'humeur vitrée ou encore l'apparition d'adipocire, une substance savonneuse blanchâtre qui se forme à la suite de l'hydrolyse et l'hydrogénation des lipides neutres du corps, peuvent aussi être indicatives d'un certain IPM (Thierauf et al., 2009; Ueland et al., 2014). Or, une méthode très utilisée pour la détermination de l'IPM est l'entomologie judiciaire (Amendt et al., 2004), une approche qui consiste à évaluer la succession des différents arthropodes qui viennent interagir et parfois coloniser le corps en décomposition. Les premiers arthropodes attirés par l'odeur de cadavres sont des mouches (diptères) de type Calliphoridae suivies par des mouches sarcophages et des coléoptères qui arrivent un peu plus tard sur les cadavres. Puisque la succession d'insectes est relativement prévisible et stable selon les régions géographiques, les saisons ou encore les habitats desdits insectes, l'identification des œufs, larves ou même d'insectes adultes sur un corps peut donner une estimation de l'IPM selon le cycle de vie des insectes présents (Amendt et al., 2004; Benbow et al., 2016). Malgré le nombre élevé de méthodes pour déterminer l'IPM, elles ont toutes leurs limites, car les marges d'erreur sont souvent très grandes. Par exemple, l'utilisation des insectes donne des valeurs imprécises lors de températures froides, dans des lieux



peu ou non accessibles aux insectes, lors de précipitations importantes ou lorsque le corps est enterré, entre autres (Amendt et al., 2011). Les recherches pour obtenir de meilleures méthodes ou des méthodes complémentaires afin de mieux déterminer l'IPM sont toujours en cours.

### **1.3.1 Le microbiome et la décomposition**

Récemment, des chercheurs se sont intéressés au microbiome humain pour explorer de nouvelles pistes de détermination de l'IPM ou simplement pour mieux comprendre le phénomène de décomposition. Prenant en considération que les microbes associés à l'hôte et les microbes environnementaux participent conjointement à la décomposition des cadavres et que différents facteurs peuvent les faire varier dans le temps, certaines études ont évalué le potentiel du microbiome comme outil de détermination de l'IPM (Carter et al., 2006; Gunn et Pitt, 2012; Pechal et al., 2013). Deux méthodes d'échantillonnage ont été proposées, soit l'échantillonnage du microbiome des organes internes (thanatomicrobiome) ou celui du microbiome des surfaces de la peau (microbiome épinécrotique) (Oliveira et Amorim, 2018). L'échantillonnage du sol est parfois utilisé pour la détermination de l'IPM, mais aussi comme marqueur géographique de localisation de corps (Adserias-Garriga et al., 2017; Thomas et al., 2017; Zhou et Brian, 2018). Les deux approches ont des avantages et des inconvénients (Javan, 2016). Notamment, le thanatomicrobiome est moins influencé par les facteurs environnementaux comme la température extérieure ou la présence d'insectes (Budowle et al., 2020). Par contre, il est beaucoup plus difficile à étudier, car l'échantillonnage invasif et destructif requiert un accès aux organes internes (Budowle et al., 2020). De plus, la plupart des communautés microbiennes du thanatomicrobiome étant anaérobiques, l'échantillonnage ne doit pas les mettre en contact avec l'oxygène pour ne pas altérer les résultats (DeBruyn et Hauther, 2017).

Le microbiome épinécrotique, quant à lui, est plus facile à échantillonner de manière non invasive. Par contre, il est grandement influencé par les facteurs environnementaux (Budowle et al., 2020). En outre, il n'y a pas de consensus entre les études au sujet des sites anatomiques les plus pertinents à échantillonner selon la méthode. Dans le cas du microbiome épinécrotique, le nez, les oreilles, la bouche, ainsi que la peau des biceps, des joues, de la tête et du torse sont tous des sites qui ont été étudiés (Belk et al., 2018; Hyde et al., 2015; Johnson et al., 2016; Metcalf et al.,

2016; Pechal et al., 2013). Pour le thanatomiobiome, les sites de prédilection sont le cœur, le foie, la rate, le cerveau ainsi que le caecum (partie du côlon) (Can et al., 2014; Hauther et al., 2015; Javan et al., 2016; Javan et al., 2017). Certains ont proposé de récolter le sang des cadavres, mais il était souvent coagulé, ce qui rendait l'échantillonnage impossible (Can et al., 2014).

### **1.3.2 Les taxons pertinents du microbiome épinérotique**

La plupart des études qui évaluent le microbiome post-mortem pour le développement d'un outil de détermination de l'IPM cherchent à quantifier l'abondance des taxons dans le temps (Belk et al., 2018; Benbow et al., 2015; Hyde et al., 2013; Hyde et al. 2015; Johnson et al. 2016). En plus du microbiome épinérotique ou du thanatomiobiome, plusieurs recherches ont été réalisées sur le microbiome du sol pour trouver des mesures complémentaires de l'IPM ( Adserias-Garriga et al., 2017; Carter et al., 2008; Finley et al., 2016; Metcalf et al., 2013; Thomas et al., 2017). Ces travaux ont révélé que les phylums à privilégier pour l'étude du microbiome épinérotique sont les Protéobactéries, les Bacterioides, les Firmicutes, les Actinobactéries et les Acidobactéries. Selon la tendance observée dans la littérature, les Protéobactéries représentent le phylum dominant sur la plupart des sites d'échantillonnage épinérotiques au temps zéro (T0) (Hyde et al., 2015; Javan et al., 2016; Pechal et al., 2014). Par la suite, on observe une augmentation des Firmicutes et une diminution des Protéobactéries (Hyde et al., 2015; Javan et al., 2016). À des stades plus avancés de décomposition, les Actinobactéries augmentent aussi en proportion (Hyde et al. 2015; Javan et al., 2016). Il est à noter que les Acidobactéries représentent un phylum dominant dans les analyses de sol, tandis que les Bacterioides se retrouvent plutôt dans le système digestif (Javan et al., 2016; Javan et al., 2017). Ce phylum est prédominant lorsque le tractus intestinal est échantillonné ou lorsque le contenu du système digestif se transfère au sol (Hyde et al., 2015; Javan et al., 2016; Javan et al., 2017; Metcalf et al., 2013; Pechal et al., 2013). Le genre *Acinetobacter* comporte également des bactéries communément trouvées dans le sol. Leur présence dans les stades avancés de la décomposition est attendue ainsi que celle de la famille des *Xanthomonadaceae* qui est associée avec la myiase d'insectes *Sarcophagidae* (Hyde et al., 2015).

### **1.3.3 Conditions d'études de la décomposition**

La plupart des études sur le microbiome épinérotique ont des périodes d'échantillonnage relativement courtes (1 à 3 semaines). Cela peut s'expliquer par le fait que la plupart des sites de recherches anthropologiques et taphonomiques qui étudient la décomposition humaine sont situés dans des climats où la température moyenne descend rarement sous les 20°C. Dans de telles conditions, le processus de décomposition est relativement rapide. De plus, de nombreuses recherches utilisent des carcasses de porcs pour quantifier la décomposition, ce qui peut biaiser les extrapolations sur des données de cadavres humains. Il est à noter que certaines études sur le microbiome épinérotique se font dans des conditions aquatiques où l'on étudie l'intervalle post-mortem de submersion (Benbow et al., 2015; Dickson et al., 2011). Comme les conditions réalistes sont extrêmement difficiles à obtenir pour ce genre d'études, les données se font rares, notamment pour comparer la décomposition de corps partiellement ou totalement immergés.

La composition du microbiome étant très influencée par la température, il est plus juste de corrélérer le changement post-mortem du microbiome avec les degrés quotidiens accumulés (ADD pour *accumulated degree-days*) qu'avec le nombre de journées. L'ADD est la somme des moyennes de température quotidiennes depuis la mise en place initiale du corps jusqu'à la fin de l'échantillonnage (Megyesi et al., 2005). Certains chercheurs utilisent aussi un score total de décomposition (TDS), bien qu'il soit difficile à évaluer objectivement sans formation préalable (Gelderman et al., 2018). De plus, le TDS étant dépendant de la température il est également corrélé avec l'ADD (Gelderman et al., 2018; Vaas, 2011).

### **1.3.4 Le microbiome environnemental en hiver**

Aucune étude sur le changement du microbiome humain post-mortem n'a encore été réalisée lors d'accumulation de neige ou pendant un hiver avec des températures semblables à celles du Québec. Il est donc impossible pour l'instant d'appliquer les résultats des travaux portant sur des cadavres en décomposition en saison hivernale pour l'estimation de l'IPM aux cadavres retrouvés pendant l'hiver au Québec. Par contre, des études sur le microbiome de la neige arctique ont révélé que les phylums dominants de cet environnement sont similaires aux phylums étudiés pour l'IPM soient les Protéobactéries, les Actinobactéries, les Bacteroides, les

Firmicutes ainsi que les Cyanobactéries (Amato et al., 2007; Cameron et al., 2015; Møller et al., 2013; Mortazavi et al., 2015). Il a aussi été démontré que des bactéries de la neige continuent leurs activités métaboliques malgré les basses températures (Zhu et al., 2020). Il est donc acceptable de penser que certaines bactéries épinérotiques continueront leurs activités de décomposition des corps pendant l'hiver. De plus, l'abondance taxonomique semble changer de façon cyclique avec les saisons, du moins au niveau du sol et des rivières, ce qui laisse croire qu'il sera possible de déterminer si un corps a été présent lors d'un hiver selon les taxons présent sur celui-ci (Cruaud et al., 2020; Jefferies et al., 2010).

## **1.4 Analyses bio-informatiques dans l'étude du microbiome**

L'étude du microbiome repose sur l'analyse des produits de séquençage de l'ARN ribosomal (ARNr) 16S par des logiciels informatiques. Il faut donc fait l'extraction d'ADN des cellules bactériennes recueillies lors de l'échantillonnage et amplifier la région de l'ARNr 16S qui nous intéresse. Parmi les neuf régions hypervariables du gènes 16S de l'ARNr, l'amplification des régions V1-V2 et V6 amène souvent une surestimation de la richesse des espèces tandis que les régions V3, V7 et V7-V8 ont tendance à sous-estimer la richesse (Graspeuntner et al., 2018; Youssef et al., 2009). La région V4 qui sert d'ancrage aux ARN de transfert (ARNt) est la plus conservée parmi les régions hypervariables (Bukin et al., 2019). De plus, cette région V4 est l'une de celles qui sont les plus comparables en termes de richesse en comparaison avec les analyses faites sur des fragments quasi complets du gène 16S (Youssef et al., 2009). De plus, la région V4 est la région privilégiée dans les études portant sur le microbiome en science judiciaire. Pour ces raisons, la région V4 semble la plus appropriée pour caractériser la composition du microbiome épinérotique et des différents fluides corporels.

### **1.4.1 Bio-informatique et statistiques**

Les séquences obtenues par l'amplification sont classifiées en ASVs (Amplicon Sequence Variant) ou en OTUs (Operational Taxonomic Unit). Cependant, la méthode basée sur les ASVs semble avoir une meilleure résolution taxonomique que les OTUs (Bharti et Grimm, 2021). De

plus, avec les OTUs, il y a une tendance à surévaluer l'abondance des séquences (Callahan et al., 2016; Edgar, 2017). Lors de l'amplification d'ADN par PCR, plusieurs erreurs peuvent se produire durant le processus et alors créer des séquences chimériques de plusieurs espèces ou encore des séquences incomplètes ou incorrectes qui biaiseraient les résultats obtenus. Pour contrecarrer l'impact de ces erreurs, il est possible de regrouper les séquences qui ont des similarités pour exclure les artéfacts d'amplification. Ainsi, la classification en OTUs regroupe les séquences qui ont une similarité au-dessus de 97% (Nearing et al., 2018). Par contre, deux espèces différentes, mais taxonomiquement près l'une de l'autre pourraient ainsi être mises dans le même groupe d'OTUs et la résolution taxonomique en serait ainsi diminuée. La méthode basée sur les ASVs tente de corriger les erreurs de séquençage sans grouper les séquences similaires. Ainsi, les espèces proches ne seront pas mises dans un même groupe et la résolution taxonomique est alors plus grande qu'avec la méthode d'OTUs (Nearing et al., 2018). Par contre, la méthode de "nettoyage des données" qui corrige les erreurs de séquençage n'est pas la même dépendamment des différents algorithmes, sa précision varie donc d'un logiciel à l'autre. Pour travailler avec les ASVs dans ce projet, l'étape de "nettoyage des données" peut être effectuée avec le logiciel DADA2 (Callahan et al., 2016). Malgré qu'il soit significativement plus long à opérer que d'autres logiciels, DADA2 est plus précis (Nearing et al., 2018). Il est à noter que l'utilisation des ASVs devrait remplacer la méthode des OTUs, car elle est plus spécifique et plus précise et plus reproductible que la précédente (Callahan et al., 2017). Il pourrait donc y avoir des différences mineures dans les résultats trouvés de cette étude avec les études précédentes faites sur le sujet pour ces raisons.

Par la suite, des analyses de diversités alpha et bêta sont réalisées pour comparer les différents échantillons. La diversité alpha mesure la diversité microbienne à l'intérieur d'un seul échantillon alors que la diversité bêta mesure la diversité entre les différents échantillons. Une raréfaction plus ou moins sévère se doit d'être effectuée pour l'analyse des données de diversité alpha afin de considérer la profondeur de séquençage (Willis, 2019). Les résultats de diversité alpha et bêta peuvent être analysés dans un modèle longitudinal linéaire pour comparer des séries temporelles.

Ces étapes d'analyses ont été choisies et adaptées pour suivre l'ébauche de lignes directrices récemment proposées dans le but de standardiser l'analyse du microbiome dans le domaine de la microbiologie judiciaire et pour faciliter son application réelle dans le système de justice (Kaszubinski et al., 2020).

## 1.5 Problématique

En ce qui concerne l'identification de fluides biologiques par le microbiome, des méthodes déjà existantes sont disponibles pour les autorités judiciaires. Il faut ainsi pouvoir tester l'efficacité et la spécificité de la méthode exploratoire qui utilise le microbiome. Les défis des sciences judiciaires sont grands. Il faut des méthodes d'identification spécifiques, qui utilisent des protocoles validés et standardisés pour qu'elles puissent être adoptées dans une cour de justice. Le but étant ultimement d'utiliser une science robuste pour éviter les erreurs judiciaires.

Plusieurs enjeux sont liés à l'utilisation du microbiome épimicrobien. Les études sur le sujet évaluent différents sites d'échantillonnage et différentes méthodes d'analyses des données de séquences. Par contre, la majorité des chercheurs s'entendent sur le séquençage de l'ARNr 16S pour obtenir des modèles plus robustes de l'IPM (Metcalf et al., 2016). De plus, la méthode d'analyse par ASVs générés par le logiciel Qiime2 est la méthode privilégiée pour standardiser les études futures du microbiome en sciences judiciaires (Kaszubinski et al., 2019). Néanmoins, il importe de quantifier l'influence des différents facteurs environnementaux qui ont un impact sur la vitesse de décomposition et sur la composition du microbiome épimicrobien pendant celle-ci. Par exemple, la grande majorité des études sont effectuées dans les sites taphonomiques extérieurs où les corps sont déposés sur le sol. On ne connaît donc pas bien le processus de décomposition de cadavres habillés retrouvés par les autorités à l'intérieur de logements où les insectes n'ont pas accès et où les températures sont stables et les intempéries inexistantes. On ne connaît pas non plus précisément le processus de décomposition de corps qui sont recouverts de plusieurs centimètres de neige. De plus, il faut prendre en considération les différentes compositions microbiennes des sols et des cadavres dans différentes régions géographiques où les conditions climatiques peuvent significativement affecter la décomposition des corps. Dans

un climat hivernal comme le Québec, lors des saisons plus froides, les résultats doivent être calculés en fonction des degrés Kelvin quotidiens accumulés (KADD pour Kelvin accumulation degree day) lorsque les températures sont inférieures au point de congélation en degrés Celsius (Roberts et Dabbs, 2015; Roberts et al., 2017).

## **1.6 Objectifs**

Afin de répondre à certaines des questions soulevées par l'utilisation du microbiome en microbiologie judiciaire, les objectifs principaux de la présente recherche consistent à, premièrement, déterminer si l'utilisation du microbiome vaginal peut servir pour développer un outil d'identification de fluides vaginaux et ainsi, pouvoir évaluer le potentiel de la méthode à des fins judiciaires concrètes. Pour ce faire, il faudra que la méthode testée ait une spécificité de 100% pour pouvoir être utilisé dans le domaine. Il sera ainsi possible de déterminer si la méthode utilisant le microbiome a un potentiel réel d'utilisation pour des enquêtes médico-légales.

Deuxièmement, caractériser le changement de la composition bactérienne du microbiome épiniécrotique dans un climat nordique comme celui du Québec et l'impact de l'hiver sur celui-ci afin de déterminer les sites d'échantillonnage anatomiques les plus pertinents et pouvoir déterminer si un corps a passé l'hiver ou non. Cela permettra d'analyser l'impact des températures hivernales québécoises et l'accumulation de neige sur le microbiome épiniécrotique des corps en décomposition.

## **1.7 Organisation du mémoire**

Les chapitres deux et trois de ce mémoire sont présentés sous forme d'articles scientifiques. Le chapitre deux présentera le premier article, soit une preuve de concept où l'on tente d'identifier la présence de fluides vaginaux dans des échantillons de mélanges de différents fluides biologiques. Le chapitre trois est constitué du deuxième article qui s'intéresse à l'impact de l'hiver sur le microbiome épiniécrotique lors de la décomposition humaine et tente de bâtir un modèle prédictif pour savoir si un corps aurait passé l'hiver ou non. En guise de conclusion, le chapitre

quatre abordera les faits importants de la recherche en microbiologie judiciaire et les perspectives de recherches futures dans ce domaine.



## **Chapitre 2 – Identification of vaginal fluids using microbiome for forensic purposes: A proof of concept**

**Audrey-Anne Leblanc<sup>1\*</sup>, Sarah Noël<sup>2</sup>, François-Joseph Lapointe<sup>1</sup>**

<sup>1</sup> Département de sciences biologiques, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, QC H3C3J7, Canada

<sup>2</sup> Laboratoire de sciences judiciaires et de médecine légale de Montréal, 1701 rue Parthenais, Montréal, H2K 3S7, Québec, Canada

\*Corresponding author: [audrey-anne.leblanc@umontreal.ca](mailto:audrey-anne.leblanc@umontreal.ca)

(Article est en préparation, non soumis)

## 2.1 Résumé

L'identification des fluides biologiques qui sont collectés sur les scènes de crimes est importante pour la reconstruction des évènements lors de l'enquête. Plus spécifiquement, identifier les fluides vaginaux peut s'avérer un être un élément clé lors de cas d'agressions sexuelles. Cependant, les tests confirmatifs pour les fluides vaginaux sont encore en développement. Pour cette recherche, nous avons testé une méthode pour l'identification des fluides vaginaux. Nous avons récolté des écouvillons de fluides vaginaux, anaux, cutanés et oraux provenant de différents individus volontaires et nous avons créé des échantillons purs ainsi que des mélanges à tester. Nous avons fait l'extraction d'ADN bactérien de ces échantillons et utilisé le séquençage de nouvelle génération pour l'identification. Nous avons testé l'exactitude des résultats du modèle de prédiction à trouver les échantillons contenant des fluides vaginaux, et ce, entraînant celui-ci avec des nombres différents de sources pures. En utilisant un seuil de 40% de bactéries vaginales pour considérer un échantillon positif, nous avons pu correctement identifier des mélanges contenant des fluides vaginaux. La méthode était améliorée chaque fois que le nombre de sources pures utilisées était augmenté. Avec le même seuil pour chaque test, nos résultats ont démontré une spécificité de 100% (aucun faux positif) et une sensibilité de 70% (30% de faux négatifs). Nous sommes confiants que cette méthode pourrait être utilisée pour l'identification de fluides vaginaux en tant que test confirmatif, considérant que des échantillons provenant des victimes peuvent être récoltés lors d'enquêtes pour agressions sexuelles.

## 2.2 Abstract

Identification of body fluids collected on crime scenes during investigations is important for the reconstruction of events. Identifying more specifically vaginal fluids can be key evidence for sexual assault cases. However, confirmatory tests for vaginal fluids are yet to be developed. Here, we tested a method for vaginal fluid identification. We collected vaginal, anal, skin and oral swabs from different individuals that volunteered, and we created mixture samples as well as pure source samples to which we extracted the bacterial DNA and used next-generation sequencing for identification. We tested the accuracy of a prediction model to correctly identify mixture samples containing vaginal fluid by training the model with different amounts of pure sources. Using a threshold of 40% of vaginal bacteria to consider the sample positive, we were able to correctly identify mixture samples containing vaginal fluid. The power of the method was improved as the number of pure sources for training the model was increased. Using a fixed threshold for all tests, our overall results exhibited a specificity for vaginal fluids of 100% (no false positive) and a sensitivity of 70% (30% false negative). We are confident that this method could be used for vaginal fluid identification as a confirmatory test, provided that source samples are collected in sexual assault cases.

## 2.3 Introduction

The collection of physical evidence, defined as evidence that can be seen, touched, extracted, and exhibited is crucial in forensic investigations (Nafte et Dalrymphe, 2016). One important type of physical evidence that greatly helped with investigations since the mid-1990s is the latent evidence of DNA that allowed, with great accuracy, the association of individuals with crime scenes and/or with pieces of evidence collected on crime scenes (Butler, 2009). DNA is very helpful for investigators to reconstitute the chain of events by linking a biological fluid to an individual. However, the DNA profiles obtained by forensic biologists do not give context on the origin of the DNA, such as time of deposit, or the body part it originated from. Additional tests, like body fluid identification (Butler, 2009), need to be run on the evidence to possibly identify the origin of the DNA.

Presumptive tests are often performed on the evidence to help identify the body fluid. Despite the rapidity and sensibility of those tests, they are not very specific and can therefore provide false positive results (Gefrides et Welch, 2011). They are often based on enzymatic reactions that more often than not, can operate on more than one body fluid, or react with common household products. For example, some presumptive tests for blood identification are based on a chemical reaction of the hemoglobin. Although hemoglobin is a molecule specific to the red blood cells, it is used for oxygen transport by a large number of vertebrates and some invertebrates (Hardison, 2012). Therefore, its presence is an indicator of blood, but not necessarily human blood. Furthermore, some cleaning agents and plant compounds can react to the components used in different presumptive tests and give false positive results (Gefrides et Welch, 2011; Hanssen et al., 2017; Petersen et Kovacs, 2014; Vennemann et al., 2014). The detection of prostate-specific antigen (PSA) to identify semen is also commonly used in forensic science (Harbison et Fleming, 2016). Here again, PSA can be found in small, but detectable amounts in urine, vaginal fluids or breast milk, making the test susceptible to false positives (Harbison et Fleming, 2016; Yu et Diamandis, 1995). Nonetheless, these tests are quick, easy to use, and can be very helpful for forensic biologists to advise them on samples to test for DNA. To have certitude on the origin of a biological fluid, one must use confirmatory tests. Yet, these are not always available. Namely, for vaginal fluids, confirmatory tests are still in development. They can also be time-consuming,

like the confirmation of seminal fluid via the microscopic observation of spermatozoa (Gefrides et Welch, 2011).

Additional methods to identify body fluids are required to help with the reconstitution of the chain of events. Currently, multiple fluid identification methods are being studied with high potential. For one, the assessment of DNA methylation was able to discriminate body fluids with high sensitivity and specificity (Choi et al., 2014). This method was used to successfully identify semen, saliva, blood and skin cells (Choi et al., 2014; Frumkin et al., 2011). However, since it uses the same DNA sample as the one used to develop DNA profiles via short tandem repeat (STR), it cannot be used with a small amount of DNA evidence. Moreover, methylation status between individuals can vary, and it changes in the methylated regions tested depending on environmental factors, or age (Harbison et Fleming, 2016). Some false positives were also found in previous studies (Choi et al., 2014), requiring further research in that area. Methods using miRNA and mRNA for identification have also been tested with success (Fleming et Harbison, 2010; Juusola et Ballantyne, 2005). However, those methods require specific scientific expertise to run and interpret and they are generally very slow to conduct (Blackman et al., 2018; Frumkin et al., 2011). Additionally, RNA is not as stable as DNA and can be harder to work with and collect from crime scenes (Setzer et al., 2008).

Another way to identify body fluids is through the microbial markers of the fluids. Different body parts offer different environments for bacteria of the microbiome colonizing it (Costello et al., 2009). Therefore, the composition of the microbiome changes depending on the body part (Consortium HMP, 2012). Some research focused on identifying specific taxa in different body fluids. The vaginal area is strongly colonized by bacteria of the genus *Lactobacillus*; it can be identified by the presence of a large proportion of this genus in the sample tested (Doi et al., 2014). It is important to note that this method could potentially only work in cases involving adults since the vaginal microflora of children is significantly different, with a reduced prevalence of *Lactobacillus* in the vaginal flora of prepubescent girls (Hammerschlag et al., 1978; Harbison et Fleming, 2016; Xiaoming et al., 2021). Other papers were looking at indicator species unique to a body part that could be an identification factor. Taxa like *Lactobacillus crispatus* and *Lactobacillus jensenii* for vaginal fluid, or *Streptococcus salivarius* for saliva, were identified as

potential markers for body fluid identification (Akutsu et al., 2012; Zhang et al., 2022). However, such microbial markers are not always colonizing the body site due to individual diversity or sometimes can be found at other body sites (e.g. *S. salivarius* can also be found in feces). Furthermore, the composition of the microbiome on a deposited stain can change with time, and microbial markers could decrease to an undetectable level with time (Yao et al., 2020). Therefore, it might be a better method to look at the composition of the microbiome as a whole for each body site instead of looking for specific individual markers.

Another consideration to account for is the fact that the collection of samples by the authorities is not always performed minutes after the crime occurred. It is important to know whether the microbial signature on pieces of evidence will last long enough to be recognized after collection. According to the literature, the microbial signature of semen was still identifiable after 15 days (Yao et al., 2020) and vaginal fluid, saliva, skin, and semen also kept their microbial signature after 30 days of indoor exposure (Dobay et al., 2019). These findings show that the identification of body fluids by the microbiome is a method that has a great potential.

Microbiome samples collected on a crime scene are unlikely to be pure since they will be deposited on a surface that could have been contaminated with other substances prior to the crime. Therefore, it is crucial to identify body fluids of importance, such as vaginal fluids, even when mixed with semen and/or saliva. Studies that have investigated mixtures of microbiome sources are rare. Random forest regressions seemed to have the discrimination power to identify body fluids in mixtures with environmental contamination when the model is strong enough (Tackmann et al., 2018). However, it seems that body fluids deposited on the skin might require an additional reference sample from the skin to obtain better identification results (Hanssen et al., 2017). Further research is necessary to assess the strength of this method on microbial mixtures.

In this paper, our objective was to test whether identification of the vaginal fluids of adults by their microbial signature would still be possible in mixtures samples. To do so, the microbial signature of mock samples made up of pure sources and mixtures of two body fluids were tested to evaluate the capability of identification of vaginal fluid with a source tracking algorithm.

## **2.4 Materials and Methods**

### **Sample collection**

15 women between the age of 32 and 59 years old participated in this study. The same homemade user manual was given to all participants for the sampling to be standardized. Sampling was executed in the same order: hands, mouth, vaginal and anal (optional) with a tampon applicator. Tampons were taken out of their packaging and needed to be kept in a compact mode within the applicator for the sampling process. For the sampling of hands, the applicator was rolled between the hands for 15 seconds. For the mouth, an applicator was inserted in the mouth and 3 back and forth movements were repeated while keeping the lips of the participant's mouth closed. For the vagina, after having washed their hands, a first applicator was inserted in the vaginal cavity for 5 seconds by the participant and a second sampling was made with another applicator in the same way. The anal site was optional for the participant to sample. Those who elected to do inserted half the applicator in the anal cavity for 5 seconds. After sampling, all the applicators were placed in an individual cardboard box (one for each site) and stored at  $-4^{\circ}\text{C}$  until DNA extraction.

Two women were menstruating on the day of the sampling, three women were postmenopausal, none of the participants were on anti-fungal medication for vaginitis and one woman finished an antibiotic treatment 7 weeks before the sampling day. The sperm microbiome was not tested for this experiment, but six participants had unprotected sexual relationships with ejaculation 7 days or less before sampling (Table S2.1). All women selected for the sampling worked at the same laboratory. The sampling protocol and consent form were approved by the Université de Montréal ethics committee (CERC-19-083-D).

### **DNA extraction and sequencing**

DNA extractions were carried out using the DNA IQ magnetic beads system (Promega, Madison, WI, USA) and eluted with 50ml of elution buffer. DNA extracts were then quantified using the 7500 Real-Time PCR System (Applied Biosystems, Waltham, MA, USA), and the Quantifilier1 Duo DNA Quantification Kit (Applied Biosystems, Waltham, MA, USA). All extractions were randomized to avoid detecting false patterns. The 515F/806R primer set was used for the amplification of the V4 hypervariable region of the 16S rRNA. To detect contamination during the

process of extraction and PCR amplification, negative controls were used. Sequencing was performed by the G enome Qu ebec facilities on a MiSeq sequencer (Illumina, San Diego, CA, USA).

### **Data analysis**

Demultiplexing was performed in Qiime2 (Bolyen et al., 2019) and the denoising process was executed by Dada2 (Callahan et al., 2016). A total of 50 samples were amplified, and 693,829 sequences were obtained, with a mean of 1,276 sequences per sample. Rarefaction was after used as a normalizing method. A sampling depth of 2500 was chosen allowing us to keep 100% of the samples in the analysis. We also used Qiime2 for generating the ASV table and for principal coordinate analysis (PCoA) based on unweighted UniFrac distances (Lozupone et al., 2011). Negative controls were contaminated by an unrecognized specie of the *Shigella* genus. Therefore, DNA sequence from this genus were removed from our ASV table. After rarefaction, a total of 16 phyla containing 97 families of bacteria were observed within the 50 samples. Also, Emperor software was used to visualize the PcoA results (V azquez-Baeza et al., 2017).

SourceTracker (Knights et al., 2011) was used in Qiime2 to assess the source composition of the samples. SourceTracker estimates the contribution percentage of each source (vaginal, oral, anal, skin) to designated sinks and applies a Bayesian framework to do so. To calculate the power of the model, we started by training the model with only one known source of vaginal fluids, and then increased the number of sources to 3, 6 and 10 (maximum number of sources) to assess the performance of the model. For each case, we tested the model 3 times with pure vaginal fluid samples that were randomly selected to be used as sources. For the other body fluids (anal, skin and oral), 5 pure sources were used for all tests. SourceTracker produces a table with relative contribution to a sample from all four putative sources (vaginal, oral, anal, skin), as well as unknown sources. Here, a conservative threshold of 40% was chosen to determine that a sample originated from a source. For comparison purposes, average numbers of predicted samples based on the different number of random replicates are reported.



## 2.5 Results

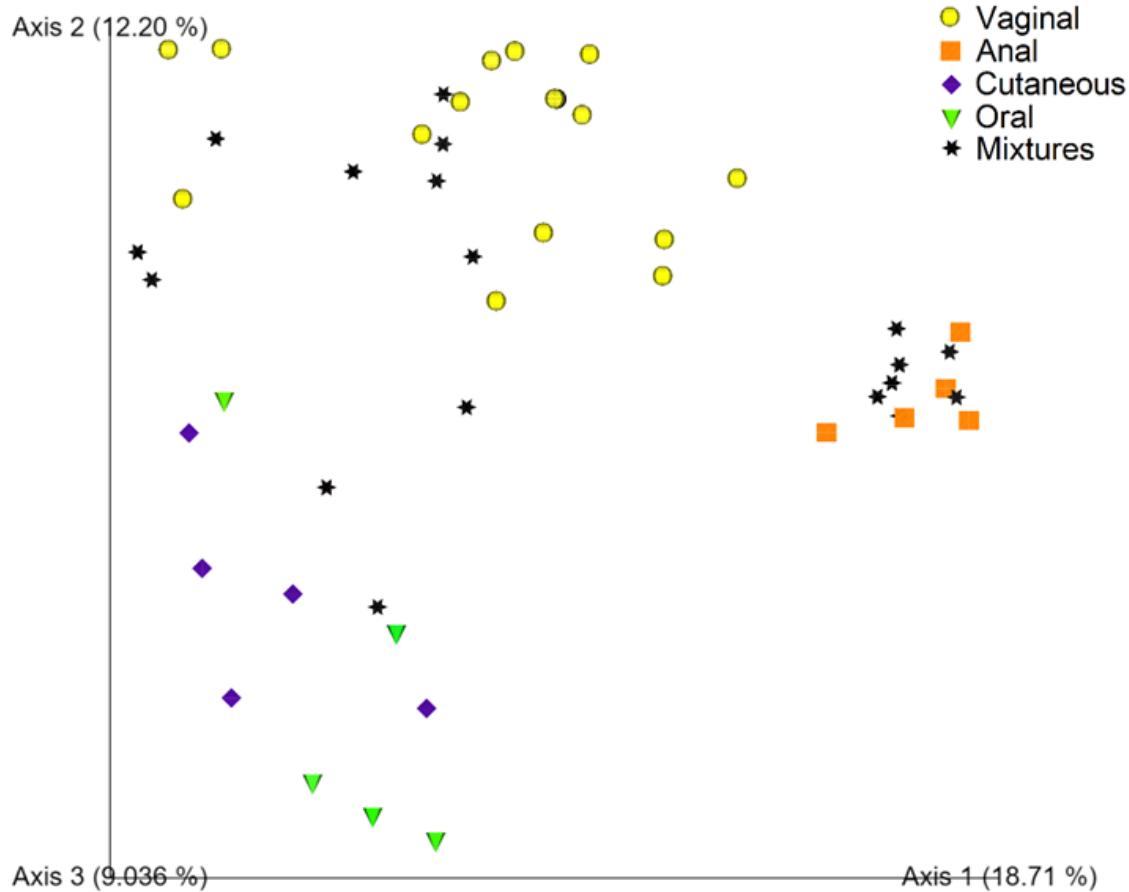
### SourceTracker accuracy

Of the 50 samples collected, 25 were originating from one source only, while the other 25 were mixture samples. There were 10 pure sources of vaginal fluids, and 5 pure source samples for anal, skin and oral sites. Among the mixture samples, 12 were mixtures of two body fluids from the same individual, whereas 8 were mixtures of two body fluids from two different individuals. The last 5 samples were mixtures of the same body fluids pertaining to two different individuals. Thereby, there was a total of 20 mixture samples containing vaginal fluids and 10 samples of pure sources of vaginal fluids, for a total of 30 samples (Table S2.2).

Using a threshold of 40% for determining the origin of a source, SourceTracker correctly identified an average of 2.67 mixture samples (13,35%) when only 1 vaginal fluid source was used to train the prediction model. When the number of vaginal fluid sources was increased to 3, an average of 10.33 mixture samples (51,67%) were correctly identified. When 6 sources were used, the average number increased at 12.67 samples found (56,65%) and that number reached 21 (70%) when the maximum number of vaginal fluid sources ( $n = 10$ ) was used to train SourceTracker. No false positive was detected by the model when using a threshold of 40%. That is, that SourceTracker predicted with perfect accuracy the samples containing vaginal fluids.

### Beta diversity analysis

Principal coordinate analysis (PCoA) based on the unweighted UniFrac distance shows discriminating clusters based on the different body sites (Fig. 2.1). The skin of the hands and the mouth samples are not well separated, unlike the anal and vaginal samples that exhibit clear clusters distant from one another. More specifically, when we focus on mixture samples, all mixtures samples within the vaginal cluster do indeed contain vaginal fluids. The same observation also applies to the anal cluster in which all mixture samples in the clusters do contain anal fluid.

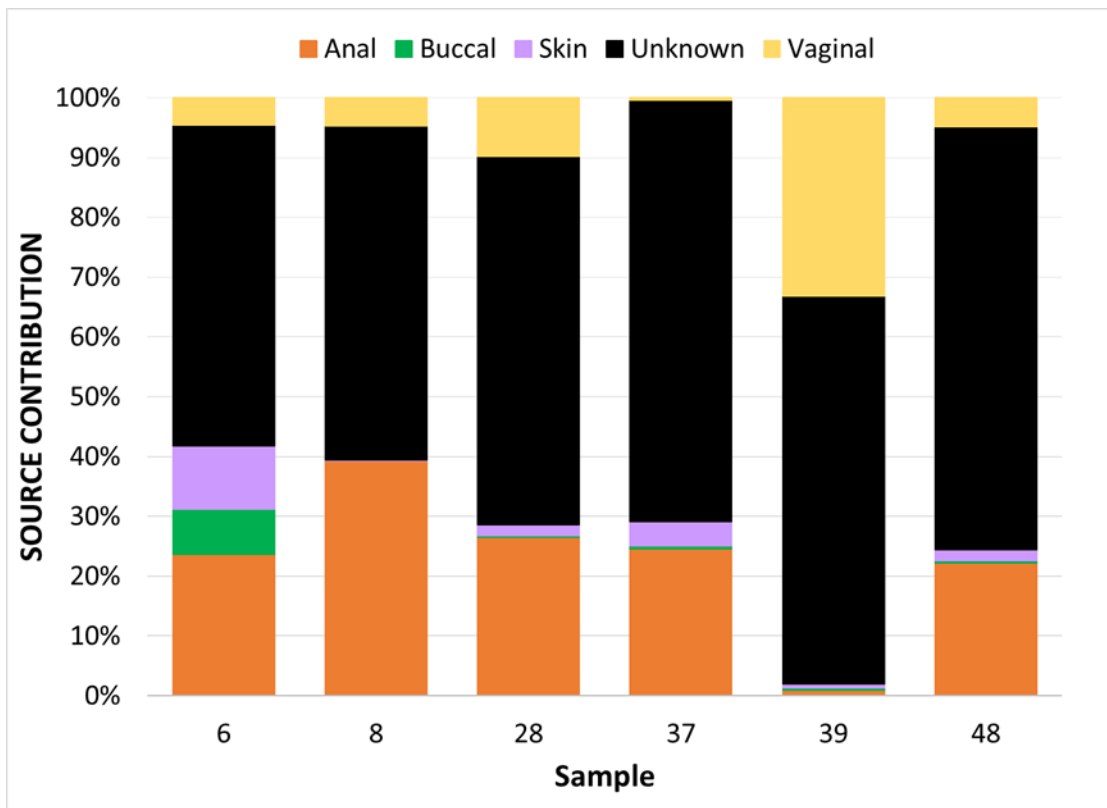


**Figure 2.1** - Principal coordinate analysis (PCoA) based on unweighted UniFrac distances generated from all 50 collected samples. The samples representing pure sources are identified by different symbols, whereas mixture samples containing two different body fluids are identified by stars (\*). The 5 samples containing a mixture of vaginal fluids from two different women are considered pure sources and have been identified in the vaginal group.

### Composition analysis

Some of the samples ( $n = 6$ ) were never identified by SourceTracker, regardless of the number of pure sources used to train the predictive model. Results for these samples (i.e., 6, 8, 28, 37, 39, and 48) illustrate that in most case, SourceTracker classified the source as unknown (Fig. 2.2). The second larger contributor was predicted as anal. For three samples (i.e., 8, 28, and 37) containing a mixture of vaginal fluids and anal fluids, the vaginal contribution was negligible (less than 15%

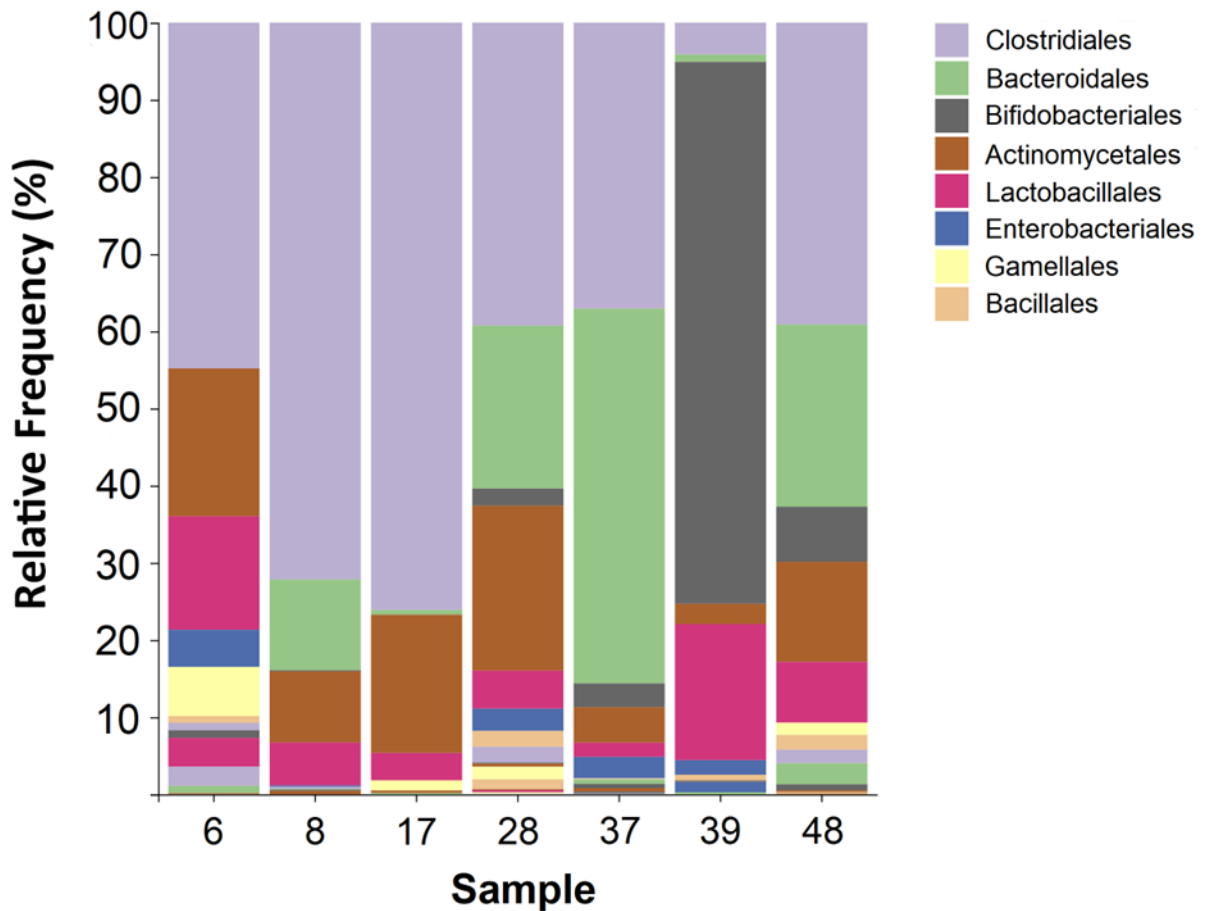
for all samples except sample 39 that was 35% which is still under the set threshold of 40%). In order to better understand such unknown sources, a taxonomic bar plot (Fig. 2.2) was generated to look at the composition of all problematic samples.



**Figure 2.2** - SourceTracker results of the source contribution including the 6 mixture samples that the model could not predict correctly the vaginal fluid contribution of it. These samples are identified by a red arrow.

Sample 6 was a mixture of vaginal fluids and a skin source, while the 2 others were only composed of vaginal fluids. Namely, sample 6 was a mixture from individual #6, whereas sample 48 also included individual #6 as a contributor. Since there was a suspicion of contamination, we looked at sample 17 (Fig. 2.3), a pure sample composed only of vaginal fluid from individual #6. We

noticed a large proportion of *Clostridiales* in this individual vaginal sample. Likewise, a large amount of *Clostridiales*, *Bacteroidales* or *Actinomycetales* were found in the other samples classified by SourceTracker as unknown. The amount of *Lactobacillales*, expected for vaginal samples, is small for all of the samples presented. Sample 39 had a large proportion of *Bifidobacteriales*, unlike the other samples.



**Figure 2.3** - Relative Frequency of bacterial orders of the samples containing vaginal fluids that were never found by the predictive model from SourceTracker2.6 Discussion

## 2.6 Discussion

The results from SourceTracker show its capacity to predict whether a body fluid comes from vaginal sources with a specificity of 100%, when the threshold is put at 40%. To our knowledge, SourceTracker has not been used to identify mixtures of body fluids in previous studies for forensic purposes. Our results also show the power of the training model since it requires just a few pure sources (as little as one!) to identify a majority of mixture samples including vaginal fluids. This is important in the context of forensics, because samples that would be collected at crime scenes would be mostly mixtures and not pure sources due to environmental contamination. The mixture samples that were not found by SourceTracker (i.e., unknown source) were always the same, regardless of the pure sources or group of pure sources that were used for the model training. Samples 8, 28, and 37 were mixtures of vaginal fluids and anal fluids. Among all mixture samples of anal and vaginal fluids, SourceTracker was able to find only one sample. This limitation is important as it implies that vaginal fluids may be undetectable in mixture samples containing anal fluids. For the other samples that did not have an anal component in them, it is possible that contamination occurred since 3 of the 4 individuals that contributed to those mixture samples also collected anal fluids. The 4th individual did not sample anal fluids, but their pure vaginal fluid sample (sample 17) was composed of *Clostridiales* for as much as 70%, an unusual proportion for vaginal fluids (Cho et al., 2021; Doi et al., 2014; Ravel et al., 2011). It has to be noted that no factors that were asked prior to the participants (eg. menopause, last sexual intercourse, antibiotic's take) could explain partially or totally the reason why those 6 samples were kept unfound by SourceTracker.

Also, when we looked more closely at the composition of the samples that were not identified by SourceTracker, we noticed that the majority of samples are composed of either *Clostridiales*, *Bifidobacteriales*, or *Bacteroidales*. These bacterial orders are known to colonize the digestive tract of humans (Manor et al., 2020; O'Callaghan et al., 2016; Rinninella et al., 2019). The *Lactobacillales* order that includes the *Lactobacillaceae* family that colonizes the vaginal area is underrepresented in most samples, and it constitutes less than 20% of samples which can be explained by contamination of the samples.

Additionally, when we looked at the PCoA results, we were able to identify distinct clusters separating pure anal samples from pure vaginal samples. On the other hand, the skin and oral samples are not distinguishable from one another and form a joint cluster. We believe that looking at the PCoA plot could provide clues about which samples to include in SourceTracker, when there is a limitation in the number of samples to analyze.

Our results show that SourceTracker could be used for the identification of vaginal fluid in forensic science. Although this was not the objective of our paper, that result implies that the identification of anal fluid in mixture samples could also be done with SourceTracker, just like vaginal fluids. This method has a specificity of 100% when the threshold of composition is set at 40%. There are multiple advantages to this novel approach to body fluid identification using microbiome data. First, it is quite simple to collect the samples using standard kits; thus the training for crime scene officers will not be difficult. Second, SourceTracker does not require a long time to run the results, and it can run multiple analyses at the same time. Finally, with a threshold that we have set at 40%, the method will not return false positives, which in turn increases its applicability in criminal cases. The sensibility of SourceTracker to identify vaginal fluid was estimated at 70%, when all pure sources ( $n = 10$ ) were used to train the model. Yet, only 15 individuals from the same working environment participated in the study. A database including a larger number of individuals from different ethnicities, ages and geographic regions would help to build a model more representative of the target population, therefore increasing the accuracy and the confidence in the results (Cho et al., 2021; Mehrabi et al., 2021; Wang et al., 2023). Although our results indicate a good potential for the identification of anal fluids, further research must be done on this subject. This novel approach could also be tested in the future for other body fluids relevant to the forensic field. For instance, semen has been shown to have a singular microbiome and it could be identified by an algorithm like SourceTracker (Dobay et al., 2019; Yao et al., 2020).

## 2.7 Supplementary files

**Table S2.1.** Information asked to participants with a potential to have an impact on the microbiome composition

No	Age	Menstruation when sampling	Menopause	Days since last ejaculation	Antibiotics
1	43	No	No	12	No
2	46	Yes	No	n/a	No
3	43	No	No	n/a	No
4	56	No	Yes	7	No
5	59	No	Yes	7	No
6	55	No	Yes	n/a	No
7	43	Yes	No	4	No
9	40	No	No	n/a	No
10	49	No	Pre-menopause	n/a	No
11	46	No	No	0	7 weeks
12	32	No	No	n/a	No
13	49	No	No	n/a	No
17	38	No	No	3	No
18	33	No	No	2	No
20	36	No	No	n/a	No

**Table S2.2.** Participant's contribution of fluids for each sample

<b>Tube</b>	<b>Individual</b>	<b>Source</b>
<b>5</b>		1 Anal
<b>14</b>		2 Anal
<b>25</b>		5 Anal
<b>3</b>		1 Oral
<b>32</b>		4 Oral
<b>41</b>		3 Oral
<b>4</b>		1 Skin
<b>15</b>		2 Skin
<b>42</b>		3 Skin
<b>1</b>		3 Vaginal
<b>18</b>		4 Vaginal
<b>19</b>		7 Vaginal
<b>20</b>		10 Vaginal
<b>21</b>		13 Vaginal
<b>2</b>		2 Vaginal
<b>17</b>		6 Vaginal
<b>22</b>		18 Vaginal
<b>31</b>		12 Vaginal
<b>40</b>		11 Vaginal
<b>6</b>	6B1-6C	vaginal + skin (1W)*
<b>7</b>	12B1-12A	vaginal + oral (1W)
<b>8</b>	9B2-9E	vaginal + anal (1W)
<b>9</b>	13C-13E	anal + skin (1W)
<b>10</b>	18C-18E	anal + skin (1W)
<b>11</b>	17B2-17E	vaginal + anal (1W)
<b>12</b>	13B1-13A	vaginal + oral (1W)
<b>13</b>	7B1-7C	vaginal + skin (1W)
<b>26</b>	11B1-12C	vaginal + skin (2W)
<b>27</b>	20B1-18A	vaginal + oral (2W)
<b>28</b>	5B2-11E	vaginal + anal (2W)
<b>29</b>	11C-13E	anal + skin (2W)
<b>30</b>	3B1-4B1	vaginal + vaginal (2W)



<b>35</b>	10B1-11C	vaginal + skin (2W)
<b>36</b>	18B1-20A	vaginal + oral (2W)
<b>37</b>	1B2-7E	vaginal + anal (2W)
<b>38</b>	12C-11E	anal + skin (2W)
<b>39</b>	1B1-2B1	vaginal + vaginal (2W)
<b>44</b>	9B1-9C	vaginal + skin (1W)
<b>45</b>	17B1-17A	vaginal + oral (1W)
<b>46</b>	20B2-20E	vaginal + anal (1W)
<b>47</b>	9C-9E	anal + skin (1W)
<b>48</b>	5B1-6B1	vaginal + vaginal (2W)
<b>49</b>	7B1-9B1	vaginal + vaginal (2W)
<b>50</b>	10B1-11B1	vaginal + vaginal (2W)
<b>34</b>	4E	Anal
<b>43</b>	3E	Anal
<b>16</b>	2A	Oral
<b>23</b>	5A	Oral
<b>24</b>	5C	Skin
<b>33</b>	4C	Skin

\* W indicates the number of women contributing to the mixture sample

## **Chapitre 3 – Impact of winter on the epinecrotic microbiome in a humid, continental (Dfb) climate in Quebec, Canada**

**Audrey-Anne Leblanc<sup>1\*</sup>, Shari L. Forbes<sup>2</sup>, François-Joseph Lapointe<sup>1</sup>**

<sup>1</sup> Département de sciences biologiques, Université de Montréal, Montréal, Québec, Canada

<sup>2</sup> Department of Chemistry and Biochemistry, University of Windsor, Windsor, Ontario, Canada

\*Corresponding author: [audrey-anne.leblanc@umontreal.ca](mailto:audrey-anne.leblanc@umontreal.ca)

(Article est en préparation, non soumis)

### 3.1 Résumé

Le processus de décomposition humaine est complexe et hautement tributaire du climat. Récemment, une nouvelle installation en recherche taphonomique appelée REST(ES) a ouvert ses portes au Québec (Canada), permettant la recherche sur la décomposition humaine dans un climat continental humide. Le but de cette étude est de caractériser le microbiome épinérotique de 2 donneurs dont les corps ont été donnés à REST(ES) qui ont été placés avant l'hiver et ont été étudiés tout au long de celui-ci. En ayant une meilleure compréhension des changements dans la composition bactérienne pendant l'hiver, nous voulions tester s'il était possible de construire un modèle prédictif qui pourrait déterminer si un corps a été placé avant ou après l'hiver. Pour ce faire, nous avons placé sur le sol, dans les installations de REST(ES), un donneur à la fin septembre, et l'autre donneur 1 semaine plus tard au début d'Octobre. Par la suite, nous avons échantillonné les donneurs sur différentes parties du corps (nez, oreille, bouche, torse, main, pied, sol) pour 256 et 249 jours respectivement. L'ADN bactérien fut extrait, un séquençage de nouvelle génération a été utilisé pour séquencer l'ADN et les données ont été analysées par des méthodes bio-informatiques. Nos résultats montrent que la diversité alpha est plus basse sur les sites de la tête (oreille, nez, bouche) que les autres sites du corps échantillonnés. La plupart des familles bactériennes qui sont différemment abondantes selon les saisons sont absentes avant l'hiver et augmente en abondance pendant ou après l'hiver. Le modèle prédictif que nous avons testé pour déterminer si un corps avait passé l'hiver ou non avait une exactitude de 80% et avait tendance à sous-estimer le temps d'échantillonnage lorsqu'il faisait erreur, prédisant des laps de temps plus court que la réalité. Ainsi, il y a assez de différences dans la composition du microbiome épinérotique avant et après l'hiver dans cette région du Québec pour bâtir un modèle prédictif pour savoir si un corps a passé l'hiver ou non.

**Mots-clés:** Microbiome épinérotique, Décomposition humaine, Science judiciaire, Microbiologie judiciaire, Nérobiome

## 3.2 Abstract

The human decomposition process is a complex continuum that is highly dependent on the climate. In 2020, a new human taphonomic facility (HTF) opened in Québec (Canada), allowing decomposition research to be performed in a humid continental (Dfb) climate. The purpose of the current study was to characterize the epinecrotic microbiome of donors (bodies) that were placed before winter and were exposed to the sub-zero temperatures and snowfall of the winter season. By better understanding the changes in bacterial composition during winter, we aimed to test if it was possible to build a prediction model that could determine whether a body was placed before or after winter. Two donors were placed on the soil surface at the HTF, one at the end of September, and the other, one week later at the beginning of October. We sampled them at different body sites, for 256 and 249 days respectively by rubbing swab on the skin of the different body sites tested (nose, ear, mouth, torso, hand, feet, soil). Bacterial DNA was extracted, next-generation sequencing was used to sequence the DNA, and microbiome data were analyzed through bioinformatics techniques. Our results show that the alpha diversity is lower on sites from the head (ear, nose, mouth) than from sites elsewhere on the body. Most of the bacterial families that are differentially abundant between seasons are absent before winter and increase during or after winter. The prediction model that we tested to determine if a body was exposed to the winter climate was 80% accurate and tended to underestimate the time of sampling when it was inaccurate. Therefore, there are sufficient differences within the epinecrotic microbiome between seasons in this region of Québec to build a model predicting whether a body was placed before or after the winter months.

**Keywords:** Epinecrotic microbiome, Human decomposition, Forensic science, Forensic microbiology, Necrobiome

### 3.3 Introduction

Forensic taphonomy is the science that studies the decomposition process of a body after death (Dirkmaat et Cabo, 2016; Ueland et al., 2014). It is important to understand the process of decomposition, especially to establish the circumstances of medico-legal death investigations. There is a multitude of intrinsic and extrinsic factors that can influence decomposition, making our understanding of this process challenging. Temperature, humidity, presence of scavengers, or the acidity of the deposition environment, have a great influence on how the body will decompose (Campobasso et al., 2001; Vass, 2011). Age, body mass and health/medical conditions also play an important role in the decomposition process (Campobasso et al., 2001; Wescott, 2018).

The presence of insects can accelerate the decomposition process, but insects are not always present at the time of death, and may take time to arrive at and colonize a body (Amendt et al., 2011; Pechal et al., 2014a). On the other hand, microbes composing the human microbiome are present antemortem and continue colonizing the body after death since they are active decomposers (Hauther et al., 2015; Pechal et al., 2018). For this reason, the study of the microbiome during decomposition is promising to explore. Two main methods involving the post-mortem microbiome have been proposed: the study of the epinecrotic microbiome and the study of the thanatomicrobiome (Can et al., 2014; Pechal et al., 2013; Pechal et al., 2018). The thanatomicrobiome (or thanatobiome) is the post-mortem microbiome of the viscera, while the epinecrotic microbiome is the community of microbes living on the surface of the body after death (Can et al., 2014; Javan et al., 2016; Pechal et al., 2014b). The epinecrotic microbiome is highly influenced by temperature, humidity, insects and other environmental factors, limitations that have less impact on the thanatomicrobiome (Budowle et al., 2019). On the other hand, the thanatomicrobiome is harder to study since it is more invasive and requires a method to collect samples without disturbing the environment of the bacteria (such as exposing them to oxygen since their habitat is anaerobic), which can be a challenge for researchers (DeBruyn et Hauther, 2017).

Most studies have looked at changes in the relative abundances of taxa through time when it comes to assessing the post-mortem microbiome (Belk et al., 2018; Hyde et al., 2015; Johnson et al., 2016; Metcalf et al., 2016, Pechal et al., 2014a). The dominant phylum of the epinecrotic microbiome at the fresh stage of decomposition has been reported to be Proteobacteria for most body sites (Hyde et al., 2015; Pechal et al., 2013; Pechal et al., 2014a). As decomposition progresses, the proportion of Proteobacteria declines as the proportion of Firmicutes increases (Hyde et al., 2015). Bacteria from the Xanthomonadaceae family will then colonize the body. These bacteria are associated with the Sarcophagidae insects during myiasis (Hyde et al., 2015).

Climate plays a major role in the decomposition process and it is important to have decomposition data for a multitude of geographic regions. Data from one geographic region cannot be extrapolated to another where the climate is significantly different (Forbes, 2017), for example using data collected in tropical regions to infer the decomposition process in temperate regions. Research in cold temperate climates has revealed that the phyla Proteobacteria, Bacteroides and Firmicutes are predominant in this environment (Cameron et al., 2015; Møller et al., 2013; Mortazavi et al., 2015). It has been demonstrated that the bacteria living in the Arctic or in snowy environments are still active and can thrive without issue (Zhu et al., 2020). Despite the high albedo of the snow that tends to more quickly cool the soil beneath it, the accumulation of thick snow can have an insulating effect on the ground beneath, thus increasing its temperature (Tan et al., 2022). It is likely that decomposition by the microbiome will continue, albeit at a slower rate in winter, and a change in the relative microbial abundance will be distinct during the winter season.

In 2020, the first human taphonomic research (HTF) facility opened in the province of Québec (Canada), making it possible to study the impact of winter on decomposing human bodies (Pecsi et al., 2020). This exploratory study was thus designed to evaluate the composition of the epinecrotic microbiome and observe its changes throughout Québec's winter. By collecting data before, during, and after the winter months, we aimed to build a model assessing the impact of sub-zero temperatures and snowfall on the composition of the epinecrotic microbiome in order to predict whether a body was deposited before or after winter.

## 3.4 Materials and Methods

### Sample collection

The research was conducted at Research in Experimental and Social Thanatology (Recherche en Sciences Thanatologiques [Expérimentales et Sociales], REST[ES]) a human decomposition facility in Canada affiliated with the Université du Québec à Trois-Rivières (UQTR) (Pecsi et al., 2020). Two bodies were sampled at multiple time points for 256 days for the first donor, and 235 days for the second one. The first body was placed at the end of September while the second body was placed 7 days later. The sampling was performed at different body sites, twice a week until November, at 5 time points during winter, and bimonthly after winter. For the purpose of this research, we established the beginning of winter by the first day of sampling where a thin layer of ice was persisting on the bodies during the day. The donors were selected by the body donation program of the anatomy Department of UQTR (ScELERA-20-13).

Microbiome samples were collected with HydraFlock® sterile flocked swabs with a plastic handle with a breakpoint. The individual collecting the samples was dressed in a Tyvek® protective suit and wore a double layer of gloves to avoid contamination. The first donor was sampled at 7 different sites: nose, right ear, mouth, torso, right hand, right foot and soil (at the abdomen level). For this donor, all sites were sampled until June except for the nose and right ear which were sampled until the end of May due to the high degree of decomposition that made sampling impossible beyond that date. During winter, the soil was not sampled because it was frozen. The second donor was only sampled at two anatomical sites (torso and mouth). Sampling started once a week, and became more extended when winter arrived on day 45. After the sampling, the swabs were immediately put into DNeasy® Powersoil® kit tubes and stored at 4°C until DNA extraction. The sampling protocol was approved by the Université du Québec à Trois-Rivières ethics committee (SCELERA-20-13) and by the Université de Montréal ethics committee (CERSES-20-124-D).

### Accumulated degree days

A weather station was installed inside the REST[ES] facility to record temperature and precipitation (Hobo U30 Station No Remote Communication, Onset Computer Corporation). The

weather station recorded the temperature every 15 minutes of a 24-hour period. To calculate the accumulated degree days (ADD), we computed the average temperature of 96 recordings of every day. The Köppen-Geiger classification places the REST[ES]facility in Bécancour in the Dfb zone, a cold continental climate with warm summers (Peel et al., 2007). Since the temperature during winter goes below 0°C, accumulated degree days were calculated in Kelvin degrees (KADD).

### **DNA extraction and next-generation sequencing**

Bacterial DNA was extracted following the manufacturer's protocol for the MoBio Powersoil DNA isolation kit. One modification was made from the original protocol, we added a 15 minutes incubation period at 70°C after adding the C1 buffer (Castelino et al., 2017). The modification was made to increase microbial cell lysis efficiency. The extracted DNA was quantified before being sent to sequencing with a Qubit 2.0 Fluorometer (Invitrogen) to obtain DNA concentration. Even when the concentration was below 10ng/μl, as recommended by the sequencing facility at Génome Québec, all samples were sent for amplification.

Sequencing was performed with Illumina MiSeq technology and targeted the V4 hypervariable region of the 16S rRNA gene. 515F/806R primer set were used for the amplification (Caporaso et al., 2011). To avoid detecting false patterns, the extractions were randomized. To detect contamination during the process of extraction and PCR amplification, negative controls were used.

### **Data analysis**

Demultiplexing was performed in Qiime2 (Bolyen et al., 2019) and Dada2 was used for the denoising process (Callahan et al., 2016). 2,643,603 sequences were amplified from 155 samples. A mean of 18,358 sequences per sample was obtained. A sampling depth of 4000 was chosen allowing us to keep 92.9% of the samples in the analysis. 8 of the 11 samples filtered out by the rarefication were the negative controls, only 3 were samples from bodies. The presence of an unrecognized specie of the *Shigella* genus was observed in the negative controls. The removal of the DNA sequences from this genus was applied to our ASV table. After rarefication, a total of 30 phyla containing 305 families of bacteria were observed within the 144 samples remaining. To assess alpha diversity (diversity of bacterial community within the sample), we used the Shannon



index due to its reduced sensitivity to the differences in sample depths (Haegeman et al., 2013; Preheim et al., 2013). Alpha diversity results were obtained using the q2-diversity function in Qiime2. Kruskal-Wallis statistical was used to test the significance of alpha diversity results.

Unweighted UniFrac distances (Lozupone et al., 2011) were computed to analyze beta diversity. Significance of the beta diversity was tested with PERMANOVA. Principle coordinates analysis (PCoA) was generated from these distances and Emperor (Vázquez-Baeza et al., 2017) was used to visualize the ordination plot. To establish bacterial families that are differentially abundant throughout the seasons, analysis of composition of microbiomes (ANCOM) was used (Mandal et al., 2015). Taxa that were significantly abundant were thereafter put in the longitudinal volatility function in Qiime2 to assess the volatility of the relative abundance of those taxa through time. Finally, a Random-Forest regression method (Breiman, 2001) was used to test the accuracy of a prediction model on whether or not a body would have gone through a winter season. To so, we trained a classifier using the “classify-sample” pipeline in Qiime2 and to be able to test every sample in the data set, we used nested cross validation.

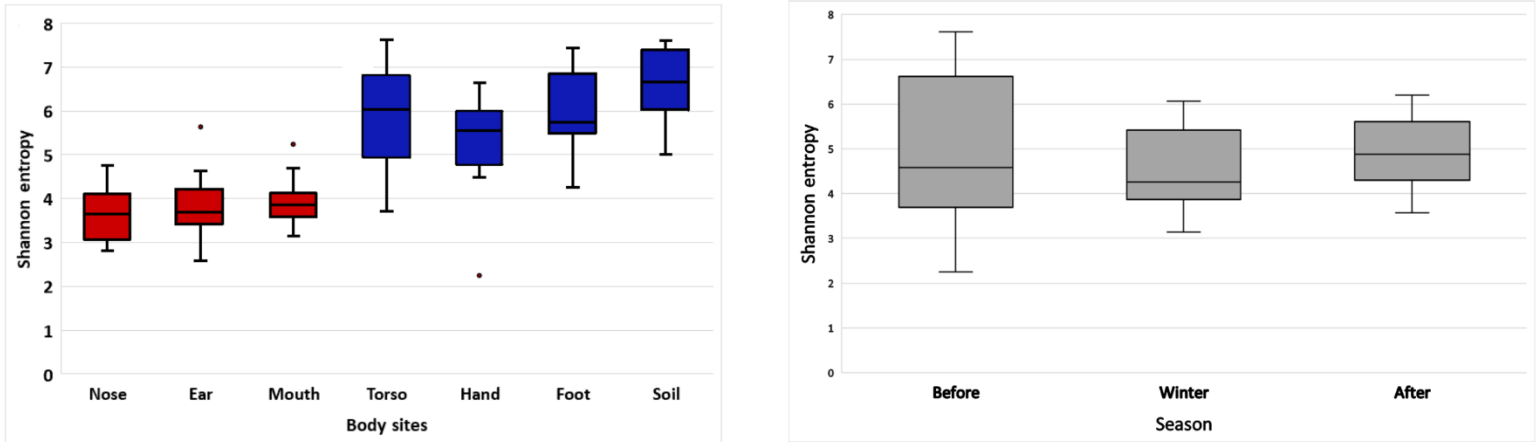
## **3.5 Results**

### **Alpha and Beta diversity analysis**

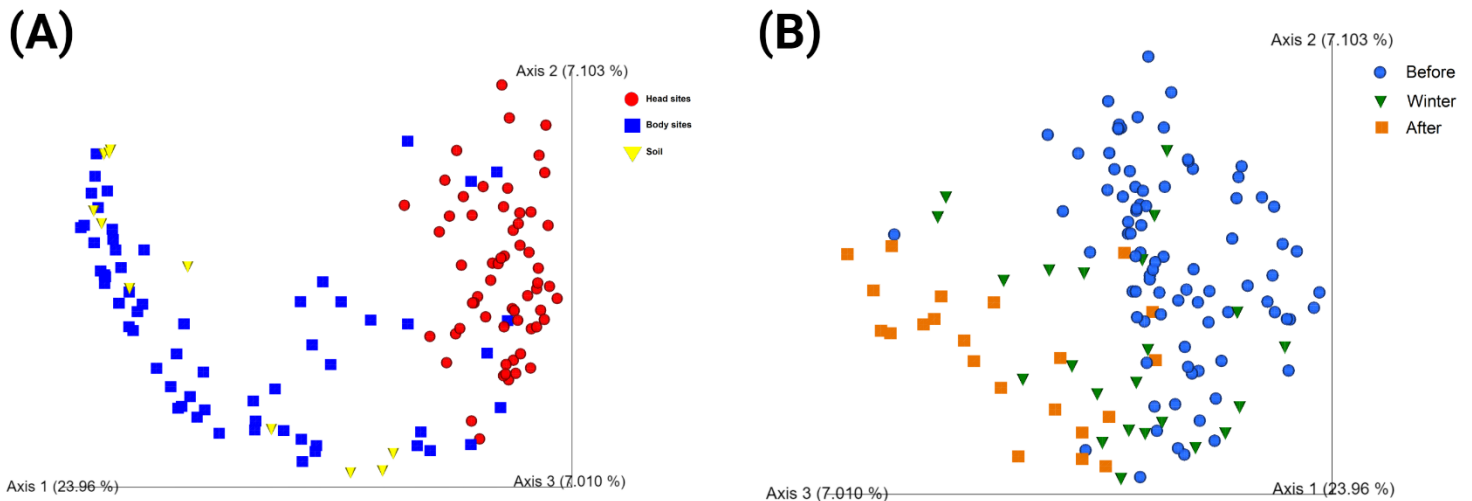
Alpha diversity of each sampling site and for each season was calculated. The results show a significant difference in diversity ( $H= 78.59$ ,  $p= 6.96e-15$ ) for samples collected from the head sites (nose, ear, mouth) compared to the rest of the body sites and soil (Fig. 3.1). Diversity within the head samples was lower than that of other samples. However, diversity did not have a significant change throughout the different seasons ( $H= 1.70$ ,  $p= 0.427$ ) when all body sites were gathered.

Unweighted UniFrac distances were computed to assess beta diversity and the effect of seasons and body sites on the epinecrotic microbiome. Using PERMANOVA, the results (Tab. 3.1) show that the microbial diversity of samples collected before winter is significantly higher than in other seasons ( $p$ -value = 0.001). The principal coordinate analysis (PCoA) confirmed the PERMANOVA

results, exhibiting a clear demarcation between each season, and between head sites and all other sites (Fig. 3.2).



**Figure 3.1** - Bacterial composition differences between seasons and between body sites. Alpha diversity differences between the different body sites sampled and between the different seasons represented by Shannon entropy. Grey dots represent outliers. The different seasons were analyzed independently of the body sites and body sites were analyzed independently of the seasons.



**Figure 3.2** - PCoA visualization of unweighted UniFrac distances based on (A) sampling sites and (B) seasons.

**Table 3.1** - PERMANOVA analysis calculating distances between samples from the three different seasons.

Group 1	Group 2	Sample size	Permutations	Pseudo-F	p-value	q-value
<b>After</b>	Before	108	999	6.417389	0.001	0.001
	Winter	45	999	3.623121	0.001	0.001
<b>Before</b>	Winter	109	999	4.140587	0.001	0.001

### Differential abundance changes throughout seasons

To identify bacterial taxa that could indicate if a body went through winter, differential abundances at the family taxonomic level was processed with ANCOM. Since diversity among different body sites is significant, they were each tested separately. ANCOM was thus used to test the difference in abundance for each body site throughout the different seasons to assess whether some families of bacteria could help predict if a body went through winter or not (Tab 3.2). ANCOM tests the null hypothesis that the abundance between season won't change between the seasons. If the abundance of a bacterial family stays within a threshold, established by the test and hidden to the user, the family will stay hidden. However, if the abundance passes the threshold established, the null hypothesis will be rejected. The *W* statistic indicates the amount of times the null hypothesis got rejected for each families that passed the threshold. The higher the *W* statistics, the more different is the abundance between the seasons and the more confident we can be regarding the results.

The results of ANCOM revealed that all sites, except torso, differed between seasons on at least for one bacterial family. Namely, the family *Clostridiaceae* was identified for all head sites as an indicative taxon between seasons.

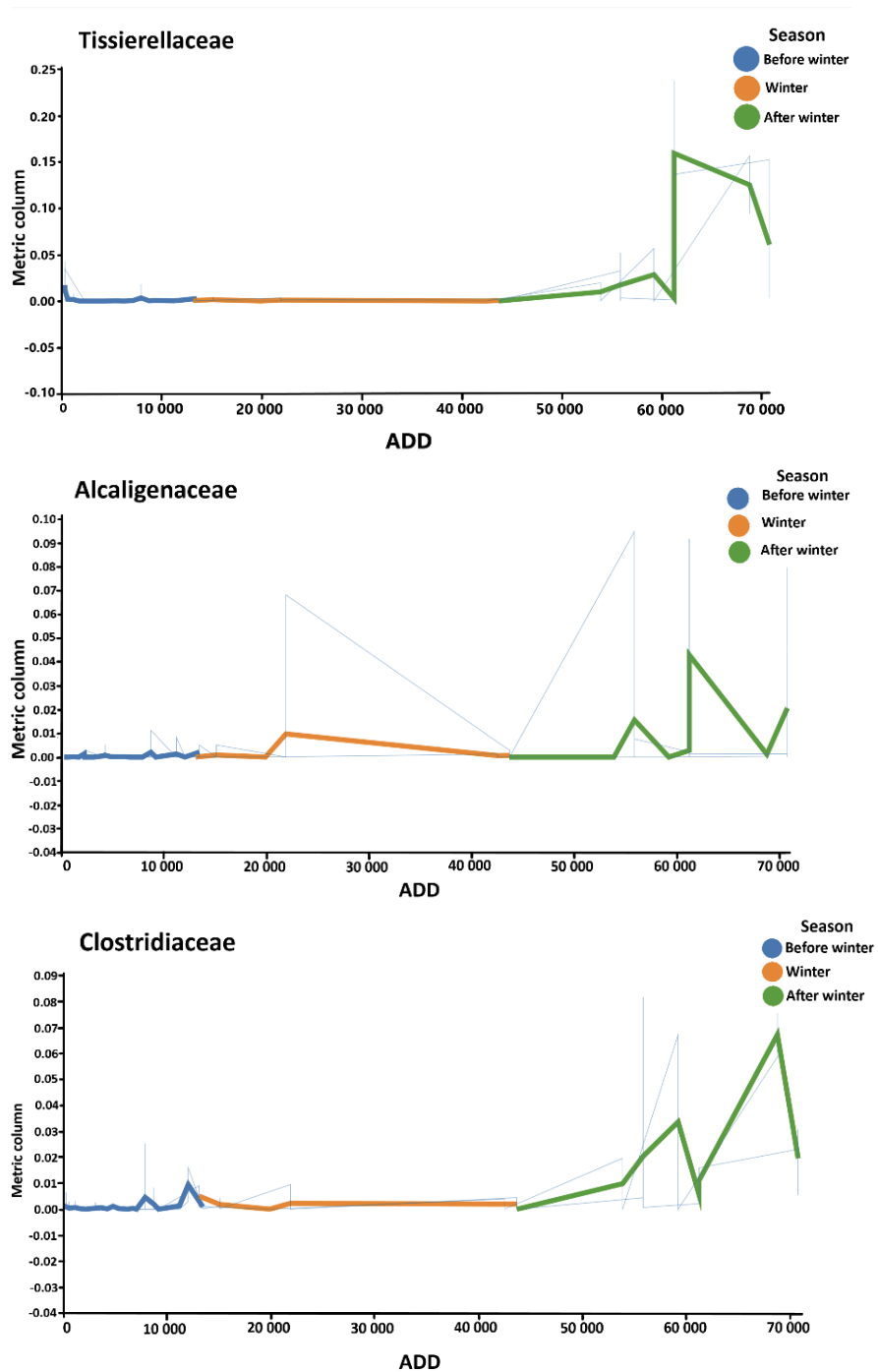
Taxa	W statistic
<b>Nose</b>	
<i>Bacillaceae</i>	35
<i>Alcaligenaceae</i>	30
<i>Clostridiaceae</i>	27
<b>Ear</b>	
<i>Tissierellaceae</i>	31
<i>Alcaligenaceae</i>	26
<i>Bacillaceae</i>	26
<i>Clostridiaceae</i>	25
<i>Rhizobiaceae</i>	21
<b>Mouth</b>	
<i>Tissierellaceae</i>	56
<i>Clostridiaceae</i>	54
<i>Pseudomonadaceae</i>	53
<b>Hand</b>	
<i>Aurantimonadaceae</i>	76
<i>Erysipelotrichaceae</i>	64
<b>Foot</b>	
<i>Bacillaceae</i>	186
<b>Soil</b>	
<i>Pseudomonadaceae</i>	149

**Table 3.2.** - Families of bacteria detected by ANCOM to be differently abundant for each body site among seasons.

### Volatility analysis of the distinctive taxa

Volatility of the 5 most important taxa in ANCOM (*Bacillaceae*, *Clostridiaceae*, *Alcaligenaceae*, *Tissierellaceae* and *Pseudomonadaceae*) was performed. We defined importance here by either the bacterial families with the strongest W statistic values or the families that appeared the most times within the different body sites. The relative abundance of *Clostridiaceae*, *Alcaligenaceae* and *Tissierellaceae* through time (ADD) is presented in Figure 3.3 and Figure S3.1 shows relative abundance of *Bacillaceae* and *Pseudomonadaceae*. This visualization reveals that most of the

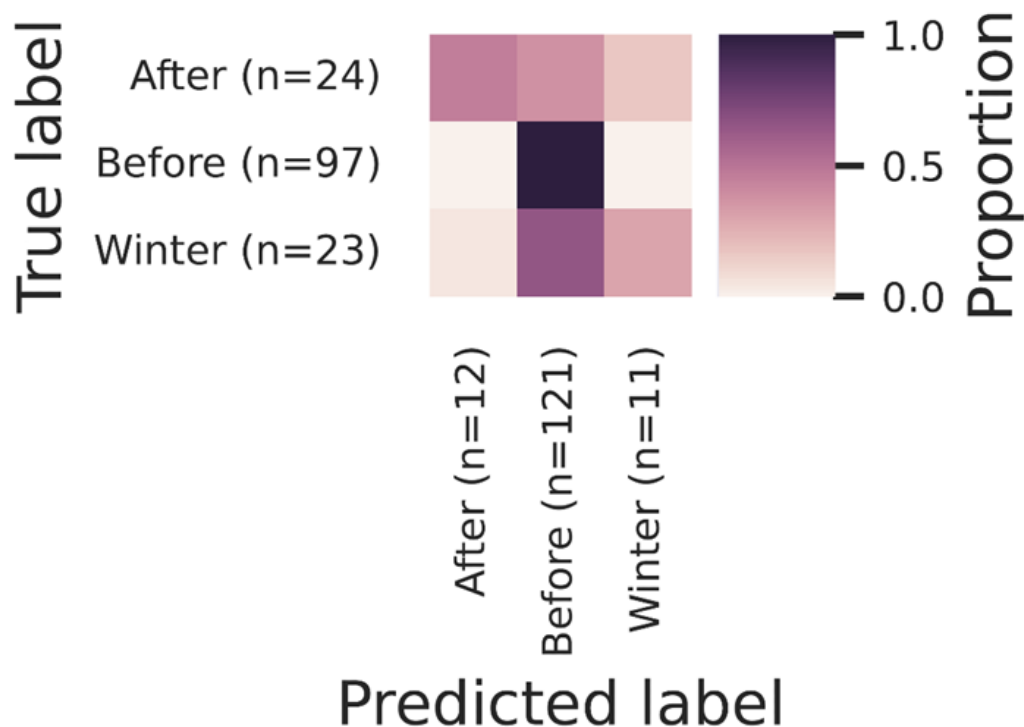
indicator taxa have a reduced abundance before winter, can increase or not during the winter season, but are relatively more abundant after winter.



**Figure 3.3** - Relative abundance through time of the 3 of the 5 most important bacterial taxa identified by ANCOM. The 3 taxa represented here show the strongest difference between “before winter” and “after winter”

### Prediction model to evaluate if a body went through winter

We build a prediction model to see if it would be possible to evaluate if a body went through winter, considering that abundance in different taxa was significantly different between seasons. We tested that prediction with a Random-Forest probability regression in Qiime2. This analysis produced predictions with an overall accuracy ratio of 80%. No false negative results were obtained for the “Before” season. That is, that no “Winter” or “After” samples were identified as a “Before” sample by the prediction model.



**Figure 3.4** - Random forest regression results of the season prediction model. The confusion matrix shows the proportion of samples that were rightly and wrongly predicted to have been sampled either before winter, during winter or after winter by the regression model versus the true number of samples collected for each season.

### 3.6 Discussion

The results of our study show that a prediction model aiming to determine if a body went through winter or not can be accurate. Indeed, the Random Forest Regression model never failed to predict that a sample came from a body that went through the winter season when it did not. However, in some instances, the model predicted that a sample was collected before winter when in reality it was collected during or after winter. The model overestimated time of collection for only one sample, but underestimated time of collection for 28 samples. Practically, this implies that the model could predict that a sample went through winter when it did not, but it would not predict that a sample did not go through winter when it actually did. Considering the reduced numbers of donors ( $n = 2$ ) that were used to build the model, further research with more donors will strengthen the confidence of the model and could improve its accuracy. Also, bodies that did not go through winter that would have been placed after winter were not included in the research. Those types of bodies could have shown if the epinecrotic microbiome observed after winter was typical of the season, or if it was that way because of winter.

Additionally, all taxa identified by the ANCOM results (Table 3.1) increased in abundance during or after winter (Fig. 3.3). Different hypotheses could explain this phenomenon. A previous study showed that some bacterial metabolic activities were more efficient at a lower temperature ( $2^{\circ}\text{C}$ ) than at room temperature ( $22^{\circ}\text{C}$ ) (Carter et al., 2006). Also, at the beginning of decomposition, there were insects present on the bodies, competing with the bacteria for resources and potentially changing the environment with some chemical secretions (Pechal et al., 2013). The low temperatures and snow killed all of the insects still present. Yet, the insects did not come back in great numbers after winter since both bodies were mummified and therefore, less attractive to them, leaving a lot of resources to the bacteria (Adserias-Garriga et al., 2017; Pechal et al., 2013). Very few studies on the impact of winter on bacterial composition were performed with temperatures below  $0^{\circ}\text{C}$ , requiring additional studies to better understand the role of the cold winter climate on decomposition. Furthermore, the bacterial families *Pseudomonadaceae* and *Alcaligenaceae* identified by ANCOM were also identified as families with abundance changes on the head of dead mice (Metcalf et al., 2013). Our results show that there are more families of bacteria that are differentially abundant between seasons on the head than on the rest of the

body sites. Yet, the W statistic was the strongest at the foot and for the soil that was collected around the abdominal regions, as suggested by past research (Adserias-Garriga et al., 2017). Despite the presence of indicator taxa, it has been suggested that to build a prediction model, it is preferable to use the entire bacterial community instead of a specific taxon or groups of taxa (Johnson et al., 2016). ANCOM and volatility results were useful to better understand the dynamic of the community across seasons and a good base for further research.

The PCoA plot (Fig. 3.2) can explain some of the results provided by the prediction model. Indeed, there was an overlap in Beta diversity between the “before” and the “winter” samples, as well as with some of the “after” samples. There were no clearly distanced clusters representing each season (Before-Winter-After). For this research, the beginning of winter was established by the first day of sampling where a layer of thin ice was detected on the bodies during the day. However, nights with temperatures below 0°C were already happening for multiple days before the establishment of the winter season, making the demarcation of Before/Winter subtle and possibly harder to determine for a prediction model. A threshold of snow precipitations might be a better factor to consider to measure the impact of winter on the epinecrotic microbiome. However, fewer samples were taken during the winter season, giving less data for the model to build predictions. For future studies, a more extensive winter sampling could help making better predictions. Another approach would be to remove the winter season from the model and build it with only two options – “before winter” and “after winter” – defining clear clusters that could strengthen the model accuracy.

Additionally, the choice of body parts sampled is a major factor in the composition of the epinecrotic microbiome. There was a major difference between the composition and diversity of samples collected from the head compared to those collected from the rest of the body or the soil. This finding is not different from other studies on the subject, making the choice of sampling site important for investigating the epinecrotic microbiome (Pechal et al., 2013; Pechal et al., 2018). This could be explained by the fact that head sites were all cavities (ear, nose, mouth) that were rapidly filled with maggots, therefore preventing environmental contamination (Campobasso et al., 2001). Body sites, on the other hand, were fully exposed to the environment

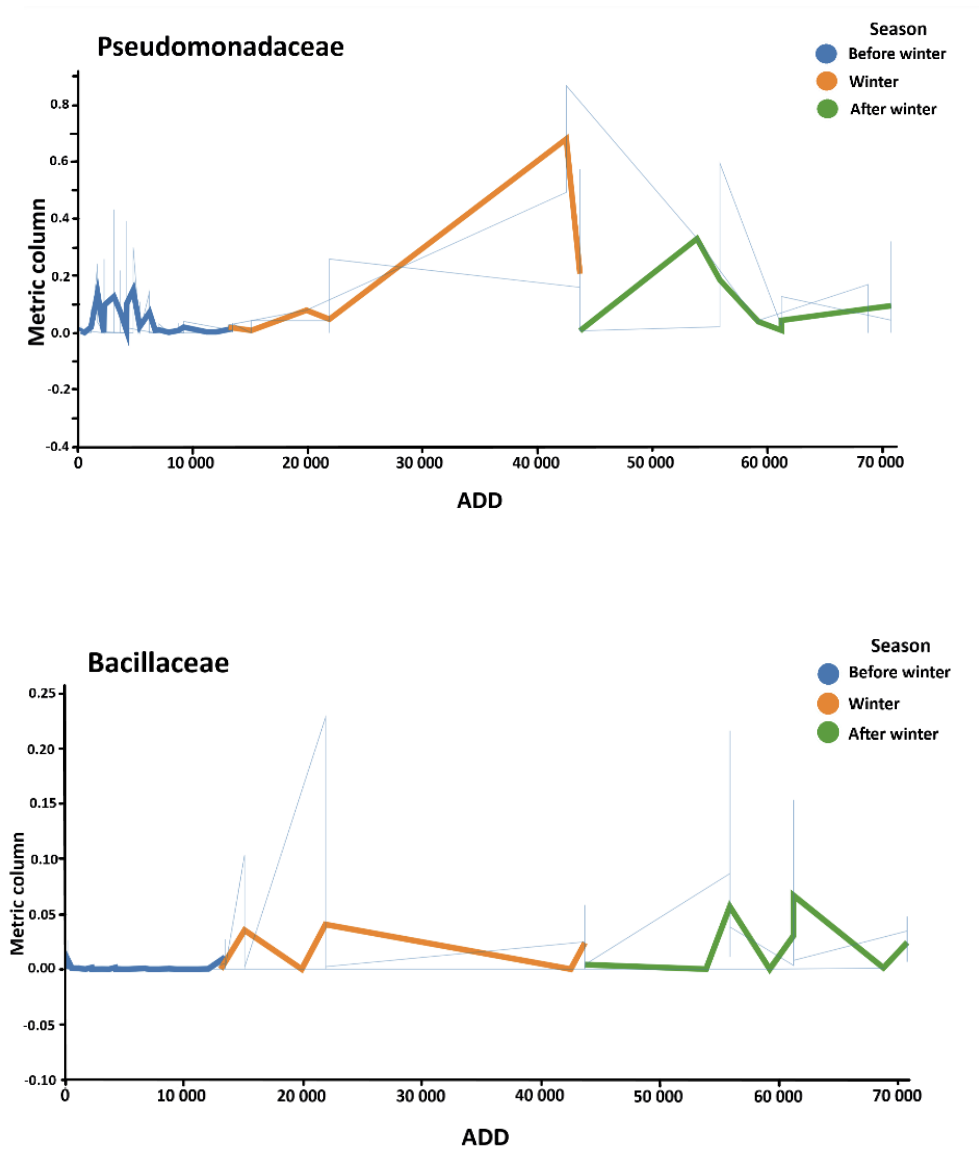


and conducive to contamination, as much by insects than by small animals, dirt, dust and plants surrounding the bodies.

### **3.7 Conclusion**

This study gives a broader view of the composition of the human epinecrotic microbiome in a cold continental climate like Canada (Ribéreau-Gayon et al., 2023). Since most of the decomposition studies are conducted in hotter climates, our results a study like this in Quebec will allow to see differences in the bacterial composition changes between the different regions. Moreover, it will allow for future progress not only in the fundamental knowledge of epinecrotic microbiome, but also to test existing predictive algorithms and improve them with the new data from this region of Canada. We provided a portrait of the epinecrotic microbiome throughout the winter season in Canada. We demonstrated the differences in diversity between sampling sites from the head that were lower than those from the rest of the body. We also put forward the difference in diversity between the different seasons through which the bodies decomposed. Diversity was slightly higher and more consistent after the winter season. The differences in bacterial communities across seasons allowed a prediction model to provide accurate estimations of whether a body went through winter or not. The model never estimated that a body was sampled after winter when it was sampled before winter, but did underestimate the moment of death for 19% of samples. Further research with a higher number of bodies should be done to improve the model. However, it is possible to believe that a winter predicting model could be used for estimating post-mortem intervals in Canada.

### 3.8 Supplementary files



**Figure S3.1.** Volatility results of *Bacillaceae* and *Pseudomonadaceae* showing the relative abundance of the 2 families of bacteria through time.

## Chapitre 4: Conclusion

La microbiologie judiciaire est une science encore en développement, mais son potentiel d'application à des enquêtes réelles ne fait pas de doute. La force des nouvelles méthodes de séquençage qui se sont développées ces dernières années nous a permis d'ajouter la microbiologie judiciaire au coffre à outils potentiel pour les enquêteurs et scientifiques judiciaires. Les deux études présentées dans ce mémoire ont permis d'élargir notre compréhension du microbiome et son application en sciences judiciaires.

Premièrement, l'objectif du premier article était de tester une méthode utilisant le microbiome pour identifier les fluides vaginaux dans des mélanges de fluides. Les résultats de l'étude montrent non seulement que cette méthode est efficace pour identifier les fluides vaginaux, mais également que pour construire un modèle fort qui identifie la majorité des mélanges, il suffit d'utiliser quelques sources pures pour l'entraînement. Ainsi, un entraînement avec des sources pures plus variées que celles utilisées dans l'étude permettraient d'améliorer les résultats du modèle et potentiellement d'augmenter sa sensibilité. Cette méthode pourrait concrètement être utilisée en sciences judiciaires avec confiance. Comme elle utilise des méthodes d'extraction d'ADN déjà présentes dans les laboratoires de biologie judiciaire, elle serait un ajout efficace et avantageux aux procédés existants. De plus, les résultats montrent qu'il s'agit d'une avenue concrète à explorer non seulement pour les fluides vaginaux, mais qu'il y a aussi un potentiel pour d'autres fluides biologiques comme les fluides anaux. La force du modèle prédictif est élevée et son utilisation réelle dans des laboratoires de sciences judiciaires pourrait se voir sous peu.

Deuxièmement, le second article visait à étudier le changement du microbiome épiniécrotique pendant l'hiver ainsi qu'à analyser les différents sites anatomiques échantillonnés. Les résultats ont montré qu'il y avait une différence significative dans l'abondance de certaines familles de bactéries et que ces différences pourraient permettre de prédire si un corps avait passé l'hiver ou non. Le modèle utilisé pour tester ces prédictions avait une exactitude de 80%. Cependant, comme aucun corps n'a été placé après l'hiver, il est difficile de dire avec certitude si les changements observés sont dus à l'hiver ou s'il s'agit simplement du comportement normal du microbiome épiniécrotique au printemps. Ainsi, d'autres recherches sur la décomposition dans un

climat comme celui du Québec doivent être réalisées pour augmenter notre compréhension du microbiome épimicrotique et pouvoir éventuellement l'utiliser dans des enquêtes judiciaires. Cependant, il ne fait pas de doute qu'il s'agit d'un domaine de recherche avec du potentiel et qu'il est pertinent de continuer à l'étudier.

## Références bibliographiques

Adserias-Garriga, J., Hernández, M., Quijada, N. M., Lázaro, D. R., Steadman, D., & Garcia-Gil, J. (2017). Daily thanatomicrobiome changes in soil as an approach of postmortem interval estimation: an ecological perspective. *Forensic Science International*, 278, 388-395.

Akutsu, T., Motani, H., Watanabe, K., Iwase, H., & Sakurada, K. (2012). Detection of bacterial 16S ribosomal RNA genes for forensic identification of vaginal fluid. *Legal Medicine*, 14(3), 160-162.

Amendt, J., Richards, C. S., Campobasso, C. P., Zehner, R., & Hall, M. J. R. (2011). Forensic entomology: applications and limitations. *Forensic Science, Medicine, and Pathology*, 7(4), 379-392.

Belk, A., Xu, Z. Z., Carter, D. O., Lynne, A., Bucheli, S., Knight, R., & Metcalf, J. L. (2018). Microbiome data accurately predicts the postmortem interval using random forest regression models. *Genes*, 9(2), 104.

Blackman, S., Stafford-Allen, B., Hanson, E. K., Panasiuk, M., Brooker, A.-L., Rendell, P., Ballantyne, J., & Wells, S. (2018). Developmental validation of the ParaDNA® Body Fluid ID System—A rapid multiplex mRNA-profiling system for the forensic identification of body fluids. *Forensic Science International: Genetics*, 37, 151-161.

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C.J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J.,... Caporasso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, 37(8), 852-857.

Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32.

Budowle, B., Connell, N. D., Bielecka-Oder, A., Colwell, R. R., Corbett, C. R., Fletcher, J., Forsman, M., Kadavy, D. R., Markotic, A., Morse, S. A., Murch, R. S., Sajantila, A., Schmedes, S. E., Ternus, K.

L., Turner, S. D., & Minot, S. (2014). Validation of high throughput sequencing and microbial forensics applications. *Investigative Genetics*, 5(1), 9.

Budowle, B., Schutzer, S. E., & Morse, S. A. (2019). *Microbial Forensics*. Elsevier Science. <https://books.google.ca/books?id=zRbBDwAAQBAJ>

Butler, J. M. (2009). *Fundamentals of Forensic DNA Typing*. Elsevier Science. <https://books.google.ca/books?id=-OZeEmqzE4oC>

Byard, R. W. (2017). Timing: the Achilles heel of forensic pathology. *Forensic Science, Medicine, and Pathology*, 13(2), 113-114.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), 581-583.

Cameron, K. A., Hagedorn, B., Dieser, M., Christner, B. C., Choquette, K., Sletten, R., Crump, B., Kellogg, C., & Junge, K. (2015). Diversity and potential sources of microbiota associated with snow on western portions of the Greenland Ice Sheet. *Environmental Microbiology*, 17(3), 594-609.

Campobasso, C. P., Di Vella, G., & Introna, F. (2001). Factors affecting decomposition and Diptera colonization. *Forensic Science International*, 120(1-2), 18-27.

Can, I., Javan, G. T., Pozhitkov, A. E., & Noble, P. A. (2014). Distinctive thanatomicrobiome signatures found in the blood and internal organs of humans. *Journal of Microbiological Methods*, 106, 1-7.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N., & Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America*, 108 (supplement\_1), 4516-4522.

Carter, D. O., & Tibbett, M. (2006). Microbial decomposition of skeletal muscle tissue (*Ovis aries*) in a sandy loam soil at different temperatures. *Soil Biology and Biochemistry*, 38(5), 1139-1145.

- Carter, D. O., Tomberlin, J. K., Benbow, M. E., & Metcalf, J. L. (2017). *Forensic Microbiology*. Wiley.  
<https://books.google.ca/books?id=AR6jDgAAQBAJ>
- Castelino, M., Eyre, S., Moat, J., Fox, G., Martin, P., Ho, P., Upton, M., & Barton, A. (2017). Optimisation of methods for bacterial skin microbiome investigation: primer selection and comparison of the 454 versus MiSeq platform. *BMC Microbiology*, 17(1).  
<https://doi.org/10.1186/s12866-017-0927-4>
- Cho, H.-W., & Eom, Y.-B. (2021). Forensic analysis of human microbiome in skin and body fluids based on geographic location. *Frontiers in Cellular and Infection Microbiology*, 11, 695191.
- Choi, A., Shin, K.-J., Yang, W. I., & Lee, H. Y. (2014). Body fluid identification by integrated analysis of DNA methylation and body fluid-specific microbial DNA. *International Journal of Legal Medicine*, 128(1), 33-41.
- Consortium, H. M. P. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402), 207-214.
- Costello, E. K., Lauber, C. L., Hamady, M., Fierer, N., Gordon, J. I., & Knight, R. (2009). Bacterial community variation in human body habitats across space and time. *Science*, 326(5960), 1694-1697.
- Cresci, G. A. M., & Izzo, K. (2019). Chapter 4 - Gut Microbiome. In M. L. Corrigan, K. Roberts, & E. Steiger (Eds.), *Adult Short Bowel Syndrome* (pp. 45-54). Academic Press.  
<https://doi.org/https://doi.org/10.1016/B978-0-12-814330-8.00004-4>
- DeBruyn, J. M., & Hauther, K. A. (2017). Postmortem succession of gut microbial communities in deceased human subjects. *PeerJ*, 5, e3437.
- Dirkmaat, D. C., & Cabo, L. L. (2016). Forensic archaeology and forensic taphonomy: Basic considerations on how to properly process and interpret the outdoor forensic scene. *Academic Forensic Pathology*, 6(3), 439-454.

- Dobay, A., Haas, C., Fucile, G., Downey, N., Morrison, H. G., Kratzer, A., & Arora, N. (2019). Microbiome-based body fluid identification of samples exposed to indoor conditions. *Forensic Science International: Genetics*, 40, 105-113.
- Doi, M., Gamo, S., Okiura, T., Nishimukai, H., & Asano, M. (2014). A simple identification method for vaginal secretions using relative quantification of *Lactobacillus* DNA. *Forensic Science International: Genetics*, 12, 93-99.
- Fleming, R. I., & Harbison, S. (2010). The development of a mRNA multiplex RT-PCR assay for the definitive identification of body fluids. *Forensic Science International: Genetics*, 4(4), 244-256.
- Forbes, S. (2017). Body farms. *Forensic Science, Medicine and Pathology*, 13(4), 477-479.
- Frumkin, D., Wasserstrom, A., Budowle, B., & Davidson, A. (2011). DNA methylation-based forensic tissue identification. *Forensic Science International: Genetics*, 5(5), 517-524.
- Haegeman, B., Hamelin, J., Moriarty, J., Neal, P., Dushoff, J., & Weitz, J. S. (2013). Robust estimation of microbial diversity in theory and in practice. *The ISME Journal*, 7(6), 1092-1101.
- Hammerschlag, M. R., Alpert, S., Rosner, I., Thurston, P., Semine, D., McComb, D., & McCormack, W. M. (1978). Microbiology of the vagina in children: normal and potentially pathogenic organisms. *Pediatrics*, 62(1), 57-62.
- Hanssen, E. N., Avershina, E., Rudi, K., Gill, P., & Snipen, L. (2017). Body fluid prediction from microbial patterns for forensic application. *Forensic Science International: Genetics*, 30, 10-17.
- Harbison, S., & Fleming, R. (2016). Forensic body fluid identification: state of the art. *Research and Reports in Forensic Medical Science*, 11.
- Hardison, R. C. (2012). Evolution of Hemoglobin and Its Genes. *Cold Spring Harbor Perspectives in Medicine*, 2(12), a011627-a011627. <https://doi.org/10.1101/cshperspect.a011627>
- Hauther, K. A., Cobaugh, K. L., Jantz, L. M., Sparer, T. E., & DeBruyn, J. M. (2015). Estimating time since death from postmortem human gut microbial communities. *Journal of Forensic Sciences*, 60(5), 1234-1240.



Hayman, J., & Oxenham, M. (2016). *Human Body Decomposition*. Elsevier Science. <https://books.google.ca/books?id=CcAPDQEACAAJ>

Hyde, E. R., Haarmann, D. P., Petrosino, J. F., Lynne, A. M., & Bucheli, S. R. (2015). Initial insights into bacterial succession during human decomposition. *International Journal of Legal Medicine*, 129, 661-671.

Javan, G. T., Finley, S. J., Abidin, Z., & Mulle, J. G. (2016). The thanatobiome: a missing piece of the microbial puzzle of death. *Frontiers in Microbiology*, 7, 225.

Johnson, H. R., Trinidad, D. D., Guzman, S., Khan, Z., Parziale, J. V., DeBruyn, J. M., & Lents, N. H. (2016). A machine learning approach for using the postmortem skin microbiome to estimate the postmortem interval. *PloS One*, 11(12), e0167370.

Juusola, J., & Ballantyne, J. (2005). Multiplex mRNA profiling for the identification of body fluids. *Forensic Science International*, 152(1), 1-12.

Kaszubinski, S. F., Pechal, J. L., Schmidt, C. J., Heather, J. R., Benbow, M. E., & Meek, M. H. (2019). Evaluating bioinformatic pipeline performance for forensic microbiome analysis. *Journal of Forensic Sciences*, 65(2), 513–525.

Knights, D., Kuczynski, J., Charlson, E. S., Zaneveld, J., Mozer, M. C., Collman, R. G., Bushman, F. D., Knight, R., & Kelley, S. T. (2011). Bayesian community-wide culture-independent microbial source tracking. *Nature Methods*, 8(9), 761-763.

Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J., & Knight, R. (2011). UniFrac: an effective distance metric for microbial community comparison. *The ISME Journal*, 5(2), 169-172.

Maisonhaute, J.-É., & Forbes, S. L. (2023). Decomposition and insect succession on human cadavers in a humid, continental (Dfb) climate (Quebec, Canada). *International Journal of Legal Medicine*, 137(2), 493-509.

Mandal, S., Van Treuren, W., White, R. A., Eggesbø, M., Knight, R., & Peddada, S. D. (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microbial Ecology in Health and Disease*, 26(1), 27663.

Manor, O., Dai, C. L., Kornilov, S. A., Smith, B., Price, N. D., Lovejoy, J. C., Gibbons, S. M., & Magis, A. T. (2020). Health and disease markers correlate with gut microbiome composition across thousands of people. *Nature Communications*, 11(1), 5206.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)*, 54(6), 1-35.

Metcalf, J. L., Wegener Parfrey, L., Gonzalez, A., Lauber, C. L., Knights, D., Ackermann, G., Humphrey, G. C., Gebert, M. J., Van Treuren, W., Berg-Lyons, D., Keepers, K., Guo, Y., Bullard, J., Fierer, N., Carter, D. O., & Knight, R. (2013). A microbial clock provides an accurate estimate of the postmortem interval in a mouse model system. *eLife*, 2, e01104

Metcalf, J. L., Xu, Z. Z., Weiss, S., Lax, S., Van Treuren, W., Hyde, E. R., Song, S. J., Amir, A., Larsen, P., & Sangwan, N. (2016). Microbial community assembly and metabolic function during mammalian corpse decomposition. *Science*, 351(6269), 158-162.

Miller, E. A., Beasley, D. E., Dunn, R. R., & Archie, E. A. (2016). Lactobacilli dominance and vaginal pH: why is the human vaginal microbiome unique? *Frontiers in Microbiology*, 7, 1936.

Møller, A. K., Søbørg, D. A., Abu Al-Soud, W., Sørensen, S. J., & Kroer, N. (2013). Bacterial community structure in high-arctic snow and freshwater as revealed by pyrosequencing of 16S rRNA genes and cultivation. *Polar Research*, 32(1), 17390.

Mortazavi, R., Attiya, S., & Ariya, P. A. (2015). Arctic microbial and next-generation sequencing approach for bacteria in snow and frost flowers: selected identification, abundance and freezing nucleation. *Atmospheric Chemistry and Physics*, 15(11), 6183-6204.

Nafte, M., & Dalrymple, B. (2015). *Crime and Measurement: Methods in Forensic Investigation*. Carolina Academic Press. <https://books.google.ca/books?id=EUHMmsgEACAAJ>

O'Callaghan, A., & Van Sinderen, D. (2016). Bifidobacteria and their role as members of the human gut microbiota. *Frontiers in Microbiology*, 7, 925.

Pechal, J. L., Benbow, M. E., Crippen, T. L., Tarone, A. M., & Tomberlin, J. K. (2014). Delayed insect access alters carrion decomposition and necrophagous insect community assembly. *Ecosphere*, 5(4), art45. <https://doi.org/10.1890/es14-00022.1>

Pechal, J. L., Crippen, T. L., Benbow, M. E., Tarone, A. M., Dowd, S., & Tomberlin, J. K. (2014). The potential use of bacterial community succession in forensics as described by high throughput metagenomic sequencing. *International Journal of Legal Medicine*, 128(1), 193-205.

Pechal, J. L., Crippen, T. L., Tarone, A. M., Lewis, A. J., Tomberlin, J. K., & Benbow, M. E. (2013). Microbial community functional change during vertebrate carrion decomposition. *PloS One*, 8(11), e79035.

Pechal, J. L., Schmidt, C. J., Jordan, H. R., & Benbow, M. E. (2018). A large-scale survey of the postmortem human microbiome, and its potential to provide insight into the living health condition. *Scientific Reports*, 8(1), 5724.

Pecsi, E. L., Bronchti, G., Crispino, F., & Forbes, S. L. (2020). Perspectives on the establishment of a Canadian human taphonomic facility: The experience of REST [ES]. *Forensic Science International: Synergy*, 2, 287-292.

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, 11(5), 1633-1644.

Petersen, D., & Kovacs, F. (2014). Phenolphthalein false-positive reactions from legume root nodules. *Journal of Forensic Sciences*, 59(2), 481-484.

Preheim, S. P., Perrotta, A. R., Friedman, J., Smilie, C., Brito, I., Smith, M. B., & Alm, E. (2013). Computational methods for high-throughput comparative analyses of natural microbial communities. *Methods in Enzymology*, 531, 353-370.

Ravel, J., Gajer, P., Abdo, Z., Schneider, G. M., Koenig, S. S. K., McCulle, S. L., Karlebach, S., Gorle, R., Russell, J., Tacket, C. O., Brotman, R. M., Davis, C. C., Ault, K., Peralta, L., & Forney, L. J. (2011). Vaginal microbiome of reproductive-age women. *Proceedings of the National Academy of Sciences of the United States of America*, 108 (supplement\_1), 4680-4687.

Ribéreau-Gayon, A., Carter, D. O., & Forbes, S. (2023). Developing a new scoring method to evaluate human decomposition in a humid, continental (Dfb) climate in Quebec. *Journal of Forensic Sciences*, 68(2), 536-548.

Rinninella, E., Raoul, P., Cintoni, M., Franceschi, F., Miggiano, G. A. D., Gasbarrini, A., & Mele, M. C. (2019). What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. *Microorganisms*, 7(1), 14.

Setzer, M., Juusola, J., & Ballantyne, J. (2008). Recovery and stability of RNA in vaginal swabs and blood, semen, and saliva stains. *Journal of Forensic Sciences*, 53(2), 296-305.

Tackmann, J., Arora, N., Schmidt, T. S. B., Rodrigues, J. F. M., & Von Mering, C. (2018). Ecologically informed microbial biomarkers and accurate classification of mixed and unmixed samples in an extensive cross-study of human body sites. *Microbiome*, 6(1), 192.

Tan, X., Luo, S., Li, H., Hao, X., Wang, J., Dong, Q., & Chen, Z. (2022). Investigating the effects of snow cover and vegetation on soil temperature using remote sensing indicators in the three river source region, China. *Remote Sensing*, 14(16), 4114.

Ueland, M., Breton, H. A., & Forbes, S. L. (2014). Bacterial populations associated with early-stage adipocere formation in lacustrine waters. *International Journal of Legal Medicine*, 128, 379-387.

Vass, A. A. (2011). The elusive universal post-mortem interval formula. *Forensic science international*, 204(1-3), 34-40.

Vázquez-Baeza, Y., Gonzalez, A., Smarr, L., McDonald, D., Morton, J. T., Navas-Molina, J. A., & Knight, R. (2017). Bringing the dynamic microbiome to life with animations. *Cell Host & Microbe*, 21(1), 7-10.

Vennemann, M., Scott, G., Curran, L., Bittner, F., & Tobe, S. S. (2014). Sensitivity and specificity of presumptive tests for blood, saliva and semen. *Forensic Science, Medicine, and Pathology*, 10(1), 69-75.

- Wang, R., Chaudhari, P., & Davatzikos, C. (2023). Bias in machine learning models can be significantly mitigated by careful training: Evidence from neuroimaging studies. *Proceedings of the National Academy of Sciences of the United States of America*, 120(6), e2211613120.
- Wescott, D. J. (2018). Recent advances in forensic anthropology: decomposition research. *Forensic Sciences Research*, 3(4), 278-293.
- Xiaoming, W., Jing, L., Yuchen, P., Huili, L., Miao, Z., & Jing, S. (2021). Characteristics of the vaginal microbiomes in prepubertal girls with and without vulvovaginitis. *European Journal of Clinical Microbiology & Infectious Diseases*, 40(6), 1253-1261.
- Yao, T., Han, X., Guan, T., Wang, Z., Zhang, S., Liu, C., Liu, C., & Chen, L. (2020). Effect of indoor environmental exposure on seminal microbiota and its application in body fluid identification. *Forensic Science International*, 314, 110417.
- Youssef, N., Sheik, C. S., Krumholz, L. R., Najar, F. Z., Roe, B. A., & Elshahed, M. S. (2009). Comparison of species richness estimates obtained using nearly complete fragments and simulated pyrosequencing-generated fragments in 16S rRNA gene-based environmental surveys. *Applied and Environmental Microbiology*, 75(16), 5227-5236.
- Yu, H., & Diamandis, E. P. (1995). Prostate-specific antigen in milk of lactating women. *Clinical Chemistry*, 41(1), 54-58.
- Zhang, J., Liu, W., Simayijiang, H., Hu, P., & Yan, J. (2022). Application of microbiome in forensics. *Genomics, Proteomics & Bioinformatics*. 21(1), 97-107
- Zhu, C., Miller, M., Lusskin, N., Bergk Pinto, B., Maccario, L., Häggblom, M., Vogel, T., Larose, C., & Bromberg, Y. (2020). Snow microbiome functional analyses reveal novel aspects of microbial metabolism of complex organic compounds. *Microbiology Open*, 9(9), e1100.